Two projects: improved prediction of metabolite retention times to aid identification

and

analysis of ammonium transporters from *Marchantia polymorpha*

A Dissertation

SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

BY

Nu Wang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisors: Dr. John M. Ward, Dr. Paul G. Boswell

September 2018

Acknowledgements

I would like to express my gratitude to advisors Dr. John Ward and Dr. Paul Boswell, for their detailed guidance, untiring moral and financial support, and enlightened supervision. As the first graduate student completing a Ph.D degree in the Boswell lab, I received limitless opportunities and training from Paul directly. I am very appreciative of all the things Paul has done for me during the first two years of my Ph.D studies. His detailed guidance helped me adapt to a new research field and a new culture. His enlightened supervision and moral support helped me publish my research results. His continuous financial support helped me finish my Ph.D. studies even after he left academia and I transitioned to a different lab. I would also like to take this opportunity to thank my advisor Dr. John Ward and my coworker Dr. Anke Reinders, for their endless patience, constructive criticism, and well-developed supervision during the last three years of my Ph.D. studies. I appreciated all the things that John and Anke have done for me, the uncountable effort and time that they spent on my project designs and editing my writing. I appreciate all the great suggestions that led me through the most difficult time when I was transitioning to a new research field, looking for an internship, and struggling to move forward to a new stage of life after graduation.

I also would like to thank my committee members Dr. Adrian Hegeman, Dr. Sue Gibson, Dr. Jane Glazebrook, and Dr. Jerry Cohen for sharing their valuable experience in their respective fields, providing support and suggestions for my research projects, and spending time in my committee meetings.

I owe a great debt to my lab fellow Praphapan Lasin for her company, Dr. Michael Wilson for his help on reviewing manuscripts, Dr. Josh Hewitt for his guidelines and help on my project, and Dr. Daniel Abate Pella for his moral support during my Ph.D. studies. Thanks to PMB cohort Erin Evans, Joseph Guhlin, Stephen Brockman, and Molly Tillmann for being there innumerable times when I needed help and support. Coming to the U of M has been a great and unforgettable experience, which not only gave me tremendous opportunities to excel in my career but also connect with great friends like Snober Ahmed, and Xin Chen.

A special thanks to all the staff members in PMB who have always been there to extend support and help whenever needed.

Funding for the projects was provided by National Institutes of General Medical Sciences [R01GM098290], the University of Minnesota Office of the Vice President for Research, the Minnesota Agricultural Experiment Station, and the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy (DE-FG02-10ER15886) to J.W.

Humans have the power to express their feelings but there are some emotions where words always fall short. I would like to thank my parents (Huitian Wang & Dongjin Li) for their unlimited effort and my sweet sister Di Wang who has always been there to help me. In addition to this, I thank very dear friends who never made me feel like I was alone here and showed their constant support, patience and love.

Dedications

I dedicate this dissertation to my advisors, my colleagues, my friends, my parents, my older sister, and my family members.

Ph.D. Products

Research Publications:

- Hanqing Guo*, **Nu Wang*,** Tami R. McDonald, Anke Reinders, and John M. Ward; MpAMT1;2 from Marchantia polymorpha is a high-affinity plasma membrane ammonium transporter, Plant & Cell Physiology; 2018. * Co-first author.
- **Nu Wang#** and Paul Boswell; Accurate Prediction of Retention in Hydrophilic Interaction Chromatography (HILIC) by Back Calculation of High Pressure Liquid Chromatography (HPLC) Gradient Profiles, Journal of Chromatography A, http://dx.doi.org/10.1016/j.chroma.2017.08.050. # Corresponding author.

Conferences:

- **Nu Wang**, Hanqing Guo, Tami R. McDonald, Anke Reinders, and John M. Ward. “High- and low-affinity ammonium transporters in Marchantia polymorpha”, International Conference of Arabidopsis Research, St Louis, MO, 2017.
- **Nu Wang**, and Paul G. Boswell. “Accurate Retention Prediction in LC-MS by "Retention Projection" and Its Prediction Accuracy in Complex Plant Matrices”, Annual meeting of the American Society of Plant Biologist, Minneapolis, MN, 2015.
- **Nu Wang,** and Paul G. Boswell. “Accurate Prediction of Retention in Hydrophilic Interaction Chromatography (HILIC) by Back-Calculation of HPLC Gradient Profiles”, Bioscience and Biotechnology, HaiKo, China, 2014.

- **Nu Wang**, and Paul G. Boswell. “Accurate Prediction of Retention in Hydrophilic Interaction Chromatography (HILIC) by Back-Calculation of HPLC Gradient Profiles”, American Society for Mass Spectrometry, Baltimore, Maryland, 2014.

Invited Talks:

- **Nu Wang**, Hanqing Guo, Tami R. McDonald, Anke Reinders, and John M. Ward. “Ammonium transporters MpAMT1;2 and Its Regulation in *Marchantia polymorpha*”, Microbial and Plant Genomics Institute (MPGI) Annual Meeting, Saint Paul, MN, 2018.
- **Nu Wang**, Hanqing Guo, Tami R. McDonald, Anke Reinders, and John M. Ward. “High- and low-affinity ammonium transporters in *Marchantia polymorpha*”, Plant and Microbial Biology Annual Meeting, Saint Paul, MN, 2017.

Abstract

Gradient retention times are difficult to project from the underlying retention factor (*k*) vs. solvent composition (***φ***) relationships. A major reason for this difficulty is that gradients produced by HPLC pumps are imperfect – gradient delay, gradient dispersion, and solvent mis-proportioning are all difficult to account for in calculations. However, we recently showed that a gradient "back-calculation" methodology, Retention Projection with Back-calculation (RPwB), can measure these imperfections and take them into account. In RPLC (Reverse Phase Liquid Chromatography), when the back-calculation methodology was used, error in projected gradient retention times is as low as could be expected based on repeatability in the *k* vs. ***φ*** relationships.

To extend the application of RPwB, we test the prediction accuracy in HILIC (Hydrophilic Interaction Liquid Chromatography) too. Compared with RPLC, HILIC presents a new challenge: the selectivity of HILIC columns drift strongly over time. Retention is repeatable in short time, but selectivity frequently drifts over the course of weeks. In this study, we set out to understand if the issue of selectivity drift can be avoided by doing our experiments quickly, and if there are any other factors that make it difficult to predict gradient retention times from isocratic *k* vs. ***φ*** relationships when gradient imperfections are taken into account with the back-calculation methodology. While in past reports, the accuracy of retention projections was >5%, the back-calculation methodology brought our error down to ~1%. This result was 6-43 times more accurate than projections made using ideal gradients and 3-5 times more accurate than the same retention projections made using offset gradients (i.e., gradients that only took gradient

delay into account). Still, the error remained higher in our HILIC projections than in RPLC. Based on the shape of the back-calculated gradients, we suspect the higher error is a result of prominent gradient distortion caused by strong, preferential water uptake from the mobile phase into the stationary phase during the gradient – a factor our model did not properly take into account. It appears that, at least with the stationary phase we used, column distortion is an important factor to take into account in retention projection in HILIC that is not usually important in RPLC.

However, this methodology has only been approved its robustness in preferred sample (solvents 73% acetonitrile and 27% water) in RPLC. Biological samples are often complicated and sample solvents vary. In this study, we set out to explore the sensitivity of RPwB under a limited set of extreme conditions. First, we selected nine sample solvents representing a wide range of polarities (methanol, ethanol, isopropanol, acetonitrile, acetone, dichloromethane, ethyl acetate, tetrahydrofuran, and toluene), and studied their effects on peak shapes, retention times, and prediction accuracy of RPwB. We found isopropanol, acetonitrile, and ethyl acetate were the top three that distorted peak shapes the most, and all these nine sample solvents shifted retention times and poorly retained compounds suffered more than well retained ones. Only ethyl acetate broke the prediction accuracy of RPwB and had a prediction error of 4.3 sec, which was more than the 3 sec that was the maximum deviation allowed for the successful application of RPwB in RPLC. Second, five plant samples (*Solanum lycopersicum* fruit, *Solanum lycopersicum* stems and leafs, *Nicotiana* flowers, *Nicotiana* leafs, and *Nicotiana forsteri* leafs) were randomly chosen and extracted using the four sample solvents (water, 70% ethanol, dichloromethane, and isopropanol) that are the most common ones and that

did not change the prediction accuracy of RPwB, and obtained the most concentrated plant extracts possible. The effect of plant matrices on retention times of nearly all analytes was negligible, despite their high concentrations. However, we observed a buildup of some plant matrix solutes in the column that lowered retention projection accuracy for two charged analytes, tetrabutylammonium and tetrapentylammonium. However, this buildup could be removed (and the accuracy of retention projections restored) by either flushing the column with a stronger mobile phase or decreasing the concentration of the injected sample.

The second part of the dissertation is in a different field- protein structure and function. Nitrogen is a limiting nutrient for plants. To understand the mechanisms that plants use to acquire nitrogen from the environment, it is useful to study diverse plants. Basal plants contain a small gene family of ammonium transporters within the AMT/MEP/Rh superfamily and transporters in the same family are used to take up ammonium in angiosperms. Here we characterized the transport activity of two ammonium transporters from *Marchantia polymorpha*, a liverwort and a representative of the most basal land plants.

Ammonium transporter MpAMT1;2 was shown to localize to the plasma membrane in *Marchantia* gametophyte thallus by stable transformation using a C-terminal citrine fusion. *MpAMT1;2* expression was studied using qRT-PCR and shown to be higher when plants were N deficient and lower when plants were grown on media containing ammonium, nitrate, or the amino acid glutamine. Expression in *Xenopus* oocytes and analysis by electrophysiology revealed that MpAMT1;2 is an electrogenic ammonium transporter with a very high affinity for ammonium (7 µM at pH 5.6 and a membrane

potential of -137 mV). A conserved inhibitory phosphorylation site identified in angiosperm AMT1s is also present in all AMT1s in Marchantia. Here we show that a phosphomimetic mutation T475D in MpAMT1;2 completely inhibits ammonium transport activity. The results indicate that MpAMT1;2 may be important for ammonium uptake into cells in the *Marchantia* thallus.

A second ammonium transporter, MpAMT1;5 was functionally expressed in both yeast and *Xenopus* oocytes. It showed a low-affinity for ammonium ($K_{0.5}$ of 0.38 mM at pH 5.6 and a membrane potential of -121 mV) when expressed in *Xenopus* oocytes and assayed using two-electrode voltage clamping. MpAMT1;5 was localized to the plasma membrane in gametophyte thalli of transgenic *Marchantia polymorpha* expressing a MpAMT1;5-citrine fusion. *MpAMT1;5* expression was studied using *M. polymorpha* transformed with a *MpAMT1;5* promoter-GUS fusion.

Table of Contents

List of Tables

List of Figures

Chapter 1

General Introduction and Objectives

1. System biology and metabolomics

System Biology was forged in the past 20 years by the advances of technologies in the biological sciences. However, our understanding of organisms at the molecular level is still very limited. The investigation of system biology at "omics" levels including genomics, transcriptomics, proteomics, and metabolomics will affect this and play a central role in biology's future (Dettmer et al., 2007). According to Dettmer *et al.*, (2006), the genome tells what can happen, the transcriptome indicates what appears to be happening, the proteome suggests what makes it happen, and the metabolome shows what has happened and is happening (Dettmer et al., 2007). Metabolomics, the identification and quantification of all metabolites in a biological system, precisely link gene functions to phenotypes and provides valid theories to explain the causes of biological effects. It is the closest block to the final "phenotypes". Metabolomics analysis can assess the substantial equivalence of genetically modified organisms. Moreover, metabolomics analysis can improve the understanding of metabolism such as novel metabolic pathway predictions and complete metabolic networks in organisms.

Metabolomics has been slowly developed for around 50 years and eventually reached its infancy stage in approximately 2000 with the development of metabolic profiling, the quantitative analysis of a set of metabolites in a selected biochemical pathway or a specific class of compounds (Trethewey et al., 1999). For example, in 1977, E. Jellum believed that diseases could be characterized if one is able to identify and determine the

concentration of all compounds inside of the human body (Jellum, 1977). However, due difficulties in developing powerful tools to analyze "total" compounds, it was impossible to study this area in depth in the last century even using the most advanced equipment: gas chromatography (GC) and mass spectrometry (MS) (Jellum, 1977). In 1999, Trethewey *et al.*, wrote that too little information was known about metabolic profiling. This situation slowed down the development of metabolic engineering since the well characterized genes, mRNAs, and proteins provided valuable but insufficient information for metabolic engineering due to the fact that changes in genes do not always cause changes in mRNA and proteins, and changes in proteins do not necessarily lead to changes in their activities (Trethewey et al., 1999).

The development of metabolic profiling in the plant kingdom was even slower, only a few studies were done before 2000. The most important study was conducted by Sauter *et al.*, in 1991, they found the changes in chromatographical patterns of herbicide-treated barley through studying metabolic profiling by using GC-MS (Sauter et al., 1991). This study was based on the well-established metabolic profiling in human disease studies from the 1970s (Jellum, 1977). In 2000, the first study of metabolic profiling in a plant was done by Fiehn *et al.* in *Arabidopsis thaliana* and quantified 326 compounds by using GC-MS from Arabidopsis leaf extracts (Fiehn et al., 2000). However, this study is far from a complete metabolic profiling. Even through the *Arabidopsis* genome has been completed sequenced, over 30% of its genes are not functionally classified, and only 9% have been experimentally characterized. Of this 9% , most of the characterizations are not based on rigid biochemical testing (The Arabidopsis Genome Initiative, 2000).

2. Metabolomics and metabolite identification

In the field of metabolomics, the biggest problem is metabolite identification due to the lack of sufficient technologies to identify metabolites with high confidence (Creek et al., 2014). For example, in plants, the estimated number of metabolites is 200,000 (Brown et al., 2009), but the reported metabolites with known structure is only 25% (De Luca and St Pierre, 2000).

The most common, accurate, and advanced technologies used to identify metabolites are Nuclear Magnetic Resonance (NMR) and MS (Lin et al., 2006). NMR is mainly used to analyze the chemical structure of small metabolites (<30 kDa) by using their stereochemistry and conformation information, whereas MS uses the information of M/Z (mass to charge ratio) and fragmentation pattern. However, MS is more popular than NMR since MS can identify metabolites with bigger mass weight (<500k Da), has higher sensitivity (10 -100 fold) and faster speed but with a cheaper price. Therefore, MS is a more generally applicable approach than NMR. In the past several decades, GC and Liquid Chromatography (LC) are employed prior to MS. The combination of GC-MS or LC-MS has more power to identify metabolites because metabolites with same molecular weight can be separated by GC or LC based on retention during the chromatography (Henion et al., 1998).

2.1 GC-MS and metabolites identification: mass, retention time, and fragmentation pattern

The application of GC-MS in metabolite identification is more common than LC-MS even though they produce same information: mass, retention time, and fragmentation pattern. However, the retention time information in GC-MS is more reproducible and calculable by Retention Indexing (RI), and the fragmentation pattern produced by

electron impact ionization (EI) in GC-MS is more consistent than the ones produced by electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) in LC-MS. Even though GC-MS has these advantages, there are still many “big holes” in using the information produced by GC-MS.

For example, Aharoni *et al.*, published metabolite identification using mass information from a super-accurate (0.0001) instrument, a Fourier Transform Ion Cyclotron Resonance Mass Sepectrometer (FTICR-MS). In that study, a metabolite was identified as vanillic acid since its extract mass is 167.0349 (Aharoni et al., 2002). However, 64 other metabolites can be found within 5ppm mass accuracy in the Metabolite and Tanden MS Database (METLIN) (Hirai et al., 2004). Unfortunately, that study has been cited 365 times and applied to metabolomics investigations (Hirai et al., 2004).

The chromatopgraphical information, retention time, can be useful to identify metabolites as long as its reproduciable. RI is a method which is used to predict retention time in GC-MS. Basically, RI predicts the identity of a metabolite by using retention information of bracketing standards on its nearest two sides (Rohrschneider, 1966). RI was derived from the prediction of retention volume of a compound of interest in a certain gas chromatographic volume in 1952. However, the accuracy of RI decreases significantly if experimental conditions are changed from original settings. For example, as shown in Fig. 1, anthracene is supposed to elute with two bracketing standards as calculated by retention indexing, but the nearest bracket compounds are n-heptadecane and n-octadecane if the temperature is lower than 150 ^{o}C, while these change to n-nonadecane and n-heneicosane if the temperature is higher than 300 ^{o}C. The accuracy of

RI will drop considerably if the measured index information for anthracene is collected at 150 ºC and used at 300 ºC. This limits the application of RI at various experimental conditions.

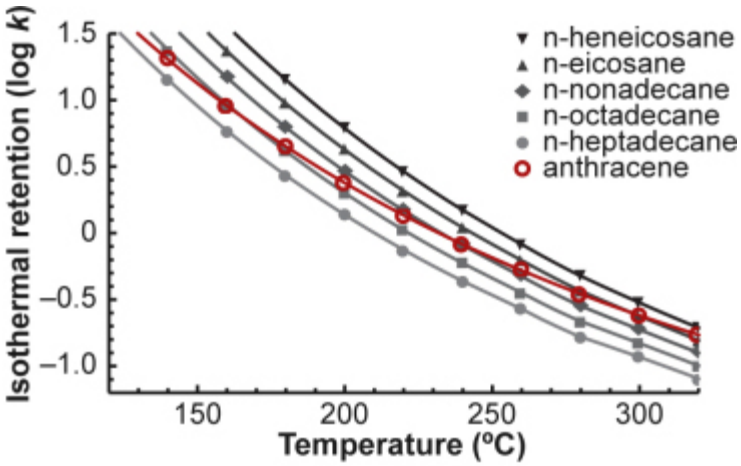


Figure 1. Isothermal retention (log *k*) vs. temperature (Tm) of four compounds. The nearest bracket compounds of anthracene are n-heptadecane and n-octadecane when the temperature lower than 150 ºC, these compounds change to n-nonadecane and n-heneicosane when temperature higher than 300 ºC. (http://www.retentionprediction.org/gc/howitworks.php).

Fragmentation pattern is the arrangement and intensity of small pieces of molecular ion, which is energetically unstable and breaks into different parts in mass spectrometry. For example, as showed in Fig. 2, the EI fragmentation pattern of pentane (C_5H_{12}). Fragmentation pattern of one molecule is different from others and believed consistent over different experimental conditions. As reported in NIST Mass Spectral Library (2014 version), there are 56,216 compounds in the GC Methods/RI Library which contains 242,466 spectra (http://chemdata.nist.gov). One can search the database and identify metabolites based on their fragmentation patterns if the compound of interest is in the

database. However, this identification can only eliminate big differences and narrow down to a level and it is unable to identify isomers.

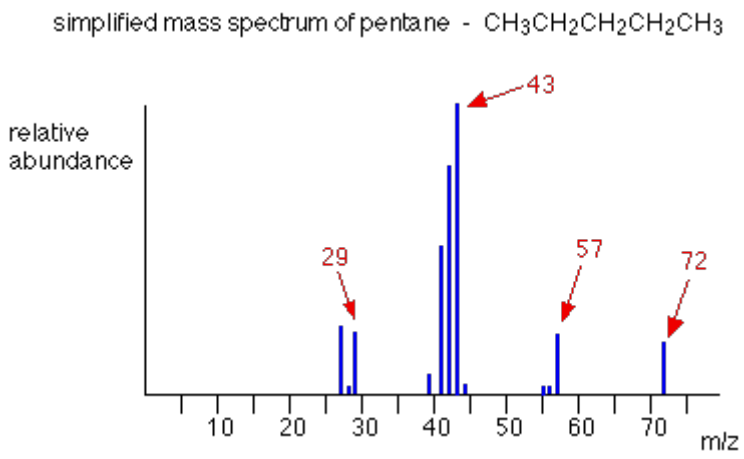


Figure 2. EI mass fragmentation pattern of pentane (C_5H_{12}) in GC-MS. the heaviest ion to pass through the mass spectrometry is m/z=72, the molecular ion. The tallest line is m/z=43 which represents the most common fragment to be formed (http://www.chemguide.co.uk/analysis/masspec/fragment.html)

2.2 LC-MS and metabolites identification: m/z and retention time

Even if GC-MS is better developed for identifying metabolites compared with LC-MS in terms of retention time information or fragmentation patterns, GC-MS is only used to identify volatile metabolites or metabolites that can be volatized after derivatization. To identify nonvolatile metabolites, LC-MS is used.

However, the information produced by LC-MS and used to identify metabolites is mass information alone. The retention information is unrepeatable due to both intentional factors (different solvents, LC-MS instruments, columns, flow rates, temperature) and unintentional factors such as columns (non-identical selectivity, sensitivity, and

repeatability), and instruments (imperfect behaviors of pumps, degasser, and auto-mixtures). Any of these factors cause unreproducible retention information (Boswell et al., 2011a, 2011b). The power of metabolite identification can be largely improved if both mass and retention information can be used together. For example, Boswell *et al*. (Barnes et al., 2013a; Boswell et al., 2011a) studied 7,307 compounds in the KEGG database (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000). Only 1.2% of the metabolites were identified by using mass information alone, while 44% (36 times more than MS alone) of them were identified by using both mass weight and retention information.

Therefore, it is important to develop a methodology to use retention information to predict metabolites using LC-MS. Retention Projection with Back-calculation (RPwB) is a novel method established by Paul. G. Boswell in 2011. The main concept of RPwB is measuring real gradient profiles produced by GC-MS and LC-MS instruments by using a set of well characterized standards, which are used to calibrate instruments and account for both intentional and unintentional factors. This real gradient profile is then applied to predict retention times of test compounds. The prediction accuracy tells how well the method will identify metabolites in the future. RPwB has been applied in GC-MS and reverse phase liquid chromatography (RPLC)-MS, the prediction errors are ±0.9 s (40 min ramp) and ±2.8 s (20 min gradient with 100 μL/min flow rate) respectively (Boswell et al., 2011a, 2011b, 2012). These represent studies with the best accuracy in the area of gradient retention prediction in GC-MS and RPLC-MS. In addition, Paul Boswell *et al.* studied the prediction accuracy of RPwB in RPLC-MS across labs and found it is 20-fold better than RI for uncharged compounds and 10-fold better for charged compounds

(Abate-pella et al., 2015). Based on these results, we believe RPwB will be a great tool to identify metabolites in the future.

2.3 Hydrophilic Interaction Chromatography (HILIC) and metabolite identification

HILIC is a type of liquid chromatography. In LC history, people first developed normal phase chromatography- hydrophilic stationary phase with hydrophobic (organic) mobile phase. Due to the unrepeatability, poor selectivity, and high control of water amount in the mobile phase, normal phase was gradually replaced by RPLC- the reverse of normal phase, which is the combination of hydrophobic stationary phase and hydrophilic mobile phase. However, there is an increasing requirement for separating polar compounds, especially in metabolomic analysis. HILIC, an alternative to RPLC, has become very popular. The biggest difference between HILIC and normal phase chromatography is their stationary and mobile phases. With very hydrophobic solvents, such as hexane and IPA, even 0.1% water in the mobile phase can change the repeatability in normal phase chromatography, whereas the mobile phase in HILIC is more diverse and more compatible with water.

HILIC has been used to separate compounds since 1975 (Hemström and Irgum, 2006; Linden and Lawhead, 1975). Actually, as early as 1951, Gregor *et al.*, described the surface of the stationary phase of HILIC as having a water enriched layer (Gregor and Collins, 1951), which can adsorb nonelectrolytes using ion-exchange resins and which was shown to separate glucose and the lyotropic series in 1954. In 1990, Alpert named the separation technique HILIC based on the hydrophilic stationary phase combined with a hydrophobic mobile phase (Alpert, 1990). However, even now, the retention mechanism of HILIC is still not totally clear. There has been a long debate on

"equilibration" or "adsorption" in HILIC. It is clear that the retention mechanism is more than liquid-liquid partitioning between analytes and the absorbed water layer on the surface of the stationary phase. Hydrogen bonding between polar functional groups and the stationary phase and electrostatic interactions between polar functional groups appear to be factors affecting retention. More importantly, the actual amount and role of water on the surface of HILIC materials should be further considered to fully analyze the elusive mechanism (Hemström and Irgum, 2006).

In HILIC, one method used to predict retention times in gradient eluents is called "equation prediction", which was published by Helen Gika. Equation prediction first measures the retention time of a mixture in several different gradient eluents. It then fits these retention times to an equation with several parameters. This equation is used to predict the retention time of the same mixture but in other gradient eluents. The prediction error of this method is less than 1%, which is more accurate than all other published results (Gika et al., 2012). However, this method could not be shared between labs since each lab would have to re-measure all the parameters to fit different HILIC systems. The second method that is used to predict gradient retention in HILIC was published by Eva Tyteca in 2014 and is more complicated. Instead of measuring gradient data only, Eva Tyteca measures both isocratic and gradient data of a set of mixtures, then fits this data in equations with different parameters. However, its overall prediction error is larger than 5%, which is too high to meet the high accuracy requirement in metabolite identification (Tyteca et al., 2014b).

3. Nitrate sources and Ammonium Transporters (AMTs) in *Marchantia polymorpha*

Nitrogen (N) is a plant macronutrient and N deficiency often limits growth in natural and agricultural environments. The amount of free N in soils is decided by mineralization, which converts organic N sources such as urea to inorganic N sources such as ammonium (through ammonification) and nitrate (through nitrification). Soil microorganisms are competitors of plants for the uptake of N. Both ammonium and nitrate uptake processes are influenced by soil conditions such as pH, temperature, and carbohydrates. However, ammonium is a more preferable N form than nitrate as nitrate has to be converted to ammonium through a series of reduction bio-reactions. Therefore, it costs less energy for plants to take up ammonium from the soil. Excessive uptake of ammonium can cause ammonium toxicity, which can inhibit plant growth and development and restrict photosynthesis. In plants, N is stored as glutamine, which is the first product of the ammonium/nitrate metabolic reaction.

Ammonium transporters (AMTs) play an essential role in taking up ammonium from the soil (U Ludewig et al., 2002; Ortiz-Ramirez et al., 2011a; Christian Sohlenkamp et al., 2000; Yang et al., 2015; Yuan et al., 2013). Ammonium transporter genes were first identified in yeast (A. M. Marini et al., 1994) and plants (Ninnemann et al., 1994a). Homologous genes were identified in all other domains of life: *AMT*s are found in plants, prokaryotes and animals (excluding vertebrates). Rh (Rhesus factor) proteins are found in vertebrates, and MEP (Methylammonia Permease) is found in prokaryotes, fungi and plants. In plants, there are two subfamilies of ammonium transporters (AMT1 and AMT2). AMT1 transporters are electrogenic, they transport a charge across the membrane, while AMT2 transporters are MEPs that were acquired by plants via horizontal gene transfer (McDonald et al., 2012) and are electroneutral.

Marchantia polymorpha is a liverwort and has 9 chromosomes (8+X or Y). Its genome size is 230 Mb and contains around 20,000 protein coding genes. Several reasons make *M. polymorpha* a good model organism. First, it produces a large number of spores. Second, it has convenient asexual reproduction using thallus tissue or gemmae, and an easy-to-control life cycle from vegetative growth to sexual reproduction. Third, the genome sequence recently became available and gene editing protocols such as homologous recombination, CRISPR/Cas9, and selectable markers have been developed for *Marchantia*. Fourth, the evolutionary position of liverworts as basal land plants makes them well suited to investigate land plant evolution (Berger et al., 2016; Bowman, 2016; Bowman et al., 2016).

M. polymorpha has 19 ammonium transporter genes, nine *AMT1* and ten *AMT2* genes (McDonald and Ward, 2016). The number of ammonium transporter genes in angiosperms varies, for example, Arabidopsis has only six ammonium transporter genes (five *AMT1* and one *AMT2*) (Neuhäuser et al., 2015) while *Populus trichocarpa* has 23 (eight *AMT1* and 15 *AMT2*). To date, most of what we know about ammonium transporters in plants comes from the study of AMTs in angiosperms. Since *AMT1* and *AMT2* genes are present in both liverworts and angiosperms, detailed characterization of ammonium transporters from a basal land plant such as Marchantia can provide more information on the structure/function relationships in this family and more generally inform us about the evolution of ammonium transporters in plants (Bowman et al., 2016). Furthermore, studying the regulation of MpAMTs will also provide more information about the evolution of signaling pathways in plants.

4. The regulation pathway of AMTs: Ca^{2+}-CBLs- CIPKs

A wide range of stresses, including biotic stresses such as pathogen infections and abiotic stresses such as salt, drought, cold-shock, touch, and UV light, can affect plant growth and development. Plants have evolved multiple signaling pathways to decode these stresses. These signaling networks involve secondary messengers and a few of them have been identified, such as cyclic nucleotides, hydrogen ions, reactive oxygen species, lipids, sugars, and calcium.

One of the most studied signal transduction messengers is Ca^{2+}. Stresses evoke changes in the cytosolic free Ca^{2+} concentration ($[Ca^{2+}]$), which can be detected by a calcium sensitive luminescent reporter protein from the coelenterate *Aequorea victoria* called aequorin. $[Ca^{2+}]$ stimulates signaling networks that turn stress response genes on or off. Stress sources such as cold-shock, touch, wounding or pathogen attack cause different changes in $[Ca^{2+}]$, and they even induced different kinetics of the $[Ca^{2+}]$ transients (Knight et al., 1991; Neuhaus et al., 1993). An interesting question raised here is how plants distinguish between the different stress sources and then accurately regulate the corresponding stress responsive genes. Several explanations are possible. First, calcium is a necessary factor but not sufficient to trigger the specific signaling pathway for a particular stress. Second, the spatial properties of the calcium signal such as its localization and source encode the specificity of the signaling pathway. Third, the dynamic of changes in $[Ca^{2+}]$ and the presence of Ca^{2+} responsive elements determine which specific signaling pathway is induced.

Following $[Ca^{2+}]$ change, the first molecules involved in a signaling network will be "sensors", which can bind calcium and transduce signals. There are two categories of calcium sensors, distinguished by whether they contain both calcium binding and

catalytic properties. Calcium-dependent protein kinases (CDPKs) are the most well studied calcium sensors. CDPKs contain both Ca^{2+}-binding and catalytic domains (Harmon et al., 2000). Two other well-known calcium sensors, calcineurin B-like proteins (CBLs) and calmodulin (CaM) lack a catalytic domain and only contain calcium binding domains, which are called EF hand motifs. Therefore, to transmit calcium signals, both CBLs and CaM need to interact with other proteins and regulate their activities. The first two CBLs to be identified were SOS3 (also called AtCBL4) in Arabidopsis (Liu and Zhu, 1998) and AtCBL1 (Kudla et al., 1999). The same year, two more CBLs, AtCBL2 and AtCBL3 were identified. In 2009, Li *et al.* identified the conserved N-myristoylation motif (MGCXXSK/T) in CBLs (Li et al., 2009).

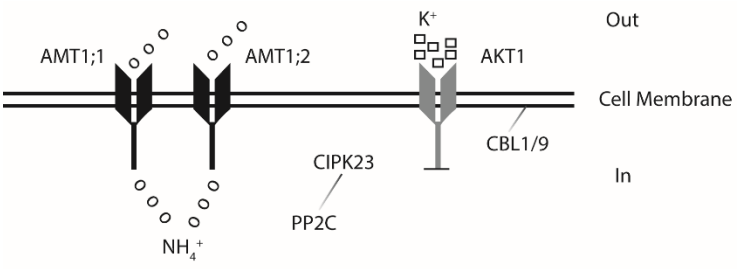
The type of protein that interacts with CBLs was identified by Shi *et al.* and named CBL-interacting protein kinase 1 (CIPK1) (Shi et al., 1999). Three additional CIPKs: CIPK2, CIPK3, and CIPK4 were identified by the same group (Shi et al., 1999). In Arabidopsis, there are 25 CIPKs and they crosstalk with 10 CBLs (Kolukisaoglu, 2004). CIPKs have a NAF domain consisting 24 amino acids, which is sufficient to interact with 10 CBLs and is used to classify members of the CIPK family (Albrecht et al., 2001). Besides, unlike most serine/threonine kinases, AtCIPKs efficiently use Mn^{2+} instead of Mg^{2+} as a cofactor and it has been suggested that CIPKs function as Mn^{2+} binding proteins in plant cells (Shi et al., 1999).

Ca^{2+}-CBLs-CIPK signaling pathways regulate their target proteins by phosphorylation. Some of their targets are ion channels or transporters. In each case, CIPKs interact with specific CBLs to regulate their targets. A number of such specific interactions and their targets have been identified. For example, SOS3 (CBL4) interacts

with SOS2 (CIPK24) to phosphorylate SOS1, a membrane-localized Na^+/H^+ antiporter to improve plant salt tolerance and maintain Na^+ homeostasis (Qiu et al., 2002). SOS2/CIPK24 also regulates the H^+/Ca^{2+} antiporter CAX1 to control intracellular Ca^{2+} homeostasis and to relieve salt stress (Cheng et al., 2004). CIPK23 interacts with CBL1/ 9 to regulate the membrane-localized potassium channel AKT1 in response to low K^+ (Li et al., 2006; Xu et al., 2006). CBL2 interacts with CIPK11 to regulate PM H^+- ATPase and CBL1/ 2/ 9 interact with CIPK11 to phosphorylate the plasma membrane H^+-ATPase in response to high extracellular pH (Li et al. 2009).

Recently, CIPK23, which regulates the NO_3^- transporter NPF6;3 and activates the K^+ channel AKT1, was also found to phosphorylate ammonium transporters (AMT1s) resulting in inhibition of AMT1 activity. Thereby, CIPK23 modulates the growth response to ammonium in Arabidopsis (Straub et al., 2017). As shown in Fig. 1, a complex of CIPK23 interacting with CBL1/ 9 phosphorylates the potassium channel AKT1 and the ammonium transporters AMT1;1 and AMT1;2, thus turning them on or off in response to ammonium stress and in order to maintain potassium and ammonium homeostasis. Protein phosphatase 2C (PP2C) was also shown to interact with the CBL1-CIPK23 complex (Lan et al., 2011).

A



B

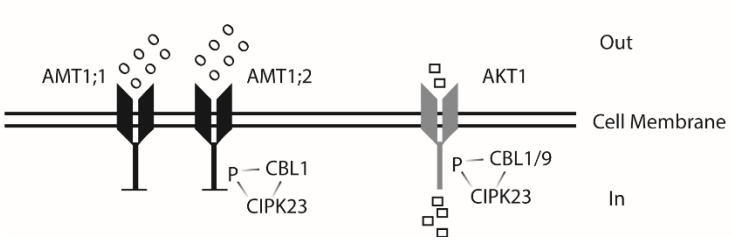


Figure 3. The phosphorylation of three transporters: AMT1;1, AMT1;2 and AKT1. A) When CIPK23 is inactive, it binds to PP2C. CBL1/9 is attached to the plasma membrane. Un-phosphorylated AKT1 is inactive and the potassium concentration is higher on the outside than on the inside of the cell. However, both ammonium transporters AMT1;1 and AMT1;2 are active when dephosphorylated, transporting ammonium from the outside to the inside of the cell. B) Active CIPK23 binds to CBL1 or CBL9 and phosphorylates the potassium transporter AKT1. Consequently, AKT1 is turned on, transporting potassium from the outside to the inside of the cell. However, when AMT1;1 and AMT1;2 are phosphorylated by CIPK23 and CBL1, their activities are inhibited.

5. Research objectives

Specific Aim 1: Apply Retention Projection with Back-calculation (RPwB) to HILIC to improve metabolite identification.

Specific Aim 2: Apply RPwB to identify metabolites in plant matrices to test the robustness of RPwB in real life cases.

Specific Aim 3: Conduct functional characterization of MpAMTs, thereby building a background for structure and function studies using Marchantia ammonium transporters.

Chapter 2

Accurate Prediction of Retention in Hydrophilic Interaction Chromatography (HILIC) by Back Calculation of High Pressure Liquid Chromatography (HPLC) Gradient Profiles

Nu Wang [1,*] and Paul G. Boswell [1]

[1]Department of Plant and Microbial Biology, University of Minnesota, 1479 Gortner Ave., St. Paul, MN 55108, USA

[*]Corresponding author, E-mail: wang4119@umn.edu, phone: 612-625-4763, fax: 612-625-1738

Abstract

Gradient retention times are difficult to project from the underlying retention factor (*k*) vs. solvent composition (***φ***) relationships. A major reason for this difficulty is that gradients produced by HPLC pumps are imperfect – gradient delay, gradient dispersion, and solvent mis-proportioning are all difficult to account for in calculations. However, we recently showed that a gradient "back-calculation" methodology can measure these imperfections and take them into account. In RPLC, when the back-calculation methodology was used, error in projected gradient retention times is as low as could be expected based on repeatability in the *k* vs. ***φ*** relationships. HILIC, however, presents a new challenge: the selectivity of HILIC columns drift strongly over time. Retention is repeatable in short time, but selectivity frequently drifts over the course of weeks. In this study, we set out to understand if the issue of selectivity drift can be avoid by doing our experiments quickly, and if there any *other* factors that make it difficult to predict gradient retention times from isocratic *k* vs. ***φ*** relationships when gradient imperfections are taken into account with the back-calculation methodology. While in past reports, the

accuracy of retention projections was >5%, the back-calculation methodology brought our error down to ~1%. This result was 6-43 times more accurate than projections made using ideal gradients and 3-5 times more accurate than the same retention projections made using offset gradients (i.e., gradients that only took gradient delay into account). Still, the error remained higher in our HILIC projections than in RPLC. Based on the shape of the back-calculated gradients, we suspect the higher error is a result of prominent gradient distortion caused by strong, preferential water uptake from the mobile phase into the stationary phase during the gradient – a factor our model did not properly take into account. It appears that, at least with the stationary phase we used, column distortion is an important factor to take into account in retention projection in HILIC that is not usually important in RPLC.

Key words

HILIC, retention projection, back calculation, gradient delay, prediction error

1. Introduction

Systematic metabolite identification in Mass Spectrometry-based metabolomics has been a long-standing challenge due to the chemical diversity of metabolites and the lack of strict chemical structural rules such as in DNA and proteins (Creek et al., 2014; Fernie et al., 2011; Fridman and Pichersky, 2005). Hydrophilic Interaction Liquid Chromatography (HILIC) combined with MS, has become a mainstay in the analysis of polar metabolites in complex mixtures (Alpert, 1990; Boguslaw Buszewski; Noga, 2012; Brown et al., 2009; Dettmer et al., 2007; Hemström and Irgum, 2006). However, in HILIC-MS, typically only mass spectral information is used to identify metabolites,

while the chromatographic retention information is rarely used because HILIC is notoriously irreproducible. HILIC retention could be affected by many experimental factors including intentional factors such as mobile phase selection, column dimensions, flow rates, and temperature, and unintentional factors such as gradient delay, solvent mis-proportioning, gradient dispersion, column drift, and column age (Boswell et al., 2011a, 2011b; Joe P. Foley et al., 1989; Knox and Kaliszan, 1985). Unfortunately, mass spectral information alone is often insufficient to identify metabolites because of the inability to distinguish among isomeric compounds, or mass signals that are too weak to measure isotope patterns and/or fragmentation patterns, or the lack of reference compounds and the inability to perform de novo structure determination (Annesley, 2003). Chromatographic retention information could be very helpful in metabolite identification if the predictability of retention information could be improved under a range of practical experimental conditions (Boswell et al., 2011b).

Several methodologies have attempted to make retention information predictable across labs and experimental conditions (Boswell et al., 2011a; Gika et al., 2012; Nikitas et al., 2011; Shinoda et al., 2008; Tyteca et al., 2014a, 2014b; Wang et al., 2009). Linear Retention Indexing (LRI) is one of these methodologies and works by calculating an index score using the retention times of bracketing standards that elute before and after the metabolite of interest (Shinoda et al., 2008; Zellner et al., 2008). LRI assumes that the metabolite will always elute at the same relative position between the two bracketing standards, which may only be true if experimental conditions are held constant. The prediction accuracy of LRI decreases significantly if experimental conditions are different from the ones where the LRIs were initially collected, especially in LC since the

relative retention times heavily depend on solvent composition (Boswell et al., 2013; Gritti and Guiochon, 2014). For example, as shown in Fig. 4, ephedrine elutes between mannitol and Phe-Gly-Gly if the percentage of water in the mobile phase is less than 30%. However, when the percentage of water is greater than 30%, not only does ephedrine elute at a different position relative to the standards, it does not even elute between them anymore. Likewise, we would expect very poor prediction accuracy from LRI if we tried to predict ephedrine's retention time at any other solvent composition than the one at which it was measured. The same is true for gradient elution, where the accuracy of LRI deteriorates as soon as the gradient slope, the flow rate, the column dimensions, or even the HPLC instrument is changed due to unintentional differences between the instruments.

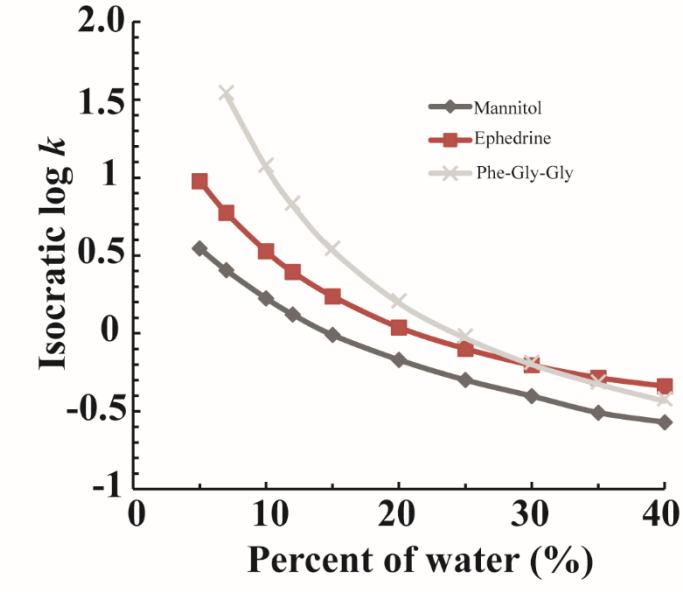


Figure 4. Retention factor (log k) vs. mobile phase water percentage (%) for three compounds: mannitol, ephedrine, and Phe-Gly-Gly. The log k of ephedrine is between mannitol and Phe-Gly-Gly when water percentage is less than 30%, but it elutes outside the two when the water percentage is greater than 30%.

A more accurate approach to predict gradient retention times is to measure the more fundamental relationships of retention factor (k) vs. solvent composition (φ) for each compound from a series of isocratic runs and then calculate, or "project", their gradient retention times (Neue and Kuss, 2010; Nikitas et al., 2011; Wang et al., 2009). To project a gradient retention time from a k vs. φ relationship, the gradient may be considered as a series of short isocratic steps that together approximate the true gradient. This is shown in Eq. (1), where t_R is the gradient retention time, t_c is the time that a solute is under the influence of a particular time slice of the gradient as it moves through the column, k_φ is the retention factor at the solvent composition of the current gradient slice, and n is the smallest integer that makes the inequality true:

$$t_R = \sum_{i=1}^{n} \delta t_c = \delta \sum_{i=1}^{n} 1 + \frac{1}{k_\varphi} \qquad (1)$$

Unfortunately, this approach is also usually inaccurate because the calculation relies on precise knowledge of the gradient, and the gradient actually produced by an LC instrument is always different from the ideal gradient it was programmed to produce (Boswell et al., 2011a, 2011b).

Retention projection with back calculation is a relatively new methodology that accounts for the effective shape of the gradient actually produced by an HPLC instrument. Instead of assuming the gradient is ideal or directly measuring the gradient, which is time consuming and can be quite difficult to measure precisely, this methodology back-calculates the effective gradient that must have been produced by the HPLC instrument to give the retention times measured for a set of back-calculation standards (Abate-pella et al., 2015; Barnes et al., 2013b; Boswell et al., 2013, 2011a,

2011b). In short, it solves for the effective gradient by iteratively making changes to an assumed gradient until the projected retention times of the back-calculation standards match their measured retention times (see Fig. 5). It then uses the back-calculated gradient to project the retention times of other compounds with known *k* vs. φ relationships.

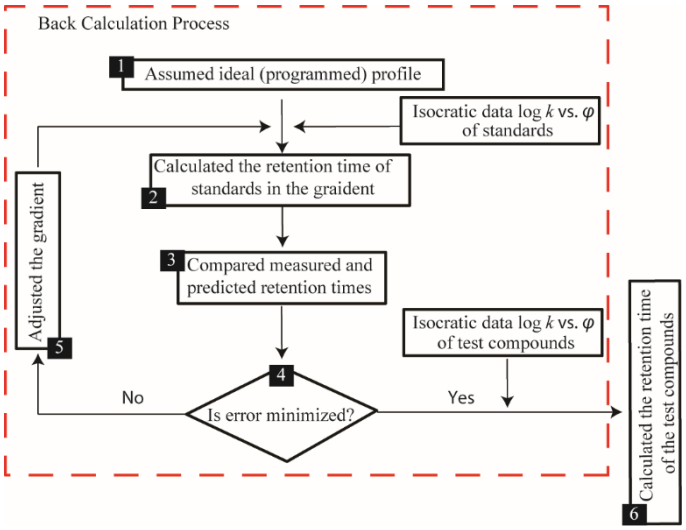


Figure 5. Flow chart of the retention projection methodology with back-calculation. Retention projection starts by assuming the gradient produced the LC instrument is ideal. Then the gradient is adjusted iteratively until the difference between the projected retention times and their experimental retention times are minimized. This back-calculated gradient is then used to project the retention times of other compounds with known k vs. φ relationships.

Retention projection with back-calculation has been applied in reversed phase liquid chromatography (RPLC) to give uniquely low prediction errors, for example, only ± 2.8 s in 20 min solvent gradient at 100 µL/min flow rate (Boswell et al., 2011a, 2011b). Using this approach, the error in our retention projections was virtually the same as our

experimental error, suggesting that the approach takes into account virtually all significant factors controlling retention. In a seven-lab study, retention projections were found to be 22-fold more accurate than LRIs under a range of experimental conditions (Abate-pella et al., 2015). Even under conditions ideal for accurate LRIs, retention projections were twice as accurate because they took into account the gradient non-idealities produced by each lab's LC instrument (Abate-pella et al., 2015). This suggests it is possible to create retention databases useful for metabolite identification.

However, this approach to retention projection using the back-calculation methodology has not been extended to HILIC, yet. In previous studies using gradient retention projection from isocratic runs in HILIC where the back-calculation methodology was not used, the prediction accuracy was much worse than in RPLC (Boswell et al., 2011a; Gritti and Guiochon, 2014; Tyteca et al., 2014a, 2014b). For example, Tyteca *et al.* predicted gradient retention times in HILIC and found the prediction error was 3.7% to 9.5%, which is roughly 30-fold less accurate than retention projections in RPLC using back-calculation (Tyteca et al., 2014b). A later publication by Tyteca *et al.* reported an improvement in the average prediction error to ~ 3% (Tyteca et al., 2014a). They attributed the improved accuracy to their use of a better equation to describe k vs. φ relationships, but it was probably also caused by their use of shallower gradients than in their previous work. Retention projections are generally more accurate in shallower gradients because the gradients more closely resemble the isocratic conditions in which the k vs. φ relationships were measured. Gika *et al.* also projected gradient retention times in HILIC and found prediction errors that were considerably more accurate, up to 0.4% (Gika et al., 2012). However, their high accuracy mainly

resulted from measuring their *k* vs. φ relationships in gradients instead of isocratic runs, and then predicting their gradient retention times in gradients similar to ones used to measure their *k* vs. φ relationships.

In this paper, we extend the back-calculation methodology to HILIC to (a) determine the level of retention projection accuracy achievable once gradient non-idealities are taken into account and (b) to determine if there are any additional factors that must be modeled in order to achieve a similar accuracy in HILIC as in RPLC. Since reproducibility in HILIC is a well-known issue (Bocian et al., 2014; Greco and Letzel, 2013; Wernisch and Pennathur, 2016), we measured both isocratic and gradient retention times within five days.

2. Material and methods

2.1 Selection of 23 compounds. 23 compounds out of 200 that we tested were selected (Fig. 6). They were selected to be chemically diverse, have different masses, retention times, and a range of pK_as. Some are hydrogen-bond donors and some are hydrogen-bond acceptors. 100 mM stock solutions of each compound were prepared in 1:1 water: acetonitrile. Stock solutions were then mixed and diluted in acetonitrile to a final injection concentration of 100 µM. All chemicals and solvents were purchased from Sigma-Aldrich® (St. Louis, MO), Alfa Aesar® (Ward Hill, MA), or TCI America (Portland, OR).

2.2 Mobile phase preparation. Mobile phase A was acetonitrile: water (95: 5 by volume) with 20 mM ammonium acetate, which contains 10 mM ammonium acetate and 10 mM acetic acid which was used to adjusted the pH to 4.7. Mobile phase B was water

with 20 mM ammonium acetate (same as in mobile phase A). To ensure a consistent pH in our mobile phase buffer from batch to batch, we measured components (ammonium acetate, acetic acid, acetonitrile, and water) gravimetrically with ±0.001 g accuracy. We made a stock buffer solution first: 1 L of 2 M stock buffer was made by adding 77.080 g ammonium acetate and 60.800 g acetic acid to 1000.000 g distilled deionized water. Then 1 L of mobile phase A was made (95: 5 acetonitrile: water) by adding 10 mL stock buffer (2 M) to 746.909 g acetonitrile and 40.000 g distilled deionized water, while 1 L of mobile phase B was made by adding 10 mL stock buffer (2 M) to 990.000 g distilled deionized water. The final concentration of buffer in mobile phase A and B was 20 mM.

2.3 Equipment and software. A HPLC-MS system that has a DIONEX Ultimate 3000 (HPG-3400 RS) pump and a Bruker Amazon SL ion trap MS (Bremen, Germany) was used. Back-calculation software was written in-house in Java 1.6 (Oracle, Redwood Shores, CA), which contains the Java OpenGL (JOGL) binding library version 2.0-rc11 (JogAmp, http://jogamp.org), the Unidata netCDF library version 4.2 (Unidata®, Boulder, CO), the Savitzky-Golay filter library version 1.2 by Marcin Rzeźnicki (http://code.google.com/p/savitzky-golay-filter/), the jmzML library, and the jmzReader library.

2.4 Chromatographic conditions. A Waters Acquity UPLC BEH HILIC column (2.1 mm × 100 mm, 1.7 µm particle size) was used in this study. To avoid the effect of temperature fluctuation on retention, the column temperature was held at 35.0 °C by a thermostatted circulating water bath (Exacal Circulating Bath EX-200DD) and an eluent pre-heater (Thermo Scientific Pre-Column heater 2 µL) in front of column. The injection volume was 5 µL in all runs.

2.5 Measurement of column repeatability. The repeatability of the column was measured by running a mixture of the 23 compounds in an isocratic run at 5% mobile phase B for 60 minutes at a 0.2 mL/min flow rate. More isocratic runs were performed 12, 24, 36, 48, 60, and 72 h later. RSD (relative standard deviation) was then calculated for each compound's retention time in these 6 isocratic runs. We found the RSDs ranged from 0.06% to 0.3% in retention time and 1.6% in *k*, which is similar to RPLC where 1% RSD in *k* is typical (Boswell et al., 2011a).

2.6 Measurement of isocratic data. After measuring repeatability, we measured the isocratic *k* vs. φ relationships of the 23 compounds at 10 isocratic solvent compositions (5%, 7%, 10%, 12%, 15%, 20%, 25%, 30%, 35% and 40% of mobile phase B) for 60 minutes at 0.2 mL/min flow rate over three replicates. To ensure system and column equilibration, we ran 12 mL (30 times the column dead volume) of eluent through the column before each isocratic run. According to the Eq. (3), where t_R is the retention time and t_0 is the column dead time:

$$\log k = \log \frac{t_R - t_0}{t_o} \qquad (3)$$

To avoid the influences of the instrument dead time (t_I) on the prediction accuracy, the equation is modified to:

$$\log k = \log \frac{(t_R - t_I) - (t_0 - t_I)}{t_o - t_I} \qquad (4)$$

The log *k* of each compound at each solvent composition was calculated by Eq. (5), where:

$$\log k = \log \frac{t_R - t_I}{t_o - t_I} \qquad (5)$$

Both t_I and t_0 were measured using our dead time marker, *n*-hexadecylbenzamide. To measure t_I we replaced the column with a narrow bore PEEK tube (1 m length with 50 μM inner diameter).

2.7 Measurement of gradient delay volume. The gradient delay volume is the solvent volume contained from the solvent proportioner to the column entrance. The gradient delay volume was measured using a 5 min gradient run at 0.2 mL/min. Solvent A was water, solvent B was water mixed with 0.1% acetone, and acetone was monitored by a UV absorbance detector at 265 nm. The gradient delay volume was measured to be 0.56 mL for our instrument.

2.8 Measurement of gradient retention times. We measured retention times in nine different gradients with three gradient slopes (3.5% mobile phase B/min, 1.75% mobile phase B/min, and 0.875% mobile phase B/min) starting from 5% mobile phase B, at three flow rates (0.2 mL/min, 0.4 mL/min, and 0.8 mL/min). The column was equilibrated before each gradient run using 12 mL of starting solvent, as when isocratic data was measured. Each gradient was held at 40% mobile phase B at the end to clear the column. In this way, both isocratic data and gradient data were measured using the same LC-MS instrument within 3 days without pause.

3. Results and discussion

The back-calculation methodology was applied to (a) study the retention projection accuracy when accounting for gradient non-idealities and (b) determine any additional factors that need to be modeled besides gradient imperfections to achieve a similar accuracy in HILIC as in RPLC. Of the 23 chemically diverse compounds we selected, 10

were used as “back-calculation standards” to calculate the actual gradients produced by our instrument and the other 13 were used as “test compounds” to probe the accuracy of our retention projections. Their isocratic k vs. φ relationships and subsequently retention times in nine gradients were measured. We then attempted to project the gradient retention times of the 13 test compounds based on their k vs. φ relationships and obtained gradient profiles by three different approaches: (1) we assumed the gradient profiles were ideal in our retention projection calculations (“ideal gradient”), (2) we assumed the gradient profiles were ideal, but simply delayed to account for the gradient delay volume we measured independently for our HPLC system (“offset gradient”), and (3) we back-calculated the effective gradient profiles from the retention times of the 10 back-calculation standards (“back-calculated gradient”). Last, we tested the prediction accuracy in each case by comparing the projected retention times of the 13 test compounds to their measured retention times.

3.1 Selection of the back-calculation standards

Out of the 23 compounds, we selected back-calculation standards based on their retention times to be evenly spaced and together, to cover a wide range of the gradient. The retention time of each standard offers information about a short section of the gradient, so it is important that the standards elute over a wide range of retention times so that there is enough information to back-calculate the entire gradient. For example, in Table 1 (retention times of the 23 compounds in a 10 min gradient), 10 compounds were used as back-calculation standards (orcinol, uracil, salicin, vanillic acid, riboflavin, mannitol, ephedrine, nicotinic acid, Phe-Gly-Gly, and γ- amino butyric acid) and 13 compounds were used as test compounds (indazole, thymidine, puromycin, patulin,

barbital, phe-phe, imidazole, nicotinamide, ribitol, caffeine, naphthalene acetic acid, 3-morpholino-2-hydroxypropanesulfonic acid (MOPSO), and *L*-proline). With these selections, the retention times of the back-calculation standards were evenly spaced and covered the entire 10 min gradient, while the retention times of the test compounds were scattered in between the back-calculation standards.

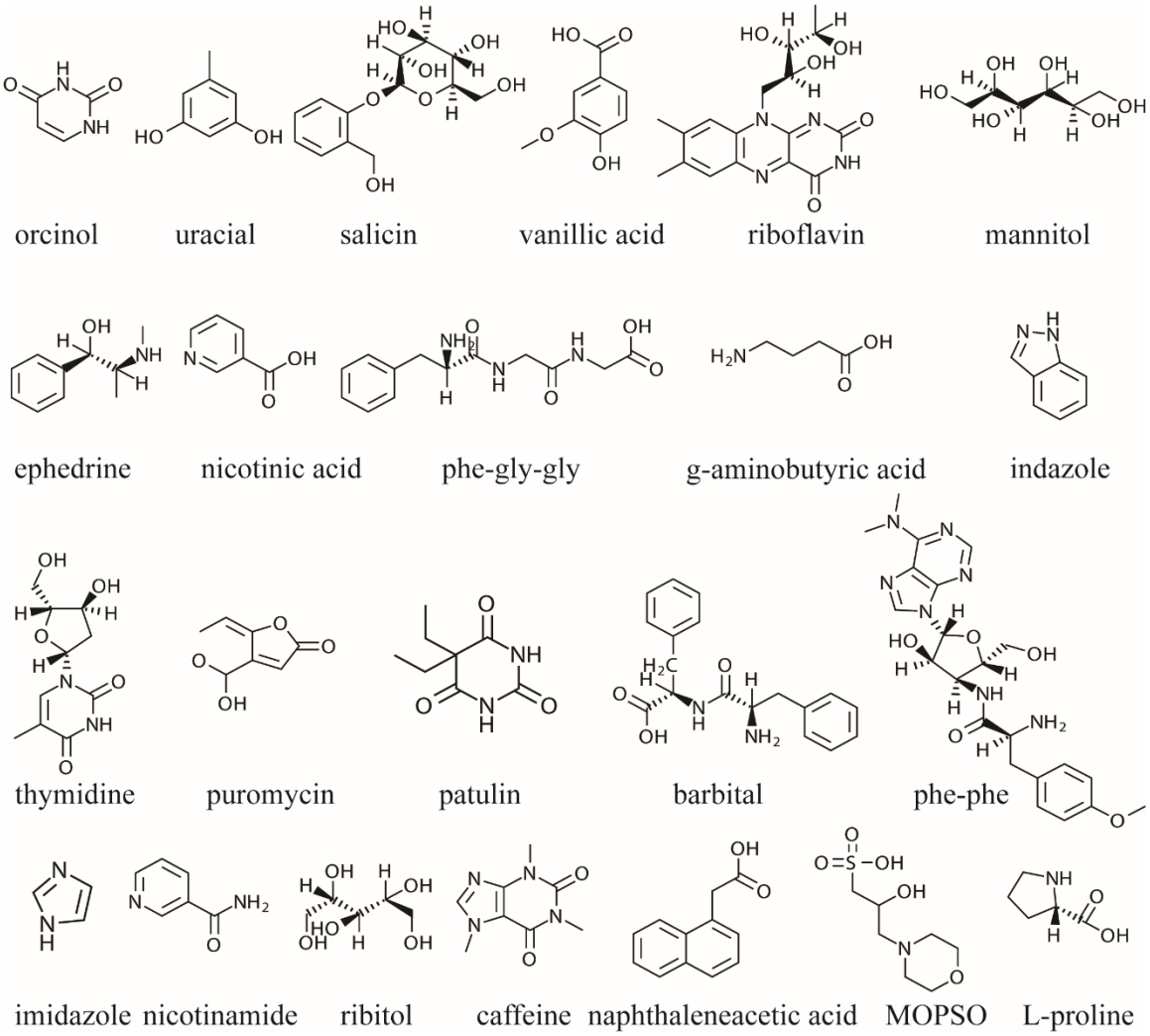


Figure 6. Chemical structures of the 23 compounds that were used in the study. These chemicals were selected because they are chemically diverse, and elute over a wide range of retention, and can be distinguished easily by mass.

Table 1. An Example of 10 min Gradient Run

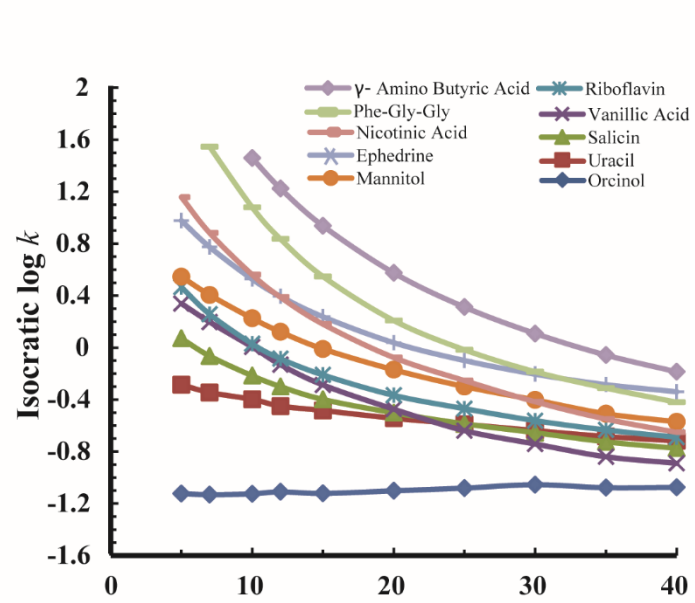
	Name	Retention time (min)	m/z
1	Orcinol	1.6438	123
2	Barbital	1.6726	183
3	Patulin	1.6835	193
4	Indazole	1.688	119
5	Caffeine	1.9957	195
6	Uracil	2.1969	111
7	Thymidine	2.3877	241
8	Nicotinamide	2.416	123
9	Imidazole	2.8904	69
10	Salicin	3.0352	285
11	Ribitol	3.8874	151
12	MOPSO	3.9721	226
13	Naphthalene Acetic Acid	4.2291	185
14	Vanillic Acid	4.2749	167
15	Puromycin	4.7923	472
16	Riboflavin	5.2287	377
17	Mannitol	5.8291	182
18	Phe-Phe	6.7509	313
19	Ephedrine	7.0336	166
20	Nicotinic Acid	7.4984	123
21	Phe-Gly-Gly	8.4105	278
22	L-proline	8.9657	115
23	γ- amino butyric acid	9.6717	103

Retention times of 23 compounds, including the 10 back-calculation standards and the 13 test compounds. Back-calculation standards are highlighted in blue. Mobile phase A: acetonitrile/water (95/5) with 20 mM ammonium acetate (pH 4.7). Mobile phase B: water with 20 mM ammonium acetate (pH 4.7). The 10 min gradient was from 5% to 40% mobile phase B, the flow rate was 0.2 mL/min, and the temperature was 35 ℃. Column: Water Acquity UPLC BEH HILIC column (2.1 mm × 100 mm) with 1.7 μm particle size.

3.2 Isocratic data (log k vs. φ) of 23 compounds

Fig. 7 shows the log *k* vs. *φ* relationships of the 10 back-calculation standards and the 13 test compounds. Retention factors ranged from –1.2 to 2 within the mobile phase B

percentage range of 5% to 40%. As is characteristic of HILIC, the retention factor of most compounds becomes smaller with the increasing percentage of water. However, of these 23 compounds, the log *k* of three test compounds (indazole, patulin, and barbital) did not change appreciably with solvent composition (their log *k* was around -1.0) because they were very poorly retained. Despite their poor retention, their retention factors were still different and repeatable and, as is often the case for compounds like these, their retention times are different enough to be identifying characteristics. Therefore, these three compounds were kept in our analysis.



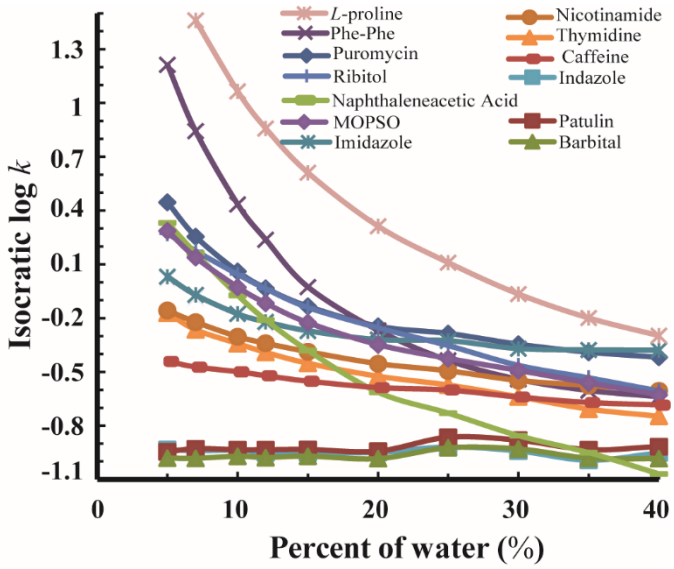


Figure 7. Isocratic data (log *k* vs. φ) measured for the 10 back-calculation standards (top) and the 13 test compounds (bottom). See section 2.6 for details.

3.3 The three types of gradient profiles: ideal, offset, and back-calculated

The accuracy of the gradient profile is key to the accuracy of projected retention times since it is used in their calculation. In this paper, we used three types of gradient profiles to project retention times: “ideal”, “offset”, and “back-calculated”. Fig. 8 shows an example of the three types of gradient profiles representing a gradient with a slope of 3.5% mobile phase B/min (starting from 5% mobile phase B) at a flow rate of 0.2 mL/min. The black dotted line in Fig. 8 is the *ideal* gradient that was programmed into the LC instrument. However, in reality, the gradient that was delivered to the inlet of the column was imperfect since it took time for the mobile phase to get from the solvent proportioner to the inlet of the column. This gradient delay is taken into account in the *offset* gradient (green line), which is the ideal gradient offset by the gradient delay (see section 2.6 for more details about measurements of the gradient delay). Though the offset

gradient accounts for the biggest source of non-idealities in the gradient, it does not account for other sources of gradient non-ideality including solvent mis-proportioning and gradient dispersion. The red line is the *back-calculated* gradient, which could potentially account for all these factors: gradient delay, solvent mis-proportioning, and gradient dispersion. This is because the back-calculated gradients take whatever shape is necessary to minimize the retention time differences between measured and calculated retention times of the 10 back-calculation standards, thus characterizing the actual gradient experienced by the test compounds.

Looking closer at these three gradient profiles in Fig. 8, the ideal gradient is the most distinct. It showed the mobile phase B fraction begin increasing right at time 0 at a rate of 3.5% mobile phase B/min, while the offset and back-calculated gradient profiles displayed a ~2.5 min delay before the onset of the gradient, which reflects the gradient delay of the LC system we used. Then, after the delay time, the two gradient profiles started increasing. Comparing just the offset and back-calculated gradients, there are a few more differences. First, the back-calculated gradient appears to be about 0.3 min more delayed than the offset gradient. Yet even though the back-calculated gradient is more delayed than the offset gradient, it reaches a steeper gradient slope and eventually catches up with the offset gradient (discussed below). Second, there appears to be a slight decline in mobile phase B fraction during the gradient delay of the back-calculated gradient. We believed this decline is just an artifact of the back-calculation process – there is not enough information from the back-calculation standards (i.e. retention times) to calculate the true mobile phase compositions in that region of the gradient profile.

In the back-calculated profile, the longer gradient delay combined with a steeper gradient slope was unexpected. One possible explanation of this behavior could be distortion of the gradient within the column ("column distortion"), caused by partitioning of water from the mobile phase into the stationary phase (Bocian et al., 2014; Greco and Letzel, 2013; Wernisch and Pennathur, 2016). At the beginning of the gradient, the amount of water in the stationary phase is relatively small, and the water concentration in the stationary phase is at equilibrium with the water concentration in the mobile phase. As the amount of water in the mobile phase increases at the beginning of the gradient, it preferentially partitions into the stationary phase and lower water concentration in the mobile phase. This would appear as an extended gradient delay (2.5 to 2.8 min). As the water fraction in the mobile phase continues to increase, eventually the stationary phase becomes saturated and the mobile phase composition starts to recover (2.8 to 3.1 min). We then see a much steeper slope as the effect of the stationary phase on the mobile phase composition becomes insignificant (i.e., it catches up). After it catches up at ~5.2 min, we expected the slope in the back-calculated gradient would match the delayed gradient. However, from 5.2 min to 6 min, the water fraction appeared to increase even higher than the offset gradient. That would not happen in reality, rather, it is an artifact of the back-calculation process – there is only one retention time in that range and the algorithm does not have enough information to converge on the true mobile phase composition.

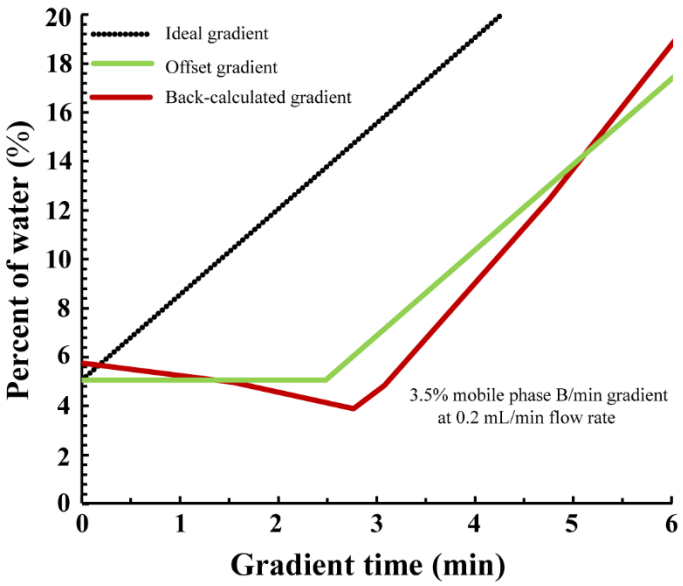


Figure 8. Example of the three types of gradient profiles: ideal, offset, and back-calculated gradient profiles in a gradient slope of 3.5% mobile phase B/min (started from 5%) at a flow rate of 0.2 mL/min. Mobile phase B: water with 20 mM ammonium acetate (pH 4.7). See section 2.8 for details.

3.4 The effects of gradient shape and flow rate on the back-calculated gradient profiles

Gradient steepness and flow rate played an important role in the degree of gradient distortion in the back-calculated gradients. To analyze this influence, we tested three gradient slopes (3.5% mobile phase B/min, 1.75% mobile phase B/min, and 0.875% mobile phase B/min) starting from 5% of mobile phase B and three flow rates (0.2 mL/min, 0.4 mL/min, and 0.8 mL/min). In total, we back-calculated nine gradients (as shown in Fig. 9).

To quantify the distortion in the back-calculated gradients, a metric was defined as "distortion delay", which represents the maximum distance in time between the offset

gradient and the back-calculated gradient. For example, gradient distortion causes the back-calculated gradient to lag behind the offset gradient (Fig. 8), but the delay between the two is greatest at the point where the offset gradient begins to rise. At that point, the distortion delay can be seen to be ~0.7 min.

We found that a slower flow rate caused the back-calculated gradient to have a greater distortion delay and a stronger influence on the shape of the back-calculated gradient (Fig. 9). For example, as showed in Fig. 13, at the same gradient slope of 3.5% mobile phase B/min, a slower flow rate of 0.2 mL/min gave a 0.44 min distortion delay while a faster flow rate of 0.4 mL/min caused a smaller distortion delay of 0.21 min, and an even faster flow rate of 0.8 mL/min caused the smallest distortion delay of 0.04 min. This observation suggests that the distortion delay is caused by column distortion, which is mitigated at faster flow rates since a greater volume of mobile phase passes through the column in a given amount of time, thereby the percentage of up taken water on the carried mobile phase is smaller, thereby lessening the effect of water uptake by the stationary phase on the mobile. We saw a similar effect when comparing gradient slopes. For example, at the flow rate 0.2 mL/min, we compared the back-calculated gradients in three gradients with slopes of 3.5% mobile phase B/min, 1.75% mobile phase B/min, and 0.875% mobile phase B/min. We found that the steepest gradient (3.5% mobile phase B/min) showed the greatest distortion delay and had the strongest influence on the shape of the back-calculated gradient.

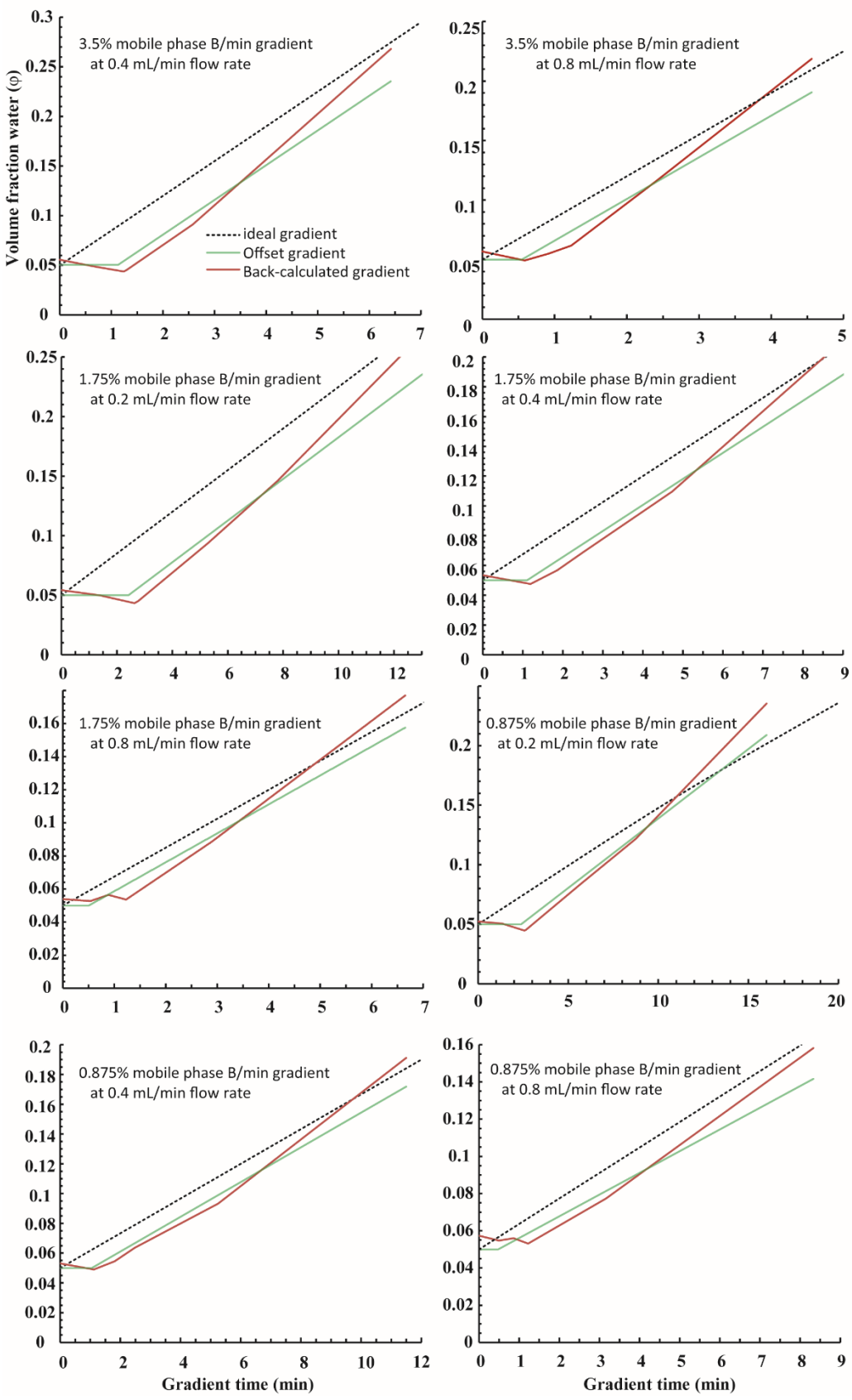

3.5% mobile phase B/min gradient
at 0.4 mL/min flow rate
ideal gradient
Offset gradient
Back-calculated gradient
Volume fraction water (φ)
3.5% mobile phase B/min gradient
at 0.8 mL/min flow rate
1.75% mobile phase B/min gradient
at 0.2 mL/min flow rate
1.75% mobile phase B/min gradient
at 0.4 mL/min flow rate
1.75% mobile phase B/min gradient
at 0.8 mL/min flow rate
0.875% mobile phase B/min gradient
at 0.2 mL/min flow rate
0.875% mobile phase B/min gradient
at 0.4 mL/min flow rate
0.875% mobile phase B/min gradient
at 0.8 mL/min flow rate
Gradient time (min)
Gradient time (min)

Figure 9. Three types of gradient profiles: ideal, offset, and back-calculated gradient profiles in three gradient slopes (3.5% solvent B/min, 1.75% solvent B/min, and 0.875% solvent B/min) starts from 5% solvent B at three flow rates (0.2 mL/min, 0.4 mL/min, and 0.8 mL/min). Solvent B: water with 20 mM ammonium acetate (pH 4.7). See section 2.8 for details.

3.5 Comparison of retention projection error using the three types of gradient profiles

Last, we tested the retention projection accuracy using the three types of gradient profiles (ideal, offset, and back-calculated), the three gradient shapes (3.5% mobile phase B/min, 1.75% mobile phase B/min, and 0.875% mobile phase B/min), and the three flow rates (0.2 mL/min, 0.4 mL/min, and 0.8 mL/min). We describe our overall prediction error in terms of the RMS (Root Mean Square) error of the 13 test compounds. As an example, Table 2 shows the individual projection errors of the 13 test compounds and the overall RMS error in one of the gradients (3.5% mobile phase B/min gradient, 0.2 mL/min flow rate) using each of the three types of gradient profiles. In this case, the overall prediction error was greatest, at ±71.5 s (28.97%), when using the ideal gradient, while the overall prediction error was considerably less, at ±7.2 s (2.25%), when using the offset gradient. The back-calculated gradient gave the smallest overall prediction error, at ±2.5 s (1.04%).

Table 3, 4, and 5 show the overall prediction errors when using the ideal, offset, and back-calculated gradient profiles in nine gradients with three gradient slopes and three flow rates. In Table 3, which shows retention projection errors using the ideal gradient profiles, the overall prediction error ranged from ±12 s to ±71 s (6.36% to 28.97%) in the

nine gradients. In Table 4, which shows retention projection errors using the offset gradient profiles, the overall prediction error was 3-10 times more accurate, ranging from ±3.0 s to ±7.2 s (1.54% to 2.73%). This prediction accuracy is similar, if not more accurate than a report that, until now, showed the most accurate retention projections in HILIC. They also used an offset gradient profile (Tyteca et al., 2014a). In Table 5, which shows retention projection errors using the back-calculated gradient profiles, the prediction error was the smallest of all, ranging from ±0.62 s to ±2.5 s (0.53% to 1.83%), which was 6-43 times more accurate than when using the ideal gradient profiles and roughly threefold more accurate than when using the offset gradient profiles. This improved accuracy was a result of the fact that the back-calculation methodology accounts not just for gradient delay, but also for gradient distortion, solvent mis-proportioning, and gradient dispersion. In addition, based on the uniquely strong distortion in the back-calculated profiles obtained here (compared to those obtained in prior publications dealing with RPLC), we suspect that column distortion is unusually pronounced in this HILIC system. The back-calculation methodology may also be *accommodating* this column distortion to some degree, even though it could not properly take it into account.

Table 2. Retention Projection Error When Using the Three Types of Gradients

Test compounds	$t_{R,\ experiment}$ [a]	$t_{ideal} - t_R$ [b]	$t_{delayed} - t_R$ [c]	$t_{back\text{-}calculated} - t_R$ [d]
indazole	101.952	0.708	0.558	-0.108
thymidine	144.072	8.19	0.768	2.184
puromycin	307.344	106.458	19.656	0.42
patulin	101.496	0.516	0.426	-0.462
barbital	100.776	0.45	0.33	-0.396
phe-phe	419.502	157.488	14.448	0.582
imidazole	174.6	18.138	1.152	2.976
nicotinamide	145.938	7.062	0.954	1.524
ribitol	236.292	44.916	3.024	1.272
caffeine	120.084	1.53	0.312	-0.132
naphthalene acetic acid	254.1	70.704	0.324	-5.532
MOPSO	241.272	57.216	2.91	1.248
L-proline	530.232	140.028	-7.17	5.598
Overall prediction error [e] (s)		±71.51	±7.17	±2.51
Overall prediction error [f] (%)		±28.98	±2.26	±1.04

3.5% mobile phase B/min (started at 5%) at flow rate 0.2 mL/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

[a] the experimental retention time (s) in the gradient

[b] the retention time differences (s) between the experimental retention time and the projected retention time, calculated using the ***ideal*** gradient profile (dotted line in Fig. 5), $t_{ideal\ gradient} - t_{R,\ Experiment}$

[c] the retention time differences (s) between the experimental retention time and the projected retention time, calculating using the ***delayed*** gradient profile (green line in Fig. 5), $t_{ielayed\ gradient} - t_{R,\ Experiment}$

[d] the retention time differences (s) between the experimental retention time and the projected retention time, calculating using the ***back-calculated*** gradient profile (red line in Fig. 5), $t_{back\text{-}calculated\ gradient} - t_{R,\ Experiment}$

[e] Overall prediction error is defined as the RMS (root-mean-square) error of the test compounds. RMS $\left(\sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 \ldots + x_n^2)}\right)$, x means the prediction error of each test compound

[f] Prediction error (%) $= \left(\sqrt{\frac{1}{n}(y_1^2 + y_2^2 + y_3^2 \ldots + y_n^2)}\right)$, y means the prediction error (%) of each test compound

Table 3. The Overall Prediction Error (s) of 13 Test Compounds When Using Ideal Gradients

Overall Prediction Error (s)			
Flow Rate / Gradient Slope	0.2 mL/min	0.4 mL/min	0.8 mL/min
3.5%/min [g]	± 71 [e] (28.97% [f])	± 30 (18.37%)	± 14 (12.57%)
1.75%/min [h]	± 64 (19.27%)	± 28 (12.27%)	± 13 (8.17%)
0.875%/min [i]	± 62 (15.22%)	± 26 (9.05%)	± 12 (6.36%)

[e] Overall prediction error is defined as the RMS (root mean square) error of the test compounds. RMS ($\sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 \ldots + x_n^2)}$), x means the prediction error of each test compound

[f] Prediction error (%) = ($\sqrt{\frac{1}{n}(y_1^2 + y_2^2 + y_3^2 \ldots + y_n^2)}$), y means the prediction error (%) of each test compound

[g] 3.5%/min is the gradient slope of 3.5% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

[h] 1.75%/min is the gradient slope of 1.75% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

[i] 0.875%/min is the gradient slope of 0.875% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

Table 4. The Overall Prediction Error (s) of 13 Test Compounds When Using Offset Gradients

Overall Prediction Error (s)			
Flow Rate / Gradient Slope	0.2 mL/min	0.4 mL/min	0.8 mL/min
3.5%/min [g]	± 7.2[e] (2.25% [f])	± 3.3 (1.63%)	± 3.0 (2.73%)
1.75%/min [h]	± 6.3 (1.63%)	± 4.2 (1.65%)	± 3.9 (2.67%)
0.875%/min [i]	± 7.0 (1.56%)	± 4.4 (1.54%)	± 4.4 (2.62%)

[e] Overall prediction error is defined as the RMS (root mean square) error of the test compounds. RMS ($\sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 \ldots + x_n^2)}$), x means the prediction error of each test compound

[f] Prediction error (%) = ($\sqrt{\frac{1}{n}(y_1^2 + y_2^2 + y_3^2 \ldots + y_n^2)}$), y means the prediction error (%) of each test compound

[g] 3.5%/min is the gradient slope of 3.5% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

[h] 1.75%/min is the gradient slope of 1.75% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

[i] 0.875%/min is the gradient slope of 0.875% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

Table 5. The Overall Prediction Error (s) of 13 Test Compounds When Using Back-Calculated Gradients

Overall Prediction Error (s)			
Flow Rate / Gradient Slope	0.2 mL/min	0.4 mL/min	0.8 mL/min
3.5%/min [g]	± 2.5[e] (1.04% [f])	± 1.01 (0.74%)	± 0.62 (1.15%)
1.75%/min [h]	± 2.34 (0.77%)	± 0.64 (0.53%)	± 1.2 (1.52%)
0.875%/min [i]	± 1.89 (0.68%)	± 1.38 (0.59%)	± 2.7 (1.83%)

[e] Overall prediction error is defined as the RMS (root mean square) error of the test compounds. RMS ($\sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 \ldots + x_n^2)}$), x means the prediction error of each test compound

[f] Prediction error (%) = ($\sqrt{\frac{1}{n}(y_1^2 + y_2^2 + y_3^2 \ldots + y_n^2)}$), y means the prediction error (%) of each test compound

[g] 3.5%/min is the gradient slope of 3.5% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

[h] 1.75%/min is the gradient slope of 1.75% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

[i] 0.875%/min is the gradient slope of 0.875% mobile phase B/min, mobile phase B: water with 20 mM ammonium acetate (pH 4.7)

4. Conclusions

In this paper, we applied the back-calculation methodology in HILIC to (a) determine the retention projection accuracy when gradient non-idealities were taken into account and (b) determine what additional factors must be considered in HILIC when projecting gradient retention times from *k* vs. φ relationships. Known sources of error were either avoided or accounted: column drift, which is especially pronounced in HILIC, was avoided by collecting all retention data (both isocratic and gradient) within a week. Gradient non-idealities (gradient delay, gradient dispersion, and solvent mis-proportioning) were taken into account by back-calculating the effective gradient using the gradient retention times of a set of standards. This is the first time that the retention projection with back-calculation has been applied to HILIC, and it enabled uniquely accurate projections of gradient retention times for compounds with known *k* vs. φ

relationships. Prediction errors were ~1% (from 0.62 s to 2.5 s), which was 6-43 times more accurate than when gradients were assumed to be ideal and 3-5 times more accurate than when only gradient delay was taken into account. To the best of our knowledge, this is the most accurate gradient retention prediction in HILIC and approximately 5-fold more accurate than previous reports (Gika et al., 2012; Tyteca et al., 2014a, 2014b).

However, the prediction error in HILIC (~1%) was still significantly worse than previously reported in RPLC (0.3%) using the same back-calculation methodology (Alpert, 1990; Boswell et al., 2011a, 2011b). Our observations suggest that this is a result of strong water uptake into the stationary phase from the mobile phase during the gradient (i.e., "column distortion") (Bocian et al., 2014; Greco and Letzel, 2013; Joe P. Foley et al., 1989; Knox and Kaliszan, 1985). Unlike RPLC, the gradients we back-calculated showed strong gradient delay beyond that caused by the instrument itself, followed by a steep recovery, which is characteristic of column distortion (Bocian et al., 2014; Joe P. Foley et al., 1989; Knox and Kaliszan, 1985). Therefore, at least with the HILIC stationary phase we used (Waters BEH HILIC, 1.8 µm particle size), it appears that column distortion is a significant factor to consider when projecting retention times that is less important in RPLC.

Acknowledgements

We thank the National Institutes of General Medical Sciences [R01GM098290], the University of Minnesota Office of the Vice President for Research, and the Minnesota Agricultural Experiment Station for funding. We appreciate the help of instrument trouble-shooting from Joshua T. Hewitt (University of Minnesota) and manuscript proofreading by Michael Wilson (University of Minnesota).

Chapter 3

How do sample solvents and matrices affect the accuracy of retention projection with back-calculation (RPwB) in high-performance liquid chromatography (HPLC)?

Nu Wang[1*], and Paul G. Boswell[1]

[1]Department of Plant and Microbial Biology, University of Minnesota, 1479 Gortner Ave., St. Paul, MN 55108, USA

[*]Corresponding author, E-mail: wang4119@umn.edu, phone: 612-625-4763, fax: 612-625-1738

Abstract

Abstract

Retention projection with back-calculation (RPwB) is a novel methodology that can improve metabolite identification by predicting their retention times in chromatography-mass spectrometry. However, this methodology has only proved its robustness in preferred sample solvents 73% acetonitrile and 27% water in RPLC instead of more real-life cases. Biological samples are often complicated and sample solvents vary. In this study, we set out to explore the sensitivity of RPwB under a limited set of extreme conditions. First, we selected nine sample solvents representing a wide range of polarities (methanol, ethanol, isopropanol, acetonitrile, acetone, dichloromethane, ethyl acetate, tetrahydrofuran, and toluene), and studied their effects on peak shapes, retention times, and prediction accuracy of RPwB. We found isopropanol, acetonitrile, and ethyl acetate were the top three that distorted peak shapes the most, and all nine sample solvents tested shifted retention times and poorly retained compounds suffered more than well retained

ones. Only ethyl acetate failed to meet the prediction accuracy required of RPwB, as it had a prediction error at 4.3 sec, which was more than the 3 sec that is the maximum deviation allowed for the successful application RPwB in RPLC. Second, five plant samples (*Solanum lycopersicum* fruit, *Solanum lycopersicum* stems and leafs, *Nicotiana* flowers, *Nicotiana* leafs, and *Nicotiana forsteri* leafs) were randomly chosen and extracted using the four sample solvents (water, 70% ethanol, dichloromethane, and isopropanol) that are the most common ones, that did not change the prediction accuracy of RPwB, and that allowed the most concentrated plant extracts to be obtained. The effect of plant matrices on retention times of nearly all analytes was negligible, despite their high concentrations. However, we observed a buildup of some plant matrix solutes in the column that lowered retention projection accuracy for two charged analytes, tetrabutylammonium and tetrapentylammonium. However, this buildup could be removed (and the accuracy of retention projections restored) by either flushing the column with a stronger mobile phase or decreasing the concentration of the injected sample.

Key words

Plant matrix, RPwB, prediction accuracy, retention time shift, peak distortion

1. Introduction

Our previous publications have shown the high retention prediction accuracy of the novel methodology “retention projection with back-calculation” (RPwB) in gas chromatography (GC) (Peng et al., 2014), reversed-phase liquid chromatography (RPLC) (Boswell et al., 2011a), and hydrophilic liquid chromatography (HILIC) (Wang and Boswell, 2017). Cross lab studies of RPwB in GC and RPLC indicated the possibility of

building a sharable retention database across labs, thus RPwB can improve metabolite identification through using both retention and mass information (Abate-pella et al., 2015). However, the high prediction accuracies of RPwB in GC, RPLC, and HILIC were obtained by injecting well characterized instrument calibration standards, testing compounds as a simple mixture, accurately controlling injection concentration and volume, and using a well-known "injection background" – only water and acetonitrile were used to dissolve these standards and testing compounds (Abate-pella et al., 2015; Boswell et al., 2013, 2011a, 2011b; Paul G. Boswell, Dwight R. Stoll, Peter W. Carr, Megan L. Nageld, Mark F. Vitha and Mabbott, 2013; Wang and Boswell, 2017).

On the one hand, sample solvents often interfere with analytes and cause peak distortions such as fronting, tailing, splitting, and broadening (Matuszewski, 2006; Matuszewski et al., 2003; Matuszewski and Constanzer, 1998; Mei et al., 2003), particularly in HILIC (Trufelli et al., 2011). In RPLC, sample solvents are generally matched to the mobile phase or at least kept less polar than the mobile phase (Castells and Castells, 1998; Jandera and Guiochon, 1991; Keunchkarian et al., 2006; Khachik et al., 1988; Layne et al., 2001; Mishra et al., 2013; Shalliker and Guiochon, 2010, 2009). Solvent polarity (P') is not quite precise, but still useful for the categorization of chromatography solvents (Bakalyar et al., 1977; Glajch et al., 1982, 1980; Snyder, 1974). One can expect peak distortions if P' for the sample solvent and mobile phase are markedly different. This requirement of solvent selection could limit the chemical diversity of extracts. For example, different sample solvents extract different compounds even from the same biological matrix, which makes it impossible to analyze all metabolites if the matrix is extracted by a limited number of sample solvents. Therefore,

it is meaningful to study the prediction accuracy of RPwB in RPLC under changing injection solvents, because the choice of injection solvents can represent a trade-off between extracted chemical diversity and analysis quality.

On the other hand, the matrix itself is complex and matrix-dependent analytical problems are also common in HPLC-MS (Jandera and Guiochon, 1991; Keunchkarian et al., 2006; Khachik et al., 1988; Zapata and Garrido, 1991). Behaviors of analytes in sample solvents are different when they are together with matrix and sample solvents (Trufelli et al., 2011). For instance, other compounds in the matrix can shift analyte retention, suppress or enhance signals, change the selectivity and repeatability of the analytical system (Pita et al., 2014; Poole, 1997). An analyte's peak shape and retention heavily depend on interactions among the analyte, matrix, mobile phase, and column (Trufelli et al., 2011). Meanwhile, since the identity and concentration of matrix components tend to be unknown, it is very difficult to predict their effects on analytes' retention.

To study the limitations of RBwP in RPLC-MS under realistic sample conditions, we first studied the effects of sample solvents on RBwP prediction accuracy and identified the best solvents among those tested. Four of the nine sample solvents analyzed were selected to dissolve five different plant tissues to test the effects of matrix concentration on RBwP. Even though the five plants used in this study were randomly selected, our data indicate the limits of RPwB when analyzing concentrated matrix. Furthermore, we preliminarily explored the causes of these limits and suggest two methods to avoid these limits.

2. Material and methods

2.1 Equipment and software. The HPLC-MS used in this study was a DIONEX Ultimate 3000 (HPG-3400 RS) LC pump and a Bruker (Bremen, Germany) Amazon SL ion trap MS. RPwB software (http://www.measureyourgradient.org/) was written in Java 1.6 (Oracle, Redwood Shores, CA). It contains the Java OpenGL (JOGL) binding library version 2.0-rc11 (JogAmp, http://jogamp.org), the Unidata netCDF library version 4.2 (Unidata®, Boulder, CO), the Savitzky-Golay filter library version 1.2 by Marcin Rzeźnicki (http://code.google.com/p/savitzky-golay-filter/), the jmzML library, and the jmzReader library.

2.2 Chromatographic conditions. The column used was an Agilent Zorbax Eclipse Plus C_{18} (2.1 × 150 mm, 3.5 μm particle size). Solvent A was water with 0.1% formic acid, while solvent B was acetonitrile with 0.1% formic acid. Both solvents were prepared (formic acid, acetonitrile, and water) gravimetrically with ±0.001 g accuracy. The gradient used to run matrices was 20 min from 5% to 95% of solvent B. The flow rate was 0.4 mL/min and the column temperature was held at 35 °C.

2.3 Standards and testing compounds. The 20 standards and 16 testing compounds used in this study were the same as used previously (Abate-pella et al., 2015; Barnes et al., 2013a; Boswell et al., 2011a) for the convenience of comparing retention prediction accuracy of RPwB at different conditions. The 20 standards used for gradient back calculation were a series of amides including (in order of increasing retention factor) N-methylacetamide, N-ethylacetamide, N, N-dimethylpropionamide, benzamide, N-methylbenzamide, N-ethylbenzamide, N-propylbenzamide, N-butylbenzamide, N-pentylbenzamide, N-hexylbenzamide, N-heptylbenzamide, N-octylbenzamide, N-nonylbenzamide, N-decylbenzamide, N-undecylbenzamide, N-dodecylbenzamide, N-

tridecylbenzamide, N-tetradecylbenzamide, N-pentadecylbenzamide, and N-hexadecylbenzamide. The final concentration of N-methylacetamide, N-ethylacetamide, N, N-dimethylpropionamide, and benzamide were 2 mM, while the concentration of the 16 other standards were 0.5 mM. We synthesized N-propylbenzamide and all other benzamides of higher molecular weight using an adaptation of the Schotten-Baumann method (Tsuda, 1963). The remaining standards were obtained from Sigma-Aldrich Company (St. Louis, MO, USA).

The sixteen testing compounds with their final injection concentrations are listed below: chlorogenic acid (3 mM), N,N-diethylacetamide (0.75 mM), N-allyl aniline (1 mM), 7-amino-4-methylcoumarin (2 mM), naphthalene acetamide (1 mM), 1,3-naphthalenediol (4 mM), abscisic acid (2 mM), prednisone (1 mM), cortisone (1 mM), hydrocortisone (1 mM), tetrabutylammonium (0.1 mM), anilinoacetaldehyde diethyl acetal (1 mM,), curcumin (2 mM), tetrapentylammonium (0.1 mM), diphenylamine (0.75 mM), di-n-pentyl phthalate (1 mM). These compounds were purchased from Sigma-Aldrich Company (St. Louis, MO, USA).

2.4 Sample solvents. Nine solvents commonly used for the extraction of analytes from biological matrices were selected to study the prediction accuracy of RPwB. The relative polarities (P') of these solvents to water (P'=1.000) cover a wide range from P' = 0.789 to 0.099 (Reichardt and Eds, 2011). They are methanol (P' = 0.764), ethanol (P' = 0.789), isopropanol (P' = 0.546), acetonitrile (P' = 0.46), acetone (P' = 0.355), dichloromethane (P' = 0.309), ethyl acetate (P' = 0.288), tetrahydrofuran (P' = 0.27), and toluene (P' = 0.099). These solvents were purchased from Sigma-Aldrich Company (St. Louis, MO, USA).

2.5 Plant samples. Five different plant tissues were randomly chosen to test the limitations of RPwB for analyzing highly concentrated and chemically diverse plant matrices. These plants were *Solanum lycopersicum* fruit, *Solanum lycopersicum* stems and leaves, *Nicotiana* flowers, *Nicotiana* leaves, and *Nicotiana forsteri* leaves. Considering the ability to extract diverse compounds from these plants and the effects of sample solvents on RPwB prediction accuracy, four sample solvents: water, 70% ethanol, isopropanol, and dichloromethane were chosen to extract these five plant tissues.

To obtain the most concentrated plant samples, we started with un-weighed fresh samples and extracted them using solvents, then collected all supernatants and air dried to residues. All residues were weighed and re-dissolved in the minimum amount of solvents and their concentrations are shown in Table 6. The final concentration of each plant extract was not normalized since this experiment did not separate the effects of different compounds and the effects of concentrations on RPwB prediction accuracy.

In more detail, fresh plant tissues were flash frozen in liquid N_2, ground using mortars and pestles, and extracted separately using 5 mL of solvents, followed by 5 min shaking, 1 hr equilibration and 9 min centrifuge at 4000 *g* at 4 °C. The supernatants were transferred to 20 new tubes. The remaining residual material was re-extracted using 3 mL of solvents, following the same procedure. This process was repeated twice, and 11 mL of supernatant of every plant extract was collected. These extract supernatants were evaporated to residue by using vacuum centrifugation and weighed. Residues were re-dissolved using a minimum amount of solvent to obtain the most concentrated plant extracts possible. Their final concentrations are shown in Table 6.

Table 6. The concentration (mg/mL)* of plant matrices

	water	70% ethanol	isopropanol	dichloromethane
S. lycopersicum fruit	483	665	1402	2676
S. lycopersicum stem/ leaf	119	113	65	65
N. tabacum flower	150	196	137	125
N. tabacum leaf	59	167	94	115
N. forsteri leaf	181	90	73	55

*$Concentrtaion\ (mg/mL) = \frac{\text{the weight of dry samples (mg)}}{\text{the minimum amount of sample solvents (mL)}}$, see section 2.4 for more details.

2.6 Injection sample preparation. To study the effects of sample solvents on RPwB prediction accuracy in RPLC, 20 standards and 16 testing compounds were used, the same as for our previous publications (Abate-pella et al., 2015; Boswell et al., 2011a, 2011b). To test the effects of sample solvents on peak shapes and retention times of compounds, only 20 standards were analyzed by comparing their behaviors in different sample solvents with their behavior in a control sample solvent of 73% water and 27% acetonitrile. 150 µL of each solution of standards and testing compounds was evaporated using N_2 and re-dissolved in 150 µL sample solvents. The injection volume of each solution was 5 µL in all runs.

To test the effects of plant matrices on the prediction accuracy of RPwB in RPLC, 20 plant extracts were mixed with 20 standards and 16 testing compounds in the injection needle by multiple injection instead of spiking standards and testing compounds in plant extracts. There were two advantages of this procedure: first, it removed the prediction error of RPwB that may be caused by possibile insoluble standards and testing compounds in plant extracts; second, the multiple injection way matches our previous

publications (Abate-pella et al., 2015; Boswell et al., 2011a, 2011b) and so it is easier to compare results at the end. The order of multiple injection was: plant extracts, standards, and testing compounds.

3. Results and discussion

To analyze the effects of sample solvents on the prediction accuracy of PRwB in RPLC, we selected nine solvents with a wide range of polarities from methanol to toluene. We studied their effects on peak shapes and retention times of the 20 standards, which were used for back calculation of the experimental gradients, and effects on the prediction accuracy of RPwB when measuring 16 testing compounds. To study the robustness of RPwB when analyzing complex plant matrices, we randomly choose five plants and extracted them using four solvents, which were common solvents used in practical cases and that did not change the prediction accuracy of RPwB. These 20 plants extracts, 20 standards, and 16 testing compounds were mixed in the auto-injector needle during chromatography to test the limits of RPwB in analyzing high concentration and chemically diverse plant matrices.

3.1 Sample solvents distort the peak shape of 20 standards

Peak shape distortions including splitting, fronting, tailing, and broadening of peaks. In this paper, we categorize all these uncommon peak shapes as distorted peaks without subdivisions. It is believed that more non-polar sample solvents distort the peak shape of analytes in RPLC due to the abrupt decrease in analyte's solubility in the mobile phase upon injection (Jandera and Guiochon, 1991). If the sample solvent is stronger than the

mobile phase ($P'_s < P'_m$), it can interfere with the adsorption of the analyte at the head of the column (Keunchkarian et al., 2006). These distorted peak shapes caused by sample solvents can potentially affect the prediction accuracy of RPwB by increasing the difficulties of deciding accurate retention times of compounds. To study this, we selected nine sample solvents, in order of decreasing polarities from methanol, ethanol, isopropanol, acetonitrile, acetone, dichloromethane, ethyl acetate, tetrahydrofuran, and toluene, and studied their effects on peak shapes of 20 standards only (see section 2.3 for more details of these 20 standards).

Instead of tetrahydrofuran and toluene, we found isopropanol, acetonitrile, and ethyl acetate distorted more analytes' peak shapes than the other six solvents. Due to the huge amount of data (9 solvents, and 20 compounds per solvent), we show Fig. 10 only as an example exhibiting the most common distorted peak shapes we observed in our study. In Fig. 10, isopropanol (right panel) caused splitting of N- ethylacetamide and fronting of N, N-dimethylpropionamide, compared to when analytes were dissolved in our preferred solvent of 73% water and 27% acetonitrile (left panel), the same as shown in our previous publications (Abate-pella et al., 2015; Boswell et al., 2011a, 2011b).

Obviously, our result did not support the idea that more non-polar solvents distort peak shapes more. However, this result agreed with a previous publication (Keunchkarian et al., 2006). Solvent strength alone failed to explain multiple uncommon peak shapes. Instead, as an important hydrodynamic property of solvents, viscosities of the sample solvent and mobile phase played an important role in peak shapes of analytes. The mismatches of viscosities could raise the instability of the interface between two solvents and this phenomenon was named 'viscous fingering' (Shalliker and Guiochon, 2009).

However, for lower mismatches of viscosities, no peak distortions would be observed (Shalliker and Guiochon, 2009). For example, there were no splitting band or shoulders when solvent contrast was up to 1.283 cP (viscosity ratio) (Shalliker and Guiochon, 2009). Overall, the severity of peak distortion depends on sample solvent strength and the viscous fingering that is caused by solvent contrast between the sample solvent and the mobile phase.

Interestingly, even though our control sample solvent of 73% water and 27% acetonitrile was close in composition to our mobile phase, acetonitrile was one of the three sample solvents that distorted peak shapes the most. This result was the opposite from what we used to believe, namely that a symmetric peak shape is produced when the sample solvent is the same as the mobile phase (Castells and Castells, 1998). We suspected that the peak shape is also dependent on the relative solubility of analytes and the amount of sample solvents. Besides, toluene caused less peak distortions in this study, which was the opposite from previous publication that carotenoids suffered more peak distortions when toluene was used as the sample solvent in RPLC (Khachik et al., 1988). This difference could be because the behaviors of benzamides (see section 2.3 for more details) in toluene are different from carotenoids in toluene. To study the causes of these effects, more diverse chemicals should be analyzed in the future.

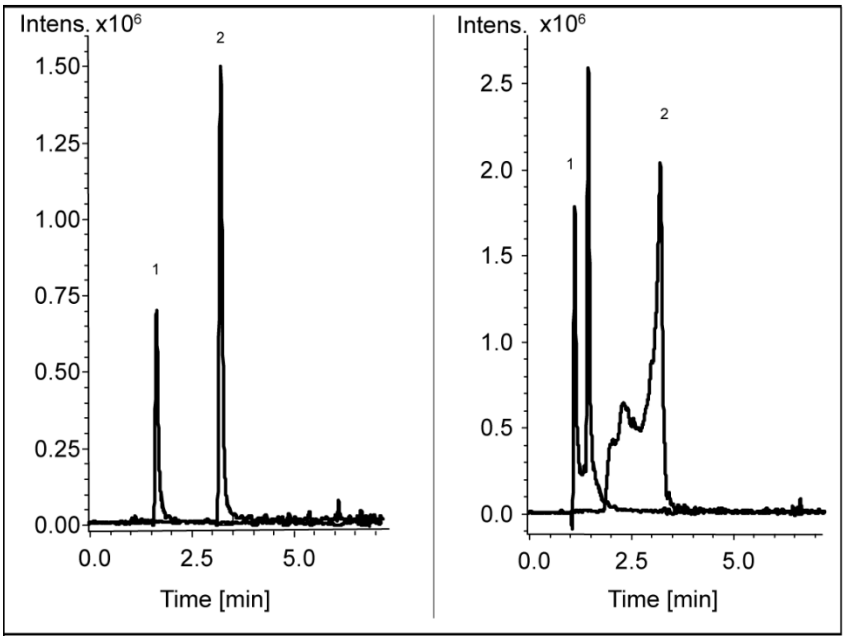


Figure 10. An example of distorted peak shapes of analytes using isopropanol as the sample solvent. Extracted ion chromatograms of ethylacetamide and N, N-dimethylpropionamide in 73% water and 27% acetonitrile (Left), and in isopropanol (Right). Peak 1 is N- ethylacetamide (m/z = 88) and peak 2 is N, N-dimethylpropionamide (m/z = 102). Peaks were smoothed using Savitzky Golay with a 2.04 smoothing width for two smoothing cycles.

3.2 Sample solvents shift the retention time of 20 standards

Retention time can drift easily due to many factors, including intentional factors such as flow rate, mobile phase, and stationary phase and unintentional factors such as imperfect behavior of the LC instrument, gradient delay, and solvent dispersion (Boswell et al., 2011a, 2011b; Boswell et al.; Wang and Boswell, 2017). There are only a few publications about the effects of sample solvents on retention time (Shalliker and Guiochon, 2009), but more on their effects on peak distortions (Castells and Castells,

1998; Keunchkarian et al., 2006; Shalliker and Guiochon, 2009). To answer how sample solvents can shift retention time and thus cause a drift in the prediction accuracy of RPwB, we tested effects of the most common nine sample solvents on retention time of 20 standards used in RPwB.

As shown in Fig. 11, poorly retained standards shift more than well retained ones in all nine sample solvents. This result agrees with previous publications that poorly retained analytes suffer more peak distortions (Castells and Castells, 1998; Jandera and Guiochon, 1991; Shalliker and Guiochon, 2009), and thus shift their retention times more as it is not possible to produce exactly the same location of peak maxima and shoulders once peak shapes are deformed. As mentioned in section 3.1, the severity of peak distortion depends on sample solvent strength and the viscous fingering that is caused by solvent contrast between the sample solvent and the mobile phase. Less strongly retained analyztes elute from sample solvent and travel through longer column lengths accompanying their sample bands, while more strongly retained analytes separate from sample solvent at the beginning of the column and have more intense interactions with the column. Thus, peak distortions of stronger retained analytes that are caused by sample solvents had more time to recover.

Furthermore, we observed (Fig. 11) large negative shifts in the retention times of less retained standards when isopropanol, acetonitrile, ethyl acetate, and tetrahydrofuran were the sample solvents. This result matches with results presented in the publication by Shalliker *et. al* that analytes elute earlier with increasing viscosity contrast (Shalliker and Guiochon, 2009). Interestingly, compared to the other eight sample solvents, we observed that ethanol causes much smaller retention time shifts. This was caused by the polarity of

ethanol, the viscosity contrast between ethanol and the mobile phase, and the properties of the 20 standards. Further information will be required to figure out the detailed causes of these effects.

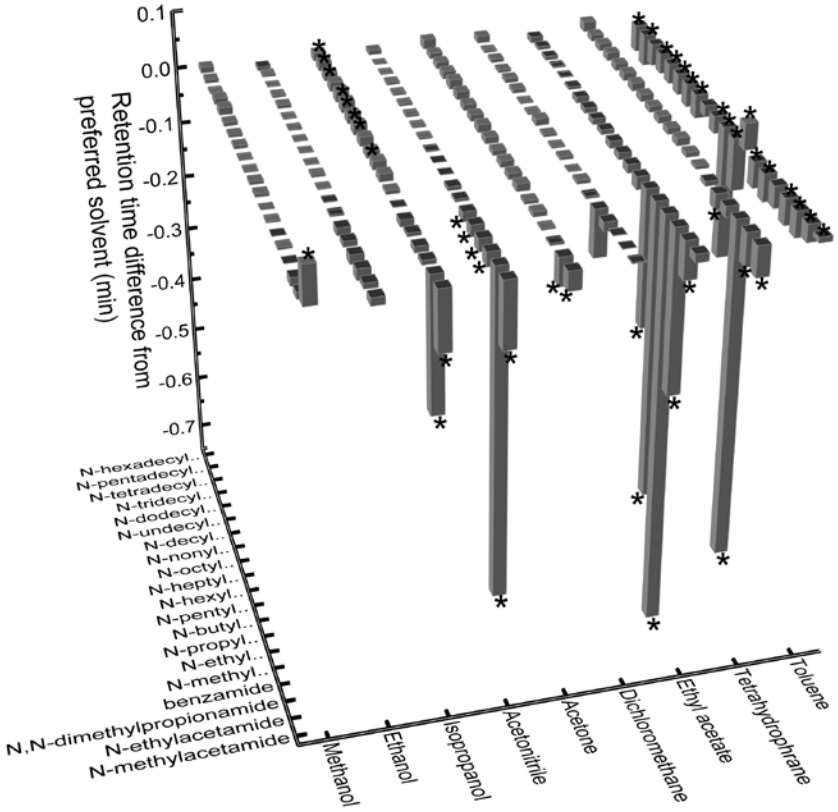


Figure 11. Sample solvents shift the retention times of some back-calculation standards. The X-axis is the nine sample solvents in order of decreasing polarity. The Y-axis is 20 back-calculation standards given in order of increasing retention time from N-methylacetamide to N-hexadecylbenzamide. The Z-axis is the retention time difference from that in the preferred sample solvent (73% water and 27% acetonitrile). Stars indicate statistical difference (Student's t-test, $p < 0.05$). There were four replicates of each run.

3.3 Sample solvents alter the prediction accuracy of RPwB in RPLC

Sample solvents distorted the peak shape and shifted the retention time of 20 standards that were used to back-calculate real behaviors of LC. Therefore, they could alter retention prediction accuracy of RPwB in RPLC. As shown in Fig. 12, the overall prediction error of RPwB in predicting retention times of 16 testing compounds was as stable as in our previous publication (Abate-pella et al., 2015), only 1.9 sec when the preferred sample solvent (73% water and 27% acetonitrile) was used. The overall prediction error was very close to that of the preferred sample solvent when methanol or ethanol was used as the sample solvent at 1.8 sec. Interestingly, the overall prediction error of acetone and dichloromethane were 1.5 sec and 1.7 sec respectively, which were even smaller than that of the preferred sample solvent. The overall prediction errors produced when isopropanol, acetonitrile, tetrahydrofuran, and toluene were used as sample solvents were 2.5, 2.0, 2.1, and 2.9 sec, respectively. Ethyl acetate produced the largest overall prediction error at 4.3 sec, which was worse than the 3 sec that was the maximum deviation allowed to determine the successful application RPwB in RPLC (Abate-pella et al., 2015; Wang and Boswell, 2017). As shown in the section 3.1, ethyl acetate was one of the three sample solvents that distorted peak shapes the most, and Fig. 11 shows that ethyl acetate was the sample solvent that shifted retention time the most. This explains the large prediction error produced by ethyl acetate. More interestingly, it indicates the robustness of RPwB, as its prediction accuracy remains high if sample solvents only slightly distort peak shapes or shift retention times, not as badly as is seen with ethyl acetate.

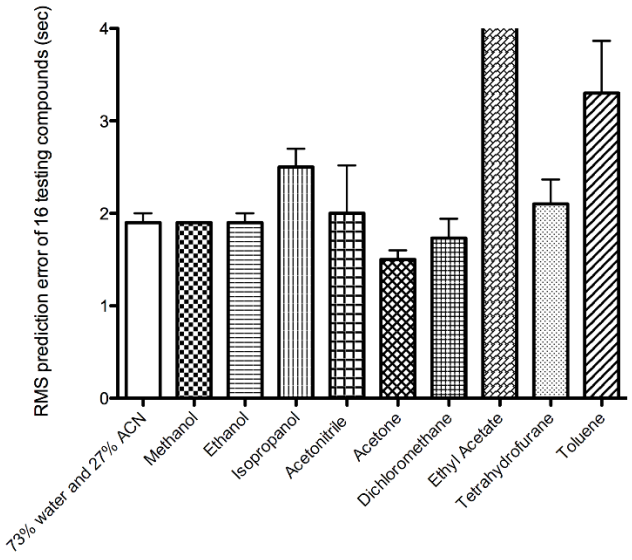


Figure 12. Sample solvents alter the prediction accuracy of RPwB in RPLC. The X-axis displays the preferred sample solvent (73% water, 27% acetonitrile) and the nine test solvents. The Y-axis shows their RMS prediction errors. RMS (Root Mean Square)=$\sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 \ldots + x_n^2)}$. The 16 test compounds (see section 2.3 for details) were the same used to test predict accuracy of RPwB in RPLC previously (Abate-pella et al., 2015). There were four replicates of each run.

3.4 Plant matrices alter the prediction accuracy of RPwB in RPLC mainly through interfering with two testing compounds: tetrabutylammonium and tetrapentylammonium

Complex samples potentially alter the prediction accuracy of RPwB as unknown interactions interfere with peak shapes and shift retention times of analytes. We randomly choose five plant tissues (*S. lycopersicum* fruit, *S. lycopersicum* stem and leaf, *N. tabacum* flower, *N. tabacum* leaf, and *N. forsteri* leaf) as examples to test the prediction

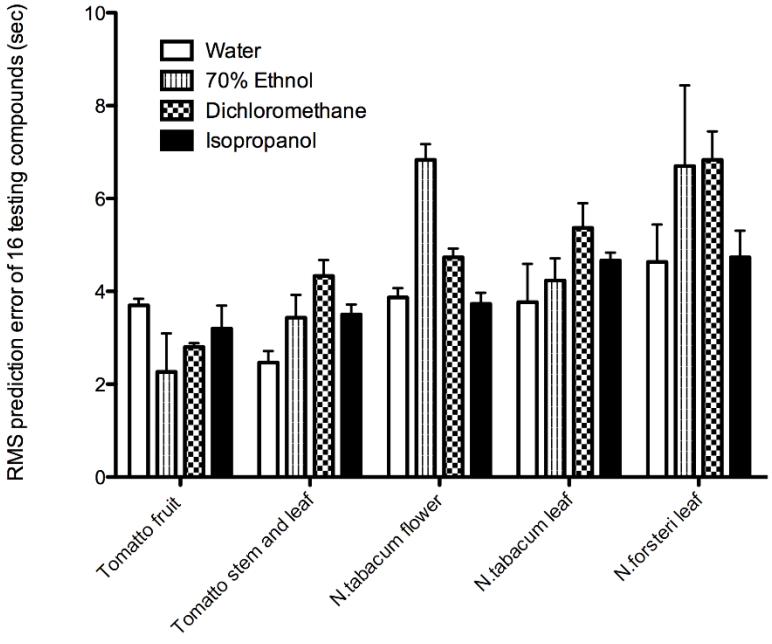
accuracy of RPwB in real cases under two limits: 1) the most concentrated samples we could obtain, see table 6 for more details; 2) complex plant matrices with diverse compounds that are extracted using four extraction solvents (water, dichloromethane, 70% ethanol, and isopropanol). Different extraction solvents extract different compounds from plant matrices. These compounds could interfere with analytes differently and thus alter prediction accuracy differently. Besides, these extraction solvents are the most common ones and did not alter the prediction accuracy of RPwB when used as sample solvents. Here, by analyzing 20 plant matrices/solvents with very diverse compounds and the most concentrated samples we could obtain, we aimed to break the prediction accuracy of RPwB and provid suggestions for RPwB application in real cases.

As shown in Fig. 13A, the overall prediction error for most samples was worse than the 3 sec that is the maximum deviation allowed in determining the success of RPwB application. Of these 20 plant extracts, three had prediction errors smaller than 3 sec: *S. lycopersicum* fruit extracted by 70 % ethanol at 2.3 sec, *S. lycopersicum* fruit extracted by dichloromethane at 2.8 sec, and *S. lycopersicum* stems/leafs extracted by water at 2.5 sec. Interestingly, looking closer at their final concentrations, shown in Table 6 (665 mg/mL, 2676 mg/mL, 119 mg/mL, respectively), *S. lycopersicum* fruit extracted using 70 % ethanol or dichloromethane were the most concentrated samples, and *lycopersicum* stems/leafs extracted using water were also highly concentrated. This indicates there is not a correlation between the concentration of plant matrices and the prediction error of RPwB.

Instead, we found that the unwanted prediction errors of these 20 plant matrices were mainly from two testing compounds: tetrabutylammonium and tetrapentylammonium. As

shown in Fig. 13B, the prediction error of RPwB in samples decreased to less than 3 sec when these two testing compounds were removed, except for samples that were extracted by isopropanol, which “struggled” to be under 3 sec. This suggests that tetrabutylammonium and tetrapentylammonium are very sensitive to some compounds that are commonly found in all plant extracts, and thus alter the prediction accuracy of RPwB. Furthermore, plant matrices extracted by 70% water, ethanol and dichloromethane show smaller prediction errors compared to when isopropanol is used as the extraction solvent (Fig. 13B). This result agrees with Fig. 12. The overall prediction error using isopropanol as the sample solvent was 2.5 sec, which was higher than the prediction error observed when using 73% water and 27% acetonitrile (1.7 sec), ethanol (1.8 sec), and dichloromethane (1.7 sec) as sample solvents.

A



B

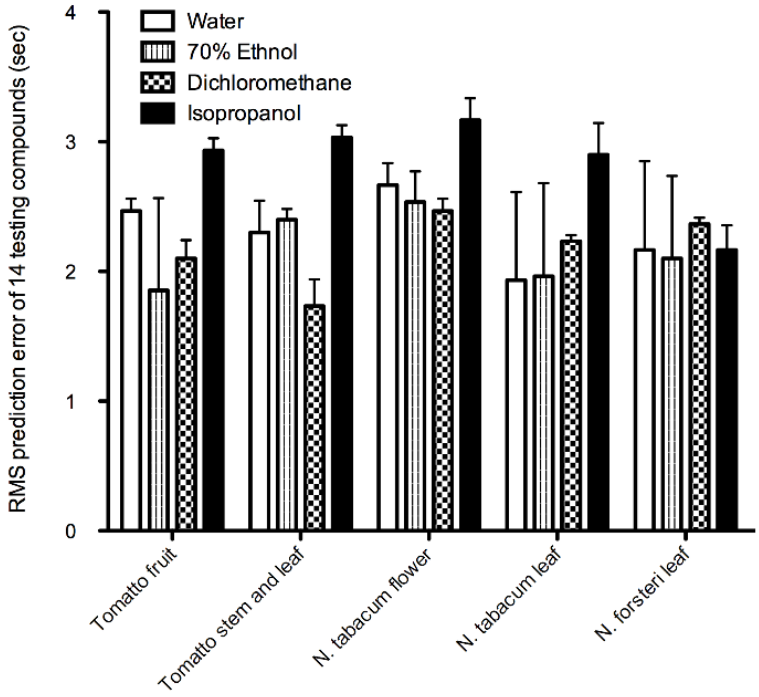


Figure 13. The overall prediction errors of RPwB in 20 plant extracts. The overall prediction errors of RPwB in 20 plant extractions. The overall prediction error was calculated by using RMS (Root Mean Square) = $\sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 \ldots + x_n^2)}$, where x was the teasing compound. (A) The overall prediction error of 16 testing compounds (see section 2.3 for more details). (B) The overall prediction error of 14 testing compounds, not including two testing compounds: tetrabutylammonium and tetrapentylammonium. There were four replicates of each run.

3.5 The prediction accuracy of RPwB in plant matrices can be rescued by diluting injection samples.

To rescue the prediction accuracy of RPwB, the easiest way was to remove the two testing compounds tetrabutylammonium and tetrapentylammonium as shown in Fig. 13B.

However, in real cases, we could not simply remove these two testing compounds to maintain prediction accuracy as we are not certain they are always the ones that decrease the prediction accuracy. Also, we still do not understand how the two compounds alter prediction accuracy. Therefore, we looked for an alternative way to rescue the prediction accuracy of RPwB.

As shown in Fig. 13A, sample concentrations contribute lots prediction errors. We took *N. tabacum* leaf extracted by 70 % ethanol as an example to study the acceptable concentration of analyzed samples. As shown in Fig. 14, three concentrations were continuously injected 25 times to mimic real cases. The three concentrations were the most concentrated we obtained, ~167 mg/mL (Table 6), a two times of dilution to ~83.5 mg/mL, and a ten times of dilution to ~16.7 mg/mL. The prediction error of the highest sample concentration (~167 mg/mL) started at 4.1 sec and increased to 11 sec. Interestingly, the prediction error did not keep increasing with the accumulation of injection times. The two times dilution (~83.5 mg/mL) brought the prediction error down to 2.1 sec regardless of the number of injections and the ten times dilution (~16.7 mg/mL) had prediction errors around 1.9 sec at all injection numbers tested. This result indicates that the prediction accuracy of RPwB can be rescued by diluting plant matrices.

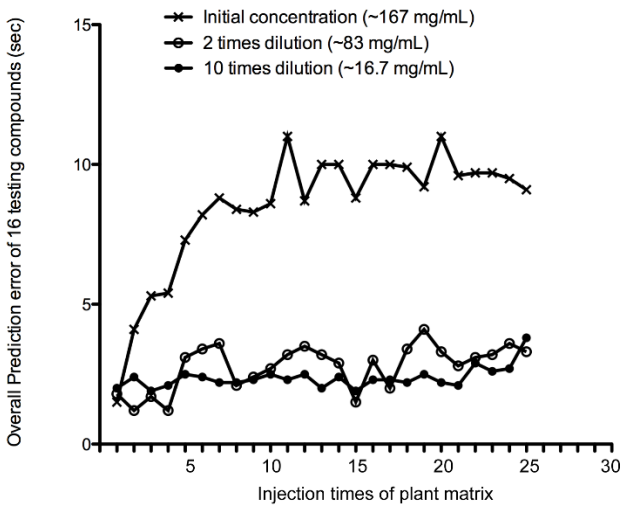


Figure 14. Analysis of the overall prediction error of RPwB using three concentrations of *N. tabacum* leaf extracts in 70 % ethanol and 25 sample injections. The overall prediction error was calculated by RMS (Root Mean Square)=$\sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 \ldots + x_n^2)}$, with x as the teasing compound. There were four replicates of each run.

4. Conclusions

RPwB is a powerful method for predicting retention information in GC, RPLC and HILIC and therefore has the potential to improve metabolite identification by combining mass information and retention information (Abate-pella et al., 2015; Boswell et al., 2011a; Peng et al., 2014; Wang and Boswell, 2017). However, this method has not been applied to real cases analyzing compounds of interest. In this work, we tested the prediction accuracy of RPwB for analyzing plant matrices using RPLC-MS, and we studied the limits of this method and propose tips to improve its prediction accuracy.

There are two parts to our study. During the first part, we selected nine sample solvents, in order of decreasing polarities from methanol, ethanol, isopropanol,

acetonitrile, acetone, dichloromethane, ethyl acetate, tetrahydrofuran, to toluene, and systematically studied their effects on the peak shape and retention time of 20 standards used to simulate real behavior of LC, and prediction accuracy of RPwB. Our results suggest that there is not a direct correlation between the polarity of sample solvents and the severity of peak distortions of analytes. Of these nine sample solvents, instead of the most non-polar sample solvents such as tetrahydrofuran and toluene, we found that isopropanol, acetonitrile, and ethyl acetate were the three that distorted peak shapes the most. Besides solvent polarity, this phenomenon was also affected by the mismatches of viscosities between sample solvent and mobile phase, which cause viscous fingering and induce peak distortions (Castells and Castells, 1998; Keunchkarian et al., 2006; Mishra et al., 2013; Shalliker and Guiochon, 2010, 2009). However, no peak distortions could be observed if the mismatches of viscosities were beneath a certain range, such as 1.283 cP (Shalliker and Guiochon, 2009). Furthermore, all nine sample solvents shifted retention times of 20 standards and poorly retained compounds suffered more than well retained ones. This was because the less retained compounds eluted at approximately the same time as the sample solvent. This means that they traveled through a greater percentage of the length of the column together with the sample solvent and thus were more strongly affected by the sample solvent. In contrast, the more highly retained analytes separated from sample solvent at the beginning of the column and thus suffered fewer peak distortions and smaller shifts in retention times (Castells and Castells, 1998; Keunchkarian et al., 2006; Mishra et al., 2013; Shalliker and Guiochon, 2010, 2009). Lastly, of these nine sample solvents, only ethyl acetate broke the prediction accuracy of RPwB and had a prediction error at 4.3 sec, which was more than 3 sec that was the

maximum deviation allowed to determine the successful application RPwB in RPLC (Abate-pella et al., 2015; Boswell et al., 2011a, 2011b). The sample solvents methanol, ethanol, acetone and dichloromethane maintained similar prediction accuracy as our preferred sample solvent of 73% water and 27% acetonitrile. The other sample solvents tested, such as isopropanol, acetonitrile, tetrahydrofuran and toluene, had worse prediction accuracy than the preferred solvent but still had prediction errors smaller than 3 sec. Overall, our result suggest the robustness of RPwB when different sample solvents are applied in real cases.

For the second part of this work, focused on studying the limit of plant extract concentrations and the effects of compound diversity in plant extracts on the prediction accuracy of RPwB, we chose five common plant tissues (*S. lycopersicum* fruit, *S. lycopersicum* stems and leafs, *Nicotiana* flowers, *Nicotiana* leafs and *Nicotiana* forsteri leafs) and extracted them using four sample solvents (water, 70% ethanol, dichloromethane and isopropanol). These four sample solvents are the most commonly used and did not change the prediction accuracy of RPwB, and allowed the most concentrated plant extractions to be obtained, as shown in Table 6. Of these 20 plant extracts, only three had prediction errors smaller than 3 sec. They were *S. lycopersicum* fruits extracted with 70% ethanol (2.3 sec), *S. lycopersicum* fruits extracted with dichloromethane (2.8 sec), and *S. lycopersicum* stems and leafs extracted with water (2.5 sec). Interestingly, of these three plant extracts, two plant matrices (*S. lycopersicum* fruits extracted by 70% ethanol and *S. lycopersicum* fruits extracted by dichloromethane) had the highest solute concentrations (665 mg/mL and 2676 mg/mL, separately). This result indicates that the concentration of plant matrices attributes the prediction errors. In

addition, we found that a substantial amount of the prediction errors of these 20 plant extracts came from two charged testing compounds, tetrabutylammonium and tetrapentylammonium. This could be because these two compounds are very sensitive to plant matrices and induce unknown interactions that interfere with prediction accuracy. For a more detailed understanding of this effect, more experiments will be required. To rescue the prediction of RPwB in plant matrices, besides removing these two sensitive testing compounds, we could also dilute injection samples.

Acknowledgements

We thank the National Institutes of General Medical Sciences [R01GM098290], the University of Minnesota Office of the Vice President for Research, and the Minnesota Agricultural Experiment Station for funding. We appreciate the help of instrument trouble-shooting from Joshua T. Hewitt (University of Minnesota) and manuscript proofreading by Michael Wilson (University of Minnesota).

Chapter 4

MpAMT1;2 From Marchantia polymorpha Is A High-Affinity, Plasma Membrane Ammonium Transporter

Nu Wang[1,3], Hanqing Guo[1,3,4], Tami R. McDonald[2], Anke Reinders[1] and John M. Ward[1]

[1]Department of Plant and Microbial Biology, University of Minnesota, USA

[2] Biology Department, St. Catherine University, St. Paul, MN, USA

[3] authors contributed equally

[4] current address: Department of Biological Sciences, Dartmouth College, Hanover, NH, USA

Abstract

Plant ammonium transporters in the AMT/MEP/Rh superfamily have only been previously characterized in flowering plants (angiosperms). Plant AMT1s are electrogenic, while plant AMT2s are electroneutral and MEP and Rh transporters in other organisms are electroneutral. We analyzed the transport function of MpAMT1;2 from the basal land plant *Marchantia polymorpha*, a liverwort. MpAMT1;2 was shown to localize to the plasma membrane in *Marchantia* gametophyte thallus by stable transformation using a C-terminal citrine fusion. *MpAMT1;2* expression was studied using qRT-PCR and shown to be higher when plants were N deficient and lower when plants were grown on media containing ammonium, nitrate, or the amino acid glutamine. Expression in *Xenopus* oocytes and analysis by electrophysiology revealed that MpAMT1;2 is an

electrogenic ammonium transporter with a very high affinity for ammonium (7 µM at pH 5.6 and a membrane potential of -137 mV). A conserved inhibitory phosphorylation site identified in angiosperm AMT1s is also present in all AMT1s in Marchantia. Here we show that a phosphomimetic mutation T475D in MpAMT1;2 completely inhibits ammonium transport activity. The results indicate that MpAMT1;2 may be important for ammonium uptake into cells in the *Marchantia* thallus.

Keywords**:** ammonium transporter, bryophyte, electrophysiology, liverwort, *Marchantia polymorpha*, *Xenopus* oocyte expression

Abbreviations: AMT - ammonium transporter
MEP - methylammonium and ammonium permease
Rh - Rhesus factor
TEVC - two-electrode voltage clamping

1. Introduction

Ammonium transporters are integral membrane proteins that transport NH_4^+ or NH_3 across the cell membrane (Andrade et al., 2005; Gruswitz et al., 2010; Zheng et al., 2004). They are present in almost all organisms, including plants, archaea, bacteria, fungi and animals (Saier et al., 1999). Ammonium transporters are classified in the AMT/MEP/Rh superfamily and fall into three related clades: AMTs (ammonium transporters), MEPs (methylammonium and ammonium permeases) and Rh (rhesus factor proteins) (A. Marini et al., 1994; Marini et al., 2000; Ninnemann et al., 1994b). MEPs are mainly found in fungi and bacteria, and Rh are found in animals, while AMTs are found both in plants and animals excluding vertebrates (McDonald and Ward, 2016). Plants contain both AMTs (called AMT1) and MEPs (called AMT2). *AMT2* genes in plants are the result of horizontal gene transfer of *MEPs* from prokaryotes (McDonald et al., 2012).

Ammonium (NH_4^+) and ammonia (NH_3) in solution are in pH-dependent equilibrium; in this paper the term 'ammonium' will be used to represent the sum of ammonium and ammonia while the chemical formula, for example NH_4^+, will be used to represent specific species. In plants, AMT1s are electrogenic (Uwe Ludewig et al., 2002), they transport a charge across the membrane as well as ammonium, and function either as NH_4^+ channels or as NH_3/H^+ symporters. Both Rh and MEP transporters are electroneutral, they do not transport a net charge (Ludewig, 2004; Mayer and Ludewig, 2006; Ripoche et al., 2004). Consistent with this, transporters in the plant AMT2 clade are MEPs and are electroneutral (Neuhäuser et al., 2009). In addition, the AmtB protein from *E. coli*, which belongs to the MEP group, is also electroneutral (Javelle et al., 2005; Khademi Shahram, O'Connell III Joseph, Remin Jonathan, Robles-Colmenares Yaneth, Miercke Larry J.W., 2004; Soupene et al., 2001, 1998; Zheng et al., 2004).

LeAMT1;1 from tomato was the first ammonium transporter to be identified as electrogenic (Uwe Ludewig et al., 2002). Electrophysiology experiments showed that LeAMT1;1 transports both ammonium and methylammonium, and both substrates induce inward currents. An inward current is defined as either cation movement into the cytoplasm or anion movement out, and in this case, the inward current could be carried by either H^+ or NH_4^+ moving into the cell. LeAMT1;1 has a high affinity for ammonium, with a K_m in the range of 10 µM, and a 25-fold lower affinity for methylammonium. K_m values for both substrates are voltage dependent with higher affinity associated with more negative membrane potentials (Uwe Ludewig et al., 2002). Subsequently, AMTs were described in other eudicots. AtAMT1;1, AtAMT1;2, and AtAMT1;4 were characterized in *Arabidopsis thaliana* (Mayer et al., 2006; Yuan et al., 2009). PvAMT1;1 from

Phasoleus vulgaris is a pH-dependent transporter a with high affinity for ammonium (Ortiz-Ramirez et al., 2011b). In venus flytrap (*Dionaea muscipula*) DmAMT1;1 was shown to be a low-affinity transporter (Scherzer et al., 2013). In monocots, AMTs have been studied in *Oryza sativa* (OsAMT1;1) (Yang et al., 2015). All AMT1s studied so far are electrogenic. However, plant AMTs have only been studied in flowering plants (angiosperms).

Very little information exists concerning AMT function in animals. AgAMT from the mosquito *Anopheles gambiae* was demonstrated to be electrogenic when expressed in *Xenopus* oocytes (R. Jason Pitts et al., 2014). This raises the question whether electrogenicity evolved with plants or earlier in a common ancestor of plants and animals. Therefore, to further elucidate whether all plant AMT1s are electrogenic we analyzed the transport activity of an AMT1 from the basal land plant, *Marchantia polymorpha*.

The liverwort *Marchantia polymorpha* is a classic model plant. It is a bryophyte and represents the earliest land plants (Bowman et al., 2017). There are several important differences in lifecycle between bryophytes and angiosperms. Bryophytes have a dominant gametophyte generation, and a dependent, diploid sporophyte generation. They have flagellate sperm and lack vascular tissue (Shimamura, 2016). A previous study showed that four ammonium transporters from *M. polymorpha* are functional when expressed in *Xenopus* oocytes (McDonald and Ward, 2016). Here we studied the transport activity of MpAMT1;2 using electrophysiology to determine substrate affinity and regulation by membrane potential and pH. We found that MpAMT1;2 protein is localized to the plasma membrane and that *MpAMT1;2* gene expression is induced under N deficiency conditions and repressed when plants are resupplied with a N source.

2. Materials and methods

2.1 MpAMT1;2 cDNA cloning from *Marchantia polymorpha.* Ammonium transporter gene *MpAMT1;2* from *Marchantia polymorpha* was identified based on genomic DNA sequence provided by Drs. Ryuichi Nishihama and Takayuki Kohchi (Kyoto University). RNA was extracted and cDNA was synthesized using the Omniscript Reverse Transcription kit (Qiagen). Primers for PCR were designed based on the coding region sequence of *MpAMT1;2.* Products were cloned into pCR8/GW/TOPO (Invitrogen, Carlsbad, CA) and recombined with the oocyte expression vector pOO2/GW (Sun et al. 2010).

2.2 Expression in oocytes. Oocytes were prepared from female *Xenopus laevis* frogs. The pOO2/GW plasmid containing *MpAMT1;2* was linearized using *Pma*CI and cRNA was synthesized using the SP6 mMessage mMachine kit (Ambion). Oocytes were injected with 50 ng of cRNA. Oocytes were incubated in Barth's solution (88 mM NaCl, 1 mM KCl, 0.33 mM $Ca(NO_3)_2$, 0.41 mM $CaCl_2$, 0.82 mM $MgSO_4$, 2.4 mM $NaHCO_3$ and 10 mM Hepes, pH 7.6, 100 μg/ml penicillin, 100 μg/ml streptomycin) containing 10 μg/ml gentamycin for 3-6 days at 15 °C before recordings.

2.3 Electrophysiology. For recording, 1.5 mm thin wall borosilicate glass pipettes (Warner Instruments Corp.) were filled with 1 M KCl. Measurements were obtained using a Dagan TEVC 200A amplifier (Dagan Corp, Minneapolis, MN). Oocytes were incubated in modified K^+-free Ringer solution (115 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$ and 5 mM MES) during recordings. Oocytes were clamped at the holding potential of -40 mV. Currents were recorded using Clampex (Axon Instruments Inc.).

Solutions of ammonium chloride in K^+-free Ringer solution were prepared with three different pH values, 5.6, 6.8 and 8.0. Ringer solution was adjusted to pH 5.6 with NaOH or to pH 6.8 or 8.0 using Tris. Solutions of methylammonium chloride in the K^+-free Ringer solution were prepared at pH 5.6. Currents were measured at membrane potentials from +37 to -137 mV and filtered at 1000 Hz. Substrate-dependent currents were obtained by subtracting an average of background currents before and after substrate application. To determine $K_{0.5}$ values, substrate-dependent currents were collected at various substrate concentrations.

Currents were normalized to Vmax to account for different levels of expression between oocytes. Currents from at least three independent oocytes were recorded for each experiment. The relationship between substrate concentration and substrate-induced currents was analyzed using the Michaelis-Menten equation.

$$I = I_{max} \times [S]/(K_{0.5} + [S])$$

where I is the substrate induced current, [S] is the substrate concentration, I_{max} is the maximum current, $K_{0.5}$ is the substrate concentration when the current is half-maximal. Non-linear regression was performed using Prism (GraphPad Software).

2.4 Expression analysis of *MpAMT1;2*. Takaragaike-1 (Tak-1) gemmae were grown under long day conditions (16 h light) on half-strength Gamborgs B5 media for two weeks. Half-strength Gamborgs B5 contains 13.4 mM total N (12.4 mM KNO_3 and 0.5 mM $(NH_4)_2SO_4$). Plants were transferred to the same media but without nitrogen and with 5 mM KCl for five days and then transferred to the same media or to media with different N sources, each with a total of 5 mM N: 5 mM KNO_3, 2.5 mM $(NH_4)_2SO_4$, 2.5 mM

NH_4NO_3 or 2.5 mM glutamine. 5 mM KCl was added to the media to replace the potassium lost due to the removal of KNO_3 (except when 5 mM KNO_3 was added as nitrogen source). Plants were grown for two days before RNA extraction. Individual thalli were harvested, flash frozen in liquid nitrogen, pulverized, and total RNA extracted using Trizol (Invitrogen). Quantitative RT-PCR was performed using the SuperScript III Platinum SYBR Green One-Step quantitative RT-PCR kit (Invitrogen) and a Lightcycler 480 Real-Time PCR system (Roche). For each 12.5 µl reaction 30 ng of total RNA and gene-specific primers were used. The cycle conditions used were 50 °C for 10 min, 95 °C for 10 min followed by 45 cycles of 95 °C for 10 sec and 60 °C for 1 min. Three biological replicates were analyzed for each growth condition, with two technical replicates each. The crossing point (Cp) was calculated using the second derivative max method provided with the Lightcycler software for each amplification curve. The Cp values for the technical replicates were averaged before further analysis. MpACTIN7 was used as the reference gene (Saint-Marcoux et al., 2015). The primers used were MpACT7qPCR-F 5'- AGGCATCTGGTATCCACGAG and MpACT7qPCR-R 5'- ACATGGTCGTTCCTCCAGAC, MpAMT1.2qPCR-F 5'- TTCTGGCTGCTCAAGAAGTT and MpAMT1.2qPCR-R 5'- GGGCTCCTCGATGTTGTAAG.

2.5 Membrane localization of MpAMT1;2. A MpAMT1;2-citrine fusion driven by dual 35S promoters was used to transform *Marchantia polymorpha* (Tak-1). The stop codon in the MpAMT1;2 cDNA in pCR8/GW/TOPO was mutated from TAA to TCA using the QuikChange II kit (Agilent) and the following primers: forward 5'- ATCCCTATGGCTTCAAAGGGCGAATTCGAC-3', and reverse 5'-

GTCGAATTCGCCCTTTGAAGCCATAGGGAT-3'. *MpAMT1;2* without the stop codon was recombined with the destination Gateway vector pMpGWB106 with C-terminal citrine fusion (Ishizaki et al., 2015). Thallus explants were transformed using Agrobacterium C58C1. As a control, an N-terminal citrine-MpSUT4 construct was generated in destination vector pMpGWB105 (Ishizaki et al., 2015). Gemmae were grown for 2 days on half-strength Gamborgs B5 and imaged using a confocal microscope.

2.6 Mutagenesis of MpAMT1;2. MpAMT1;2 (T475D), and MpAMT1;2 (T475A) were generated using the QuikChange II site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA). The sequence encoding threonine (ACG) at position 475 was changed to either GAT to encode aspartic acid or GCG to encode alanine. The PCR primers used to generate MpAMT1;2 (T475D) were: forward 5'- GTCGGGCATGGATCTCGATCGCCA and reverse 5'-TGGCGATCGAGATCCATGCCCGAC. The PCR primers used to generate MpAMT1;2 (T475A) are: forward 5'- GTCGGGCATGGATCTCGCGCGCCA and reverse 5'-TGGCGCGCGAGATCCATGCCCGAC. Mutated MpAMT1;2 clones in pCR8/GW/TOPO (Invitrogen, Carlsbad, CA) were sequenced, then recombined into the oocyte expression vector pOO2/GW (Sun et al., 2010).

3. Results

MpAMT1;2 (Genbank KU975452) is a membrane protein containing 515 amino acids and 11 predicted membrane spans. MpAMT1;2 was previously expressed in Xenopus oocytes and shown to be electrogenic (McDonald and Ward, 2016) but its transport activity was not characterized further. AMT1 transporters are highly conserved from bryophytes through angiosperms, for example MpAMT1;2 shares 72% identity at the

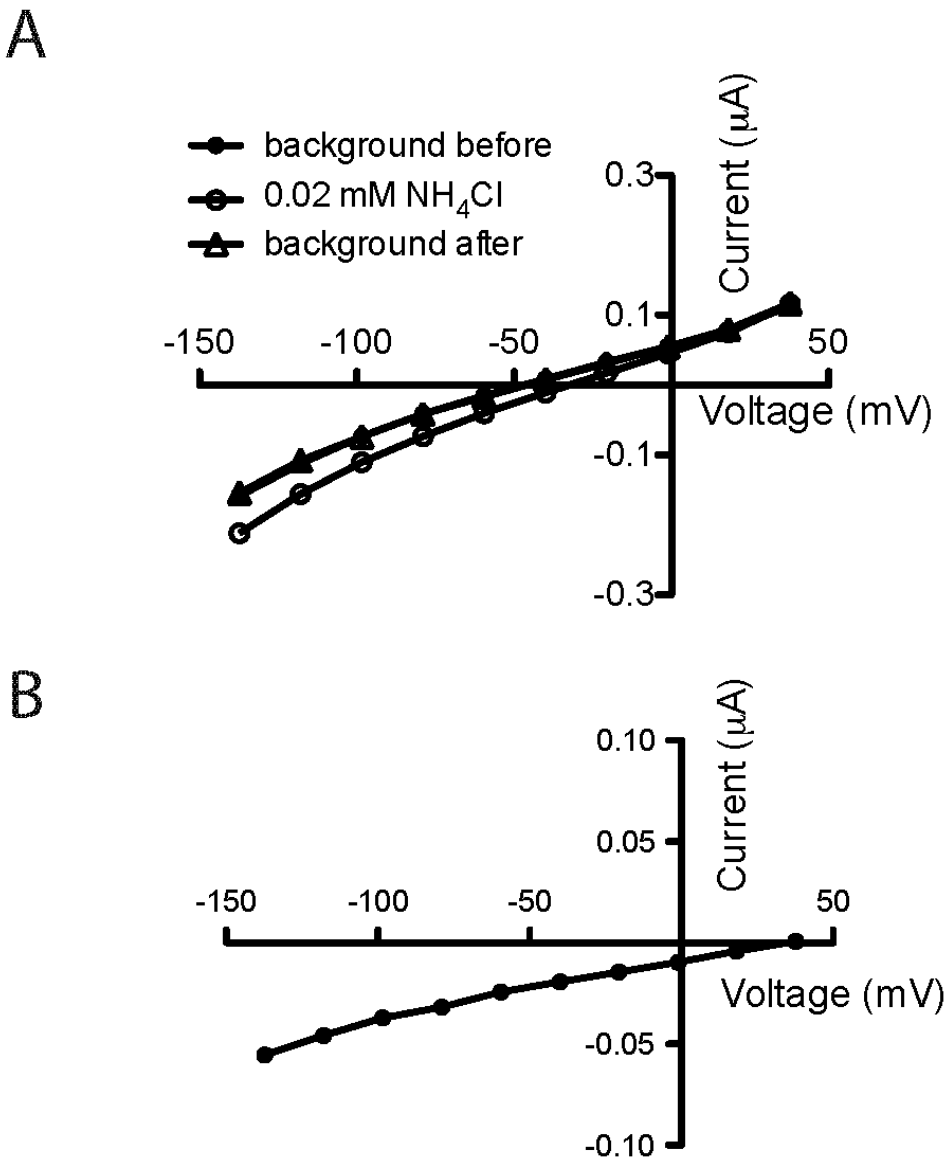
amino acid level with AtAMT1;1. MpAMT1;2 was expressed in *Xenopus laevis* oocytes and ammonium-induced currents were recorded using two-electrode voltage clamping. Potassium-free Ringer solution was used for the oocyte electrophysiology experiments so we could test for transport of K^+ by MpAMT1;2 which was not observed.

3.1 Kinetic analysis of MpAMT1;2

Currents across the plasma membrane of oocytes expressing MpAMT1;2 were recorded at a range of membrane potentials from -137 mV to +38 mV. Representative currents from one oocyte bathed in K^+-free Ringer solution at pH 5.6 are shown in Fig. 15A. Currents were recorded first with no ammonium in the solution. Then the solution was exchanged to K^+-free Ringer solution containing 0.02 mM ammonium chloride and currents recorded again. Finally, the oocyte was recorded again in bathing solution without ammonium. Background currents before and after application of ammonium overlap, indicating a stable baseline. At all membrane potentials the addition of ammonium caused a downward deflection in currents compared to the baseline. Ammonium-induced currents were calculated by taking an average of the background currents before and after the ammonium addition and subtracting it from currents recorded in the presence of ammonium. Ammonium-induced currents were inward at all membrane potentials tested (Fig. 15B).

To study the affinity of MpAMT1;2 for ammonium, ammonium-induced currents were recorded at different ammonium concentrations. Currents from different oocytes were normalized to Vmax and those recorded at -137 mV are shown in Fig. 15C. In this paper, $K_{0.5}$ values are presented at different extracellular pH values and membrane potentials. The conditions used for Fig. 15C, a membrane potential of -137 mV and pH

5.6, are presented because, of the conditions used in this study, they most closely represent conditions found in plant cells. A $K_{0.5}$ value was calculated by fitting collected data to the Michaelis-Menten equation. The calculated $K_{0.5}$ for ammonium at pH 5.6 for MpAMT1;2 was 0.007 mM (7 µM) indicating that MpAMT1;2 has very high affinity.



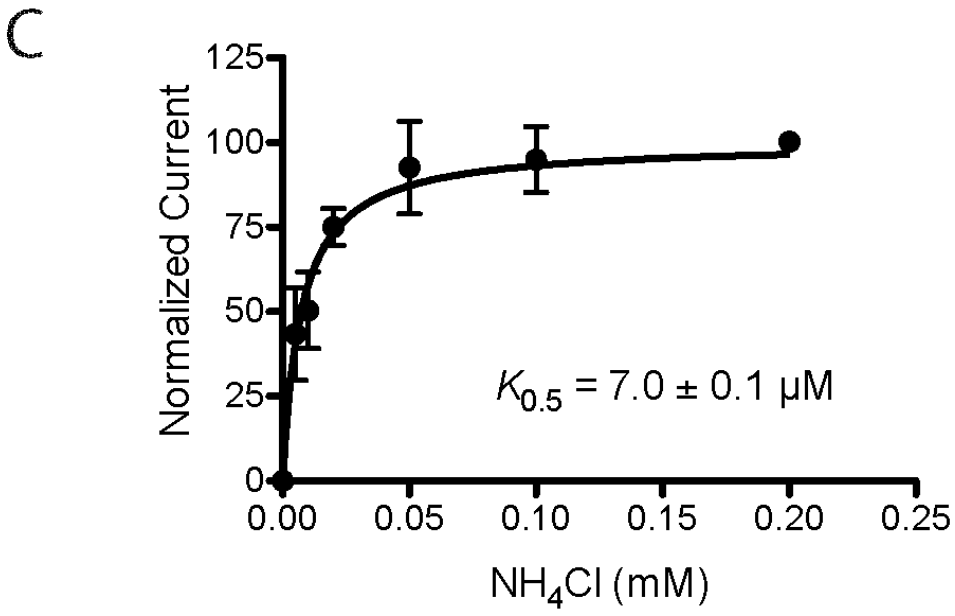
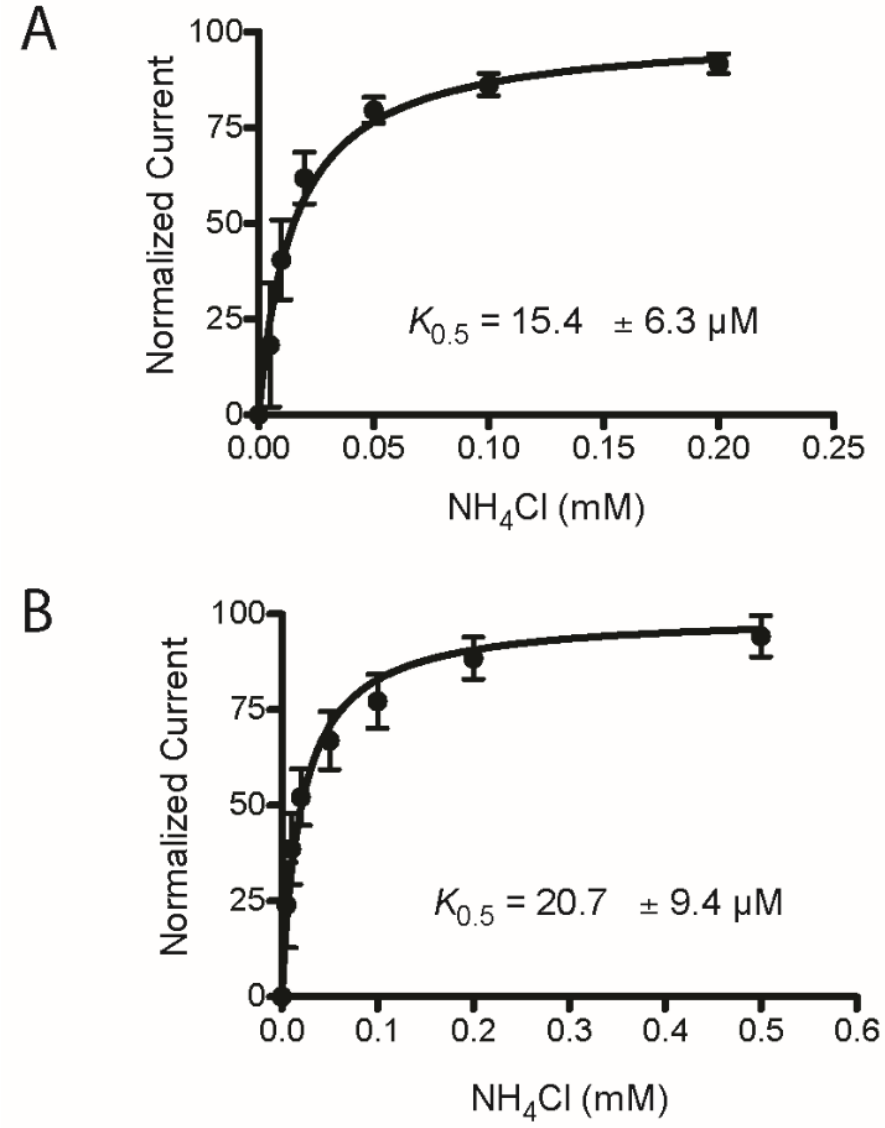


Figure 15. MpAMT1;2 is a high affinity ammonium transporter. Ammonium-induced currents were recorded using oocytes expressing MpAMT1;2. The bath solution was K+-free Ringer at pH 5.6. (A) Current-voltage relation. Background currents before and after application of 0.2 mM NH_4Cl overlap. Recording was done at membrane potentials from +38 to -137 mV. (B) Ammonium-induced (background subtracted) currents in response to different concentrations of NH_4Cl at a membrane potential of -137 mV. Currents for each oocyte were normalized to Vmax. Line indicates a fit of the Michaelis-Menten equation to the data. Data are presented as mean ± SD (n = 3 oocytes).

The effect of extracellular pH on the affinity of MpAMT1;2 for ammonium was tested. Normalized currents plotted against ammonium chloride concentrations at pH 6.8 (Fig. 16A) and pH 8.0 (Fig. 16B) are shown. The $K_{0.5}$ value at pH 6.8 was 15 µM and at pH 8.0 was 20 µM ammonium chloride, both were determined at -137 mV (Fig. 16A and 16B), so $K_{0.5}$ values increase as pH increases above pH 5.6. NH_4^+ and NH_3 are in pH-

dependent equilibrium with a p*K*a of 9.24. Plant AMT1 proteins are thought to bind NH_4^+ and transport NH_3 (Neuhauser and Ludewig 2014), therefore the $K_{0.5}$ for transport is related to the affinity of the transporter for NH_4^+ and the measured $K_{0.5}$ values can be corrected using the Henderson–Hasselbalch equation as shown in Fig. 15C. At pH 5.6 the correction is not significant since only 0.03% of total ammonium is present as NH_3 but at higher pH a greater proportion of ammonium is present as NH_3. At pH 8.0, 5.7% of total ammonium is present as NH_3 and the $K_{0.5}$ value at pH 8.0 was corrected from 20.7 to 19.5 (Fig. 16C) assuming that NH_4^+ binds to the transporter and NH_3 does not. To show the voltage dependence of $K_{0.5}$ for NH_4^+, corrected $K_{0.5}$ values measured at three pH values were plotted for membrane potentials between -40 and -137 mV. The $K_{0.5}$ for NH_4^+ was both pH-dependent and voltage-dependent; MpAMT1;2 showed higher affinity at more negative membrane potentials and at lower extracellular pH (Fig. 16D).


A
Normalized Current
100
75
50
25
0
$K_{0.5}$ = 15.4 ± 6.3 µM
0.00
0.05
0.10
0.15
0.20
0.25
NH_4Cl (mM)
B
Normalized Current
100
75
50
25
0
$K_{0.5}$ = 20.7 ± 9.4 µM
0.0
0.1
0.2
0.3
0.4
0.5
0.6
NH_4Cl (mM)

C

pH	$K_{0.5}$ $NH_4^+ + NH_3$ (μM)	$K_{0.5}$ NH_4^+ (μM)
5.6	7.3 ± 0.1	7.3 ± 0.1
6.8	15.4 ± 6.3	15.3 ± 6.3
8.0	20.7 ± 9.4	19.5 ± 8.9

D

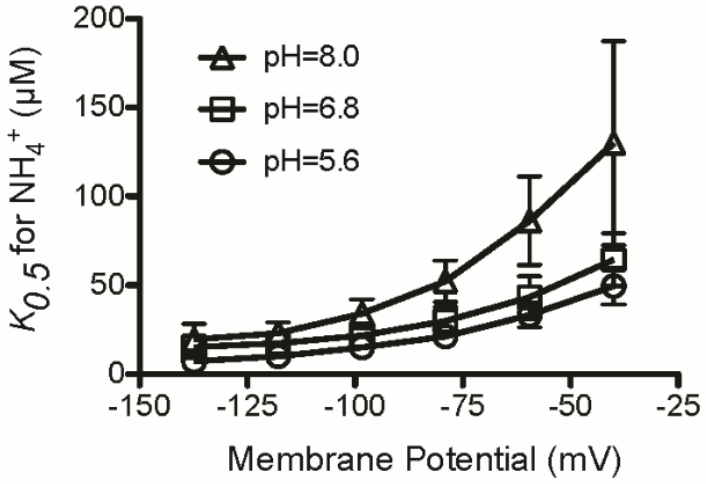
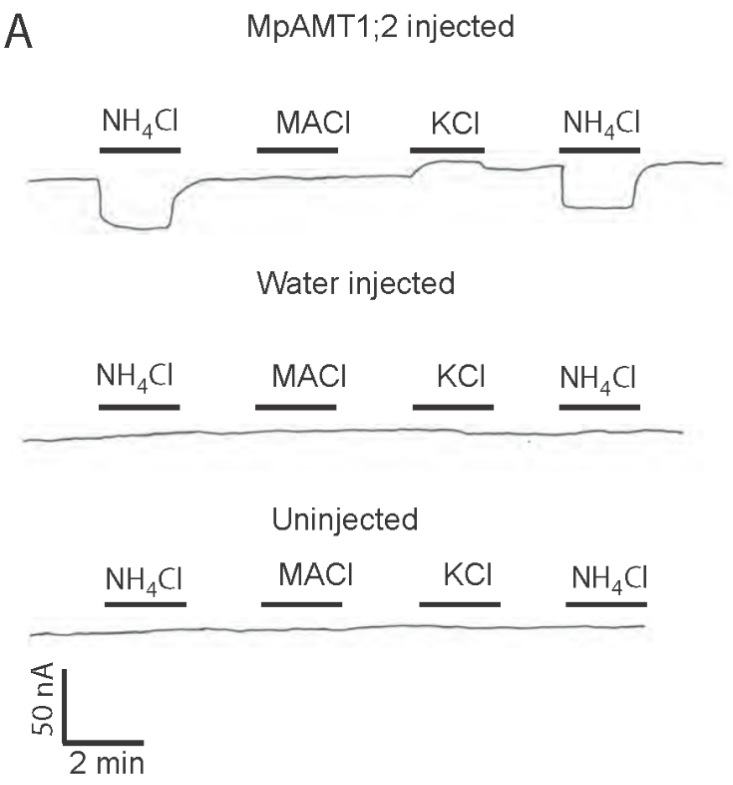


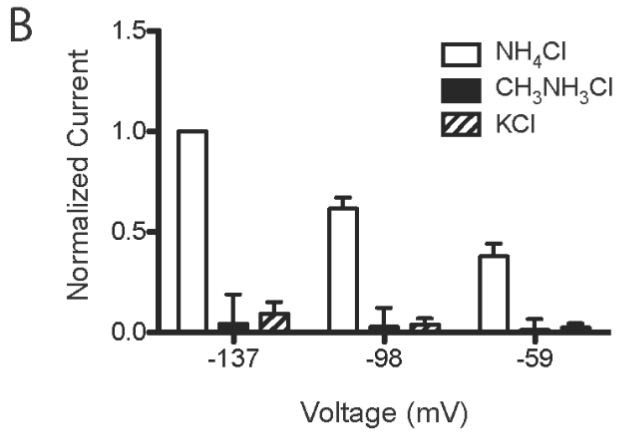
Figure 16. MpAMT1;2 affinity for ammonium is regulated by extracellular pH and membrane potential. MpAMT1;2 was expressed in *Xenopus* oocytes and ammonium-induced currents were recorded using two-electrode voltage clamping. (A) Kinetic analysis at pH 6.8. (B) Kinetic analysis at pH 8.0. In A and B currents were recorded at a membrane potential of -137 mV, normalized to Vmax, and plotted against the ammonium concentration as in Fig. 20B indicates a fit of the Michaelis-Menten equation to the data. Data are means of 4 oocytes at pH 6.8 and 3 oocytes at pH 8.0 and error bars are standard deviation. (C) Recalculation of K0.5 values. The Henderson-Hasselbalch equation was

used to calculate K0.5 values when the substrate is limited to NH4+ using the measured K0.5 values (for NH4+ + NH3) and a pKa value of 9.24. (D) Voltage dependence of K0.5 values of MpAMT1;2 for NH4+. Mean K0.5 values ± SD were recalculated as in Fig. 13C and plotted against membrane potentials from -137 mV to -40 mV.

3.2 Substrate specificity of MpAMT1;2

We tested the ability of MpAMT1;2 to transport methylammonium ($CH_3NH_3^+$) and potassium. AMT1 ammonium transporters from angiosperms are known to transport methylammonium as well as ammonium (De Michele et al., 2012). KCl was also tested because K^+ has a similar size as NH_4^+. Substrates were applied at 0.2 mM and the currents recorded at -40 mV showed that inward currents only appeared when ammonium chloride was applied (Fig. 17A). There was no obvious current when methylammonium chloride (MACl) or potassium chloride (KCl) was applied (Fig. 17A). Water-injected and uninjected oocytes showed no current changes for any of the substrate tested (Fig. 17A). Also at more negative membrane potentials, methylammonium or potassium applied at 0.2 mM did not induce significant currents (Fig. 17B). Experiments using higher concentrations showed that MpAMT1;2 could transport methylammonium. A $K_{0.5}$ of 4.04 mM for methylammonium was measured at pH 5.6 and a membrane potential of -137 mV (Fig. 17C), which is 577-fold higher than the $K_{0.5}$ for ammonium (7 µM) measured under the same conditions.


A
MpAMT1;2 injected
NH4Cl
MACl
KCl
NH4Cl
Water injected
NH4Cl
MACl
KCl
NH4Cl
Uninjected
NH4Cl
MACl
KCl
NH4Cl
50 nA
2 min


B
Normalized Current
1.5
1.0
0.5
0.0
NH4Cl
CH3NH3Cl
KCl
-137
-98
-59
Voltage (mV)

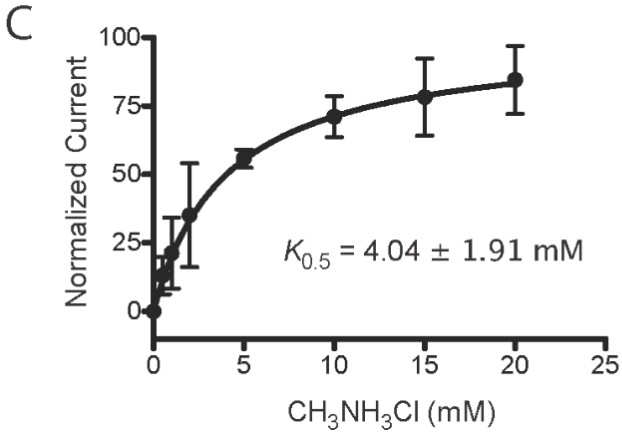

C
Normalized Current
100
75
50
25
0
K0.5 = 4.04 ± 1.91 mM
0
5
10
15
20
25
CH3NH3Cl (mM)

Figure 17. MpAMT1;2 is selective for ammonium over methylammonium and potassium. (A) Currents recorded at pH 5.6 in K^+-free Ringer solution at -40 mV. Substrates were applied at 0.2 mM. MACl is methylammonium chloride (CH_3NH_3Cl). Upper recording was from an oocyte expressing MpAMT1;2. Middle and lower recordings were from water-injected and uninjected control oocytes, respectively. (B) Substrate-dependent currents with 0.2 mM NH_4Cl were recorded at different membrane potentials and normalized to currents recorded at -137 mV. The average current with 0.2 mM NH_4Cl at -137 mV was -0.087 ± 0.053 μA. Mean currents ± SD are presented ($n \geq 4$). (C) K0.5 value for methylammonium chloride. Normalized currents were recorded at -137 mV and pH 5.6. Line indicates a fit of the Michaelis-Menten equation to the data. Data represents mean ± SD (n=3 oocytes).

3.3 Expression and membrane localization of MpAMT1;2

A timeline for experiments designed to analyze expression of *MpAMP1;2*, Takaragaike-1 (Tak-1) is shown in Fig. 18A. Gemmae were grown on half-strength Gamborgs B5 media for two weeks and then grown on the same media without N for 5 days to induced N deficiency. Then they were transferred to media with different sources of N at 5 mM total N or media lacking N. RNA was extracted after 2 days and quantitative RT-PCR was performed. As shown in Fig. 18B, the highest expression of *MpAMP1;2* was found when plants were grown in the absence of N in the media. N supplied as 2.5 mM $(NH_4)_2SO_4$ or 2.5 mM NH_4NO_3 suppressed *MpAMP1;2* expression to the lowest levels (Fig. 18B). 5 mM KNO_3 also suppressed expression of *MpAMP1;2* but not as strongly. N supplied as 2.5 mM glutamine suppressed *MpAMP1;2* expression to an intermediate level (Fig. 18B). The results indicate that *MpAMP1;2* showed the highest

expression in Marchantia thalli when the plants experienced N deficiency. All of the N sources tested were able to suppress expression indicating that repression of *MpAMT1;2* expression is not specific to ammonium.

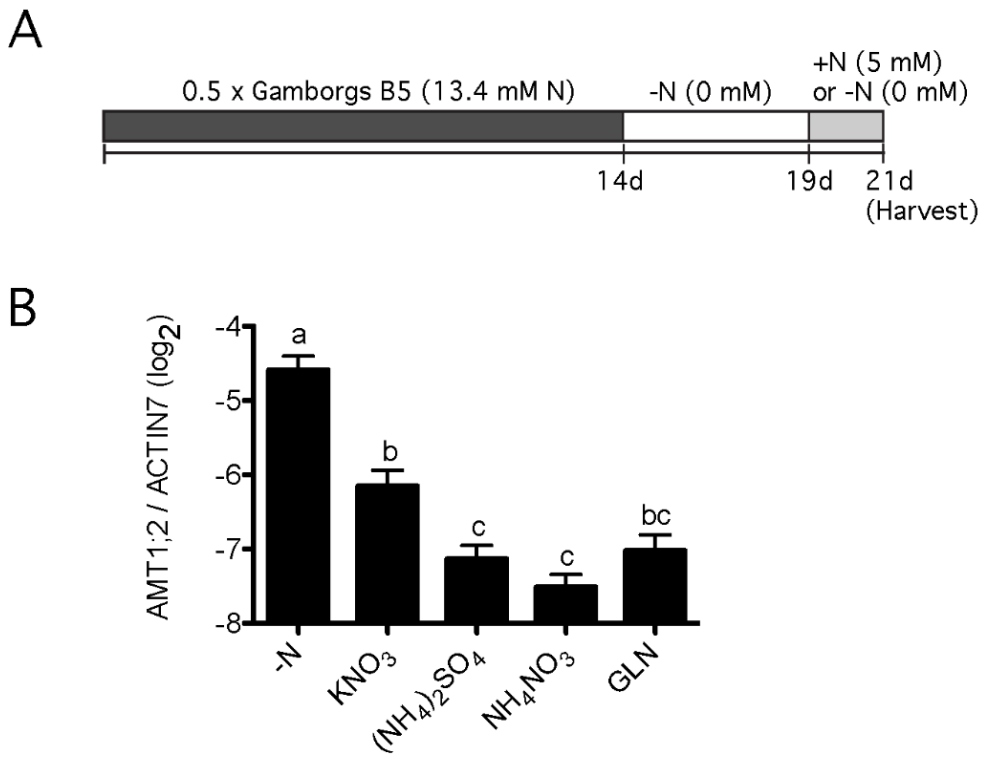


Figure 18. MpAMT1;2 expression is negatively regulated by nitrogen concentration. Expression of MpAMT1;2 in Tak-1 grown on different nitrogen sources. (A) Time course for expression study. Plants were grown for two weeks on half-strength Gamborgs B5, then transferred to the same media without N with the addition of 5 mM KCl for 5 days, then transferred to media containing no N or different N sources at 5 mM N final concentration (see Materials and methods). (B) Expression was measured using qRT-PCR two days after resupply of nitrogen and normalized to the expression of MpACTIN7. The data shown are the mean log2 ratios to MpACTIN7 ± SEM of three

biological replicates. Different letters above the bars indicate significant differences with $p>0.05$.

A C-terminal MpAMT1;2-citrine fusion was used to study the localization of MpAMT1;2 in *Marchantia polymorpha* gametophytes (Fig. 19). Fluorescence appeared peripheral (Fig. 19 top panel) and when overlaid with chlorophyll fluorescence it was clear that citrine fluorescence was external to chloroplast fluorescence indicating a plasma membrane localization for MpAMT1;2. An N-terminal citrine fusion to MpSUT4, a sucrose transporter in the vacuolar clade (Reinders et al., 2008; Weise, 2000) from *Marchantia* was used as a comparison to observe vacuolar localization (Fig. 19, right side). When citrine-MpSUT4 fluorescence was overlaid with chlorophyll fluorescence, it was evident that fluorescence was towards the inside of cells relative to chloroplasts, indicating a vacuolar localization for MpSUT4. The controls (Fig. 19, lower panels) were untransformed *Marchantia* with background citrine fluorescence overlaid with chlorophyll fluorescence using the same microscope and camera settings. The results indicate that MpAMT1;2 specifically localizes to the plasma membrane in the *Marchantia* gametophyte thallus.

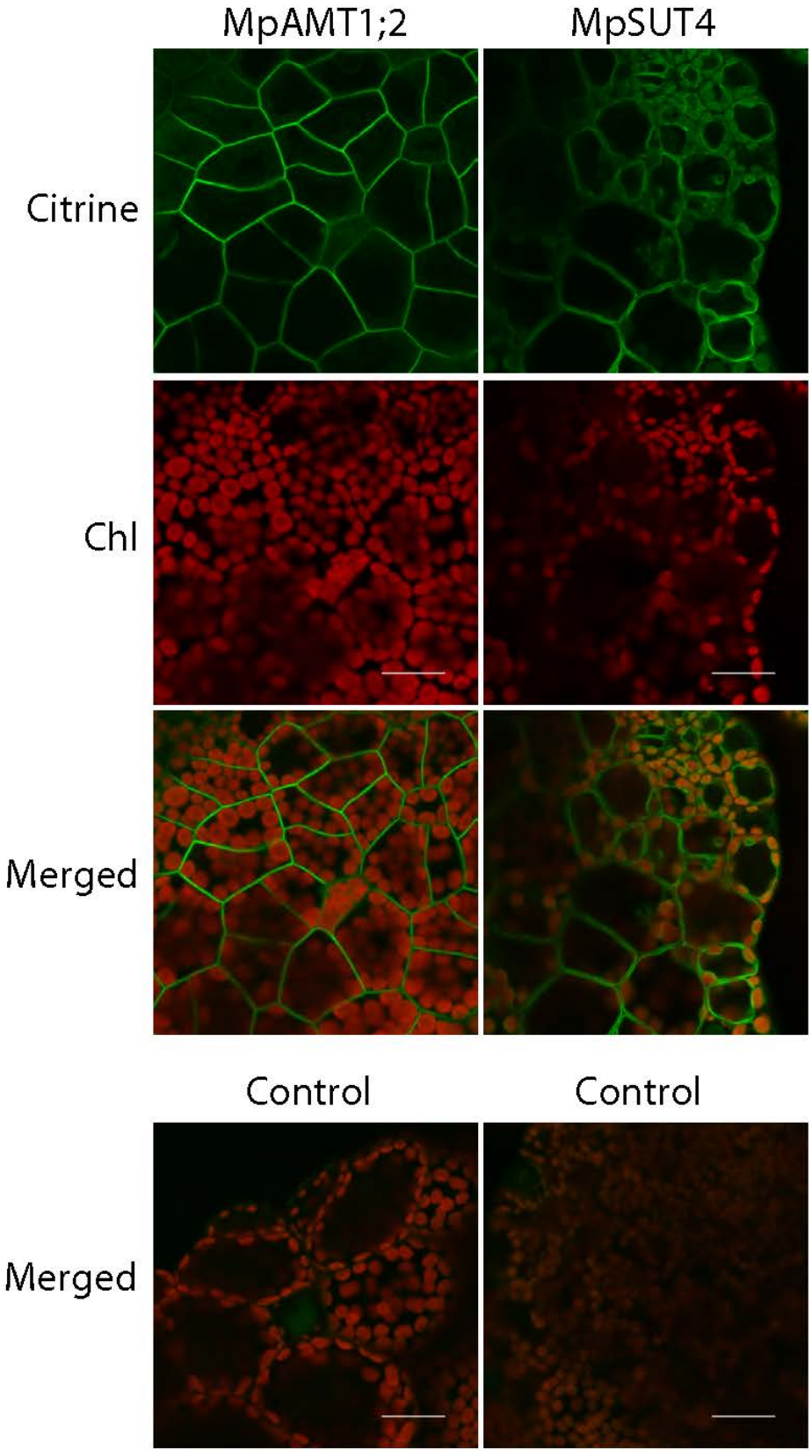

MpAMT1;2
MpSUT4
Citrine
Chl
Merged
Control
Control
Merged

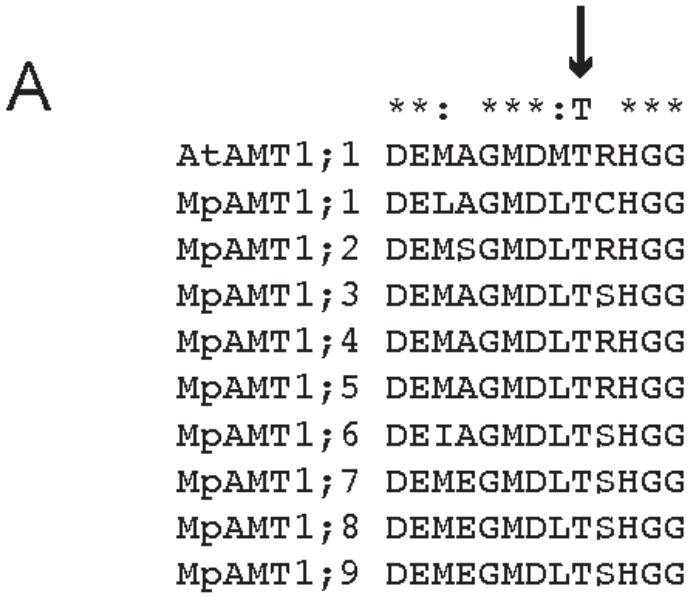
Figure 19. MpAMT1;2 localizes to the plasma membrane in *Marchantia polymorpha* gametophytes. Confocal images of the membrane localization of either the ammonium transporter MpAMT1;2 fused to citrine at the C-terminus or the sucrose transporter MpSUT4 fused to citrine at the N-terminus under control of the 35S promoter in Tak-1 Marchantia polymorpha gemmae. Optical sections were collected at the appropriate wavelengths for citrine (488 nm/ 525 nm) and autofluorescence of chlorophyll (chl, 405 nm/ 595 nm). Merged images show an overlay of both citrine and chlorophyll fluorescence. Controls were collected from untransformed Tak-1 gemmae at the same settings as the corresponding MpAMT1;2 or MpSUT4 images, only the merged image is shown for the controls. Bars shown are 24 μm.

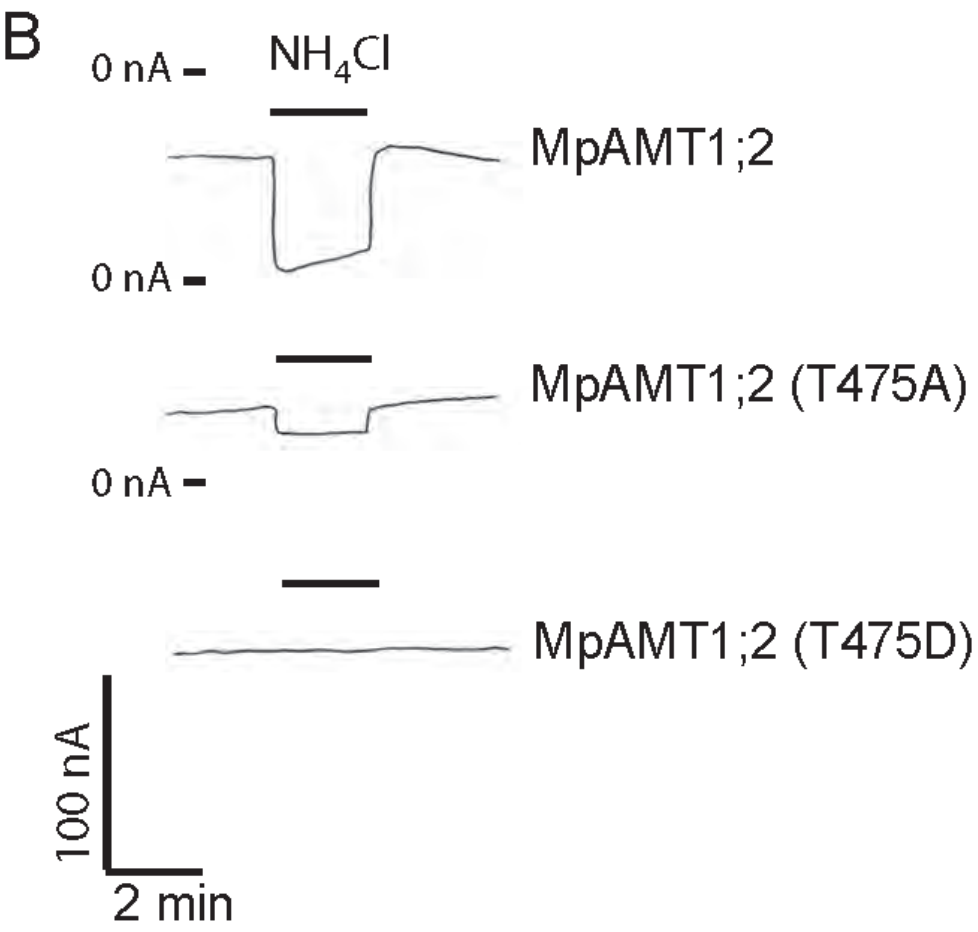
3.4 MpAMT1;2 regulatory phosphorylation site

Angiosperm AMT1 ammonium transporters contain a conserved inhibitory phosphorylation site (Loqué et al., 2007; Neuhauser et al., 2007). Extracellular ammonium induces rapid phosphorylation of this site (Lanquar et al., 2009) which is T460 in AtAMT1;1 leading to inhibition of transport activity. All nine AMT1s in Marchantia contain this phosphorylation site (Fig. 20A). In MpAMT1;2 the corresponding position is T475 indicted by the arrow in Fig. 24A. To test whether phosphorylation of this site is inhibitory, we compared the transport activity of wild type MpAMT1;2 with a phosphomimetic mutant T475D and a non-phosphorylatable mutant T745A. When currents were recorded at -40 mV, ammonium (20 μM) application to *Xenopus* oocytes expressing MpAMT1;2 resulted in inward current (Fig. 20B). Under the same conditions, oocytes expressing MpAMT1;2 (T475A) showed smaller inward currents and oocytes expressing showed no detectable currents (Fig. 20B). Replicated

experiments done at a membrane potential of -121 mV show that MpAMT1;2 (T475A) produced significantly less current compared to wild type and that MpAMT1;2 (T465D) showed no response to applied ammonium (Fig. 20C). The results indicate that phosphorylation of conserved T475 in MpAMT1;2 inhibits transport activity as it does in Arabidopsis AtAMT1;1 (Loqué et al., 2007; Neuhauser et al., 2007).

A




B
0 nA
NH4Cl
MpAMT1;2
0 nA
MpAMT1;2 (T475A)
0 nA
MpAMT1;2 (T475D)
100 nA
2 min

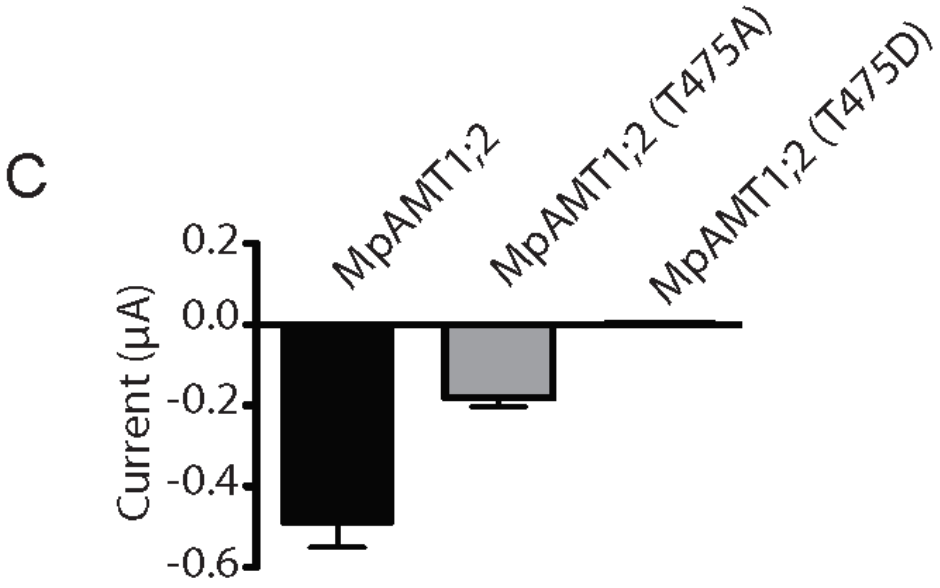

C
MpAMT1;2
MpAMT1;2 (T475A)
MpAMT1;2 (T475D)
Current (µA)
0.2
0.0
-0.2
-0.4
-0.6

Figure 20. Phosphomimetic mutation of MpAMT1;2 inhibits transport activity. (A) A partial multiple alignment of protein sequences from AtAMT1;1 and nine ammonium transporters from Marchantia is shown. The regulatory phosphorylation site of AtAMT1;1 (T460) is conserved in all AMT1 transporters from Marchantia polymorpha. The corresponding site in MpAMT1;2 is T475. An arrow indicates the threonine (T) residues conserved with T460 from AtAMT1;1. (B) MpAMT1;2, (T475D), and MpAMT1;2 (T475A) were expressed in Xenopus oocytes and recording was done using two-electrode voltage clamping. The oocytes were bathed in K+-free ringer solution at pH 5.6, voltage clamped at -40 mV and NH4Cl was applied at 20 µM (three times higher than the K0.5 of MpAMT1;2). Downward deflections during substrate application represent inward currents. (C) Ammonium-induced (background subtracted) currents in response to 20 µM NH4Cl at a membrane potential of -121 mV. Data are presented as mean ± SD (n = 3 oocytes).

4. Discussion

Marchantia encodes 9 AMT1s and 10 AMT2s (MEPs) (Bowman et al., 2017). The only other completely sequenced bryophyte genome is from *Physcomitrella patens* which encodes a similar number of ammonium transporters (6 AMT1s and 7 MEPs) (De Michele et al., 2012). Some angiosperms encode a smaller number of ammonium transporters, for example Arabidopsis has 5 AMTs and 1 MEP while rice has 3 AMTs and 6 MEPs. The angiosperm *Populus tricocarpa* encodes 6 AMT1 and 17 MEPS and the large number of ammonium transporters may be related to its ability to obtain N through mycorrhizal interactions. A recent study found that expression of 5 *MEP* and 1 *AMT1* genes are induced in *Populus tricocarpa* during mycorrhizal interaction with

Rhizophagus irregularis (Calabrese et al., 2017). This suggests that ammonium transporters have a role in ammonium uptake from the fungus but it is not known if any of the ammonium transporters localize to the periarbuscular membrane. The large number of ammonium transporter genes in bryophytes may also be related to nitrogen uptake through symbiotic interaction. Bryophytes, including *Marchantia* (Adams and Duggan, 2008), are known to interact with nitrogen-fixing cyanobacteria (Adams, 2002). Another possibility is that since *Marchantia polymorpha* lacks vascular tissue and plasmodesmata have not been observed, cell-to-cell transfer of nitrogen may be more dependent on transmembrane transport thus requiring more ammonium transporters.

4.1 MpAMT1;2 is electrogenic and has a high affinity for ammonium

The electrophysiological analysis of MpAMT1;2 showed that ammonium induced an inward current in MpAMT1;2-expressing oocytes. This result indicates that MpAMT1;2 is electrogenic, it transports a charge across the membrane, which is consistent with a previous report (McDonald and Ward, 2016). The mechanism of ammonium transport for the AMT/MEP/Rh superfamily remained unclear until the crystal structures for transporters from bacteria and archaebacteria were obtained (Javelle et al., 2005; Khademi Shahram, O'Connell III Joseph, Remin Jonathan, Robles-Colmenares Yaneth, Miercke Larry J.W., 2004; Zheng et al., 2004).The structures indicated that the pore region of ammonium transporters is hydrophobic, suggesting that NH_3 is translocated rather than NH_4^+. Therefore, electrogenic ammonium transporters such as MpAMT1;2 likely cotransport H^+ and NH_3 rather than function as NH_4^+ uniporters. Analysis of mutants of AtAMT1;2, an electrogenic ammonium transporter from Arabidopsis is consistent with a H^+/NH_3 cotransport mechanism (Neuhäuser and Ludewig, 2014).

Kinetic analysis revealed that MpAMT1;2 is a high-affinity ammonium transporter with a $K_{0.5}$ of 7 μM. High-affinity ammonium transporters have also been characterized in angiosperms. For example, Arabidopsis AtAMT1;1 has a $K_{0.5}$ of 2.7 μM (Mayer and Ludewig, 2006), ZmAMT1;3 from maize has a $K_{0.5}$ of 33 μM (Gu et al., 2013) and PvAMT1;1 from bean has a $K_{0.5}$ of 28 μM (Ortiz-Ramirez et al., 2011b). The affinity of MpAMT1;2 for ammonium was pH-dependent, lower pH values produced a lower $K_{0.5}$, indicating a higher substrate affinity (Fig. 16). This is the opposite pH regulation compared to PvAMT1;1 from bean, for which $K_{0.5}$ decreases as pH increases (Ortiz-Ramirez et al., 2011b). More negative membrane potential also leads to a decrease in $K_{0.5}$ values, which represents an increase in affinity for ammonium (Fig. 16C). Therefore, MpAMT1;2 is regulated by both extracellular pH and membrane potential, both of which are controlled in plants by plasma membrane H^+-ATPases which are highly regulated enzymes. Activation of H^+-ATPases makes the extracellular pH more acidic and makes the membrane potential more negative, both of which cause MpAMT1;2 to have a higher affinity for ammonium.

4.2 MpAMT1;2 substrate specificity

MpAMT1;2 is a highly selective ammonium transporter (Fig. 17A). At 0.2 mM concentrations, ammonium chloride induced inward current while methylammonium chloride and potassium chloride did not (Fig. 17A). Experiments using higher concentrations of methylammonium showed that MpAMT1;2 can transport methylammonium (Fig. 17C). The $K_{0.5}$ for methylammonium was determined to be 4.04 mM at pH 5.6 and -137 mV. Based on this analysis, MpAMT1;2 has approximately a 577-fold higher affinity for ammonium compared to methylammonium.

NH_4^+ and NH_3 are in equilibrium in solution so it is reasonable to ask whether NH_4^+ or NH_3 is transported. Since the pK_a for NH_4^+ is 9.24, very little NH_3 is present in the typically acidic extracellular space of plant cells. For example at the $K_{0.5}$ of MpAMT1;2 of 7 µM at pH 5.6 the concentration of NH_3 is approximately 0.00157 µM. This low concentration of NH_3 is unlikely to support the observed high rate of transport, therefore it is most likely that ammonium transporters bind NH_4^+. According to predominant models, transporters in the AMT/MEP/Rh superfamily bind NH_4^+, deprotonate it, and transport NH_3. The high substrate specificity of ammonium transporters has been used to support the idea that deprotonation of NH_4^+ and transport of NH_3 is part of the transport mechanism (Neuhäuser and Ludewig, 2014). It is a reasonable conclusion since cations of similar size as NH_4^+, such as K^+, are not transported yet a deprotonatable substrate that is larger, methylammonium ($CH_3NH_3^+$), is transported. This model also has support based on the hydrophobicity of the pore of AmtB and RhCG (Gruswitz et al., 2010; Khademi Shahram, O'Connell III Joseph, Remin Jonathan, Robles-Colmenares Yaneth, Miercke Larry J.W., 2004).

4.3 Expression and membrane localization of MpAMT1 in *Marchantia polymorpha*

The expression of *MpAMT1;2* in *Marchantia polymorpha* was studied using quantitative RT-PCR and plants grown under N deficient conditions showed higher expression. All sources of N including ammonium, nitrate or the amino acid glutamine, when added to media at 5 mM N concentration, caused a decrease in *MpAMT1;2* expression. This indicates that MpAMT1;2 expression is repressed under N sufficiency. This is similar to several *AMT* genes from Arabidopsis that are induced under N

deficiency (Gazzarrini et al., 1999; C Sohlenkamp et al., 2000) and opposite from the expression of *LeAMT1* from tomato (Lauter et al., 1996).

AMTs are also regulated at the protein level. AMTs are trimeric and transactivated: the C-terminus of each subunit binds and activates a neighboring subunit (Loqué et al., 2007). This transactivation is controlled by protein phosphorylation in response to the presence of extracellular ammonium (Lanquar et al., 2009). Phosphomimetic replacement of a conserved T in the C-terminus (T460 in AtAMT1;1) transinactivates AMT complexes (Loqué et al., 2007; McDonald and Ward, 2016; Neuhauser et al., 2007). In Arabidopsis, AtAMT1;1 and AtAMT1;2 were shown to be phosphorylated and inactivated by the kinase CIPK23 (Calabrese et al., 2017). MpAMT1;2 has the conserved phosphorylation site in the C-terminus (Fig. 20A) at T475 and as shown here, phosphomimetic mutation of MpAMT1;2 (T475D) resulted in complete loss of ammonium transport activity (Fig. 20B and C). This regulatory mechanism is known to have evolved prior to plants since ammonium transporter AfAmt-2 from the archaea *Archaeoglobus fulgidus* is regulated by the same mechanism (Loqué et al., 2009). Our results suggest that inhibition of AMT transporters in plants via phosphorylation has been maintained through evolution from basal plants such as Marchantia through angiosperms.

MpAMT1;2 was localized to the plasma membrane in *Marchantia polymorpha* gametophyte thallus using a C-terminal citrine fusion. This is consistent with observed localizations of AMTs in angiosperms. In Arabidopsis, AtAMT1;1 was localized to the plasma membrane in root hairs and other parts of the root and shoot (Mayer and Ludewig, 2006). AtAMT1;4 has been localized to the plasma membrane in pollen (Yuan et al., 2009).

5. Conclusions

MpAMT1;2 is the first ammonium transporter from a bryophyte that has well-characterized transport activity. It is electrogenic and has high affinity and high selectivity for ammonium. The plasma membrane localization of MpAMT1;2 indicates that it functions in cellular uptake of ammonium into the cytoplasm. *MpAMT1;2* showed higher expression when plants were grown under N deficiency. Several different N sources, including ammonium, caused a decrease in *MpAMT1;2* expression indicating that general N sufficiency suppressed expression of the ammonium transporter. The presence of a conserved phosphorylation site in the C-terminus indicates that transport activity of the MpAMT1;2 protein may be down-regulated under high ammonium concentrations similar to angiosperm ammonium transporters (Loqué et al., 2007; Neuhauser et al., 2007).

Acknowledgements

We thank Drs. Ryuichi Nishihama and Takayuki Kohchi (Kyoto University) for providing Marchantia genome sequence prior to publication and for sharing the binary vectors as well as the Tak-1 strain used in this study. We also thank Drs. Takayuki Kohchi, Takeshi Araki, Ryuichi Nishihama and members of their labs for sharing protocols and providing invaluable advice on working with Marchantia. We acknowledge Grant Bartel at the University Imaging Centers at the University of Minnesota for help with confocal imaging. This work was supported by a grant from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy (DE-FG02-10ER15886) to J.W.

Chapter 5

MpAMT1;5 from *Marchantia polymorpha* is a low-affinity ammonium transporter in the plasma membrane

Nu Wang[1], Tami R. McDonald[2], Anke Reinders[1] and John M. Ward[1]

[1]Department of Plant and Microbial Biology, University of Minnesota, USA

[2] Biology Department, St. Catherine University, St. Paul, MN, USA

Abstract

Basal plants contain a small gene family of ammonium transporters within the AMT/MEP/Rh superfamily and transporters in the same family are used to take up ammonium in angiosperms. Here we characterized the transport activity of ammonium transporter MpAMT1;5 from *Marchantia polymorpha*, a liverwort and a representative of the most basal land plants. MpAMT1;5 was functionally expressed in both yeast and *Xenopus* oocytes. It showed a low-affinity for ammonium ($K_{0.5}$ of 0.38 mM at pH 5.6 and a membrane potential of -121 mV) when expressed in *Xenopus* oocytes and assayed using two-electrode voltage clamping. MpAMT1;5 was localized to the plasma membrane in gametophyte thalli of transgenic *Marchantia polymorpha* expressing a MpAMT1;5-citrine fusion. The results indicate that MpAMT1;5 could function in ammonium uptake under conditions where ammonium is abundant.

Keywords: ammonium transporter, bryophyte, electrophysiology, liverwort, *Marchantia polymorpha*, Xenopus oocyte expression

1. Introduction

Nitrogen (N) is a plant macronutrient and N deficiency often limits plant growth in natural and agricultural environments. Ammonium, along with nitrate, is a main form of inorganic N taken up by plants from the soil. Ammonium is in pH-dependent equilibrium with ammonia (pKa = 9.25); in this paper we refer to the total ammonium and ammonia as "ammonium" and use the chemical formulas NH_4^+ and NH_3 to refer to individual chemical species.

In plants, ammonium transporters (AMTs) play an essential role in taking up ammonium from the soil (Sohlenkamp et al., 2000; Ludewig et al., 2002; Ortiz-Ramirez et al., 2011; Yuan et al., 2013; Yang et al., 2015). Ammonium transporter genes were first identified in yeast (Marini et al., 1994) and plants (Ninnemann et al., 1994). Homologous genes were identified in all other domains of life: *AMT*s are found in plants, prokaryotes and animals (excluding vertebrates), Rh (Rhesus factor) are found in vertebrates, and MEP (Methylammonia Permease) in prokaryotes, fungi and plants. In plants, there are two subfamilies of ammonium transporters (AMT1 and AMT2) that differ in transport function. AMT1 transporters are electrogenic, they transport a charge across the membrane, while AMT2 transporters are MEPs, they are electroneutral and acquired by plants via horizontal gene transfer (McDonald et al., 2012).

Marchantia polymorpha is a liverwort and has been used as a model organism that represents the earliest land plants (Berger et al., 2016; Bowman et al., 2017). Since *AMT1* and *AMT2* genes are present in both liverworts and angiosperms, detailed characterization of ammonium transporter function from the basal plant provides more information for structure/function analysis in this family and informs us about evolution of ammonium transporters in plants. To date, most of what we know about ammonium transporters in

plants comes from the study of AMTs in angiosperms. Limited information is available concerning ammonium transporter function in bryophytes. Four AMT1s from Marchantia were demonstrated to be electrogenic (McDonald and Ward, 2016), that is consistent with the function of AMT1s in angiosperms. MpAMT1;2, has a very high affinity for ammonium, is localized to the plasma membrane, and shows higher expression under N-deficiency conditions (Guo et al., 2018).

Marchantia has 19 ammonium transporter genes, nine *AMT1* and ten *AMT2* genes (McDonald and Ward, 2016). The number of ammonium transporter genes in angiosperms varies widely by species. For example, Arabidopsis has six ammonium transporter genes (five *AMT1* and one *AMT2*) while *Populus trichocarpa* has twenty-three (eight *AMT1* and fifteen *AMT2*). Fig. 21, shows a phylogenetic tree with ammonium transporters from Marchantia and Arabidopsis. The AMT1 and AMT2 transporters formed distinct clades. We know more about the physiological function of transporters within the AMT1 clade. AMT1 transporters have been characterized in both eudicots (Gazzarrini et al., 1999; Ludewig et al., 2002; Mayer and Ludewig, 2006; Yuan et al., 2009; Ortiz-Ramirez et al., 2011; Scherzer et al., 2013) and monocots (Gu et al., 2013; Yang et al., 2015; Hao et al., 2016), they are localized to the plasma membrane and show a wide range of affinities for ammonium. Outside of angiosperms, the only plant ammonium transporter has been characterized in detail; MpAMT1;2 from Marchantia is electrogenic, localized to the plasma membrane, and has a very high affinity for ammonium (Guo et al., 2018). In this paper, we studied the transport function and membrane localization of another ammonium transporter from Marchantia, MpAMT1;5,

and found that it is also localized to the plasma membrane but that it has a low affinity for ammonium.

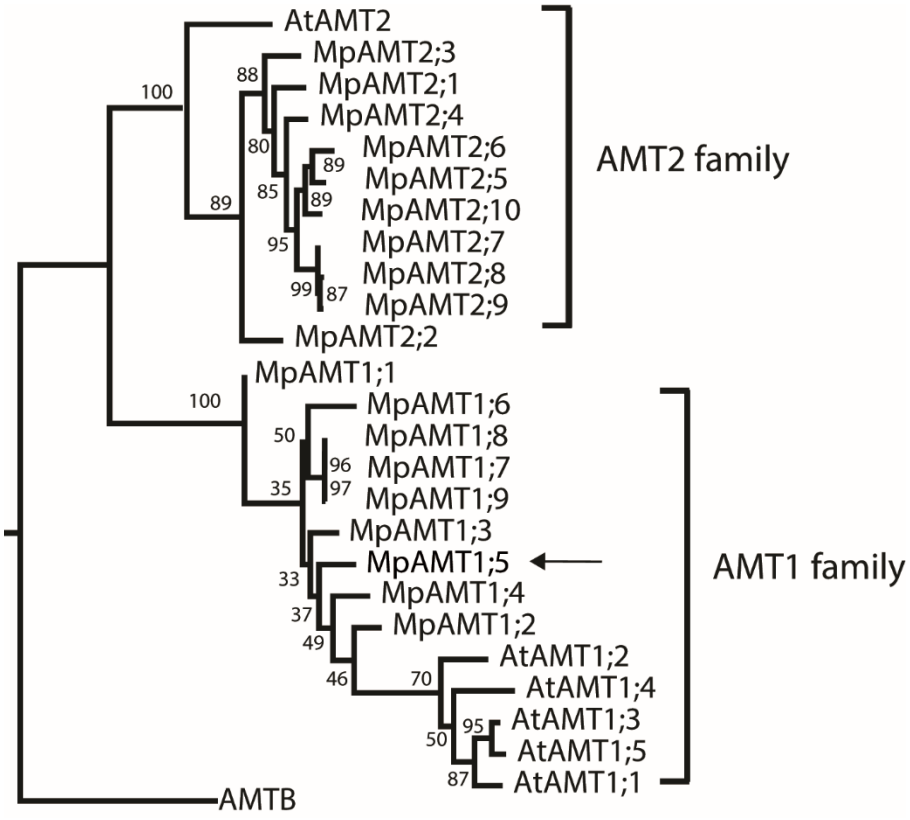


Figure 21. Phylogenetic tree of ammonium transporters from *M. polymorpha* and *A. thaliana*. Protein sequences were aligned using Clustal X and a Maximum Parsimony was made using RAxML. TAIR accession numbers of *A. thaliana* ammonium transporters and ID numbers of M. polymorpha (annotation- version 3.1) are shown. The tree was rooted using amtB from *E. coli* (NCBI accession number NP_414985). The ID number of *M. polymorpha* (annot- version=3.1) ammonium transporters MpAMT 1;1–9 and MpAMT 2;1–10 are Mapoly0023s0005.1, Mapoly0060s0090.1, Mapoly0081s0032.1, Mapoly0081s0029.1, Mapoly0081s0031.1, Mapoly0081s0034.1,

Mapoly0077s0030.1, Mapoly0077s0028.1, Mapoly0077s0031.1, Mapoly0028s0091.1, Mapoly0080s0061.1, Mapoly0174s0008.1, Mapoly0192s0010.1, Mapoly0192s0011.1, Mapoly0192s0012.1, Mapoly0192s0015.1, Mapoly0213s0002.1, Mapoly0213s0001.1, Mapoly0192s0013.1. The NCBI accession number of AMTB is NP_414985. And add NCBI numbers for MpAMT1;2 - 1;5

2. Materials and Methods

2.1 *Marchantia polymorpha* culture conditions. Male *Marchantia polymorpha* Takaragaike-1 (Tak-1) was obtained from Dr. Takayuki Kohchi lab in Kyoto, Japan. Plants were grown on ½ B5 medium (87.5 mg NaH_2PO_4-$2H_20$, 1.25 g KNO_3, 67 mg $(NH_4)_2SO_4$, 125 mg $MgSO_4$-$7H_2O$, 75 mg $CaCl_2$-$2H_2O$, 20 mg EDTA-NaFe(III), 0.5 mL Gamborg's B5 micro-elements, 0.5 mL 0.075% KI solution, 0.5 g MES, 10 g agar, water to 1 L, pH 5.5). Plants were grown under a 16h light and 8h dark cycle at 22 °C.

2.2 Cloning *MpAMT1.5* from Tak-1 and heterologous expression in oocytes. RNA was isolated from Tak-1 thalli using the RNeasy kit (Qiagen) and cDNA was made using the Omniscript reverse transcription kit (Qiagen). The *MpAMT1;5* coding region was cloned by PCR from Tak-1 cDNA. PCR products were cloned into the entry vector pCR8/GW/TOPO (Invitrogen) and sequenced, then recombined with the oocyte expression vector pOO2/GW.

MpAMT1.5 cRNA was synthesized by using the SP6 mMessage mMachine kit (Ambion) after the construct pOO2/GW- *MpAMT1.5* was linearized using MluI. Oocytes were injected with 50 ng of *MpAMT1;5* cRNA per oocyte. After injection, oocytes were incubated in Barths solution (88 mM NaCl, 1 mM KCl, 0.33 mM $Ca(NO_3)_2$, 0.41 mM

$CaCl_2$, 0.82 mM $MgSO_4$, 2.4 mM $NaHCO_3$, 10 mM HEPES, pH 7.6, 100 µg ml^{-1} penicillin, and 100 µg ml^{-1} streptomycin) with 10 µg ml^{-1} gentamycin for 3–4 d at 15°C.

2.3 Electrophysiology. Electrophysiology data were collected by two electrode voltage clamping (TEVC) using a Dagan TEVC 200 amplifier (Dagan Corp.). Glass pipettes (Warner Instruments Corp.) used to impale oocytes were filled with 1 M KCl. *MpAMT1.5*-injected oocytes were perfused in K^+-free Ringer solution (115 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 5 mM MES, pH 5.6 with NaOH). The holding potential was -40 mV. Electrophysiology data were analyzed using Clampex (Axon Instruments Inc.).

2.4 Promoter of *MpAMT1.5* with GUS vector construction. Tak-1 *M. polymorpha* genomic DNA was extracted using the CTAB method (CTAB extraction buffer: 0.7 M NaCl, 50mM Tris, pH 8, 10mM EDTA, 1% CTAB), and used as template to amplify 4996 bp 5' of the coding sequence of the gene *MpAMT1.5*. The forward primer was 5'-CACCATACATATGAGGTGGGATGCAGGAT-3', and the reverse primer was 5'-GCATCCTGCCATGTCTACCTTTGAAT-3'. The PCR product was cloned into the entry vector PENTR-D-TOPO (Invitrogen). After sequencing, the fragment was cloned into the destination Gateway vector No pro/C pMpGWB104 containing the GUS reporter gene (Ishizaki et al., 2015).

2.5 Subcellular localization of MpAMT1;5. The stop codon in pCR8/GW/TOPO-*MpAMT1.5* was changed from TGA to TCA using site-directed mutagenesis. The new construct was then recombined with destination vector pMpGWB106 (Ishizaki et al., 2015) to yield *MpAMT1;5* with a C-terminal fusion to citrine with expression driven by dual 35S promoters (35s-35S-*MpAMT1.5*-citrine). This construct was used to transform

agrobacterium (C58C1) by electroporation which was grown in LB containing gentamicin.

For plant transformation, Tak-1 gemmae were grown on ½ B5 medium under 24 h continuous white light for 2 weeks until a "butterfly" shape with 4 basal fragments was developed. Then the apical parts of these basal fragments were cut, the remaining parts harvested, and cultured on ½ B5 medium containing 1% sucrose for 3 days under 24 h continuous white light. The harvested thallus and the agrobacterium were co-cultured in 0M51C (0.275 g KH_2PO_4, 2 g KNO_3, 0.4 g NH_4NO_3, 0.37 g $MgSO_4$-$7H_2O$, 0.3 g $CaCl_2$-$2H_2O$, 0.04 g EDTA-NaFe(III), 1 mL Gamborg's B5 micro-elements, 1 mL Gamborg's B5 vitamin mix, 1 mL 0.075% KI solution, 20 g sucrose, 0.5 g MES, 1 g casamino acids, 0.03 g *L*-glutamine, water to 1 L, pH 5.5) and shaken at 130 rpm, at 23°C, under 24 h continuous white light for 3 days. Lastly, the co-cultured thallus was washed with water containing cefotaxime and cultured on the ½ B5 medium containing both hygromycin (10 µg/mL) and cefotaxime (100 µg/mL).

The G1 gemmae of transformed Tak-1 *M. polymorpha* with 35s-35s-*MpAMT1.5*-citrine were cultured on ½ B5 medium plus 1% sucrose for 2-3 days, and then collected to analyze by using a confocal microscope Nikon A1si. Optical sections were collected at the appropriate wavelengths for citrine (488 nm/ 525 nm) and autofluorescence of chlorophyll (405 nm/ 595 nm).

2.6 The treatment of different nitrogen sources. The transformed Tak-1 *M. polymorpha* with promoter of *MpMAT1.5*- GUS was cultured on ½ B5 medium plus hygromycin (10 µg/mL) and cefotaxime (100 µg/mL) at 23 °C, under 24 h continuous white lights until gemmae (G1) were grow up and produce new gemmae (G2), then

cultured the G2 gemmae for 2 weeks under the same conditions. Then move grown G2 thallus to 3 different media for one week under 16h/8h light/darkness at 22 °C. These 3 different media were: ½ B5 medium, ½ B5 without any nitrogen sources (KNO_3 and $(NH_4)_2SO_4$), and ½ B5 containing same concentration of N (13.38 mM) as common ½ B5 medium, but the nitrogen source is from L-Alanine instead of KNO_3 or $(NH_4)_2SO_4$.

2.7 Gus staining. The G2 thallus of transformed Tak-1 *M. polymorpha* with promoter of *MpMAT1.5*- GUS were cut and put in GUS solution (1 mL 100 mM X-GlucA that dissolved in dimethyl formamide, 10 mL 0.5 M sodium phosphate buffer, 500 µL triton X-100, add water to 100 mL). Followed by infiltration by pulling the vacuum twice for 5 min each until thallus were solution soaked, and incubated them at 37 °C overnight. Thallus were de-stained by 70% ethanol at 65 °C.

3. Results

3.1 *MpAMT1.5* is an ammonium transporter

MpAMT1;2 and MpAMT1;5 from Marchantia were demonstrated to be electrogenic ammonium transporters (McDonald and Ward, 2016) and detailed analysis of MpAMT1;2 using electrophysiology showed that it is a high affinity transporter for ammonium (Guo et al., 2018). To test whether MpAMT1;2 and MpAMT1;5 function in ammonium uptake into cells, they were expressed in yeast strain MLY131a (Lorenz and Heitman, 1998) which is defective in three MEP ammonium transporters and requires expression of an ammonium transport activity to allow growth on ammonium. As shown in Fig. 22, expression of either MpAMT1;2 or MpAMT1;5 allowed yeast to grow on 0.5 mM $(NH_4)_2SO_4$ as a nitrogen source while yeast transformed with the empty vector

(pDR196/GW) did not grow on ammonium. As a control, MLY131a transformed with MpAMT1;2, MpAMT1;5, or the empty vector all grew on proline as the nitrogen source (Fig. 22).

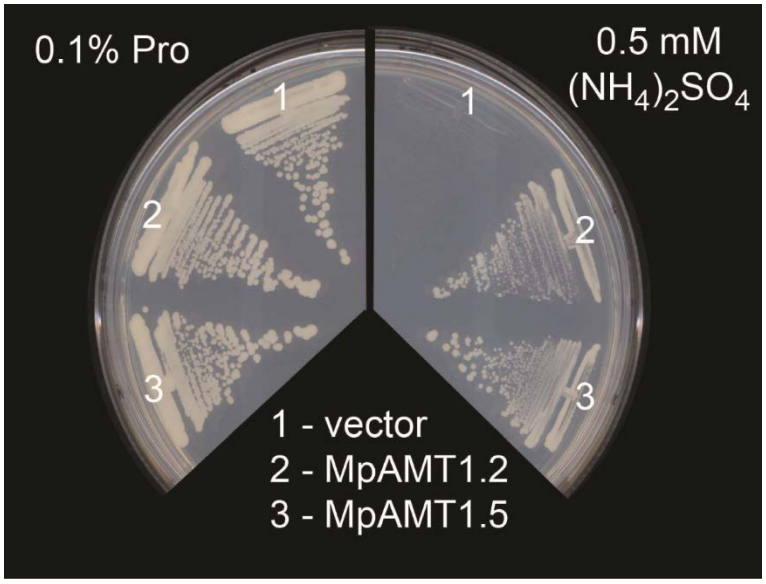


Figure 22. MpAMT1;5 complements triple-MEP mutant yeast strain MLY131a. Yeast was transformed with MpAMT1;2 or MpAMT1;5 in pDR196/GW or with the empty vector. Transformants were streaked out on media (1.7 g/L yeast nitrogen base, 20 g/L glucose, 20 g/L agar) containing either 0.1% proline or 0.5 mM $(NH_4)_2SO_4$ as nitrogen source and grown for 4 d at 30 °C.

3.2 Ammonium induced inward currents in *MpAMT1;5*- expressing oocytes

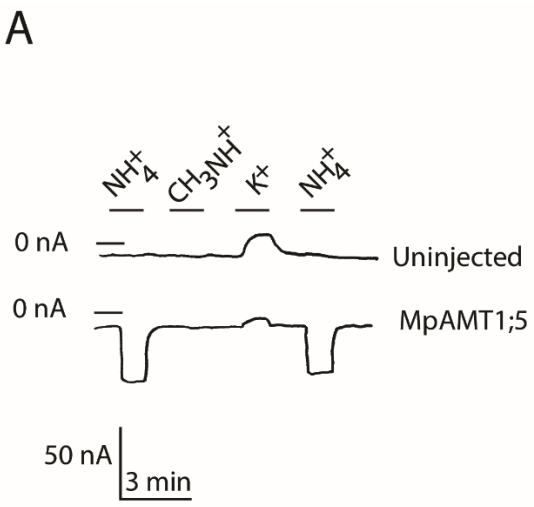
The ability of MpAMT1;5 to transport ammonium, methyl ammonium, and potassium was tested in MpAMT1;5 expressing *Xenopus* oocytes. These substrates were applied at a concentration of 2 mM in K^+-free ringer solution (pH 5.6), at a holding potential of -40 mV. As shown in Fig. 23, uninjected oocytes showed no currents except when potassium was applied, an outward current was measured consistent with activation of the

endogenous NaK-ATPase (Sobczak et al., 2010). Ammonium induced inward currents in MpAMT1;5-expressing oocytes while methylammonium at the same concentration did not (Fig. 23A). This indicates that either the affinity of MpAMT1;5 for methylammonium is much lower than for ammonium or that methyl ammonium is not transported. This experiment was replicated and the results are shown in Fig. 23B. Substrate-induced currents (background subtracted) are shown for MpAMT1;5-expressing oocytes at three voltages: -142 mV, -121 mV and -40 mV. Ammonium-induced currents were normalized to those measured at -142 mV (average was -100.7 µA). Small outward currents induced by potassium can be attributed to activation of the endogenous Na/K ATPase in the oocyte plasma membrane. Methyl ammonium currents were not detected under these conditions.

Xenopus oocytes expressing MpAMT1;5 were voltage clamped and currents were recorded at a range of voltages from 42 to -142mV. Fig. 24A shows an example of currents before, after, and during ammonium (1 mM) application. The background currents (before and after ammonium was applied) were similar indicating that the baseline was stable. Inward currents were measured at all potentials when 1mM ammonium was applied. Fig. 24B shows ammonium-induced currents (background subtracted) for a range of ammonium concentrations.

The affinity of MpAMT1;5 for ammonium is related to the $K_{0.5}$ value for transport. At a pH of 5.6 and a membrane potential of -120 mV, the $K_{0.5}$ of MpAMT1;5 for ammonium was 380 µM (Fig. 25A), indicating that MpAMT15 has a low affinity for ammonium. $K_{0.5}$ values were pH dependent with the lowest $K_{0.5}$ (200 µM) occurring at pH 6.8. This indicates that the affinity of MpAMT1;5 for ammonium is highest at neutral pH and

lower at both acidic and alkaline extracellular pH (Fig. 25A). Previous work has shown that the affinity of many ammonium transporters for their substrate is pH-dependent, such as Af-Amt1 and Af-Amt3 from *Archaeoglobus fulgidus* (Wacker et al., 2014) and PvAMT1;1 from bean plant *Phaseolus vulgaris* (Ortiz-Ramirez et al., 2011a). However, there are also examples of ammonium transporters with pH-independent $K_{0.5}$ values for ammonium such as LeAMT1;1 from tomato. (Ludewig et al., 2002). As shown in Fig 25B, the $K_{0.5}$ of MpAMT1;5 was measured at different membrane potentials from 0 to -142 mV. At more negative potentials (-142 mV), $K_{0.5}$ of MpAMT1;5 was lower, indicating that MpAMT1;5 has a higher affinity for ammonium when the membrane potential is polarized.



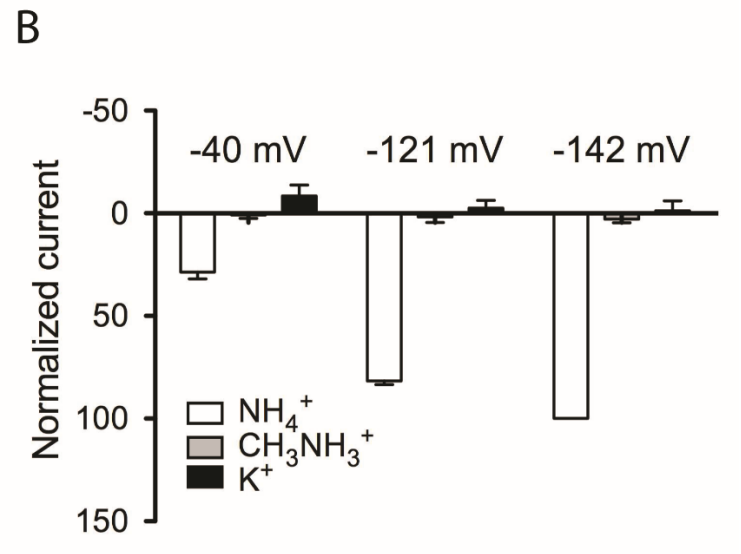


Figure 23. MpAMT1;5 is an electrogenic transporter. *Xenopus* oocytes were perfused with K^+-free Ringer solution at pH 5.6 under voltage clamp conditions at a holding potential of -40 mV. Substrates were applied at 2 mM in the same solution. (A) Representative currents are shown for oocytes expressing MpAMT1;5 and uninjected oocytes. Bars above the traces indicate the time when substrates were applied, the downward deflections indicate inward current. (B) Substrate-induced currents (background subtracted) were recorded at -40 mV, -121 mV, and -142 mV. Currents were normalized to ammonium-induced cur.rents at -142 mV (mean = -100.7 nA). Error bars are SD (n=3 oocytes).

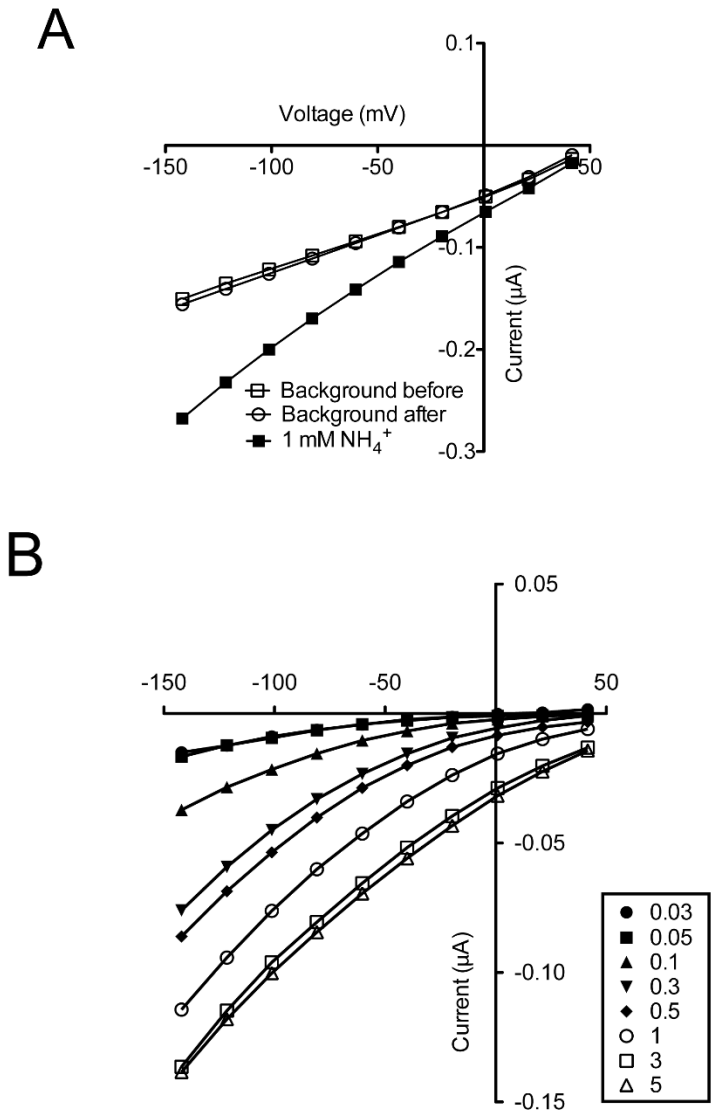


Figure 24. Kinetic analysis of ammonium transporter MpAMT1;5 expressed in *Xenopus* oocytes. Voltage clamped oocytes were perfused with K^+ free Ringer solution at pH 5.6 at a holding potential of -40 mV. Currents were recorded during voltage pulses from 42 to -142 mV. (A) Currents recorded before, during and after application of 1 mM ammonium. (B) Ammonium-dependent currents recorded at different ammonium concentrations from 0.03 mM to 5 mM. Background currents, before and after

ammonium application (as in A) were averaged and subtracted from currents recorded during ammonium application.

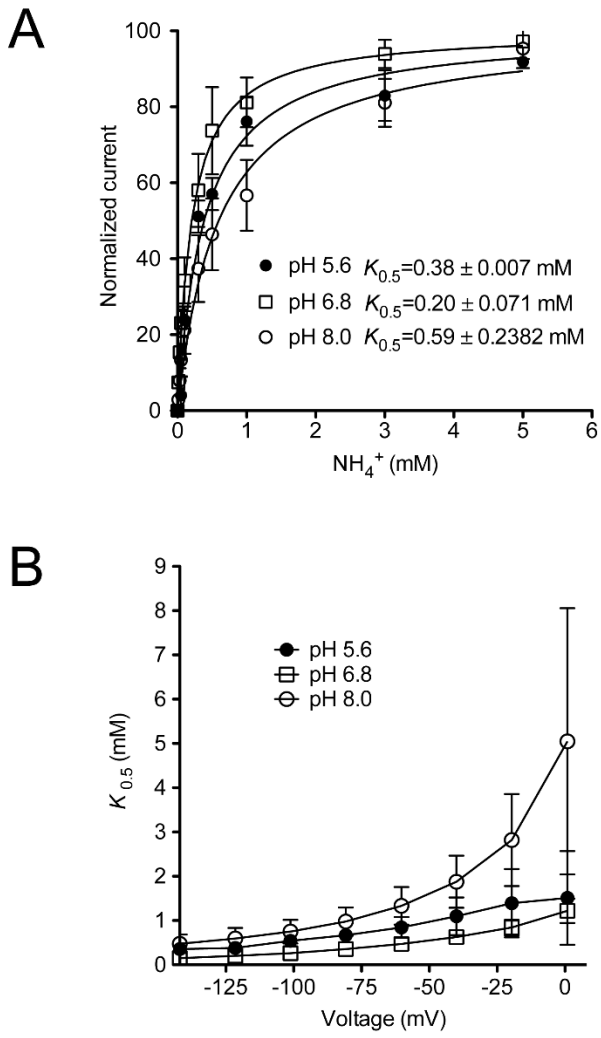
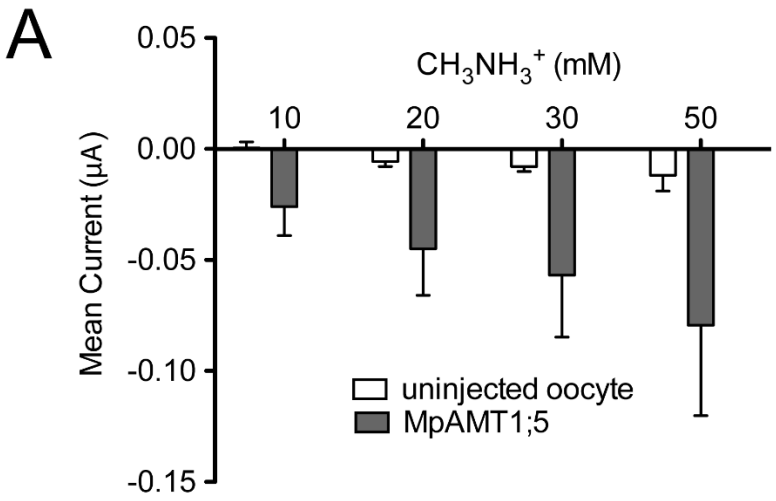


Figure 25. Voltage-dependence of ammonium affinity for MpAMT1;5 at three pHs (pH 5.6, 6.8, and 8.0). $K_{0.5}$ values for MpAMT1;5 were measured at three pHs (pH 5.6, 6.8, and 8.0). Oocytes were perfused with K^+-free Ringer solution at a holding potential of -40 mV. (A) Kinetic analysis of MpAMT1;5 at three pH values at a membrane potential of -121 mV. The data were fitted to the Michaelis-Menten equation and normalized to V_{max}

which were -0.1307 µA (pH 5.6), -0.4719 µA (pH 6.8), and -0.2973 µA (pH 8.0). Error bars are SD (n = 3 oocytes). (B) $K_{0.5}$ values for ammonium were measured at different membrane potentials from 0 to -142 mV.

3.3 Kinetic analysis of methyl-ammonium transported by MpAMT1;5

Although no currents were measured in response to 2 mM methylammonium (Fig. 23), the application of high concentrations of methylammonium to MpAMT1;5-expressing oocytes indicated that methylammonium is transported by MpAMT1;5. Fig. 26A shows that MpAMT1;5-expressing oocytes had larger methylammonium-induced currents than uninjected oocytes at concentrations of methylammonium from 10 to 50 mM. Fig.26B shows that methylammonium-induced currents did not saturate, indicating that the $K_{0.5}$ of MpAMT1;5 for methylammonium is higher than 20 mM. Even though we could not measure the the $K_{0.5}$ for methylammonium it was at least 52 times higher than the $K_{0.5}$ of ammonium. This indicates the affinity of MpAMT1;5 for methylammonium was very low.



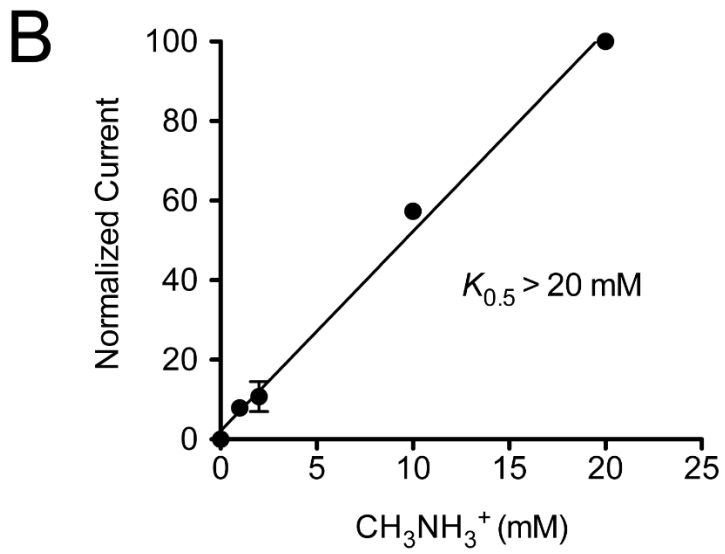
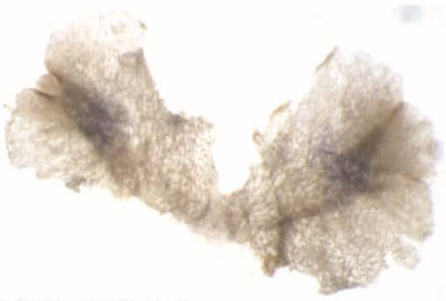


Figure 26. Kinetic analysis of methyl ammonium transported by MpAMT1;5. Oocytes were perfused with K^+-free Ringer solution at pH 5.6. Representative results are presented at a membrane potential of -80 mV. Methylammonium-dependent currents are presented for uninjected or MpAMT1;5-injected oocytes. (A) Inward methylammonium-induced currents are represented as negative values (downward bars). (B) Dependence of methyl ammonium-induced current on substrate concentration. Currents were normalized to methyl ammonium-induced currents at 20 mM (mean = -0.0449 µA) and plotted against methylammonium concentration. The line represents results of linear regression. Error bars are SD (n = 3 oocytes).

3.4 Gene MpAMT1;5 expression is regulated by N sources

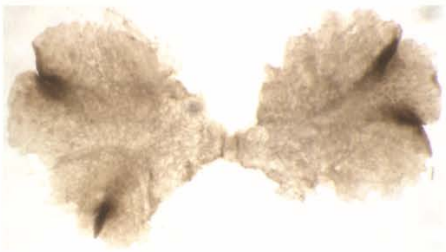
A. 13.38 mM [N]: 12.37 mM KNO_3, 0.51 mM $(NH_4)_2SO_4$

B. 0 mM [N]



C. 13.38 mM [N]: Alanine

D. 13.38 mM [N]: NH_4Cl



E. Control, *MpAMT1.2*
13.38 mM N]: 12.37 mM KNO_3, 0.51 mM $(NH_4)_2SO_4$

Figure 27. MpAMT1;5 expression is very low. M. polymorpha (Tak-1) was transformed with a MpAMT1;5 promoter :: GUS construct. Transformants were cultured on four different types of media and their thallus were GUS stained: (A and E) ½ B5 medium (contains KNO_3 and $(NH_4)_2SO_4$) , (B) ½ B5 medium without N, (C) ½ B5 medium with L-Alanine as the only N source (D) ½ B5 medium with NH_4Cl as the N source. Marchantia transformed with MpAMT1;2 promoter:: GUS (E) is included as a positive control.

To analyze the gene expression of *MpAMT1.5* and its regulation, we amplified 4977 bp upstream of the coding region of *MpAMT1.5*, and fused the fragment with GUS gene. In total, 19 lines of transgenic male *M. polymorpha* (Tak-1) were obtained by using agrobacterium- mediated transformation, and 5 lines of them were chosen to analyze the regulation of *MpAMT1.5* by culturing them from gemmae on ½ B5 media for two weeks and then transferred to ½ B5 containing different N sources but with same concentration of N (13.38 mM) (½ B5 medium, ½ B5 medium without N, ½ B5 medium with L-Alanine as the N source, or ½ B5 medium with NH_4Cl as the N source). The thallus was collecting and stained using X-gluc. As showed in Fig. 27, the expression of *MpAMT1.5* was too low to observe on any of the N sources. As a control, Marchantia transformed with *MpAMT1;2* promoter ::GUS was grown on ½ B5 medium and stained for GUS activity. The thallus shows expression throughout.

3.5 Protein MpAMT1.5 localized at the plasma membrane

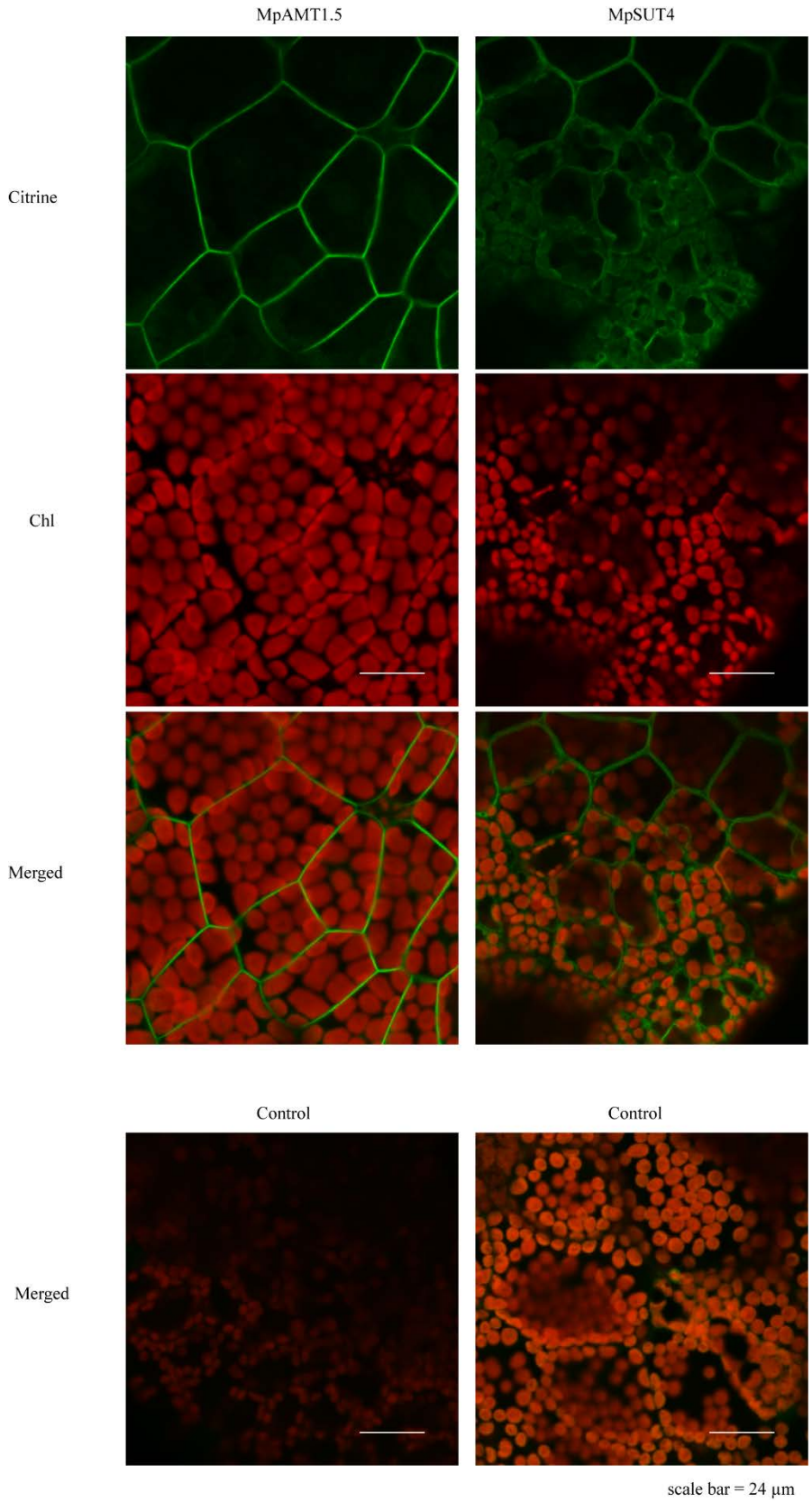

MpAMT1.5
MpSUT4
Citrine
Chl
Merged
Control
Control
Merged
scale bar = 24 μm

Figure 28. MpAMT1;5 localized to the plasma membrane in Marchantia. Confocal images are shown of either the ammonium transporter MpAMT1.5 fused to citrine at the C-terminus or the vacuolar sucrose transporter MpSUT4 fused to citrine at the N-terminus under control of the 35S promoter in Tak-1 *Marchantia polymorpha* gemmae. Optical sections were collected at the appropriate wavelengths for citrine (488 nm/ 525 nm) and autofluorescence of chlorophyll (405 nm/ 595 nm). Merged images show an overlay of both citrine and chlorophyll fluorescence. Controls were collected from untransformed Tak-1 gemmae at the same settings as the corresponding MpAMT1.5 or MpSUT4 images, only the merged image is shown for the controls. Bars shown are 24 μm.

AMT proteins have 11 TMS (trans-membrane spans) with the N- terminus outside of the cell and the C- terminus inside. AtAMT1;1 from Arabidopsis is localized at the plasma membrane using a C-terminal green florescent protein (GFP) fusion (Mayer and Ludewig, 2006). Here, a C-terminal citrine fusion, driven by a dual 35S promoter, was expressed in Marchantia thalli. MpAMT1.5 localized to the plasma membrane as shown in Fig. 28. In comparison with chloroplast autofluorescence, citrine fluorescence can be seen toward the outside of the cells indicating a plasma membrane localization (Fig. 28). As a control, an N-terminal citrine fusion with the vacuolar sucrose transporter MpSUT4 was used (Guo et al., 2018). Chlorophyll autofluorescence in the cytosol can be observed toward the outside of MpSUT4-citrine fluorescence which appears at the vacuole membrane.

4. Discussion

Marchantia polymorpha is a good model organism to study transporters in the AMT/MEP/Rh superfamily because it is evolutionarily diverged from angiosperms, the genome sequence is available (Bowman et al., 2017), and it is easily transformed. Marchantia encodes 19 ammonium transporters: nine *AMT1* and ten *AMT2* (McDonald and Ward, 2016). MpAMT1;2 has been characterized in detail as an electrogenic plasma membrane transporter with high affinity for ammonium (Guo et al., 2018). In this paper, we characterized MpAMT1;5, which is also an electrogenic plasma membrane transporter, but with much lower affinity and lower expression.

Transporters are classified as electrogenic if they transport a charge across the membrane and electroneutral if they do not transport a charge. The AMT/MEP/Rh superfamily contains both electrogenic and electroneutral ammonium transporters. MEP and Rh clades are thought to be electroneutral while those in the AMT clade in plants are electrogenic. It is not known if all AMTs are electrogenic or if AMTs acquired electrogenicity at some point before land plant evolution. Considering that all AMT1s studied in angiosperms and in Marchantia are electrogenic, it is likely that all land plant AMT1s are electrogenic (McDonald and Ward, 2016).

MpAMT1;5 is 77% identical to MpAMT1;2 at the level of protein sequence. MpAMT1;5 was previously shown to function when expressed in *Xenopus* oocytes as an electrogenic ammonium transporter but was not further characterized (McDonald and Ward, 2016). Both MpAMT1;5 and MpAMT1;2 were able to complement the ability of "triple MEP" mutant yeast strain MLY131a to grow on ammonium as the sole N source. This is a confirmation that MpAMT1;2 and MpAMT1;5 transport ammonium across the membrane as well as a charge, and is important because two-electrode voltage clamping

of oocytes expressing MpAMT1;2 (Guo et al., 2018) or MpAMT1;5 only shows that ammonium induces an inward current. Showing that MpAMT1;5 allows ammonium uptake provides additional confidence that ammonium-induced inward currents are coupled to ammonium uptake in MpAMT1;5.

In addition to ammonium, MpAMT1;5 transports methylammonium but not K^+. This is consistent with a model for the transport mechanism in which NH_4^+ is deprotonated and NH_3 is transported across the membrane. This model is supported by the 3D structures of bacterial members of the AMT/MEP/Rh superfamily (Zheng et al., 2004; Li et al., 2007; Lupo et al., 2007). If NH_4^+ was the transported substrate then we would expect MpAMT1;5 to also transport K^+ which has a similar size and charge as NH_4^+ rather than $CH_3NH_3^+$ which is larger than NH_4^+. $CH_3NH_3^+$ is similar to NH_4^+ in that it can be deprotonated leaving CH_3NH_2 to be transported.

The major difference in transport activity observed between MpAMT1;5 and MpAMT1;2 is a strong difference in substrate affinity. The affinity of ammonium transporters for their substrate varies widely. AMTs can be classified as high affinity transporters (HATs) and low affinity transporters (LATs) (Wang et al., 1993). The $K_{0.5}$ of HATs ammonium transporters ranged from 10 to 170 µM (Kronzucker et al., 1996). For example, the $K_{0.5}$ of LeAMT1.1 for ammonium is 7 µM (pH 7.5, -140 mV) (Ludewig et al., 2002), PvAMT1.1 is 28 µM (pH 5.5, -160 mV) (Ortiz-Ramirez et al., 2011), and TaAMT1.1 is 54 µM (pH 5.5, -140 mV) (Sogaard et al., 2009). LATs have also been characterized, for example, the $K_{0.5}$ of DmAMT1 is 820 µM (pH 5.6, -140 mV) (Scherzer et al., 2013).

The $K_{0.5}$ of MpAMT1;5 for ammonium was 380 µM (pH 5.6, -140 mV). This indicates that MpAMT1;5 is a low affinity ammonium transporter. By comparison, the $K_{0.5}$ for ammonium of MpAMT1;2 is 7 µM measured under similar conditions (pH 5.6, -137 mV) (Guo et al., 2018). It is interesting that *Marchantia* has two related ammonium transporters with $K_{0.5}$ values that differ by 54-fold. And it is reasonable to ask why *Marchantia* requires low affinity ammonium transporters, especially, when it has very high affinity transporters. One possible explanation is that some cells are used to take up ammonium when the extracellular concentration is very high. Transport activity is controlled by substrate concentration and the $K_{0.5}$ indicates the concentration at which the transport activity is most responsive. At concentrations above the $K_{0.5}$, transport activity is high but plateaus and at concentrations below $K_{0.5}$ the activity responds to substrate concentration but is low. The $K_{0.5}$ of transporters and enzymes is often is very similar to the physiological concentration of the substrate, allowing the transporter or enzyme activity to respond to substrate concentration changes. More information on where and when MpAMT1;5 is expressed would be useful to better answer the question of why *Marchantia* has both low and high affinity ammonium transporters.

The $K_{0.5}$ of MpAMT1;5 for ammonium was pH-dependent, and showed the highest affinity for ammonium at pH 6.8 ($K_{0.5}$ = 200µM). Most ammonium transporters have pH-dependent $K_{0.5}$ values for ammonium. Most show the highest affinity at the most acidic pH and it is unusual that MpAMT1;5 shows the highest affinity at a neutral pH. Previous publications have shown that affinities of many ammonium transporters are pH-dependent. For MpAMT1;2, the $K_{0.5}$ for ammonium is the lowest at the lowest pH (Guo et al., 2018) and this is the same for LeAMT1;1 from tomato (Ludewig et al., 2002). For

PvAMT1;1 from bean plant, the relation between pH and $K_{0.5}$ is opposite, the lowest $K_{0.5}$ occurs at alkaline pH (pH 8). And this is the same for Amt1 from the euryarchaeon *Archaeoglobus fulgidus* (Wacker et al., 2014). The $K_{0.5}$ of LeAMT1;1 from tomato for ammonium is pH independent (Ludewig et al., 2002). The implication is that some ammonium transporters are regulated though the effect of pH on their affinity for their substrate.

Methylammonium is a methyl analogue of ammonium, which has been commonly used to study ammonium transport for a long time. Hackette et al. (1970) reported that the ammonia permease in *Penicillium chrysogenum* can uptake methylamine as a substrate, but with greatly reduced affinity. Roon et al. (1975) found that methylammonium and ammonia share the same transport system in *S. cerevisiae* and showed that ^{14}C-methylammonium can be used as a convenient isotopic substrate to study ammonia transport. Subsequently, methylammonium was used to study many ammonium transporters such as MEP1 and MEP2 from yeast *Saccharomyces cerevisiae* (Marini et al., 1994), AgAMT from mosquito *Anopheles gambiae* (Pitts et al., 2014), and AtAMT1;1-3 from *Arabidopsis thaliana* (Gazzarrini et al., 1999).

The affinity of MpAMT1;5 for methylammonium ($K_{0.5}$ >20 mM) was much lower than for ammonium. The result is similar to a previous study of ammonia permease in *Penicillium chrysogenum*, which transports methylammonium but with greatly reduced affinity (Hackette et al., 1970). Most ammonium transporters can transport both methylammonium and ammonium. However, the affinity of transporting methyl-ammonium is usually much lower than for transporting ammonium.

The sub-cellular location of MpAMT1;5 was studied by using C-terminal citrine fusion with MpAMT1;5 protein in *Marchantia* Tak-1 thallus. MpAMT1;5 was localized on the plasma membrane, which was same as the sub-cellular location of MpAMT1;2 that was analyzed by using the same method (Guo et al., 2018). This result suggests that MpAMT1;5 functions to transport ammonium from the extracellular space into the cytoplasm.

Acknowledgements

We thank Drs. Ryuichi Nishihama and Takayuki Kohchi (Kyoto University) for providing *Marchantia* genome sequence prior to publication and for sharing the binary vectors as well as the Tak-1 strain used in this study. We also thank Drs. Takayuki Kohchi, Takeshi Araki, Ryuichi Nishihama and members of their labs for sharing protocols and providing invaluable advice on working with *Marchantia*. We acknowledge Grant Bartel at the University Imaging Centers at the University of Minnesota for help with confocal imaging. This work was supported by a grant from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy (DE-FG02-10ER15886) to J.W.

General conclusions

In conclusion, this thesis describes two distinct research projects across the field of analytical chemistry, biochemistry, and molecular biology. This research has been successful in achieving the outlined objectives in five years even with having difficulties due to a lab transition. Research products include seven posters, two invited talks, two papers, and other two manuscripts that are on the way.

The first chapter is a review of metabolite identification by using chromatography, mass spectrometry and the methodology development in retention prediction. Of course, this chapter also contains the review of *Marchantia polymorpha,* ammonium cell membrane transporters and their regulation pathways, which are mediated by CBLs and CIPKs.

The second and third chapters have been done with my first supervisor Dr. Paul Boswell. The motivation of these two chapters is to develop the methodology Retention Projection with Back-calculation (RPwB) to predict retention time of metabolites in gradient elution by using chromatography and mass spectrometry. The prediction accuracy of RPwB in RPLC indicates its potential to improve metabolite identification with higher confidence by using both retention information and mass information. However, RPLC mainly is used to separate non-polar compounds, therefore it is difficult to apply RPwB to predict all compounds if using RPLC alone. In the second chapter, we extend RPwB to HILIC, which is mainly used to analyze polar compounds but has a notorious reputation for causing retention time shifts due to complex mechanisms. Through the study of RPwB in HILIC, we found its prediction error was around 1%,

which was 6-43 times more accurate than projections made using ideal gradients and 3-5 times more accurate than the same retention projections made using offset gradients (i.e., gradients that only took gradient delay into account). RPwB improved the prediction accuracy to 5-25 times higher than previous studies in HILIC. Still, the error remained higher in our HILIC projections than in RPLC. Based on the shape of the back-calculated gradients, we suspect the higher error is a result of prominent gradient distortion caused by strong, preferential water uptake from the mobile phase into the stationary phase during the gradient – a factor our model did not properly take into account. It appears that, at least with the stationary phase we used, column distortion is an important factor to take into account in retention projection in HILIC that is not usually important in RPLC.

Even though the prediction accuracy of RPwB has been demonstrated in both RPLC and HILIC, there is still a long way to apply this method to real life cases such as biological samples since its nature of complexity, such as unknown chemical interactions, over concentrated extractions, and diverse chemical structures. To study the robustness of RPwB when applying it to biological samples, we choose five plant samples and nine sample solvents to study their effects on peak shapes, retention time shifts, and prediction accuracy. Of these nine sample solvents methanol, ethanol, isopropanol, acetonitrile, acetone, dichloromethane, ethyl acetate, tetrahydrofuran, and toluene, we found isopropanol, acetonitrile, and ethyl acetate were the top three that distorted peak shapes the most, and all these nine sample solvents shifted retention times and poorly retained compounds suffered more than well retained ones. However, ethyl acetate was the only one that broke the prediction accuracy of RPwB and had a prediction error at 4.3 sec, which was more than 3 sec that was the maximum deviation allowed to determine the

successful application RPwB in RPLC. Second, five plant samples (*Solanum lycopersicum* fruit, *Solanum lycopersicum* stems and leafs, *Nicotiana* flowers, *Nicotiana* leafs, and *Nicotiana forsteri* leafs) were randomly chosen and extracted using four sample solvents (water, 70% ethanol, dichloromethane, and isopropanol) that are the most common ones and that did not change the prediction accuracy of RPwB, and obtained the most concentrated plant extractions. The effect of plant matrices on retention times of nearly all analytes was negligible, despite their high concentrations. However, we observed a buildup of some plant matrix solutes in the column that lowered retention projection accuracy for two charged analytes, tetrabutylammonium and tetrapentylammonium. However, this buildup could be removed (and the accuracy of retention projections restored) by either flushing the column with a stronger mobile phase or decreasing the concentration of the injected sample.

Based on results we found that are described in the second and third chapters, some additional research could be done. First, to extend the online database of retention projection, we should add isocratic data in HILIC to the database of GC and RPLC website site: http://www.retentionprediction.org/hplc/database/index.php. Furthermore, to apply the online database to identify metabolites in biological samples, we should collaborate with companies and labs to test the prediction accuracy of RPwB across labs and experimental conditions.

The fourth and fifth chapters describe research projects in biochemistry and molecular biology focusing on studying plasma membrane ammonium transporters in *Marchantia polymorpha*. Nitrogen is a main required nutrient for plant growth and development. There are two main forms of nitrogen sources in soils, one is nitrate and another is

ammonium. However, plants prefer using ammonium since there is less energy cost during the assimilation process. Plants uptake ammonium mainly through ammonium transporters (AMTs). So far, ammonium transporters have been studied only in angiosperms. As a model organism, *M. polymorpha* occupies a special position in the evolutionary tree in the plant kingdom. Function characterizations of ammonium transporters in *M. polymorpha* give insights of AMT evolution. Plants encode both electrogenic (AMT1) and electroneutral (AMT2) ammonium transporters. Genome sequence for the bryophyte *Marchantia polymorpha* was recently obtained and showed that Marchantia contains 9 *AMT1* and 10 *AMT2* genes. We have successfully cloned four *MpAMTs* including *MpAMT1;2*, *MpAMT1;3*, *MpAMT1;4*, and *MpAMT1;5*. In this thesis, we studied transport activity, gene expression and membrane localization of *MpAMT1;2* and *MpAMT1;5*. By expression in *Xenopus* oocytes and two-electrode voltage clamping, we determined that *MpAMT1;2* and *MpAMT1;5* are both electrogenic ammonium transporters. *MpAMT1;5* has a low affinity for ammonium ($K_{0.5}$=0.38 mM at pH 5.6) while MpAMT1;5 is a high affinity ammonium transporter ($K_{0.5}$=0.007 mM at pH 5.6). *Marchantia polymorpha* gametophytes were transformed with C-terminal citrine fusions for both transporters and both clearly localize to the plasma membrane. Expression of both transporters was regulated by nitrogen supply. Using qRT-PCR, *MpAMT1;2* showed low expression when the plants were grown on half-strength B5 medium (12 mM KNO_3, 0.5 mM $(NH_4)_2SO_4$) or with sufficient N supplied as the amino acid alanine, but high expression when grown under N-deficiency conditions. Expression of *MpAMT1;5* was lower and using promoter-GUS analysis expression was not detected in the thallus.

In Arabidopsis, AMT1 transporters are down-regulated by phosphorylation by the complex of CIPK23-CBL1. Sequence alignment revealed a conserved phosphorylation site in all Marchantia AMT1 homologs and none of the AMT2 homologs suggesting similar regulation in Marchantia. Activity analysis showed MpAMT1;2 and MpAMT1;4 lost their activities when the conserved threonine (T475 of MpAMT1;2, and T472 of MpAMT1;4) was mutagenized to phosphorylation-mimic amino acid aspartic acid, while only part of their activity was lost when the conserved amino acid was mutagenized to un-phosphorylatable amino acid alanine. Furthermore, using Arabidopsis CBL1 protein sequence blast against Marchantia genome, it indicates Marchantia has three CBLs: MpCBL1 (Mapoly0134s0040.1), MpCBL2 (Mapoly0066s0053.1), and MpCBL3 (Mapoly0015s0061.1). Meanwhile, using Arabidopsis CIPK23 protein sequence blast against Marchantia genome, only two candidates in Marchantia: MpCIPK1 (Mapoly0025s0017) and MpCIPK2 (Mapoly0025s0039) contain a NAF domain which is necessary and enough to category CIPKs and interact with CBLs. Overall, the results indicate that Marchantia is a good model to study ammonium transporter activity and regulation.

To understand mechanisms of uptake of nitrogen from soils, more work should be done through studying the model system *Marchantia polymorpha*. First, identifying which and how MpCIPKs and MpCBLs regulate MpAMTs, and therefore turn on or off ammonium transporters. This project can be done either in a heterologous system by co-injecting cRNAs of MpCIPKs, MpCBLs, and MpAMTs in oocytes, or *in planta* by creating gene deletion/knockdown mutants via CRISPER. Second, as another big family of ammonium transporter in Marchantia, MpAMT2 family has not been studied yet. To

understand the mechanism of MpAMTs, we should explore MpAMT2 by studying their transport function, gene expression and regulation, and protein regulation pathways.

Bibliography

Abate-pella, D., Freund, D.M., Ma, Y., Simón-manso, Y., Hollender, J., Corey, D., et al. (2015) Realiable calculation of high performance liquid chromatography retention times across labs and methods by back-calculation of system non-idealities. *J Chromatogr A*. 1412: 43–51.

Adams, D.G. (2002) Cyanobacteria in symbiosis with hornworts and liverworts. *Cyanobacteria in Symbiosis*. 117–135.

Adams, D.G., and Duggan, P.S. (2008) Cyanobacteria-bryophyte symbioses. *J Exp Bot*. 59: 1047–1058.

Aharoni, A., Vos, C.H.R.I.C.D.E., Verhoeven, H.A., Maliepaard, C.A., Kruppa, G., and Bino, R. (2002) Nontargeted Metabolome Analysis by Use of Fourier Transform Ion Cyclotron Mass Spectrometry. *A J Integr Biol*. 6: 217–234.

Albrecht, V., Ritz, O., Linder, S., Harter, K., and Kudla, J. (2001) The NAF domain defines a novel protein-protein interaction module conserved in Ca2+-regulated kinases. *EMBO J*. 20: 1051–1063.

Alpert, A.J. (1990) Hydrophilic interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journnl Chromatogr*. 499: 177–196.

Andrade, S.L. a, Dickmanns, A., Ficner, R., and Einsle, O. (2005) Crystal structure of the archaeal ammonium transporter Amt-1 from Archaeoglobus fulgidus. *Proc Natl Acad Sci U S A*. 102: 14994–14999.

Annesley, T.M. (2003) Ion Suppression in Mass Spectrometry. *Clin Chem*. 1044: 1041–1044.

Bakalyar, S.R., McIlwrick, R., and Roggendorf, E. (1977) Solvent selectivity in reversed-phase high-pressure liquid chromatography. *J Chromatogr A*. 142: 353–365.

Barnes, B.B., Wilson, M.B., Carr, P.W., Vitha, M.F., Broeckling, C.D., Heuberger, A.L., et al. (2013a) Retention projection enables reliable use of shared gas chromatographic retention data across laboratories, instruments, and methods. *Anal Chem*. 85: 11650–11657.

Barnes, B.B., Wilson, M.B., Carr, P.W., Vitha, M.F., Broeckling, C.D., Heuberger, A.L., et al. (2013b) Retention projection enables reliable use of shared gas chromatographic retention data across laboratories, instruments, and methods. *Anal Chem*. 85: 11650–11657.

Berger, F., Bowman, J.L., and Kohchi, T. (2016) Marchantia. *Curr Biol*. 26: R186–R187.

Bocian, S., Rychlicki, G., Matyska, M., Pesek, J., and Buszewski, B. (2014) Study of hydration process on silica hydride surfaces by microcalorimetry and water adsorption. *J Colloid Interface Sci*. 416: 161–166.

Boguslaw Buszewski; Noga, S. (2012) Hydrophilic interaction liquid chromatography (HILIC) — a powerful separation technique. *Anal Bioanal Chem*. 402: 231–247.

Bostick, D.L., and Brooks, C.L. (2007) On the equivalence point for ammonium (de)protonation during its transport through the AmtB channel. *Biophys J*. 92: L103–L105.

Boswell, P.G., Schellenberg, J.R., Carr, P.W., Cohen, J.D., and Hegeman, A.D. (2011a) Easy and accurate high-performance liquid chromatography retention prediction with different gradients, flow rates, and instruments by back-calculation of gradient and flow rate profiles. *J Chromatogr A*. 1218: 6742–9.

Boswell, P.G., Schellenberg, J.R., Carr, P.W., Cohen, J.D., and Hegeman, A.D. (2011b) A study on retention 'projection' as a supplementary means for compound identification by liquid chromatography-mass spectrometry capable of predicting retention with different gradients, flow rates, and instruments. *J Chromatogr A*. 1218: 6732–41.

Boswell, P.G., Stoll, D.R., Carr, P.W., Nagel, M.L., Vitha, M.F., and Mabbott, G.A. (2013) An advanced, interactive, high-performance liquid chromatography simulator and instructor resources. *J Chem Educ*. 90: 198–202.

Bowman, J.L. (2016) A brief history of marchantia from Greece to genomics. *Plant Cell Physiol*. 57: 210–229.

Bowman, J.L., Araki, T., and Kohchi, T. (2016) Marchantia: Past, present and future. *Plant Cell Physiol*. 57: 205–209.

Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017) Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell*. 171: 287–304.e15.

Brown, M., Dunn, W.B., Dobson, P., Patel, Y., Winder, C.L., Francis-McIntyre, S., et al. (2009) Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst*. 134: 1322–1332.

Calabrese, S., Kohler, A., Niehl, A., Veneault-Fourrey, C., Boller, T., and Courty, P.E. (2017) Transcriptome analysis of the populus trichocarpa-rhizophagus irregularis mycorrhizal symbiosis: Regulation of plant and fungal transportomes under nitrogen starvation. *Plant Cell Physiol*. 58: 1003–1017.

Castells, C.B., and Castells, R.C. (1998) Peak distortion in reversed-phase liquid chromatography as a consequence of viscosity differences between sample solvent and mobile phase. *J Chromatogr A*. 805: 55–61.

Cheng, N.H., Pittman, J.K., Zhu, J.K., and Hirschi, K.D. (2004) The Protein Kinase SOS2 Activates the Arabidopsis H+/Ca 2+ Antiporter CAX1 to Integrate Calcium Transport and Salt Tolerance. *J Biol Chem*. 279: 2922–2926.

Creek, D.J., Dunn, W.B., Fiehn, O., Griffin, J.L., Hall, R.D., Lei, Z., et al. (2014) Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics*. 10: 350–353.

Dettmer, K., Aronov, P. a, and Hammock, B.D. (2007) Mass spectrometry-based metabolomics. *Mass Spectrom Rev*. 26: 51–78.

Fernie, A.R., Aharoni, A., Willmitzer, L., Stitt, M., Tohge, T., Kopka, J., et al. (2011) Recommendations for reporting metabolite data. *Plant Cell*. 23: 2477–2482.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol*. 18: 1157–1161.

Frank, K., Garr, L., Li, J., and Cerra, F. (1987) Overcoming Retention Time Shifts as a

Source of Error in HPLC Analysis. *J Chromatogr Sci*. 25.

Fridman, E., and Pichersky, E. (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Curr Opin Plant Biol*. 8: 242–248.

Gazzarrini, S., Lejay, L., Gojon, A., Ninnemann, O., Frommer, W.B., and von Wirén, N. (1999) Three functional transporters for constitutive, diurnally regulated, and starvation-induced uptake of ammonium into Arabidopsis roots. *Plant Cell*. 11: 937–948.

Gika, H., Theodoridis, G., Mattivi, F., Vrhovsek, U., and Pappa-Louisi, A. (2012) Hydrophilic interaction ultra performance liquid chromatography retention prediction under gradient elution. *Anal Bioanal Chem*. 404: 701–9.

Glajch, J.L., Kirkland, J.J., and Snyder, L.R. (1982) Practical optimization of solvent selectivity in liquid-solid chromatography using a mixture-design statistical technique. *J Chromatogr A*. 238: 269–280.

Glajch, J.L., Kirkland, J.J., Squire, K.M., and Minor, J.M. (1980) Optimization of solvent strength and selectivity for reversed-phase liquid chromatography using an interactive mixture-design statistical technique. *J Chromatogr A*. 199: 57–79.

Greco, G., and Letzel, T. (2013) Main Interactions and Influences of the Chromatographic Parameters in HILIC Separations. *J Chromatogr Sci*. 51: 684–693.

Gregor, H.P., and Collins, F.C. (1951) Studies on ion-exchange resins. III. Diffusion of neutral molecules in a sulfonic acid cation-exchange resin. *J Colloid Sci*. 6: 1951.

Gritti, F., and Guiochon, G. (2014) Calculated and experimental chromatograms for distorted gradients and non-linear solvation strength retention models. *J Chromatogr A*. 1356: 96–104.

Gruswitz, F., Chaudhary, S., Ho, J.D., Schlessinger, A., Pezeshki, B., Ho, C.M., et al. (2010) Function of human Rh based on structure of RhCG at 2.1 A. *Proc Natl Acad Sci U S A*. 107: 9638–9643.

Gu, R., Duan, F., An, X., Zhang, F., Von Wirén, N., and Yuan, L. (2013) Characterization of AMT-mediated high-affinity ammonium uptake in roots of maize (Zea mays L.). *Plant Cell Physiol*. 54: 1515–1524.

Guo, H., Wang, N., Mcdonald, T.R., Reinders, A., and Ward, J.M. (2018) MpAMT1;2 from Marchantia polymorpha is a High-Affinity, Plasma Membrane Ammonium Transporter. 0: 1–9.

Hackette, S.L., Skye, G.E., Burton, C., and Segel, I.H. (1970) Characterization of an ammonium transport system in filamentous fungi with methylammonium-14C as the substrate. *J Biol Chem*. 245: 4241–4250.

Harmon, A.C., Gribskov, M., and Harper, J.F. (2000) CDPKs–a kinase for every Ca 2+ signal? *Trends Plant Sci*. 5: 154–159.

Hemström, P., and Irgum, K. (2006) Hydrophilic interaction chromatography, Journal of Separation Science.

Henion, J., Brewer, E., and Rule, G. (1998) Peer Reviewed: Sample Preparation for LC/MS/MS: Analyzing Biological and Environmental Samples. *Anal Chem News*

Featur. 650–656.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., et al. (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*. 101: 10205–10210.

Hu, H.C., Wang, Y.Y., and Tsay, Y.F. (2009) AtCIPK8, a CBL-interacting protein kinase, regulates the low-affinity phase of the primary nitrate response. *Plant J*. 57: 264–278.

Ishizaki, K., Nishihama, R., Ueda, M., Inoue, K., Ishida, S., Nishimura, Y., et al. (2015) Development of gateway binary vector series with four different selection markers for the liverwort marchantia polymorpha. *PLoS One*. 10: 1–13.

Jandera, P., and Guiochon, G. (1991) Effect of the sample solvent on band profiles in preparative liquid chromatography using non-aqueous reversed-phase high-performance liquid chromatography. *J Chromatogr A*. 588: 1–14.

Javelle, A., Thomas, G., Marini, A.-M., Krämer, R., and Merrick, M. (2005) In vivo functional characterization of the Escherichia coli ammonium channel AmtB: evidence for metabolic coupling of AmtB to glutamine synthetase. *Biochem J*. 390: 215–222.

Jellum, E. (1977) Profiling of human body fluids in healthy and diseased states using gas chromatography and mass spectrometry, with special reference to organic acids. *J Chromatogr B*. 143: 427–462.

Joe P. Foley, Crow, J.A., Thomas, B.A., and Zamora, M. (1989) Unavoidable flow-rate errors in high-performance liquid chromatography. *J Chromatogr*. 478: 287–309.

Kanehisa, M., and Goto, S. (2000) KEGG : Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 28: 27–30.

Keunchkarian, S., Reta, M., Romero, L., and Castells, C. (2006) Effect of sample solvent on the chromatographic peak shape of analytes eluted under reversed-phase liquid chromatogaphic conditions. *J Chromatogr A*. 1119: 20–28.

Khachik, F., Beecher, G.R., Vanderslice, J.T., and Furrow, G. (1988) Liquid chromatographic artifacts and peak distortion: sample-solvent interactions in the separation of carotenoids. *Anal Chem*. 60: 807–811.

Khademi, S. (2004) Mechanism of Ammonia Transport by Amt/MEP/Rh: Structure of AmtB at 1.35 A. *Science (80-)*. 305: 1587–1594.

Khademi Shahram, O'Connell III Joseph, Remin Jonathan, Robles-Colmenares Yaneth, Miercke Larry J.W., S.R.M. (2004) Mechanism of Ammonia Transport by Amt / MEP / Rh : Structure of AmtB at 1 . 35 Å Author (s): Shahram Khademi , Joseph O â€™ Connell III , Jonathan Remis , Yaneth Robles-. *Science*. 305: 1587–1594.

Knight, M.R., Campbell, A.K., Smith, S.M., and Trewavas, A.J. (1991) . Nature.

Knox, J.H., and Kaliszan, R. (1985) Theory of solvent disturbance peaks and experimental determination of thermodynamic dead-volume in column liquid chromatography. *Chromatographia*. 349: 211–234.

Kolukisaoglu, U. (2004) Calcium Sensors and Their Interacting Protein Kinases:

Genomics of the Arabidopsis and Rice CBL-CIPK Signaling Networks. *Plant Physiol*. 134: 43–58.

Kronzucker, H.J., Siddiqi, M.Y., and Glass, A. (1996) Kinetics of NH4+ Influx in Spruce. *Plant Physiol*. 110: 773–779.

Kudla, J., Xu, Q., Harter, K., Gruissem, W., and Luan, S. (1999) Genes for calcineurin B-like proteins in Arabidopsis are differentially regulated by stress signals. *Proc Natl Acad Sci U S A*. 96: 4718–4723.

Lamoureux, G., Klein, M.L., and Berneche, S. (2007) A stable water chain in the hydrophobic pore of the AmtB ammonium transporter. *Biophys J*. 92: L82-4.

Lan, W.Z., Lee, S.C., Che, Y.F., Jiang, Y.Q., and Luan, S. (2011) Mechanistic analysis of AKT1 regulation by the CBL-CIPK-PP2CA interactions. *Mol Plant*. 4: 527–536.

Lanquar, V., Loque, D., Hormann, F., Yuan, L., Bohner, A., Engelsberger, W.R., et al. (2009) Feedback Inhibition of Ammonium Uptake by a Phospho-Dependent Allosteric Mechanism in Arabidopsis. *Plant Cell*. 21: 3610–3622.

Lauter, F.R., Ninnemann, O., Bucher, M., Riesmeier, J.W., and Frommer, W.B. (1996) Preferential expression of an ammonium transporter and of two putative nitrate transporters in root hairs of tomato. *Proc Natl Acad Sci U S A*. 93: 8139–8144.

Layne, J., Farcas, T., Rustamov, I., and Ahmed, F. (2001) Volume-load capacity in fast-gradient liquid chromatography: Effect of sample solvent composition and injection volume on chromatographic performance. *J Chromatogr A*. 913: 233–242.

Li, L., Kim, B.-G., Cheong, Y.H., Pandey, G.K., and Luan, S. (2006) A Ca2+ signaling

pathway regulates a K+ channel for low-K response in Arabidopsis. *Proc Natl Acad Sci*. 103: 12625–12630.

Li, R., Zhang, J., Wei, J., Wang, H., Wang, Y., and Ma, R. (2009) Functions and mechanisms of the CBL-CIPK signaling system in plant response to abiotic stress. *Prog Nat Sci*. 19: 667–676.

Lin, C.Y., Viant, M.R., and TJeerdema, R.S. (2006) Metabolomics : Methodologies and applications in the environmental sciences. *Pestic Sci*. 31: 245–251.

Linden, J.C., and Lawhead, C.L. (1975) Liquid Chromatography of Saccharides. *Chromatographia*. 105: 125–133.

Liu, J., and Zhu, J.-K.J. (1998) A calcium sensor homologue required for plant salt tolerance. *Sci*. 280: 1943–1945.

Loqué, D., Lalonde, S., Looger, L.L., Von Wirén, N., and Frommer, W.B. (2007) A cytosolic trans-activation domain essential for ammonium uptake. *Nature*. 446: 195–198.

Loqué, D., Mora, S.I., Andrade, S.L.A., Pantoja, O., and Frommer, W.B. (2009) Pore mutations in ammonium transporter AMT1 with increased electrogenic ammonium transport activity. *J Biol Chem*. 284: 24988–24995.

Lorenz, M.C., Heitman, J., Alspaugh, J., Perfect, J., Heitman, J., Andre, B., et al. (1998) The MEP2 ammonium permease regulates pseudohyphal differentiation in Saccharomyces cerevisiae. *EMBO J*. 17: 1236–1247.

De Luca, V., and St Pierre, B. (2000) The cell and developmental biology of alkaloid

biosynthesis. *Trends Plant Sci*. 5: 168–173.

Ludewig, U. (2004) Electroneutral ammonium transport by basolateral rhesus B glycoprotein. *J Physiol*. 559: 751–759.

Ludewig, U., von Wiren, N., and Frommer, W.B. (2002) Uniport of NH4+ by the root hair plasma membrane ammonium transporter LeAMT1;1. *J Biol Chem*. 277: 13548–13555.

Ludewig, U., Von Wiren, N., and Frommer, W.B. (2002) Uniport of NH4+ by the root hair plasma membrane ammonium transporter LeAMT1;1. *J Biol Chem*. 277: 13548–13555.

Marini, A., Vissers, S., Urrestarazu, A., and Andre, B. (1994) Cloning and expression of the MEP1 gene encoding an ammonium transporter in Saccharomyces cerevisiae. *EMBO J*. 13: 3456–3463.

Marini, A.M., Matassi, G., Raynal, V., André, B., Cartron, J.P., and Chérif-Zahar, B. (2000) The human Rhesus-associated RhAG protein and a kidney homologue promote ammonium transport in yeast. *Nat Genet*. 26: 341–344.

Marini, A.M., Vissers, S., Urrestarazu, A., and Andre, B. (1994) Cloning and expression of the MEP1 gene encoding an ammonium transporter in Saccharomyces cerevisiae. *Embo J*. 13: 3456–3463.

Matuszewski, B.K. (2006) Standard line slopes as a measure of a relative matrix effect in quantitative HPLC-MS bioanalysis. *J Chromatogr B Anal Technol Biomed Life Sci*. 830: 293–300.

Matuszewski, B.K., and Constanzer, M.L. (1998) Matrix Effect in Quantitative LC / MS / MS Analyses of Biological Fluids : A Method for Determination of Finasteride in Human Plasma at Picogram Per Milliliter Concentrations Matrix Effect in Quantitative LC / MS / MS Analyses of Biological Fluids : A Met. *Anal Chem*. 70: 882–889.

Matuszewski, B.K., Constanzer, M.L., and Chavez-Eng, C.M. (2003) Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. *Anal Chem*. 75: 3019–3030.

Mayer, M., Dynowski, M., and Ludewig, U. (2006) Ammonium ion transport by the AMT/Rh homologue LeAMT1;1. *Biochem J*. 396: 431–437.

Mayer, M., and Ludewig, U. (2006) Role of AMT1;1 in NH4+ acquisition in Arabidopsis thaliana. *Plant Biol*. 8: 522–528.

McDonald, T.R., Dietrich, F.S., and Lutzoni, F. (2012) Multiple horizontal gene transfers of ammonium transporters/ammonia permeases from prokaryotes to eukaryotes: Toward a new functional and evolutionary classification. *Mol Biol Evol*. 29: 51–60.

McDonald, T.R., and Ward, J.M. (2016) Evolution of Electrogenic Ammonium Transporters (AMTs). *Front Plant Sci*. 7: 352.

Mei, H., Hsieh, Y., Nardo, C., Xu, X., Wang, S., Ng, K., et al. (2003) Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: Application to drug discovery. *Rapid Commun Mass Spectrom*. 17: 97–103.

De Michele, R., Loqué, D., Lalonde, S., and Frommer, W.B. (2012) Ammonium and urea transporter inventory of the selaginella and physcomitrella genomes. *Front Plant Sci*. 3: 62.

Mishra, M., Rana, C., De Wit, A., and Martin, M. (2013) Influence of a strong sample solvent on analyte dispersion in chromatographic columns. *J Chromatogr A*. 1297: 46–55.

Neue, U.D., and Kuss, H.J. (2010) Improved reversed-phase gradient retention modeling. *J Chromatogr A*. 1217: 3794–3803.

Neuhaus, G., Bowler, C., Kern, R., and Chua, N.H. (1993) Calcium/calmodulin-dependent and -independent phytochrome signal transduction pathways. *Cell*. 73: 937–952.

Neuhäuser, B., Dynowski, M., and Ludewig, U. (2009) Channel-like NH3 flux by ammonium transporter AtAMT2. *FEBS Lett*. 583: 2833–2838.

Neuhäuser, B., Dynowski, M., and Ludewig, U. (2015) Switching substrate specificity of AMT/MEP/ Rh proteins. *Channels*. 8: 496–502.

Neuhauser, B., Dynowski, M., Mayer, M., and Ludewig, U. (2007) Regulation of NH4+ Transport by Essential Cross Talk between AMT Monomers through the Carboxyl Tails. *Plant Physiol*. 143: 1651–1659.

Neuhäuser, B., and Ludewig, U. (2014) Uncoupling of ionic currents from substrate transport in the plant ammonium transporter atamt1;2. *J Biol Chem*. 289: 11650–11655.

Nikitas, P., Pappa-Louisi, a., Agrafiotou, P., and Mansour, a. (2011) Multilinear gradient elution optimization in reversed-phase liquid chromatography based on logarithmic retention models: Application to separation of a set of purines, pyrimidines and nucleosides. *J Chromatogr A*. 1218: 5658–5663.

Ninnemann, O., Jauniaux, J.C., and Frommer, W.B. (1994a) Identification of a high affinity NH4+ transporter from plants. *EMBO J*. 13: 3464–3471.

Ninnemann, O., Jauniaux, J.C., and Frommer, W.B. (1994b) Identification of a high affinity NH4+ transporter from plants. *EMBO J*. 13: 3464–3471.

Ortiz-Ramirez, C., Mora, S.I., Trejo, J., and Pantoja, O. (2011a) PvAMT1;1, a highly selective ammonium transporter that functions as H +/NH 4 + symporter. *J Biol Chem*. 286: 31113–31122.

Ortiz-Ramirez, C., Mora, S.I., Trejo, J., and Pantoja, O. (2011b) PvAMT1;1, a highly selective ammonium transporter that functions as H +/NH 4+ symporter. *J Biol Chem*. 286: 31113–31122.

Paul G. Boswell, Dwight R. Stoll, Peter W. Carr, Megan L. Nageld, Mark F. Vitha, A., and Mabbott, G.A. (2013) An Advanced, Interactive, High-Performance Liquid Chromatography Simulator and Instructor Resources. *J Chem Educ*. 90: 198–202.

Peng, B., Kuo, M.-Y., Yang, P., Hewitt, J.T., and Boswell, P.G. (2014) A practical methodology to measure unbiased gas chromatographic retention factor vs. temperature relationships. *J Chromatogr A*. 1374: 207–215.

Pita, J.C.L.R., Gomes, I.F., dos Santos, S.G., Tavares, J.F., da Silva, M.S., de F??tima

Formiga Melo Diniz, M., et al. (2014) Matrix effect and optimization of LC-MSn determination of trachylobane-360 in mice blood. *J Pharm Biomed Anal*. 100: 262–270.

Pitts, R.J., Derryberry, S.L., Pulous, F.E., and Zwiebel, L.J. (2014) Antennal-expressed ammonium transporters in the malaria vector mosquito Anopheles gambiae. *PLoS One*. 9.

Pitts, R.J., Derryberry, S.L., Pulous, F.E., and Zwiebel, L.J. (2014) Antennal-expressed ammonium transporters in the malaria vector mosquito Anopheles gambiae. *PLoS One*. 9: e111858.

Poole, D.B.C.E. (1997) Influence of Solvent Effects on Retention for a Porous Polymer Sorbent in Reversed Phase Liquid Chromatography. *Chromatographia*. 46: 381–398.

Qiu, Q.-S., Guo, Y., Dietrich, M.A., Schumaker, K.S., and Zhu, J.-K. (2002) Regulation of SOS1, a plasma membrane Na+/H+ exchanger in Arabidopsis thaliana, by SOS2 and SOS3. *Proc Natl Acad Sci*. 99: 8436–8441.

Reichardt, C., and Eds, A. (2011) Solvents and Solvent Effects in Organic Chemistry.

Reinders, A., Sivitz, A.B., Starker, C.G., Gantt, J.S., and Ward, J.M. (2008) Functional analysis of LjSUT4, a vacuolar sucrose transporter from Lotus japonicus. *Plant Mol Biol*. 68: 289–299.

Ripoche, P., Bertrand, O., Gane, P., Birkenmeier, C., Colin, Y., and Cartron, J.P. (2004) Human Rhesus-associated glycoprotein mediates facilitated transport of NH(3) into

red blood cells. *Proc Natl Acad Sci U S A*. 101: 17222–17227.

Rohrschneider, L. (1966) Eine Methode Zur Charakterisierung Von Gaschromatographischen Trennflussigkeiten. *Chromatographia*. 22: 6–22.

Roon, R.J., Even, H.L., Dunlop, P., and Larimore, F.L. (1975) Methylamine and ammonia transport in Saccharomyces cerevisiae. *J Bacteriol*. 122: 502–503.

Saier, M.H., Eng, B.H., Fard, S., Garg, J., Haggerty, D.A., Hutchinson, W.J., et al. (1999) Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim Biophys Acta - Rev Biomembr*. 1422: 1–56.

Saint-Marcoux, D., Proust, H., Dolan, L., and Langdale, J.A. (2015) Identification of reference genes for real-time quantitative PCR experiments in the liverwort marchantia polymorpha. *PLoS One*. 10: 1–14.

Sauter, H., Lauer, M., and Fritsch, H. (1991) Metabolic profiling of plants--a new diagnostic technique. *ACS Symp Ser Am Chem Soc*. 288–299.

Scherzer, S., Krol, E., Kreuzer, I., Kruse, J., Karl, F., von Rüden, M., et al. (2013) The Dionaea muscipula ammonium channel DmAMT1 provides NH_4^+ uptake associated with Venus flytrap's prey digestion. *Curr Biol*. 23: 1649–57.

Shalliker, R.A., and Guiochon, G. (2009) Understanding the importance of the viscosity contrast between the sample solvent plug and the mobile phase and its potential consequence in two-dimensional high-performance liquid chromatography. *J Chromatogr A*. 1216: 787–793.

Shalliker, R.A., and Guiochon, G. (2010) Solvent viscosity mismatch between the solute

plug and the mobile phase: considerations in the applications of two-dimensional HPLC. *Analyst*. 135: 222–9.

Shi, J., Kim, K.N., Ritz, O., Albrecht, V., Gupta, R., Harter, K., et al. (1999) Novel protein kinases associated with calcineurin B-like calcium sensors in Arabidopsis. *Plant Cell*. 11: 2393–2405.

Shimamura, M. (2016) Marchantia polymorpha: Taxonomy, phylogeny and morphology of a model system. *Plant Cell Physiol*. 57: 230–256.

Shinoda, K., Tomita, M., and Ishihama, Y. (2008) Aligning LC peaks by converting gradient retention times to retention index of peptides in proteomic experiments. *Bioinformatics*. 24: 1590–1595.

Snyder, L.R. (1974) Classification of the solvent properties of common liquids. *J Chromatogr A*. 92: 223–230.

Søgaard, R., Alsterfjord, M., MacAulay, N., and Zeuthen, T. (2009) Ammonium ion transport by the AMT/Rh homolog TaAMT1;1 is stimulated by acidic pH. *Pflugers Arch Eur J Physiol*. 458: 733–743.

Sohlenkamp, C., Shelden, M., Howitt, S., and Udvardi, M. (2000) Characterization of Arabidopsis AtAMT2, a novel ammonium transporter in plants. *FEBS Lett*. 467: 273–278.

Sohlenkamp, C., Shelden, M., Howitt, S., and Udvardi, M. (2000) Characterization of Arabidopsis AtAMT2, a novel ammonium\ntransporter in plants. *FEBS Lett*. 467: 273–278.

Soupene, E., He, L., Yan, D., and Kustu, S. (1998) Ammonia acquisition in enteric bacteria: Physiological role of the ammonium/methylammonium transport B (AmtB) protein. *Proc Natl Acad Sci*. 95: 7030–7034.

Soupene, E., Ramirez, R.M., and Kustu, S. (2001) Evidence that Fungal MEP Proteins Mediate Diffusion of the Uncharged Species NH3 across the Cytoplasmic Membrane. *Mol Cell Biol*. 21: 5733–5741.

Straub, T., Ludewig, U., and Neuhäuser, B. (2017) The Kinase CIPK23 Inhibits Ammonium Transport in Arabidopsis thaliana. *Plant Cell*. 29: 409–422.

Sun, Y., Reinders, A., Lafleur, K.R., Mori, T., and Ward, J.M. (2010) Transport activity of rice sucrose transporters OsSUT1 and OsSUT5. *Plant Cell Physiol*. 51: 114–122.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*. 408: 796–815.

Trethewey, R.N., Krotzky, A.J., and Willmitzert, L. (1999) Metabolic profiling: a rosetta stone for genomics? *Curr Opin Plant Biol*. 2: 83–85.

Trufelli, H., Palma, P., Famiglini, G., and Cappiello, A. (2011) an overview of matrix effects in liquid chromatography-mass spectorometry. *Mass Spectrom Rev*. 30: 491–509.

Tsuda, M. (1963) Schotten-Baumann Esterification of Poly (viny1 alcohol). I . *From Gov Chem Ind Res Institute, Tokyo, Hiratsuka, Kanagawa, Japan*.

Tyteca, E., Guillarme, D., and Desmet, G. (2014a) Use of individual retention modeling for gradient optimization in hydrophilic interaction chromatography : Separation of

nucleobases and nucleosides. *J Chromatogr A*. 1368: 125–131.

Tyteca, E., Périat, A., Rudaz, S., Desmet, G., and Guillarme, D. (2014b) Retention modeling and method development in hydrophilic interaction chromatography. *J Chromatogr A*. 1337: 116–27.

Wacker, T., Garcia-Celma, J.J., Lewe, P., and Andrade, S.L. a (2014) Direct observation of electrogenic NH4+ transport in ammonium transport (Amt) proteins. *Proc Natl Acad Sci U S A*. 111: 9995–10000.

Wang, M., Mallette, J., and Parcher, J.F. (2009) Interconversion of gradient and isocratic retention data in reversed-phase liquid chromatography: Effect of the uptake of eluent modifier on the retention of analytes. *J Chromatogr A*. 1216: 8630–8635.

Wang, M.Y., Siddiqi, M.Y., Ruth, T.J., and Class, A.D.M. (1993) Ammonium Uptake by Rice Roots 1. *Plant Physiol*. 103: 1259–1267.

Wang, M.Y., Siddiqi, M.Y., Ruth, T.J., and Class, A.D.M. (1994) Ammonium Uptake by Rice Roots ’. *Plant Physiol*. 104: 899–906.

Wang, N., and Boswell, P.G. (2017) Accurate prediction of retention in hydrophilic interaction chromatography by back calculation of high pressure liquid chromatography gradient profiles. *J Chromatogr A*.

Weise, A. (2000) A New Subfamily of Sucrose Transporters, SUT4, with Low Affinity/High Capacity Localized in Enucleate Sieve Elements of Plants. *Plant Cell Online*. 12: 1345–1356.

Wernisch, S., and Pennathur, S. (2016) Evaluation of coverage , retention patterns , and

selectivity of seven liquid chromatographic methods for metabolomics. *Anal Bioanal Chem*. 408: 6079–6091.

Xu, J., Li, H.D., Chen, L.Q., Wang, Y., Liu, L.L., He, L., et al. (2006) A Protein Kinase, Interacting with Two Calcineurin B-like Proteins, Regulates K+Transporter AKT1 in Arabidopsis. *Cell*. 125: 1347–1360.

Yang, S., Hao, D., and Cong, Y. (2015) The rice OsAMT1 ; 1 is a proton-independent feedback regulated ammonium transporter. *plant cell Rep*. 34: 321–330.

Yuan, L., Graff, L., Loqué, D., Kojima, S., Tsuchiya, Y.N., Takahashi, H., et al. (2009) AtAMT1;4, a pollen-specific high-affinity ammonium transporter of the plasma membrane in arabidopsis. *Plant Cell Physiol*. 50: 13–25.

Yuan, L., Gu, R., Xuan, Y., Smith-Valle, E., Loqué, D., Frommer, W.B., et al. (2013) Allosteric regulation of transport activity by heterotrimerization of Arabidopsis ammonium transporter complexes in vivo. *Plant Cell*. 25: 974–84.

Zapata, M., and Garrido, J.L. (1991) Influence of injection conditions in reversed-phase high-performance liquid chromatography of chlorophylls and carotenoids. *Chromatographia*. 31: 589–594.

Zellner, A., Bicchi, C., Dugo, P., Rubiolo, P., Dugo, G., Mondello, L., et al. (2008) Linear retention indices in gas chromatographic analysis : a review. *Flavour Fragr J*. 23: 297–314.

Zheng, L., Kostrewa, D., Berneche, S., Winkler, F.K., and Li, X.D. (2004) The mechanism of ammonia transport based on the crystal structure of AmtB of

Escherichia coli. *Proc Natl Acad Sci U S A.* 101: 17090–17095.

Appendix: the regulation of MpAMTs by phosphorylation

1. CIPKs in *Marchantia polymorpha*

To understand the regulation of MpAMTs (ammonium transporters in *M. polymorpha*) and to test whether MpAMTs are also regulated by CIPK-CBL complexes, similar to the situation in Arabidopsis, we initially set out to identify candidate sequences in the *Marchantia* genome. Blasting the CIPK23 protein sequence against the genome of *M. polymorpha* returned 101 significant hits. Of these hits, most of them are protein kinases. Since the NAF domain is sufficient for interaction with CBLs it is used to identify members of the CIPK family (Loqué et al., 2007) Only two candidates in Marchantia (MpCIPK1, MpCIPK2) contain a NAF domain. Fig. 29 is an alignment of the NAF domains of all 25 CIPKs from *Arabidopsis* and the two putative CIPKs from *Marchantia*. Compared against CIPKs in *A. thaliana*, the two putative CIPKs in *M. polymorpha* shared 16 conserved amino acids of the NAF domain (24 amino acids).

```
 CIPK1: (313) DSPTIINAFQLIG-MSSFLDLSGFF (336)
 CIPK2: (309) PRLATLNAFDIIA-LSTGFGLAGLF (332)
 CIPK3: (307) EQPAAINAFEIIS-MSRGLNLENLF (330)
 CIPK4: (302) KSSNAITAFDLIS-LSSGLDLSGLF (325)
 CIPK5: (307) ISPKFFNAFEFISSMSSGFDLSSLF (330)
 CIPK6: (310) EETETLNAFHIIA-LSEGFDLSPLF (333)
 CIPK7: (302) SSVNSITAFDLIS-LSSGLDLSGLF (325)
 CIPK8: (302) TGPLTLNAFDLII-LSQGLNLATLF (325)
 CIPK9: (314) EKPVSMNAFELIS-SSSEFSLENLF (337)
CIPK10: (322) DEPTNLNAFDLIA-LSAGFDLAGLF (345)
CIPK11: (302) EAVKSLNAFDLIS-YSSGLDLSGLF (325)
CIPK12: (336) PRPASLNAFDIIS-FSQGFDLSGLF (359)
CIPK13: (366) PRPASLNAFDILS-FS---DLSGLF (386)
CIPK14: (305) MGARRMNAFDIIS-GSPGFNLSGLF (328)
CIPK15: (299) KRCINLNAFEIIS-LSTGFDLSGLF (322)
CIPK16: (317) ASPRSFNAFQFITSMSSGFDLSNLF (340)
```

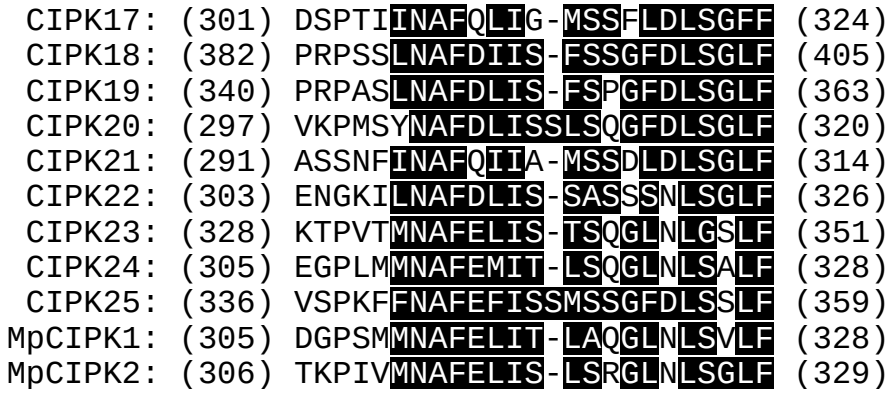


Figure 29. The alignment of NAF domain of CIPKs from Arabidopsis and Marchantia. The NAF domain consists of 24 amino acids, it was first characterized for CIPK1 in Arabidopsis. CIPK1-25 are from Arabidopsis and MpCIPK1-2 are from Marchantia. Black shading indicates conserved amino acids (identical or similar) in the NAF domain based on an alignment done using the MegAlign program of the DNASTAR Suite. NCBI protein accession numbers: CIPK1_AAG28776.1, CIPK2_AAF86506.1, CIPK3_AAF86507.1, CIPK4_AAG01367.1, CIPK5_AAF86504.2, CIPK6_AAF86505.1, CIPK7_AAK16682.1, CIPK8_AAK16683.2, CIPK9_AAK16684.1, CIPK10_AAK16685.1, CIPK11_AAK16686.1, CIPK12_AAK16687.1, CIPK13_AAK16688.1, CIPK14_AAK16689.1, CIPK15_AAK16692.1, CIPK16_AAK50348.1, CIPK17_AAK64513.1, CIPK18_AAK59695.1, CIPK19_AAK50347.1, CIPK20_AAK61493.1, CIPK21_AAK59696.1, CIPK22_AAL47845.1, CIPK23_AAK61494.1, CIPK24_AAK72257.1, CIPK25_AED93402.1, MpCIPK1_Mapoly0025s0017, MpCIPK2_Mapoly0005s0039.

Besides using the NAF domain to classify CIPKs in *M. polymorpha*, we also created a phylogenetic tree to further analyze similarities of CIPKs in *M. polymorpha* and *A.*

thaliana. As shown in Fig. 30, the unrooted phylogenetic tree indicates MpCIPK1 (Mapoly0025s0017) and MpCIPK2 (Mapoly0005s0039) are more similar to CIPK24 and CIPK8 than to CIPK23. Several functions of CIPK24 and CIPK8 have been previously identified. A complex of CIPK24-CBL1/4/10 maintains Ca^{2+} homeostasis, ROS detoxification, and Na^{+} homeostasis through regulation of the Ca^{2+}/ H^{+} antiporter CAX1, the plasma membrane Na^{+}/H^{+} antiporter SOS1, and the tonoplast Na^{+}/H^{+} antiporter NHX1. CIPK8 is involved in early nitrate signaling through regulating nitrate transporter genes (Hu et al., 2009). It will be interesting to characterize the functions of the two putative CIPKs, MpCIPK1 and MpCIPK2 in *M. polymorpha*.

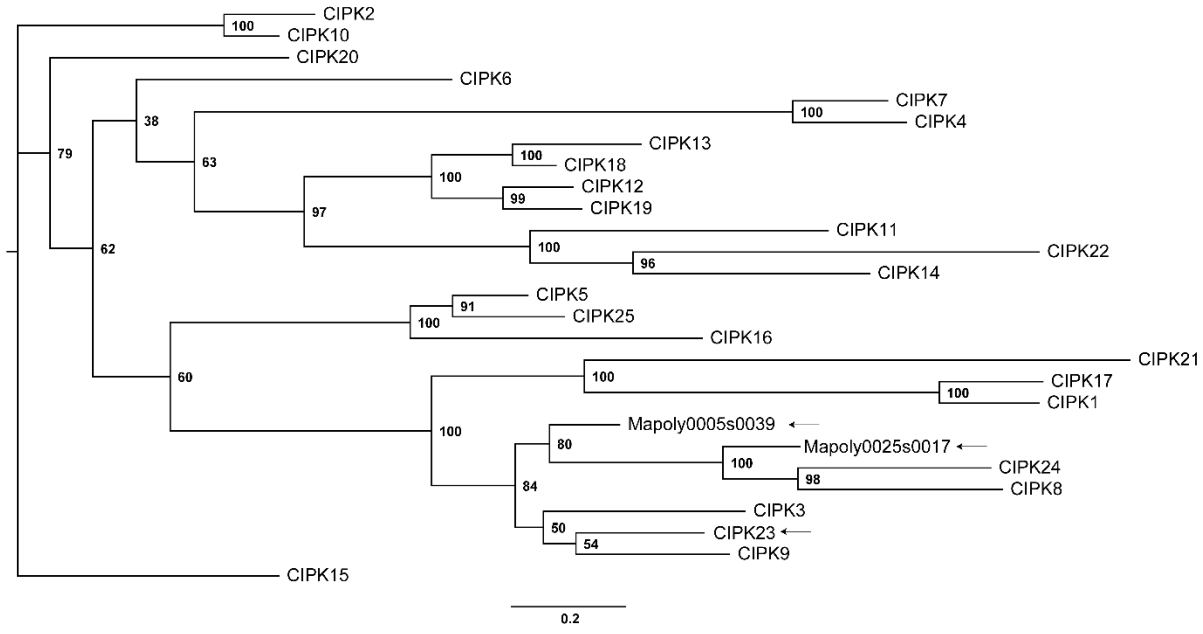


Figure 30. Phylogenetic tree of CIPKs from *M. polymorpha* and *A. thaliana*. Protein sequences were aligned using Clustal X and a tree was created using the Maximum Parsimony algorithm in Clustal X. See Fig. 29 for the NCBI accession numbers of the CIPKs.

2. CBLs in *Marchantia polymorpha*

Based on the BLAST results from *M. polymorpha*, the presence of the conserved NAF domain and the phylogenetic tree, *M. polymorpha* contains two potential CIPKs. In order to determine if there are also CBLs in *M. polymorpha* and how many, we used the protein sequence of AtCBL1 as a query in BLAST; 20 hits were returned. Of these hits, only three candidates showed high similarity to AtCBL1. They are Mapoly0134s0040.1, Mapoly0066s0053.1, and Mapoly0015s0061.1, they were designated MpCBL1, MpCBL2, and MpCBL3.

It is well known that CBLs share a conserved N-myristoylation motif at their N-terminus (MGCXXSK/T) (Li et al., 2009). We therefore analyzed the protein sequence of the three CBL candidates from *M. polymorpha*. As shown in Fig. 31, MpCBL1 contains the conserved N-myristoylation motif (MGCXXSK/T) just like AtCBL1. However, the predicted protein sequence of MpCBL2 contains additional amino acids upstream of the N-myristoylation motif. It is likely that this is due to an error in predicting the MpCBL2 protein sequence and that instead of starting at the first M, MpCBL2 actually starts at the second M, at position 35. Therefore, we decided to clone MpCBL2 starting at the second M, resulting in a sequence of 642 bp instead of 744 bp. Interestingly, the predicted sequence of MpCBL3 is completely lacking the N-myristoylation motif. However, it shares significant amino acid sequence similarities with other CBLs further downstream from the predicted start codon. To assure that this discrepancy is not due to an error in the reported genome sequence, we will re-sequence a portion of the genomic sequence upstream from the sequence shared with the other CBLs.

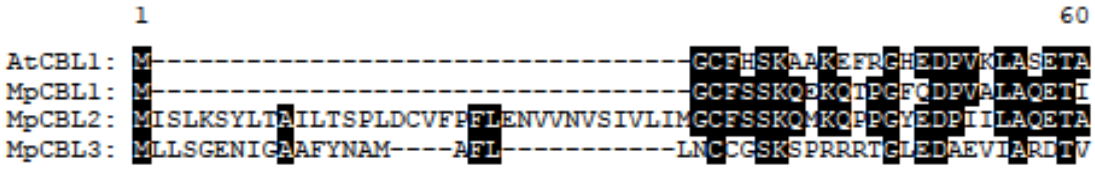


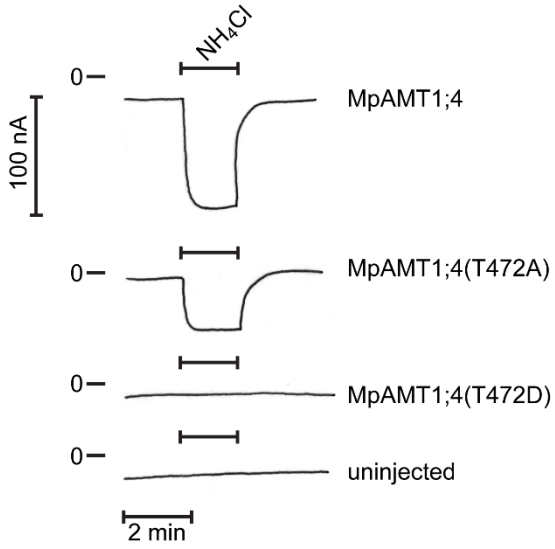
Figure 31. Protein alignment of CBL1 in Arabidopsis and CBL1-3 in Marchantia. Black shading indicates conserved amino acids within the first 60 amino acids of the CBLs. NCBI protein accession numbers for the CBLs: AtCBL1_AAC26008.1, MpCBL1_ Mapoly0134s0040.1, MpCBL2_Mapoly0066s0053.1, and MpCBL3_ Mapoly0015s0061.1.

3. The activity of MpAMTs after phosphorylation

In Arabidopsis, it has been shown that ammonium transporters are regulated by phosphorylation through a complex of CIPK23/CBL1. The activity of AMT1;1 and AMT1;2 in Arabidopsis is inhibited after they become phosphorylated (Straub et al., 2017). In 2007, Loque *et al.* identified a conserved amino acid (Thr 360 in AtAMT1;1) near the C-terminus of AMTs (Loqué et al., 2007). In *M. polymorpha,* based on our BLAST results, there are two putative CIPKs and three putative CBLs. These CIPKs and CBLs may work together to phosphorylate AMTs to turn on or off their activity, similar to the situation in Arabidopsis. To find out whether this is the case, we first determined if MpAMT1 activity would be changed by phosphorylation. We selected MpAMT1;2 and MpAMT1;4 for this study because of their higher affinities compared with other two MpAMTs we have characterized (MpAMT1;5, MpAMT1;3).

We created point mutants in MpAMT1;2 (T475A, D) and MpAMT1;4 (T472A, D) where Thr (T) was either replaced by phosphorylation mimic amino acid Asp (D) or by

Ala (A), which cannot be phosphorylated. These AMT variants were then expressed in Xenopus oocytes and their transport activities determined by TEVC. The results obtained from expressing MpAMT1;2 (T475A, D) have been published (Guo et al., 2018) and are included in Chapter 4, therefore here only the result from MpAMT1;4 (T472A, D) are shown. Replacing T474 with Ala, resulted in lower ammonium transport activity than the wild type, while the mutant MpAMT1;4 (T472D) was inactive (Fig. 32). These results are consistent with the interpretation that phosphorylation of the conserved threonine near the C-terminus of MpAMTs turns off the activity of MpAMT1;2 and MpAMT1;4. These findings are similar to those for AMT1;1 in *Arabidopsis* (Loqué et al., 2007).



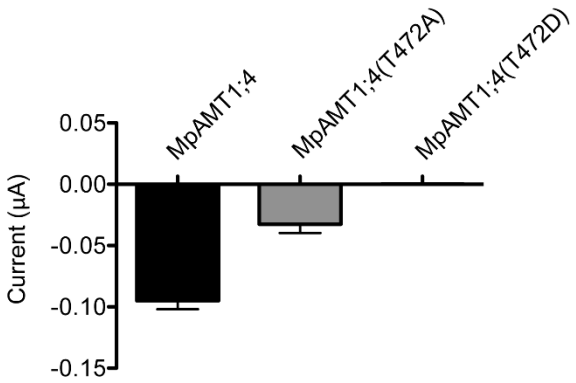


Figure 32. The activity of ammonium transporter MpAMT1;4, and its point mutants MpAMT1;2 (T475A, D), and MpAMT1;4 (T472A, D). *Xenopus* oocytes were perfused with K^+ - free Ringer solution at pH 5.6. The concentration of applied NH_4^+ was 500 µM. The membrane potential was held at –40 mV, oocytes were injected with MpAMT1;4 cRNA. Bars above each trace indicate the time when substrates were applied, the downward deflections indicate inward current. Error bars are SD (n=3 oocytes).