An integrated approach reveals how lipo-chitooligosaccharides interact with the lysin motif receptor-like kinase MtLYR3

Younes Bouchiba, Jérémy Esque, Ludovic Cottret, Maude Maréchaux, Mégane Gaston, Virginie Gasciolli, Jean Keller, Nico Nouwen, Djamel Gully, Jean-françois Arrighi, et al.

FULL-LENGTH PAPER

An integrated approach reveals how lipo-chitooligosaccharides interact with the lysin motif receptor-like kinase MtLYR3

Younes Bouchiba[1] | Jérémy Esque[1] | Ludovic Cottret[2] | Maude Maréchaux[2] | Mégane Gaston[2] | Virginie Gasciolli[2] | Jean Keller[3] | Nico Nouwen[4] | Djamel Gully[4] | Jean-François Arrighi[4] | Clare Gough[2] | Benoit Lefebvre[2] 🄖 | Sophie Barbe[1] 🄖 | Jean-Jacques Bono[2]

[1]TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

[2]LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France

[3]Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Castanet-Tolosan, France

[4]IRD, Laboratoire des Symbioses Tropicales et Méditerranéennes (LSTM), UMR IRD/SupAgro/INRAE/UM/CIRAD, Montpellier, France

Correspondence
Benoit Lefebvre, LIPME, INRAE, 24 Chemin de Borde Rouge, Auzeville, CS 52627, 31326 Castanet Tolosan cedex, France.
Email: benoit.lefebvre@inrae.fr

Sophie Barbe, TBI, Université de Toulouse, CNRS, INRAE, INSA, 135 avenue de Rangueil, 31077 Toulouse CEDEX 04. France.
Email: sophie.barbe@insa-toulouse.fr

Funding information
This work was supported by the ANR projects "WHEATSYM" (ANR-16-CE20-0025-01), "NICE CROPS" (ANR-14-CE18-0008-01) and "DUALITY" (ANR-20-CE20-0017-01). This study is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-10-LABX-41) and of the "École Universitaire de Recherche (EUR)" TULIP-GS (ANR-18-EURE-0019).

Review Editor: Aitziber Cortajarena

Abstract

N-acetylglucosamine containing compounds acting as pathogenic or symbiotic signals are perceived by plant-specific Lysin Motif Receptor-Like Kinases (LysM-RLKs). The molecular mechanisms of this perception are not fully understood, notably those of lipo-chitooligosaccharides (LCOs) produced during root endo-symbioses with nitrogen-fixing bacteria or arbuscular mycorrhizal fungi. In *Medicago truncatula*, we previously identified the LysM-RLK LYR3 (MtLYR3) as a specific LCO-binding protein. We also showed that the absence of LCO binding to LYR3 of the non-mycorrhizal *Lupinus angustifolius*, (LanLYR3), was related to LysM3, which differs from that of MtLYR3 by several amino acids and, particularly, by a critical tyrosine residue absent in LanLYR3. Here, we aimed to define the LCO binding site of MtLYR3 by using molecular modelling and simulation approaches, combined with site-directed mutagenesis and LCO binding experiments. 3D models of MtLYR3 and LanLYR3 ectodomains were built, and homology modelling and molecular dynamics (MD) simulations were performed. Molecular docking and MD simulation on the LysM3 identified potential key residues for LCO binding. We highlighted by steered MD simulations that in addition to the critical tyrosine, two other residues were important for LCO binding in MtLYR3. Substitution of these residues in LanLYR3-LysM3 by

those of MtLYR3-LysM3 allowed the recovery of high-affinity LCO binding in experimental radioligand-binding assays. An analysis of selective constraints revealed that the critical tyrosine has experienced positive selection pressure and is absent in some LYR3 proteins. These findings now pave the way to uncover the functional significance of this specific evolutionary pattern.

KEYWORDS

Aeschynomene spp, docking, free energy landscape, lipo-chitooligosaccharide, *Lupinus angustifolius*, lysin motif receptor-like receptor, *Medicago truncatula*, plant endosymbiosis, receptor ligand binding, steered molecular dynamics

1 | INTRODUCTION

Lipo-chitooligosaccharides (LCOs), called Myc-LCOs when they are produced by arbuscular mycorrhizal (AM) fungi, and nod factors (NFs) when they are produced by rhizobia, are involved in two agronomically and ecologically important symbioses: the AM and the legume nitrogen-fixing root nodule (RN) symbioses, respectively.[1] LCOs consist of four or five β-1,4-linked N-acetylglucosamine (GlcNAc) units, N-acylated by a fatty acid on the non-reducing terminal sugar. Chemical substitutions on the chitin backbone, length, and degree of unsaturation of the fatty acid chain characterize the NFs produced by each rhizobial strain and are important determinants of host legume specificity.[2] Myc-LCOs also show diversity of chemical substitutions although roles of these substitutions are not known. Other GlcNAc containing compounds such as chitin and peptidoglycan are important plant defense elicitors.[3] Recently, short- and long-chain chitooligosaccharides (COs) were proposed to play a role in the AM symbiosis, despite being considered for a long time as a plant defense elicitor.[4] The perception of these GlcNAc-based symbiotic or pathogenic signals is controlled by members of the plant-specific Lysin Motif Receptor-Like Kinase (LysM-RLK) family.[5] Perception by distinct LysM-RLKs results in appropriate cell responses leading to symbiont accommodation or plant immunity. LysM-RLKs consist of an ectodomain with three LysMs, each consisting of 40–50 amino acids, a transmembrane domain followed by an intracellular kinase domain. Therefore, chemically-related microbial signals are perceived by structurally related plant receptors, leading to distinction between friends and foes. One of the outstanding questions pointed out by Chiu and Paszkowski in a recent review is now to understand "What are the mechanisms/ligands/receptors for achieving signalling specificities-in immunity vs symbiosis; and AM symbiosis vs nitrogen-fixing RN symbiosis?".[6] To answer this complex and ambitious question, we first need to determine the molecular bases of ligand/receptor interaction.

The most advanced studies concern the interaction between COs and the LysM-RLK AtCERK1 in *Arabidopsis thaliana*, or the GPI-anchored LysM protein OsCEBiP in rice. Crystal structures of the ectodomains of these two receptors have been solved in complex with chitopentaose.[7,8] These were important breakthroughs and showed that, in both cases, the binding involves the second LysM and conserved positions of residues that bind the ligand. The crystal structure of a close homolog of AtCERK1 in *Lotus japonicus* (LjLYS6) has also been solved, but no co-crystallisation with COs was reported.[9]

For LCOs, and notably NFs, genetic approaches in the model legumes *Medicago truncatula* and *L. japonicus* have identified pairs of LysM-RLKs controlling the RN symbiosis: they are called MtNFP and MtLYK3 in *M. truncatula*, LjNFR5 and LjNFR1 in *L. japonicus*.[10–13] Genome sequencing revealed that these LysM-RLKs belong to a family of more than 20 members in these species, divided almost equally into two sub-families: the LYRs and the LYKs, to which belong MtNFP/LjNFR5 and MtLYK3/LjNFR1 respectively. LYRs differ from LYKs by their kinase domains, shown or predicted to be enzymatically inactive.[11] Mutation of either *LjNFR5* or *LjNFR1* completely abolishes nodulation, and binding to LjNFR5 and LjNFR1 of the NF of the symbiont of *L. Japonicus* (*Mesorhizobium loti*) has been shown.[14] Taken together, these results favor a model where these two LysM-RLKs form a hetero-dimeric receptor complex that perceives and transduces the NF signal.[15] Recently, the structure of MtLYK3 was solved[16] and helped to identify motif signatures in the first LysM of this class of LysM-RLKs that could discriminate LCOs from COs and could account for the specific perception of NFs by host plants. However, formal identification of the binding site(s) for NFs is still lacking since crystal structures of symbiotic LysM-RLKs in complex with NFs have not been reported.

In *M. truncatula*, we identified the LysM-RLK MtLYR3 and its orthologs in different legumes as high-affinity NF- and Myc-LCO binding proteins.[17,18] These

proteins belong to another phylogenetic group of the LYR subfamily compared with MtNFP/LjNFR5 (called LYRIII-A, for MtLYR3 and LYRI-A for MtNFP[19]). A clear phenotype for *Mtlyr3* mutants has so far been elusive, but a *L. japonicus* mutant (*LjLys12*) affected in the orthologous gene is more sensitive to the pathogenic oomycete *Phytophthora palmivora*.[20] Interestingly, among the legume MtLYR3 orthologs, only those of lupin species (*Lupinus angustifolius* and *Lupinus atlanticus*), rare examples of legume species unable to establish AM symbiosis, did not bind LCOs.[18] This peculiarity allowed us to identify the crucial role played by the third LysM in LCO binding to MtLYR3, and particularly that of a tyrosine residue (Y228), which is highly conserved among LCO-binding LYR3 proteins, but which is replaced by a Glutamine residue (Q224) in the *L. angustifolius* LYR3 protein (LanLYR3). Whereas the Y228Q exchange in MtLYR3-LysM3 resulted in a complete loss of LCO binding, the reverse exchange in LanLYR3-LysM3 did not produce any gain of binding, suggesting that other residues are also important.[18] The aim of the present work was to identify these additional residues to better understand LCO binding to LYR3 proteins and how this could have been lost in some plant species. By combining molecular modelling and molecular dynamics (MD) simulations with site directed mutagenesis and LCO binding experiments, we identified, in addition to the critical tyrosine, other key residues for LCO recognition. Furthermore, an analysis of selective constraints showed that residues under positive selection pressure

are mainly located in the LysM3, and include the tyrosine. The fact that this tyrosine is absent in LYR3 proteins of several unrelated legume and non-legume plant species raises questions about the possible convergent evolution of this loss, and the functions of tyrosine-containing and tyrosine-lacking LYR3 proteins.

2 | RESULTS

Due to the lack of structural data on the LYR3 protein family, molecular modelling, docking and MD simulation approaches were used in this study to better understand LCO recognition by MtLYR3.

2.1 | 3D-models of MtLYR3 and LanLYR3

3D structural models of the ectodomains of MtLYR3 and LanLYR3 (Figure 1a,b) were built using the I-TASSER server and three structural templates corresponding to the ectodomains of OsCEBiP (PDB 5JCE), AtCERK1 (PDB 4EBY) and LjLYS6 (PDB 5LS2). Although MtLYR3 and LanLYR3 share on average only 52% sequence similarity and 25% sequence identity with the sequences of the structural templates used, the models built by I-TASSER keep an overall LysM-RLK fold as indicated by the TM score above 0.8 (Figure S1). On average 90% of the backbone torsion angles of the generated LYR3
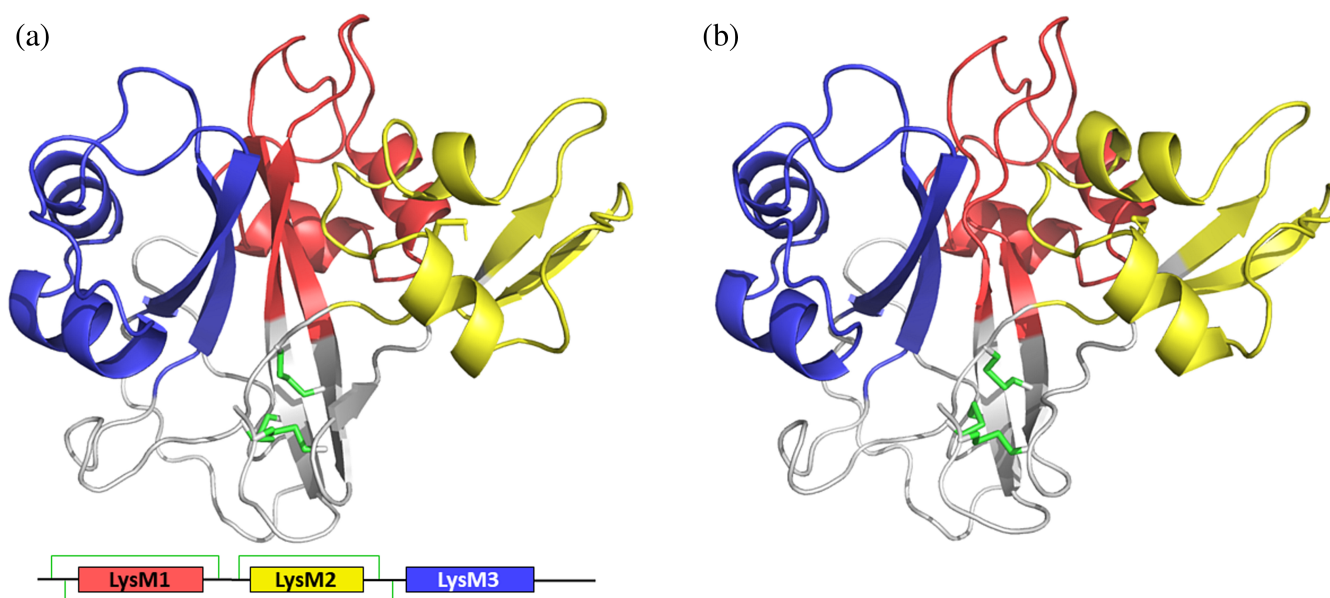
FIGURE 1 LysM-RLK structural models. 3D structural models of MtLYR3 (a) and LanLYR3 (b) receptor ectodomains, composed of three LysMs: LysM1 (colored in red), LysM2 (colored in yellow), and LysM3 (colored in blue). The three LysMs are linked by flexible regions (colored in gray). The overall protein architecture is stabilized by three disulfide bridges located in the linker regions (shown in green sticks in the 3D model and as connected lines in the schematic representation).
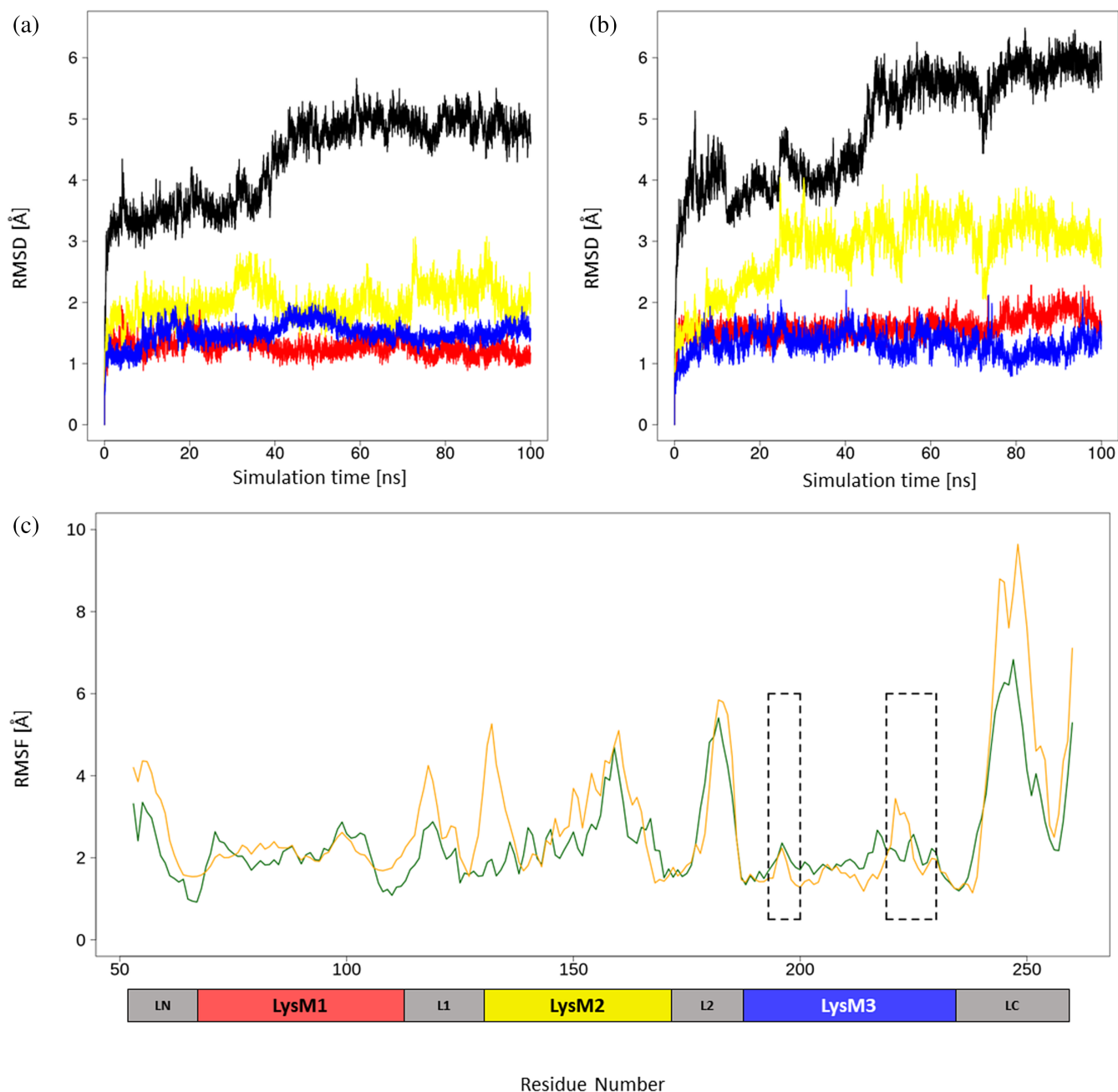
FIGURE 2 Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analysis from molecular dynamics (MD) simulations of MtLYR3 and LanLYR3 receptor ectodomains. RMSD of backbone atoms as a function of time were computed for MtLYR3 (a) and LanLYR3 (b) with respect to the starting conformation of the MD production phase. For each plot, the black curve corresponds to the RMSD of the whole protein while the red, yellow, and blue curves correspond to the RMSD of the LysM1, LysM2, and LysM3, respectively. RMSF based on backbone atoms (c) were computed for MtLYR3 (in green) and LanLYR3 (in orange) over the course of the simulation after RMS-fitted the MD snapshots onto the average structure of the MD simulation. Loops 1 and 2 are marked with dashed black rectangles.

models are located in favored regions of the Ramachandran plot, which indicates an overall good geometrical quality of the models (Figure S2a,b). Moreover, after structural superposition of MtLYR3 and LanLYR3 models on the templates, the minimum root mean square deviation (RMSD) values from the backbone were 0.56 and 1.15 Å, respectively, compared with the structure of

the ectodomain of OsCEBiP (PDB 5JCE). When only backbone atoms of the LysM3s are considered for the superposition and RMSD calculations, the smallest RMSD values were 0.41 Å for MtLYR3 and 0.62 Å for LanLYR3, compared with the ectodomain structure of AtCERK1 (PDB 4EBY). All these analyses indicated a good confidence on the quality of the 3D models

compared with the available structural information. As expected since the ectodomains of MtLYR3 and LanLYR3 share 86% and 73.9% of sequence similarity and identity, respectively, the structural models of MtLYR3 and LanLYR3 ectodomains are close in terms of fold. Indeed, the backbone RMSD values between MtLYR3 and LanLYR3 are 0.33 Å for the whole ectodomain and 0.74 Å for the LysM3s.

To relax and sample more stable conformations, MD simulations of 100 ns length were performed with the MtLYR3 and LanLYR3 ectodomain 3D models. To follow the model stability during these MD simulations, the backbone RMSD was computed as a function of time and taking the starting conformation as the reference (Figure 2a,b). For both ectodomains, the RMSD plots showed an evolution in two stages. During the first 3 ns of the MD simulations, the RMSD values quickly increased until 3.5 Å for MtLYR3 and by 4 Å for LanLYR3, and then fluctuated around this value for up to 40 ns of the simulations. A second jump of RMSD values was observed after 40 ns, leading to a second plateau until 100 ns at 5 Å for MtLYR3 (Figure 2a) and around 6 Å for LanLYR3 (Figure 2b). The variations of backbone RMSD around the mean values of each plateau were slightly larger for the ectodomain of LanLYR3 than that of MtLYR3. These results indicate that the protein structures underwent conformational changes during the course of the simulation. A detailed analysis of the simulations at a molecular level showed a particularly high mobility of linker regions between the LysMs that is clearly illustrated by the lower RMSD values when only considering LysMs in the calculations (Figure 2a,b). The backbone RMSD values of LysM1 and LysM3 never exceeded 2 Å and remained constant around 1.5 Å, suggesting that no important conformational change occurs. A higher increase of backbone RMSD values is observed for the LysM2, mostly in the LanLYR3 structure, reaching a final value of about 2.8 Å. These results are consistent with the Root Mean Square Fluctuation (RMSF) calculated on protein backbone atoms from simulation frames (Figure 2c). As expected, the highest mobility was observed for the linker regions that connect the LysMs in the MtLYR3 and LanLYR3 ectodomains. RMSF analysis also showed that LysM1 and LysM3 are the most stable in MtLYR3 and LanLYR3 amongst the three LysMs. Regarding RMSD and RMSF analyses, the LanLYR3 model appears to be globally more flexible than MtLYR3. Finally, these results are supported by detailed analyses through RMSD per residue along the MD simulation (Figure S3a,b). Indeed, linkers for both protein models and LysM2 only in the LanLYR3 model showed RMSD values higher than 9 Å. Interestingly, in spite of these high fluctuations, the main fold of LysMs (β-α-α-β)
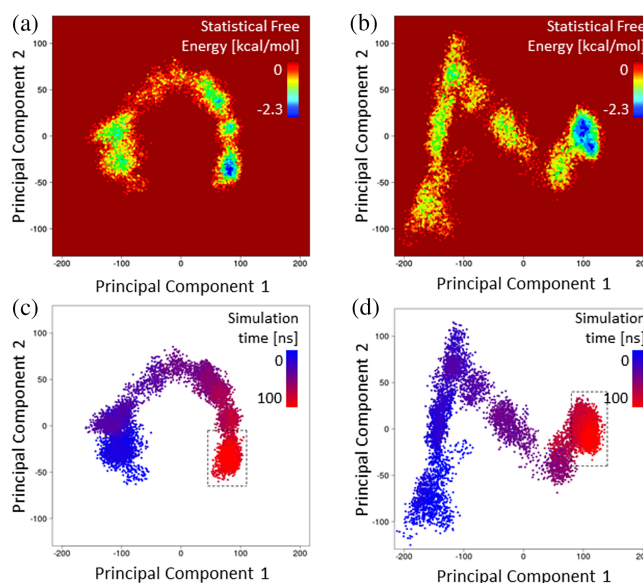

FIGURE 3 Free energy landscapes of MtLYR3 and LanLYR3 ectodomains. FEL of MtLYR3 (a, c) and LanLYR3 (b, d) were determined using the projection of the first and second principal components as reaction coordinates. MtLYR3 simulation first and second principal components account for respectively, 82% and 11% of the motions. LanLYR3 first two principal components account for, respectively, 86% and 10% of the motions. The free energy defined in Equation (1) is represented in the colour gradient shown in (a, b) and the simulation time is depicted in the color gradient shown in (c, d). Both simulations showed an evolution of the initial conformation in the bottom-left side of the landscapes toward a more densely sampled region around coordinates (PC1 [45,110] and PC2 [−65,-5]) for MtLYR3 and (PC1 [80,140] and PC2 [−40,40]) for LanLYR3. These regions are indicated by black rectangles in (c, d) and this basin is sampled through the 27 last ns of simulation for MtLYR3 and 16 last ns for LanLYR3.

is maintained during the whole simulation for both MtLYR3 and LanLYR3 models (Figure S3c,d).

To have a better insight into the stability and the convergence of MD simulations, free energy landscape was derived from principal component analysis (PCA) on the backbone atom coordinates for MtLYR3 and LanLYR3 models. The first two principal components (PCs) describe overall 93% and 96% of the motions for MtLYR3 and LanLYR3, respectively. The free energy landscape is related to the logarithm of the sampling probability (Equation 1). Interestingly, both structural models (MtLYR3 and LanLYR3) move away from the basins in which they were originally located to reach another, more stable, one (lowest energy) (Figure 3a,b). Regarding these results, the free energy landscape of LanLYR3 seems to be more complex, with two stable sub-basins compared with one for MtLYR3. The most stable regions, shown as rectangles in Figure 3c,d, were sampled continuously during the 27 and 16 last ns of MD simulations for
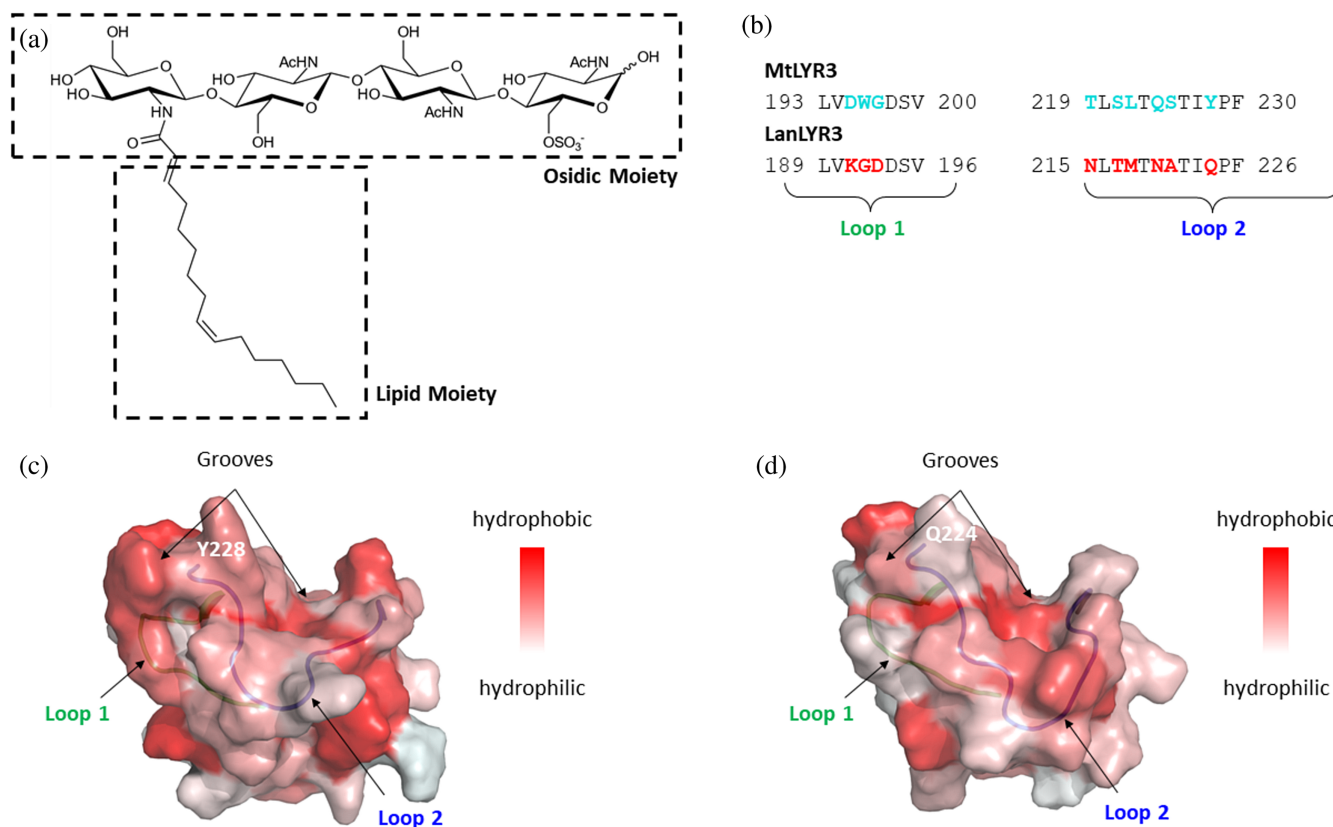
FIGURE 4 Lipo-chitooligosaccharide (LCO) structure and LCO binding site. 2D chemical structure of the studied LCO (a). Difference in sequence identity of loops 1 and 2 between MtLYR3 and LanLYR3 (b). Structural models extracted from molecular dynamics simulations of MtLYR3-LysM3 (c) and LanLYR3-LysM3 (d) colored according to their hydrophobic/hydrophilic profiles. Loops 1 and 2 backbone atoms are represented in cartoon.

MtLYR3 and LanLYR3, respectively. One representative structure was extracted from the stable regions described in Figure 3c,d and used in further analyses. It has to be noticed that this MD simulation sampling improved the Ramachandran plot compared with the initial models, the new plots showing an average of 96.5% versus 90% of residues occupying favorable regions, respectively (Figure S2). All these results are also in agreement with the global observation from RMSD analyses. These stable conformations picked-up from each minimum energy basin were used for the docking assays.

2.2 | Lipo-chitooligosaccharide binding site prediction

Previous swapping experiments between each of the three LysMs of the ectodomains of MtLYR3 and LanLYR3 proteins revealed the crucial role of the third LysM in LCO binding to MtLYR3.[18] To explore the binding mode of LCOs (Figure 4a) on the MtLYR3-LysM3, 100 docking poses of LCO-IV(S,C16:2Δ2E,9Z), an LCO known to bind to MtLYR3 with high affinity,[17] were generated using vina-carb by

targeting the stable conformation of MtLYR3-LysM3. A statistical analysis of all docking poses was performed, and six clusters were determined in terms of spatial distribution (Figure S4). Amongst these clusters, cluster 1 showed a proximity with Tyr228 of LysM3-MtLYR3 and contained the two poses with the lowest energies (−5.6 and −5.5 kcal/mol). For these reasons, these latter were picked up for further analysis. Moreover, these poses exhibited close contacts between the LCO and hydrophobic regions of LysM3-MtLYR3 (Figure 5a,b). These regions are mainly defined by two loops which bear 9 of the 15 amino acid residues that differ between MtLYR3-LysM3 and LanLYR3-LysM3 (Figure 4b). We refer to loop 1 as the region corresponding to residues from L193 to V200 in MtLYR3-LysM3 (from L189 to V196 in LanLYR3-LysM3), and loop 2 as the region from T219 to F230 in MtLYR3-LysM3 (from N215 to F226 in LanLYR3-LysM3). As shown in Figure 4c,d, MtLYR3-LysM3 is more hydrophobic than LanLYR3-LysM3. According to previous RMSF analysis, loop 1 is more flexible in LanLYR3-LysM3 than in MtLYR3-LysM3 (Figure 2c). In the first binding pose (Figure 5a), the lipid moiety of the LCO is buried in the groove bordered by loop 2, while in the second pose the
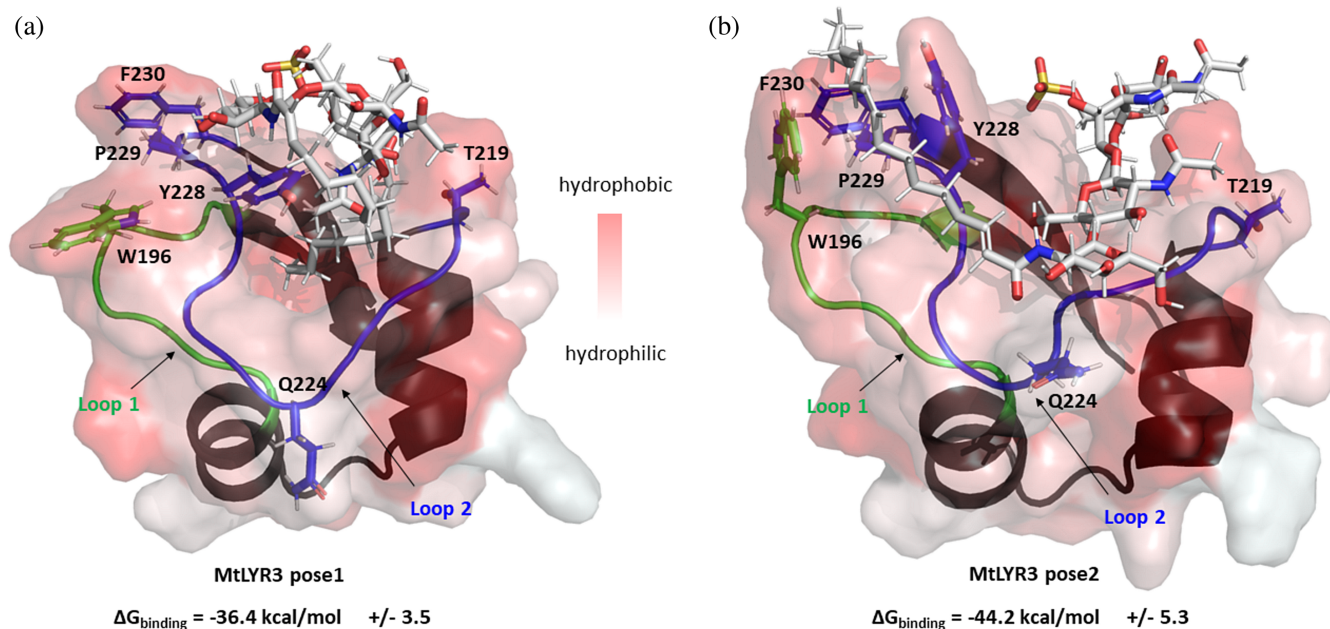
FIGURE 5 Comparison of two best lipo-chitooligosaccharide (LCO) docking poses on MtLYR3-LysM3. The two poses were selected according to their docking scores. The loops 1 and 2 are colored in blue and green, respectively. The LCO is shown in stick representation. The LysM3 is shown in surface mode and colored according to its hydrophobic/hydrophilic profile. The binding free energy calculated with the MM/GBSA method is indicated below each pose.

lipidic part is located between loops 1 and 2. For the second docking pose (Figure 5b), the osidic moiety of the LCO shows a closer interaction with loop 2 compared with the first docking pose. Regarding Y228, two behaviors are observed: (i) a pi-stacking of the sugar ring of the LCO in pose 1 and (ii) an interaction with the unsaturation at C9 of the acyl chain in pose 2.

To investigate the dynamics of the molecular systems, MD simulation of 50 ns length was performed on each LCO-MtLYR3 complex. For each docking pose, detailed analyses of interactions (distances, contacts, hydrogen bonds) with LCO were performed and shown in Figure S5. Analysis of the distance between the LCO and Tyr228 shows that LCO in pose 1 moves away after 10 ns whereas LCO in pose 2 stays in its initial binding site during the first 30 ns of MD simulation (Figure S5a,b respectively). This behavior is displayed on the Figure S6 where the LCO lipid part of pose 1 is more solvent exposed at 30 ns (Figure S6b) compared with the LCO in pose 2 (Figure S6d). Interestingly, when looking at contact analysis during MD time, LCO in pose 1 seems to lose some initial contacts in favor of non-specific ones, such as contacts with the neighboring LysM2 (Figure S5c). Moreover a higher number of residues involved in hydrogen bonds along the MD simulation is found for the LCO in pose 2 (Figure S5e). Altogether, these results show that the LCO binding pose 2 seems to be more stable compared with pose 1.

Finally, to quantify the binding strength and discriminate the most stable docking pose, MM/GBSA calculations were performed. To reduce the standard deviation in the binding free energy calculations, only the first 10 ns of the MD simulations were analyzed and 100 conformations were uniformly picked-up during this time. The results showed that the binding free energy of pose 2 (-44.2 kcal/mol $\pm$ 5.3) is lower than that of pose 1 (-36.4 kcal/mol $\pm$ 3.5, Figure 5). These calculations are in agreement with the stability of the LCO during MD simulations from pose 2, described above. Based on these observations, the pose 2 was considered the most likely binding pose and was therefore used for further analysis.

2.3 | Identification of key residues for lipo-chitooligosaccharide binding to MtLYR3

From previous work[18] it is known that LysM3 Y228 is crucial for LCO binding to MtLYR3, but it is not sufficient for conferring LCO binding to LanLYR3 when it substitutes the corresponding glutamine. From sequence and structural information of the most probable pose described previously, we searched for other potential key residues. Interestingly, in loop 2, a threonine (T219) and a glutamine (Q224) in MtLYR3 are replaced by two potentially glycosylated asparagines (N215 and N220) in

LanLYR3, as predicted by the NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/), due to the presence of the N-X-S/T motif (Figure 4b).[21] The presence of this glycosylation pattern in LanLYR3 LysM3 could explain the LCO binding differences with MtLYR3. In addition, the corresponding MtLYR3 residues are slightly more hydrophobic than the LanLYR3 asparagines and could thus be important for molecular recognition.
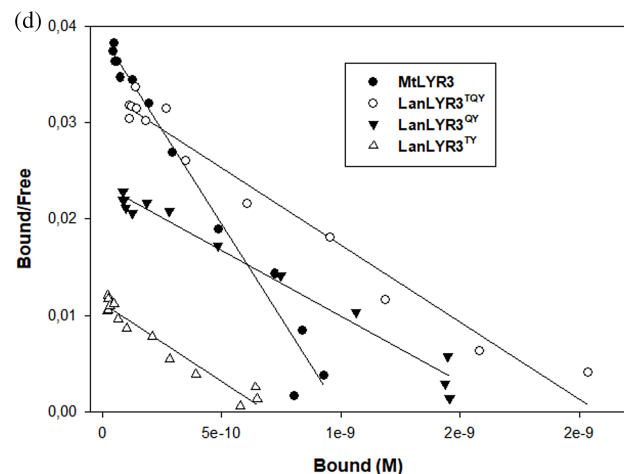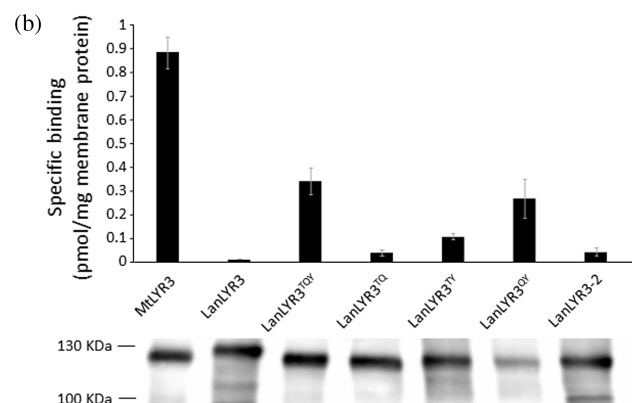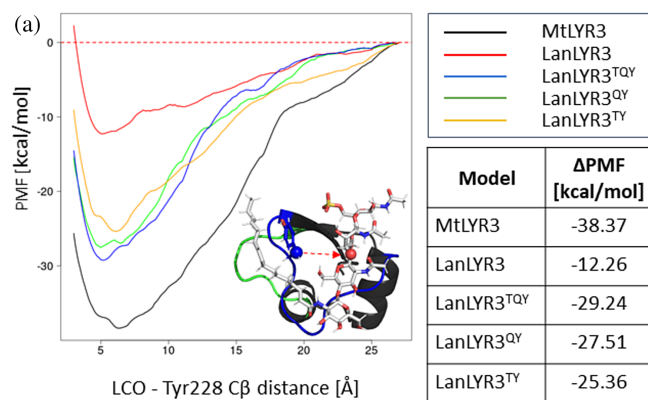
2.4 | In silico binding studies

To estimate the involvement of these two residues in the molecular recognition of LCOs, three mutants of LanLYR3 were modelled. For the first LanLYR3 mutant, N215 and N220 were substituted by the corresponding residues of MtLYR3, in addition the substitution at the position of the critical tyrosine in MtLYR3, that is, N215T-N220Q-Q224Y (LanLYR3^{TQY}). For the two other LanLYR3 mutants, just one asparagine together with Q224 were mutated, that is, N215T-Q224Y (LanLYR3^{TY}), and N220Q-Q224Y (LanLYR3^{QY}). For in silico binding affinity estimation, the LanLYR3 model, built without N-glycosylation, was used, since the integration of glycan chains would have required structural determination, which in plants is particularly complex. Mutant models were built from the most stable conformation of LanLYR3 picked-up from the PCA (see section *3D-Models of MtLYR3 and LanLYR3*). All complexes (wild-type LanLYR3 and mutants) were docked according to pose 2 identified for MtLYR3-LysM3. For all complexes, the equilibration procedure was the same as classical MD simulations described previously. Then, all equilibrated complexes were studied through Steered MD simulations, which is an approach to estimate the relative binding affinity from non-equilibrium simulations along a reaction coordinate as a distance for example.[21] Here, the reaction coordinate is the distance between the beta carbon (C_β) of the Y228 and the O of the second osidic linkage of the LCO, which seems to correspond to an anchoring point. The potential of mean force (PMF) describes the free energy landscape of the binding along the reaction coordinate. For wild-types, MtLYR3-LysM3 and LanLYR3-LysM3, and the three LanLYR3-LysM3

FIGURE 6 Affinity of LYR3 receptors toward Lipochitooligosaccharide (LCOs). The plot shows the potential mean force landscape obtained by steered (MD) simulations (a). The reaction coordinates are represented as spheres, in blue for Y228 C_β and in red for LCO second osidic linkage oxygen atom. The table on the right shows the difference of PMF between the lowest and highest point in the landscape. (a). Binding activity and immunodetection using anti-GFP antibodies of MtLYR3, LanLYR3, LanLYR3-2, and mutated forms of LanLYR3 (b). Membrane extracts were incubated as described in the experimental section, with the radiolabeled ligand in either the absence or the presence of a large excess of the corresponding unlabelled ligand to calculate the amount of specific binding. Results are means ± SEM. LysM3 sequence alignment of *Medicago truncatula* and *Lupinus angustifolius* LYR3 proteins (c). Residues mutated in LanLYR3 are highlighted by red asterisks. Scatchard plot of saturation experiments using LCO-IV(^{35}S,C16:2Δ2,9) and increasing concentrations of the corresponding cold ligand (d).

mutants, PMF curves are shown in Figure 6a and the binding free energy values (ΔPMF) were estimated by computing the depth of the corresponding wells with LCOs in solution as a reference (PMF around 0 kcal/mol). PMF curves showed optimal distances between 5 and 6 Å along the reaction coordinate axis. In agreement with the literature and the non-recognition of LCO by LanLYR3,[18] MtLYR3 (ΔPMF∼−38 kcal/mol) has a better affinity for LCOs than LanLYR3 (ΔPMF∼−12 kcal/mol). It has to be again mentioned that the LanLYR3 wild-type model was not glycosylated, which could affect its affinity for LCOs. Interestingly, all LanLYR3 mutants showed a gain of binding affinity toward LCOs compared with the LanLYR3 wild-type, with ΔPMF values in the same range −29.24 kcal/mol for LanLYR3^{TQY}, −27.51 kcal/mol for LanLYR3^{QY} and −25.36 kcal/mol for LanLYR3^{TY}. Thus, these results show that the double mutants, LanLYR3^{TY} and LanLYR3^{QY}, seem to have the same impact on the LCO binding affinity. However, the triple mutant LanLYR3^{TQY} does not show a cumulative effect. Altogether, our in silico study predicts that either N215T or N220Q combined with Q224Y, would be beneficial for LanLYR3 to recognise LCOs.

2.5 | Experimental binding studies

A radioligand binding assay showed that when N215, N220, and Q224 in the LysM3 sequence of LanLYR3 were substituted by the corresponding residues of MtLYR3 (T219, Q224, Y228), the mutated protein, LanLYR3^{TQY}, exhibited a significant gain of LCO-binding in comparison with LanLYR3 (Figure 6b). This gain of binding was lost if the Q224 was not substituted by the corresponding Y228 (LanLYR3^{TQ}). This is reminiscent of the loss of LCO binding previously reported for MtLYR3 when Y228 was substituted by Q.[18]

When only N215 or N220 together with Q224 were mutated, the resulting proteins (LanLYR3^{TY} and LanLYR3^{QY}) also showed a gain of LCO binding in comparison with LanLYR3. SDS-PAGE analysis of the membrane proteins followed by immunodetection of the different LYR3 proteins, revealed slight shifts in the apparent molecular masses of MtLYR3, LanLYR3 and its different mutated forms (Figure 6b). These shifts are consistent with the presence of N-Glycans on N215 and N220, since the molecular masses of LanLYR3 decreased when these predicted glycosylation sites were mutated into threonine and glutamine.

Lupin and some other legumes have a LYR3 paralog we named LYR3-2. The amino acid sequence of LanLYR3-2 LysMs share 67% identity with those of LanLYR3 (Table S1). LysM3 sequences differ between the

two proteins notably by N215 and N220, which are serine and glutamine in LanLYR3-2 LysM3 (Figure 6c). However, like LanLYR3, the LysM3 of LanLYR3-2 does not contain a tyrosine at position 228 of MtLYR3, but instead an asparagine, in the corresponding position (224). Like LanLYR3, LanLYR3-2 did not bind LCOs (Figure 6b), showing that it is not a functional ortholog of MtLYR3, and reinforcing the crucial role of tyrosine in LCO binding. The affinity of LCOs for the three LanLYR3 mutants was then determined by saturation experiments using the radioligand binding assay. As shown in Figure 6d, the Scatchard plots deduced from the saturation plots revealed that they had similar affinities (Kd): 62 nM for LanLYR3^{TQY}, 72 nM for LanLYR3^{QY}, and 60 nM for LanLYR3^{TY}. These values were only slightly lower than that of MtLYR3 (Kd = 25 nM). Therefore, either the introduction of tyrosine and glutamine or tyrosine and threonine in LanLYR3, was sufficient to make it a high-affinity LCO binding protein.

Among the other legume species that can have two LYRIII-A genes (*LYR3* and *LYR3-2*), are the Dalbergioid group of legumes, which includes *Arachis* and *Aeschynomene* species. *Aeschynomene* species differ from other legumes because some species are able to establish the RN symbiosis with rhizobia independently of NF perception.[22] Interestingly, *Aeschynomene* species have only one of the two *LYR3* genes depending on their nodulation properties. Hence, the NF-independent *A. evenia* has *LYR3*, but no *LYR3-2* in its genome.[23] In line with this observation, only *LYR3* was detected in transcriptomic data available for other NF-independent *Aeschynomene* species. In contrast, only *LYR3-2* was found in transcriptomic data for the NF-dependent *A. afraspera*. Another NF-dependent species, *A. patula*, was recently proposed as a model in this legume group.[24] We attempted to amplify LYRIII-A genes in *A. patula*. A *LYR3-2* gene was successfully sequenced, but we failed to detect any trace of *LYR3*. After sequence analysis we predicted that LYR3-2 from NF-dependent *Aeschynomene* species would not bind LCOs, while LYR3 from NF-independent *Aeschynomene* species would. These predictions were confirmed in LCO binding assays performed on *A. patula* LYR3-2 and *A. evenia* LYR3 (Figure S7).

2.6 | Phylogenetic analysis of LYR3 proteins and identification of residues under selection pressure

In the MtLYR3-LysM3, our molecular modelling study identified two loops defining a potential LCO binding site that differed in LanLYR3. In loop 2, two residues were found to intervene in LCO binding, in addition to the
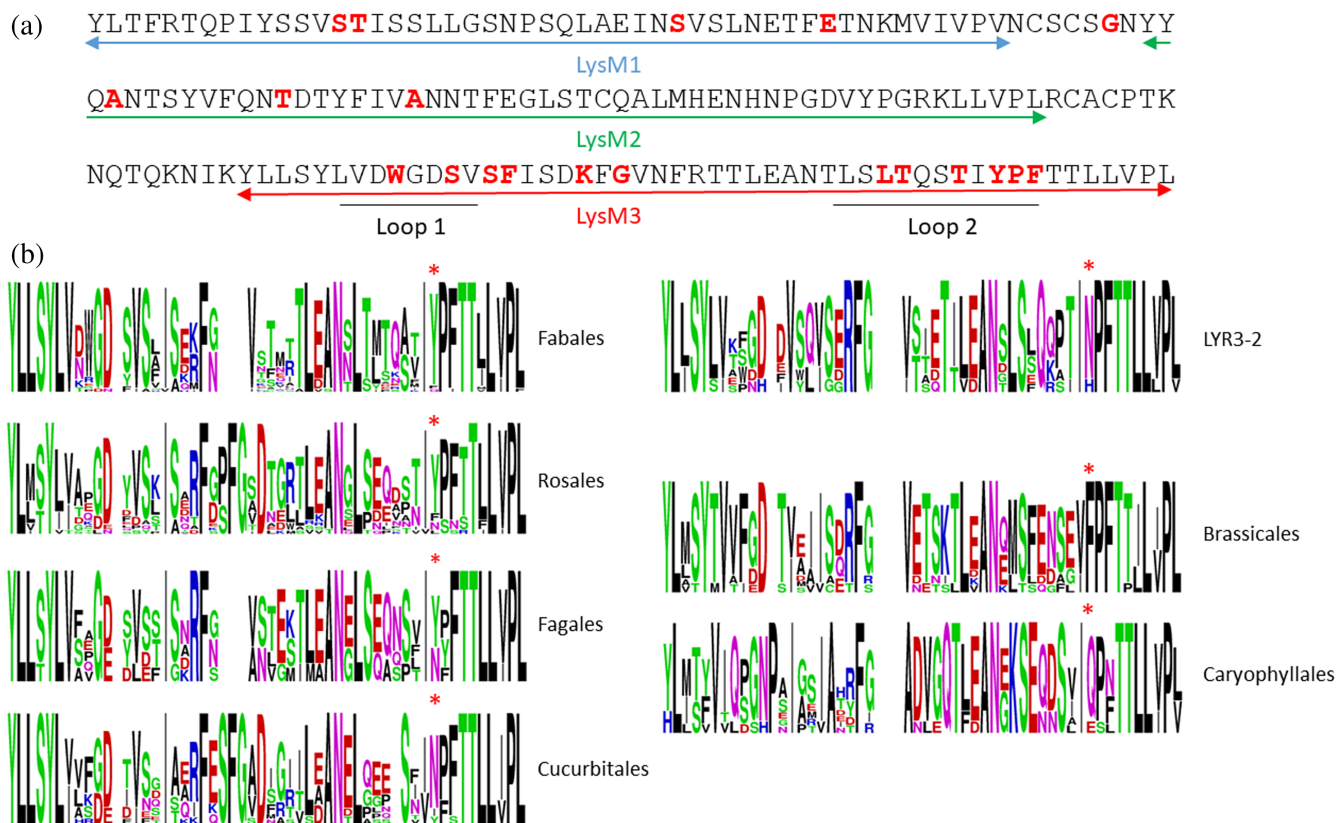
(a)

(b)

FIGURE 7 Sites under selection pressure. Sites under positive selection pressure were identified by MEME analysis are reported in the sequence of the LysMs of MtLYR3 as red and bold characters (a). Loops 1 and 2 identified in MtLYR3-LysM3 are indicated. Sequence logos showing the sequence conservation of LysM3 in the LYR3 proteins of plant species belonging to the different orders (b). Sequence logos generated using Clustal W[59] and WebLogo software.[60,61] The position corresponding to the critical tyrosine in MtLYR3 is marked with the red asterisk.

critical tyrosine. As a complementary approach to point out residues potentially involved in LCO binding, we searched for sites under relaxed selective pressure with MEME. For this purpose we used a set of LYR3 sequences from plant species able to establish both mycorrhizal and rhizobial symbioses, just one of these interactions or neither. Actinorhizal plants that host N fixing bacteria other than Rhizobia in RNs, were also included, resulting in a total of 113 analyzed sequences, including seven LYR3-2 sequences (Table S2; Figure S8).

The MEME analysis identified 20 residues under positive selection pressure (Figure S9). Interestingly, in the ectodomain, which contains the LysMs, 12 of the 20 identified residues under positive selection are in the LysM3 (Figure 7a). Moreover, these 12 residues were mainly located in loop 1 and loop 2. The residue corresponding to the Y228 in MtLYR3, was found to be under positive selection pressure, but not the residues corresponding to T219 and Q224 in MtLYR3 (N215 and N220 in LanLYR3). Due to the important role of the tyrosine in LCO binding, we therefore analyzed in more detail how widespread this residue is in different LYR3 proteins

belonging to plants with different symbiotic status. This deep inspection in the different clades revealed that the tyrosine was mainly present in the Fabales and Rosales that include species able to establish the two root (RN and AM) endosymbioses (Figure 7b). In contrast, in the Brassicales and Caryophyllales, which establish neither of these symbioses, the tyrosine is substituted by phenylalanine and glutamine, respectively. Compared with the Brassicales and Caryophyllales, lupin has also lost the ability to establish the AM symbiosis, but can nodulate. Interestingly, in LanLYR3, it is also glutamine that replaces the tyrosine. However, in the Cucurbitales, which are mycorrhizal plants, asparagine predominantly substitutes tyrosine, as in all the LYR3-2 paralogs present in the Fabales, suggesting non-binding proteins.

3 | DISCUSSION

The molecular mechanisms by which microbial LCO symbiotic signals bind to their receptors in plants are far from being elucidated. In NF-dependent RN symbiosis,

two LysM-RLKs belonging to the LYRI-A and LYKI phylogenetic groups,[19] of which the best characterized are MtNFP/LjNFR5 and MtLYK3/LjNFR1, respectively, are predicted to be assembled in a heterodimeric complex that controls the perception of LCOs. It is assumed that the LysM2 of a LYRI-A and the LysM1 of a LYKI play a major role in host specificity by discriminating the different chemical decorations of NFs important for the specific recognition of the host plant by its symbiont.[16,25] However, because the physical interaction between the heterodimer and its cognate ligand has never been shown, the topology of the LCO binding site made by this LysM-RLK complex is still hypothetical. The mechanism of LCO binding to LysM-RLKs is even more complex, since in LYRIII-A phylogenetic group (which includes MtLYR3), the LysM3 is critical. In the LYRIII-A phylogenetic group, there are closely related orthologs able or not to bind LCOs (such as MtLYR3 and LanLYR3, respectively). This has given us a unique opportunity to better understand how LCOs are perceived by LysM-RLKs from this phylogenetic group.

Molecular modelling and simulation studies helped us propose a binding mode of LCOs on LysM3 of MtLYR3. The binding mode resulting from our study involves mainly two loops at the surface of MtLYR3-LysM3. Previous work of Malkov and co-workers[18] proposed a binding mode where the lipid moiety of an LCO molecule would be located in the groove between the loops 1 and 2, while the osidic part would be located in a groove near the V200-F207 alpha helix. However, here, using a molecular modelling study taking into account the three LysMs and their conformational rearrangements through MD simulations, we propose another binding mode. In this new binding mode the lipid tail is located in the groove between the loops 1 and 2, similar to the binding mode proposed by Malkov and co-workers, whereas the osidic moiety is in the groove defined by loop 2. This binding mode is stable over MD simulation time. Finally, a sequence-structure analysis focused on LCO binding regions of this pose combined with steered MD simulations has allowed us to point out key residues for molecular recognition of the LCO that were confirmed by experimental radioligand binding assays. Indeed, the substitution of N215 and Q224 or N220 and Q224 residues in LanLYR3 by the amino acid types of the corresponding residues in MtLYR3 was sufficient to make LanLYR3 a high-affinity LCO binding protein.

Most of residues under positive selection, identified by the MEME analysis, were found in the LysM3 of LYR3 proteins, and located in loops 1 and 2. Among these residues, some might also be involved in the LCO binding site identified by molecular modelling. Indeed,

the position 228 occupied by the tyrosine in MtLYR3, for which we have shown the critical role in LCO binding, has experienced positive selection pressure. This is not the case for the positions 219 and 224, for which we found a role in the LCO binding site, showing the limitations of phylogenetic analyses to identify functional residues. However, the combination of molecular modelling and phylogenetic analyses highlighted a few residues in loops 1 and 2 that would be interesting to study in more detail to evaluate their involvement in the binding site.

The biological role of MtLYR3 remains unknown, but its biochemical properties (lack of specificity for the decorations of LCOs[17] and the absence of LCO binding for LanLYR3[18]) suggest that it could be involved in mycorrhizal symbiosis. Because molecular modeling and phylogenetic analyses pointed out the central role of the residue corresponding to Y228 in MtLYR3, we used the presence of a tyrosine at this position as a predictor of the ability of the LYRIII-A proteins to bind LCOs and analyzed how this was distributed in species able or not to establish endosymbioses with microorganisms producing LCOs. Although well conserved in LYRIII-A proteins in the Fabales and the Rosales, and partly in Fagales, in which most species are able to form AM symbiosis, the tyrosine was completely absent in the Brassicales and the Caryophyllales that are unable to form AM symbiosis, suggesting both a potential role for LYR3 in mycorrhizal symbiosis and a mechanism of convergent evolution leading to the loss of LCO binding ability in plants that have lost the mycorrhizal symbiosis. However, In Cucurbitales, where all species of this order used for the phylogenetic analysis are mycorrhizal plants, asparagine predominantly substitutes tyrosine, as in all the LYR3-2 paralogs present in the Fabales. Based on our results showing that LanLYR3-2 is not an LCO binding protein, it can be hypothesized that LYR3 proteins of the Cucurbitales are also non-binding proteins. Moreover, *Aeschynomene* species are also mycorrhizal (JFA, unpublished data), and we have shown that some *Aeschynomene* species have only a LYR3-2 protein that cannot bind LCOs. Therefore, the absence of LCO binding is not always associated with loss of the AM symbiosis, and the LCO binding property of LYR3 proteins is not a requirement for colonization by AM fungi.

The presence of asparagine instead of tyrosine in some LYR3 proteins, as in LYR3-2 proteins, is only found in orders in which nodulation has evolved (Fabales, Fagales, Cucurbitales, Rosales). Moreover, our data showing that among *Aeschynomene* legumes, NF-independent species have an LCO-binding LYR3 while NF-dependent species only have a non LCO-binding LYR3-2 protein, suggest that presence and/or the LCO binding property of LYRIII-A could have evolved to

accommodate particular types of RN symbioses. A similar reasoning can be made for actinorhizal RN species, most of those analyzed being predicted to have a non-binding LYR3 protein. Genomic and genetic research on these species will be useful to better specify and understand this differential distribution of LYR3 paralogs, and binding/non-binding versions of LYRIII-A, and their evolution in the context of RN symbioses.

In the Fabales, we predict that none of the LYR3-2 proteins can bind LCOs because they lack the critical tyrosine, and it was experimentally confirmed for LanLYR3-2 and ApLYR3-2. A duplication event from a common LYR3 and LYR3-2 ancestor is probably associated with the whole genome duplication event described in legumes.[26] Since we predict that LYRIII-A in other orders can bind LCOs, we hypothesize that the LYR3 and LYR3-2 ancestor was an LCO binding protein, but that all LYR3-2 proteins have lost their ability to bind LCOs while most of the legume LYR3 proteins have conserved it. In addition to a suspected role in the AM symbiosis, LYRIII-A proteins have a demonstrated role in defense against pathogens as shown for the *Arabidopsis thaliana* and the *Lotus japonicus* members.[20,27] This role might have been conserved in the legume LYR3 and LYR3-2 proteins independently of their ability to bind LCOs.

In conclusion, molecular docking and MD simulations, combined with experimental radioligand binding assays and mutagenesis, have identified a potential LCO binding site on MtLYR3-LysM3 and key residues for LCO binding. Our data also indicate that LCO binding to LYR3 proteins is an ancient property, often, but not always, associated with the ability of plants to establish the AM symbiosis and could have evolved to accommodate particular types of RN symbioses.

4 | METHODS

4.1 | 3D model construction of LYR3 ectodomains

3D structural models of the MtLYR3 and LanLYR3 ectodomains (the whole ectodomain containing 208 amino acid residues) were constructed using the iterative threading assembly refinement (I-TASSER) server.[28] The X-ray structures of ectodomains of AtCERK1 (PDB 4EBY), OsCEBiP (PDB 5JCE), and LjLYS6 (PDB 5LS2; see Figure S1 for percentage of similarity with targeted protein sequences) were used as threading templates for building both LYR3 3D models. Model quality was assessed through Ramachandran plots analysis using ProCheck.[29] PROPKA[30] was used to determine the protonation state of ionisable residues, that is, aspartate,

glutamate, histidine, arginine, lysine, cysteine, tyrosine, at pH 5.5 corresponding to an intracellular pH.

4.2 | Molecular dynamics simulations of LYR3 ectodomains

MD simulations of MtLYR3 and LanLYR3 ectodomains were performed using the AMBER ff14SB force-field.[31] Each LYR3 ectodomain was neutralized with a number of counter-ions and solvated with explicit TIP3P water molecules, using a rectangular box with a minimum distance of 0.10 nm between the solute and the simulation box edge. The molecular systems were first energy minimized using consecutively steepest descent and conjugate gradient algorithms, and gradually removing positional restraints starting from 25.0 kcal/mol/Å^2. Then, the systems were heated from 100 to 298 K over 100 ps in NVT ensemble and using harmonic positional restraints of 25.0 kcal/mol/Å^2 on the solute atoms. A short equilibration of 150 ps, in NPT ensemble (pressure at 1 bar and temperature at 298 K), was performed for each system with a gradual release of positional restraints. Finally, the production phases in the NPT ensemble were then carried out over 100 ns. The temperature (298 K) and the pressure (1 bar) were controlled using the Berendsen algorithms.[32] Long-range electrostatic forces were handled by using the Particle-Mesh Ewald method.[33] A 9 Å space cut-off for non-bonded interactions was used. The integration time-step of the simulations was 2.0 fs and the SHAKE algorithm[34] was used to constrain the lengths of all chemical bonds involving hydrogen atoms to their equilibrium values. The *sander* program of the AMBER16 suite[35] was used for the first stages (minimization, heating, and equilibration) while the *pmemd.CUDA* program[36] was used for the production phases. Atomic coordinates were saved every 10 ps for each simulation. MD trajectory analyses, including RMSD, RMSF, secondary structures, and PCA were performed using the *cpptraj*[37] module from Ambertools17 suite.[35] PCA from the protein heavy atoms was performed on the whole MD simulation trajectories. The coordinates of the two first PCs were projected on a map of 100 × 100 bins based on the minimal and maximal value of each PC. The free-energy landscape was constructed along the two first PCs for each bin i following the equation:

$$E_i = -k_B T \ln(p_i), \qquad (1)$$

where E_i is the free energy of each bin i, k_B is the Boltzmann constant, T is the temperature of the simulation and p_i is the probability of a state i defined as the number of frames contained in the considered bin

divided by the total number of frames of the trajectory. For each system, the bin with the lowest energy was used to extract the most stable conformations sampled by MD simulations. Then, for each system, all these stable conformations were structurally aligned on the corresponding average conformation and RMSD calculations were performed. Finally, the closest conformation to the median RMSD was considered the most representative and stable conformation and used as starting conformation for molecular docking.

4.3 | Parametrization of the lipo-chitooligosaccharide

The LCO-IV(S,C16:2Δ2E,9Z) model was built and geometrically optimized using avogadro software.[38] The charges for the lipid part were calculated using gaussian09 software[39] and the *Hartree-Fock* method with 6-31G* Pople basis sets. ESP charges were then fitted to get RESP charges with a weight of 0.01 to be consistent with GLYCAM-06 force-field.[40] Thanks to the adaptation of GLYCAM for the parametrization of lipids, both carbohydrates and lipids were parameterized according to the GLYCAM-06j-1 force field parameters. The linkage between carbohydrate and lipid part was parameterized using GAFF bonded parameters. AMBER formatted parameter files are given in supplementary material in prepi format (for charges and atom connectivity) and in frcmod format (for assigned GAFF bonded parameters). The partial charge of each atom is also given in Figure S10.

4.4 | Molecular docking, molecular dynamics simulations and binding free energy calculations of LYR3-LCO complex

Molecular docking experiments of the LCO-IV(S,-C16:2Δ2E,9Z) were carried out using the vina-carb software.[41] The search space on the receptor was defined as a box of 1.89 nm³ centered on the geometric center of the LysM3 of the MtLYR3 stable conformation extracted from the MD simulation. The box was adjusted to avoid contacts with the two other LysMs or inter-LysM regions. The LCO unsaturations, osidic cycles and N-acetyl linkages were restrained during the docking sampling while all other bonds of the ligand were considered flexible. The protein was considered as rigid during the docking sampling, except the side chains of the following amino acid residues that were considered as flexible: D195, W196, F202, T219, L220, S221, L222, T223, Q224, S225, T226, I227, Y228, and F230. One hundred poses were generated and clustered based on the Euclidean distances between the mass centers of the LCOs using hierarchical clustering and average distance for partition. The average silhouette coefficient was computed on a range of cluster numbers to identify the best partition that is, that would maximize silhouette coefficient. The docking energy distribution was computed for each cluster. The two poses with the lowest scores, which belong to the same cluster, were selected and subjected to MD simulations using the AMBER ff14SB force-field for the receptors and GLYCAM-06j-1 for the ligand as described in the Parametrization section. Same protocol as free-ligand receptor simulations was used to prepare the complexed systems. The production of MD simulations was performed in the NPT ensemble over 50 ns. Atomic coordinates were saved every 10 ps for each simulation. MD trajectory analyses, including distances and percentage of occurrence of hydrogen bonds were performed using the *cpptraj*[37] module from Ambertools17 suite.[35] To estimate the binding free energy between LCO and LYR3 receptors, MM/GBSA calculations were performed on 100 snapshots taken from every 0.1 ns of the first 10 ns of each MD trajectory, using the MMPBSA.py.MPI module of the AMBER program. Generalized Born (GB) was performed with the model developed by the Case group[42] using igb = 5 control command and a salt concentration of 0.1 M.[43] The surface area (SA) for the nonpolar solvation term was calculated using Linear Combination of Pairwise Overlaps algorithm,[44] and the surface tension was set to default value *that is*, 0.0072 kcal/mol Å². Based on MD simulation analyses, binding free energy estimations and visual inspection of MD trajectories, the most relevant docking pose was selected. From this given pose, the conformation of LCO was docked on LanLYR3 by structural superposition of LysM3 and the resulting complex was energy minimized.

4.5 | Mutant building

This 3D complex model was used for building LanLYR3 mutants. An in-house computational protein design method[45] was used for amino acid side chain substitution and the resulting 3D models were energy minimized and equilibrated using the same procedure as previously described.

4.6 | Steered molecular dynamics simulations

Steered MD simulations of wild-type (MtLYR3 and LanLYR3) and LanLYR3 mutant molecular complexes

were performed using pmemd.CUDA of AMBER16[36] from the equilibrated conformations. Steered MD was performed along the reaction coordinate corresponding to the distance between the beta carbon (C_β) of the Y228 and the O of the second osidic linkage of the LCO. 12 stages of 2 Å, ranging from 3 to 27 Å, were defined. For each stage, 25 independent trajectories of 500 ps were performed and the last conformation of the closest simulation to the Jarzynski average was selected as a starting structure to sample the next stage.[46] The sampling consisted of a Langevin dynamics simulation using a collision frequency γ of 5 ps and a time step of 2 fs. The SHAKE algorithm was used to constrain the covalent bonds involving hydrogens.

4.7 | Production of LYRIII-A WT and variant proteins

LYR3 constructs were obtained as previously described by using the Golden Gate technology[18] DAN fragments coding the LysM3 containing point mutations were obtained by custom DNA synthesis. The mutated LysM3 of LanLYR3 was then assembled with its LysM1 and LysM2 to reconstitute the ectodomain. Fragments coding AeLYR3 and ApLYR3-2 ectodomains (Table S3) were also obtained by custom DNA synthesis. As previously established, we used the MtNFP sequence to provide the trans-membrane and the intracellular domain tagged with YFP to facilitate the transient expression in *Nicotiana benthamiana* leaves and the detection of the protein.[17]

4.8 | Binding experiments

NF-binding assays were performed as described in,[18] using 10–30 μg proteins from membrane fractions prepared from LYR3 and its mutated forms expressed in *N. benthamiana* leaves, and 0.4–2 nM LCO-IV(^{35}S,-C16:2Δ2E,9Z). LCO-IV(S,C16:2Δ2E,9Z) and LCO-IV(C16:2Δ2E,9Z) were obtained as described in.[47] The radioactive LCO-IV(S,C16:2Δ2E,9Z) was synthesized by enzymatic sulfation of LCO-IV(C16:2Δ2E,9Z) with 35S according to.[48]

4.9 | Phylogenetic analysis and identification of residues under positive selection pressure

LYRIII-A orthologs (Table S2) were obtained using MtLYR3 sequence as query for tBLASTn v2.9.0+ search[49]

with default parameters and an e-value threshold of 1e-10 against a database containing representative species of Eudicots and covering AM and RN symbioses. LYRIII-A homologs were aligned with Mafft v6.861b[50] with default parameters and trimmed to remove positions with more than 20% of gaps using trimAl v1.4.[51] Phylogenetic analysis was conducted with IQ-TREE v1.5.5.[52] Prior to maximum likelihood tree reconstruction, best fitting evolutionary model was defined using ModelFinder[53] according to the Bayesian Information Criteria (BIC). Branch support were estimated 10,000 replicates of both SH-aLRT and ultrafast bootstrap.[54] LYRIII-A orthologs were confirmed based on the tree vizualised and annotated with the iTOL v4 platform.[55]

For *Aeschynomene* species, LYR3 was identified in the *A. evenia* annotated genome and used to identify orthologs in the transcriptomes of other NF-independent species (namely *A. ciliata*, *A. deamii*, *A. rudis*, *A. scabra*, *A. sensitiva*, *A.* sp 328, and *A. tambacoundensis*) using previously developed OrthoFinder gene orthogroups.[23] Blast searches in the NF-dependent *A. afraspera* transcriptome (DG, unpublished data) retrieved a *LYR3-2* paralog. Series of primers were designed using the different *Aeschynomene LYR3* and *LYR3-2* sequences to amplify LYRIII-A genes in *A. patula*. Partial sequence of *ApLYR3*, encompassing the whole ectodomain and part of the kinase domain, was successfully PCR amplified from leaf-extracted genomic DNA using the primers TTCAACAATTTTGCAGCCATC and ACCCCAAATGCAAAGACATC. For selection pressure analysis, the codon alignment was performed from the nucleic acid sequences and the amino acid sequences with the ete3 framework[56] using Mafft v6.861b.[50] The last 40 codons were removed considering the poor quality of the alignment at these positions. The phylogenetic tree was built with IQ-tree 1.5.5.[52] IQ-tree used ModelFinder[53] to identify the best-fit evolutionary model according to Bayesian information criterion (BIC). Branches were tested by SH-like aLRT[57] with 1,000 replicates. The list of positions under positive selection was inferred from the nucleic acid alignment and the phylogenetic tree thanks to the MEME method (Hyphy package) with default parameters.[58]

and editing (lead). **Ludovic Cottret:** Formal analysis (lead); investigation (lead); methodology (lead). **Maude Maréchaux:** Formal analysis (supporting); investigation (supporting). **Mégane Gaston:** Investigation (supporting). **Virginie Gasciolli:** Investigation (lead). **Jean Keller:** Investigation (supporting). **Nico Nouwen:** Investigation (supporting). **Djamel Gully:** Investigation (supporting). **Jean-François Arrighi:** Conceptualization (supporting); supervision (supporting); writing – review and editing (supporting). **Clare Gough:** Conceptualization (supporting); funding acquisition (lead); writing – review and editing (supporting). **Benoit Lefebvre:** Conceptualization (lead); funding acquisition (lead); project administration (lead); supervision (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Sophie Barbe:** Conceptualization (lead); formal analysis (lead); funding acquisition (lead); methodology (lead); project administration (lead); supervision (lead); validation (lead); visualization (lead); writing - original draft (lead); writing - review and editing (lead). **Jean-Jacques Bono:** Conceptualization (lead); funding acquisition (lead); supervision (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead).

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ORCID

Benoit Lefebvre https://orcid.org/0000-0002-3810-605X
Sophie Barbe https://orcid.org/0000-0003-2581-5022

REFERENCES

1. Maillet F, Poinsot V, André O, et al. Fungal lipochitooligo saccharide symbiotic signals in arbuscular mycorrhiza. Nature. 2011;469:58–63.
2. Dénarié J, Debellé F, Rosenberg C. Signaling and host range variation in nodulation. Annu Rev Microbiol. 1992;46:497–531.
3. Zipfel C, Oldroyd GED. Plant signalling in symbiosis and immunity. Nature. 2017;543:328–336.
4. Feng F, Sun J, Radhakrishnan GV, et al. A combination of chitooligosaccharide and lipochitooligosaccharide recognition promotes arbuscular mycorrhizal associations in *Medicago truncatula*. Nat Commun. 2019;10: 5047.
5. Gust AA, Willmann R, Desaki Y, Grabherr HM, Nürnberger T. Plant LysM proteins: Modules mediating symbiosis and immunity. Trends Plant Sci. 2012;17:495–502.
6. Chiu CH, Paszkowski U. How membrane receptors tread the fine balance between symbiosis and immunity signaling. Proc Natl Acad Sci U S A. 2021;118:e2106567118.
7. Liu T, Liu Z, Song C, et al. Chitin-induced dimerization activates a plant immune receptor. Science. 2012;336: 1160–1164.
8. Liu S, Wang J, Han Z, Gong X, Zhang H, Chai J. Molecular mechanism for fungal Cell Wall recognition by Rice chitin receptor OsCEBiP. Structure. 2016;24:1192–1200.
9. Bozsoki Z, Cheng J, Feng F, et al. Receptor-mediated chitin perception in legume roots is functionally separable from nod factor perception. Proc Natl Acad Sci U S A. 2017;114: 8118–8127.
10. Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R. LysM domain receptor kinases regulating rhizobial nod factor-induced infection. Science. 2003;302:630–603.
11. Arrighi J-FO, Barre A, Ben Amor B, et al. The *Medicago truncatula* lysin motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. Plant Physiol. 2006;142: 265–279.
12. Radutoiu S, Madsen LH, Madsen EB, et al. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature. 2003;425:585–592.
13. Madsen EB, Madsen LH, Radutoiu S, et al. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature. 2003;425:637–640.
14. Broghammer A, Krusell L, Blaise M l, et al. Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. Proc Natl Acad Sci U S A. 2012;109: 13859–13864.
15. Krönauer C, Radutoiu S. Understanding nod factor signalling paves the way for targeted engineering in legumes and non-legumes. Curr Opin Plant Biol. 2021;62:102026.
16. Bozsoki Z, Gysel K, Hansen SB, et al. Ligand-recognizing motifs in plant LysM receptors are major determinants of specificity. Science. 2020;369:663–670.
17. Fliegmann J, Canova S, Lachaud C, et al. Lipochitooligosaccharidic symbiotic signals are recognized by LysM receptor-like kinase LYR3 in the legume *Medicago truncatula*. ACS Chem Biol. 2013;8:1900–1906.
18. Malkov N, Fliegmann J, Rosenberg C, et al. Molecular basis of Lipo-chitooligosaccharide recognition by the Lysin motif receptor-like kinase LYR3 in legumes. Biochem J. 2016;473: 1369–1378.
19. Buendia L, Girardin A, Wang T, Cottret L, Lefebvre B. LysM receptor-like kinase and LysM receptor-like protein families: An update on phylogeny and functional characterization. Front Plant Sci. 2018;9:1531.
20. Fuechtbauer W, Yunusov T, Bozsoki ZN, et al. LYS12 LysM receptor decelerates *Phytophthora palmivora* disease progression in *Lotus japonicus*. Plant J. 2018;93:297–310.
21. Gomord V, Fitchette A-C, Menu-Bouaouiche L, et al. Plant-specific glycosylation patterns in the context of therapeutic protein production. Plant Biotech J. 2010;8:564–587.

22. Giraud E, Moulin L, Vallenet D, et al. Legumes symbioses: Absence of nod genes in photosynthetic Bradyrhizobia. Science. 2007;316:1307–1312.

23. Quilbé J, Lamy LO, Brottier L, et al. Genetics of nodulation in *Aeschynomene evenia* uncovers mechanisms of the rhizobium-legume symbiosis. Nat Commun. 2021;12:829.

24. Brottier L, Chaintreuil C, Simion P, et al. A phylogenetic framework of the legume genus Aeschynomene for comparative genetic analysis of the nod-dependent and nod-independent symbioses. BMC Plant Biol. 2018;18:333.

25. Gysel K, Laursen M, Thygesen MB, et al. Kinetic proofreading of lipochitooligosaccharides determines signal activation of symbiotic plant receptors. Proc Natl Acad Sci U S A. 2021;118: e2111031118.

26. Cannon SB, McKain MR, Harkess A, et al. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Mol Biol Evol. 2015;32:193–210.

27. Wan J, Tanaka K, Zhang X-C, et al. LYK4, a Lysin motif receptor-like kinase, is important for chitin signaling and plant innate immunity in Arabidopsis. Plant Physiol. 2012;160: 396–406.

28. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction. Nat Methods. 2015;12:7–8.

29. Anderson RJ, Weng Z, Campbell RK, Jiang X. Main-chain conformational tendencies of amino acids. Proteins. 2005;60: 679–689.

30. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. Proteins. 2005;61: 704–721.

31. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput. 2015;11:3696–3713.

32. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phy. 1984;81:3684–3690.

33. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. J Chem Phy. 1993;98:10089–10092.

34. van Gunsteren WF, Berendsen HJC. Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys. 1977;34: 1311–1327.

35. Case, D. A., R.M. Betz, D.S. Cerutti, T.E. Cheatham, I., T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, Xiao, L., P.A. Kollman (2016), AMBER 2016, University of California, San Francisco.

36. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem Theory Comput. 2013;9: 3878–3888.

37. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9:3084–3095.

38. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. J Chem. 2012;4:17.

39. Frisch, M. J., Trucks, G. W., Schlegel, H. B., et al. (2016) Gaussian 16 rev. C.01 in Wallingford, CT.

40. Kirschner KN, Yongye AB, Tschampel SM, et al. GLYCAM06: A generalizable biomolecular force field. Carbohydrates, J Comput Chem. 2008;29:622–655.

41. Nivedha AK, Thieker DF, Hu H, Woods RJ. Vina-carb: Improving Glycosidic angles during carbohydrate docking. J Chem Theory Comput. 2016;12:892–901.

42. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins. 2004;55:383–394.

43. Anandakrishnan R, Drozdetski A, Walker RC, Onufriev AV. Speed of conformational change: Comparing explicit and implicit solvent molecular dynamics simulations. Biophys J. 2015;108:1153–1164.

44. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). J Comput Chem. 1999;20:217–230.

45. Vucinic J, Simoncini D, Ruffini M, Barbe S, Schiex T. Positive multistate protein design. Bioinformatics. 2019;36: 122–130.

46. Jarzynski C. Nonequilibrium equality for free energy differences. Phys Rev Lett. 1997;78:2690–2693.

47. Rasmussen MO, Hogg B, Bono J-J, Samain E, Driguez H. New access to lipo-chitooligosaccharide nodulation factors. Org Biomol Chem. 2004;2:1908–1910.

48. Gressent F, Cullimore JV, Ranjeva R, Bono J-J. Radiolabeling of lipo-chitooligosaccharides using the NodH sulfotransferase: A two-step enzymatic procedure. BMC Biochem. 2004;5:4.

49. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: Architecture and applications. BMC Bioinform. 2009;10:421.

50. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 2008;9: 286–298.

51. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–1973.

52. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32: 268–274.

53. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–589.

54. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–522.

55. Letunic I, Bork P. Interactive tree of life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 2019;47: 256–259.

56. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol. 2016;33:1635–1638.

57. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 2006;55:539–552.

58. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8:e1002764.

59. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–2948.

60. Schneider TD, Stephens RM. Sequence logos: A new way to display consensus sequences. Nucleic Acids Res. 1990;18:6097–6100.

61. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. Genome Res. 2004;14:1188–1190.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.