

STAGE DE MASTER 2

Maëva ABADIE--LIMA

Master Bioinformatique et Biologie des Systèmes

2024 - 2025

Impact de faibles concentrations d'éthanol sur la conformation de membranes cellulaires et de protéines transmembranaires

Christian Chervin - Juan Cortès

Matthieu Chavent - Marc Guérout

Abstract FR	3
Abstract EN	4
Remerciements	5
Introduction	6
Matériel et Méthodes	11
Modélisation de VvETR2	11
Identification peptide signal et du domaine transmembranaire	11
Outils de prédictions tridimensionnel tout atome	11
Alignement multiple	12
Intégration structurale des ponts disulfures non-prédit	13
Intégration membranaire et dynamique structurale	14
Simulation de l'environnement de VvETR2	14
Dynamique de VvETR2	14
Ajout des ligands	18
Etude des trajectoires	20
Résultats	23
VvETR2 structuré selon les contraintes biologiques	23
Définition d'un domaine transmembranaire de 120 acides aminés	23
Plusieurs régions sont très conservées à travers l'évolution	25
Les ions cuivres stabilisent le modèle	27
Chai1 comme modèle de référence pour VvETR2	29
Croisement des chaînes N-terminale pour la création de ponts disulfures	31
Analyse des trajectoires dans les différentes conditions	32
La stabilité de la protéine varie selon les conditions	32
L'éthylène diffuse à travers la membrane	35
L'éthanol influence le comportement de l'éthylène	36
L'éthylène et l'éthanol interagissent avec le récepteur	39
Discussion	42
Références	46
Annexe	50

Abstract FR

Le récepteur à l'éthylène VvETR2 de *Vitis vinifera* localisé à la membrane du réticulum endoplasmique joue un rôle central dans la perception de l'éthylène, phytohormone gazeuse impliquée dans de nombreux processus de développement comme la germination des plantes. L'éthanol qui est une molécule produite en conditions hypoxiques, est suspecté de moduler cette signalisation. L'objectif de cette étude est de modéliser le domaine transmembranaire de VvETR2 et de simuler par dynamique moléculaire tout atome l'interaction de ce récepteur avec l'éthylène, l'éthanol ou des deux simultanément.

Une comparaison rigoureuse des modèles tridimensionnels issus de divers outils bioinformatiques basés sur l'intelligence artificielle a permis d'identifier le modèle de Chai1 comme référence pour cette étude.

Le récepteur a pu être inséré dans une bicouche lipidique à l'aide de CHARMM-gui. Afin de simuler le système, sur un temps long (une microseconde), au plus proche des conditions biologiques et physico-chimiques, nous avons utilisé le logiciel GROMACS.

Les analyses structurales effectuées (RMSD, RMSF et les temps de diffusion) et l'étude des interactions révèlent que l'éthanol seul n'interagit pas de manière stable avec le récepteur mais il augmente le temps que l'éthylène passe dans la protéine. L'éthanol perturbe également légèrement la stabilité conformationnelle du squelette protéique du récepteur, tandis que l'éthylène seul stabilise les hélices transmembranaires. Le comportement observé en présence conjointe des deux ligands pourrait ainsi résulter d'un équilibre entre ces deux effets.

Cette étude suggère que l'éthanol pourrait moduler l'effet de l'éthylène en influençant la dynamique structurale du récepteur, ce qui pourrait expliquer les observations expérimentales de modulation dose-dépendante.

Ce travail ouvre des perspectives incluant la réalisation de simulations plus longues grâce à des modèles gros grains, sur des simulations avec différentes compositions membranaires et pour approfondir les connaissances sur le comportement dose-dépendant de l'éthanol, des simulations supplémentaires faisant varier les concentrations d'éthanol et d'éthylène.

Abstract EN

The ethylene receptor VvETR2 in *Vitis vinifera* is a key component in the perception of this gaseous phytohormone, which regulates several developmental processes. Ethanol, a molecule produced under hypoxic conditions, is suspected to modulate this signaling pathway. This study aims to model the transmembrane domain of VvETR2 and simulate, using all-atom molecular dynamics, its interactions with ethylene, ethanol, or ethylene and ethanol simultaneous.

The receptor was inserted into a lipid bilayer using CHARMM-gui. In order to simulate the system over a long period of time (one microsecond), as closely as possible to the biological and physico-chemical conditions, we used the GROMACS software.

The structural analyses carried out (RMSD, RMSF and diffusion times) and the study of interactions reveal that ethanol alone does not interact stably with the receptor, but it does increase the time ethylene spends in the protein. Ethanol also slightly disrupts the conformational stability of the protein backbone of the receptor, whereas ethylene alone stabilises the transmembrane helices. The behaviour observed in the joint presence of the two ligands could therefore be the result of a balance between these two effects.

This study suggests that ethanol could modulate the effect of ethylene by influencing the structural dynamics of the receptor, which could explain the experimental observations of dose-dependent modulation.

This study opens up new prospects, including longer simulations using coarse-grained models, simulations with different membrane compositions and, in order to gain a better understanding of the dose-dependent behaviour of ethanol, additional simulations varying the concentrations of ethanol and ethylene.

Remerciements

Je tiens à exprimer ma sincère gratitude envers mes trois encadrants pour leur bienveillance et leurs encouragements.

À Christian Chervin, pour son implication de longue date dans le projet, sa vision biologique, et pour n'avoir pas hésité à mobiliser ses contacts afin de m'aider à avancer.

À Juan Cortès, pour son accueil au sein de son laboratoire et pour m'avoir offert un cadre de travail aussi rigoureux que stimulant.

À Matthieu Chavent, pour son expertise technique , et dont les conseils et retours m'ont constamment poussée à donner le meilleur de moi-même.

Je souhaite adresser un remerciement tout particulier à Marc Guérout. Bien qu'il ne soit pas officiellement mon encadrant, ses explications, ses conseils techniques et sa grande disponibilité m'ont été d'une grande aide tout au long de ce stage.

Introduction

L'éthylène est une hormone végétale régulant un grand nombre de métabolismes, dont la germination des graines, la croissance et la sénescence des organes, exemples : les tiges, les feuilles et les fruits (Binder, 2020; Dubois et al. 2018; Giovannoni, 2007).

Malgré sa production dans le cytosol, l'éthylène est une phytohormone gazeuse dont la diffusion est possible à travers les différents compartiments cellulaires, car elle diffuse très rapidement dans les membranes végétales (Dubois et al. 2018).

La perception de l'éthylène par les plantes et d'autres organismes, dont des champignons et des algues (Binder, 2020) est décrite depuis une vingtaine d'années, suite à la découverte des protéines réceptrices spécifiques (Chang et al. 1993).

La famille des récepteurs éthylène (ETRs) se divise en deux sous-familles, la première comme AtETR1 qui ont un domaine histidine kinase fonctionnel et la seconde comme AtETR2 qui ont un domaine histidine kinase non fonctionnel (Bleecker, 1999; Qu et al, 2007).

La perception de l'éthylène par les ETRs permet le déclenchement d'une cascade de signalisation qui contient de nombreux composants : les ETRs qui se situent dans la membrane du réticulum endoplasmique, des protéines kinases appelées CTR1 “constitutive triple response 1”, des protéines transmembranaires, aussi du réticulum endoplasmique, appelées EIN2 “ethylene insensitive 2”, des facteurs de transcriptions comme EIN3 ou EIN3-like qui sont des intermédiaires du signal au niveau nucléaire.. Lorsqu'un ETR est en présence d'éthylène, celui-ci est “désactivé”, car son état de phosphorylation baisse, et cette perte de phosphorylation va se transmettre via le CTR1 au EIN2. La partie C-term de ce dernier est alors clivée et se déplace dans le noyau où elle active l'expression de nombreux gènes (Binder, 2020; Chang, 2016).

La phylogénie des récepteurs éthylène a été étudiée, particulièrement chez certains fruits (Chen et al 2020). L'arbre phylogénétique des ETR (figure 1) montre l'homologie de séquences protéiques des ETRs de tomate, pomme, vigne et mandarine, en utilisant les ETRs d'*Arabidopsis* comme points de comparaison, cette plante étant un des modèles végétaux les plus étudiés.

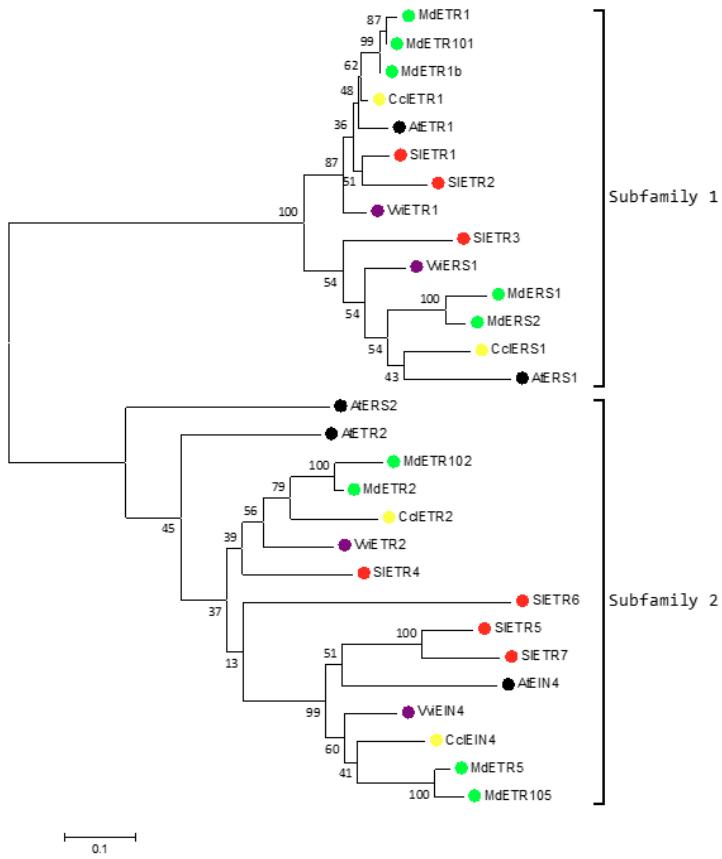


Figure 1. Arbre phylogénétique des ETRs, construit par un alignement MUSCLE par la méthode de maximum de vraisemblance. Les nombres sur les branches sont les valeurs de confiance basée sur une méthode de bootstrap ($B= 500$ réplications). S1 désigne *Solanum lycopersicum* (rouge), Md désigne *Malus domestica* (vert), Vvi désigne *Vitis vinifera* (violet), Ccl désigne *Citrus clementina* (jaune) et At désigne *Arabidopsis thaliana* (noir). Reproduit de *Ethylene receptors and related proteins in climacteric and non climacteric fruits*, Chen et al 2018.

Certains auteurs ont observé des liens entre la signalisation de l'éthylène et d'autres régulateurs, dont l'éthanol, lequel est proposé dans un article récent comme un signal que les organismes végétaux pourraient se servir afin de s'adapter à divers stress (Diot et al., 2024). Une partie de ces auteurs avaient précédemment observé que l'éthanol module la germination de pépins de tomate et que cette modulation est affectée par une mutation sur un ETR (Chen et al., 2020). Les modifications de perception sont effectives à de très faibles concentrations

d'éthanol, 0,01mM à 1mM. Il y aurait donc un lien entre perception éthylène et perception d'éthanol.

L'éthanol est produit par de nombreuses plantes à faible concentration, par exemple suite à des épisodes hypoxiques (Diot el al., 2024).

Ethylene Receptor 2 *Vitis Vinifera*

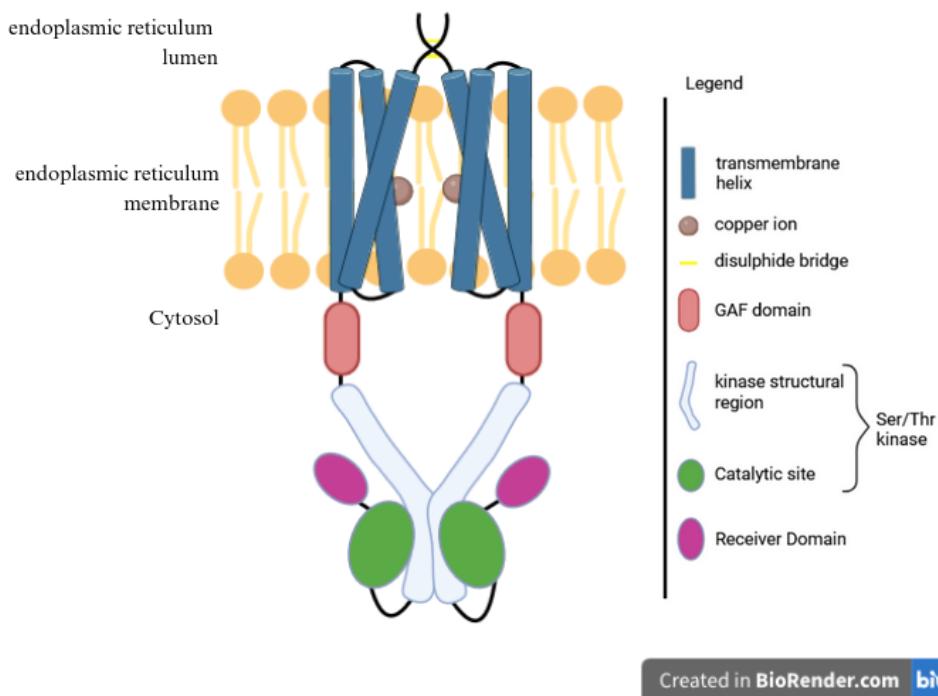


Figure 2. Représentation du récepteur à l'éthylène 2 chez *Vitis vinifera*. Le récepteur ETR2 est un homodimère. Chaque dimère est composé de trois hélices transmembranaires où se trouve un ion cuivre Cu²⁺ impliqué dans la fixation de l'éthylène. Les chaînes N-terminales des monomères se situent dans le lumen du réticulum endoplasmique. Sur ces chaînes N-ter deux ponts disulfures permettent la stabilisation du dimère. Dans le cytosol, chaque monomère va posséder un domaine GAF qui permet la signalisation en aval de la l'interaction entre l'éthylène et le récepteur. Chaque monomère possède également une kinase avec un site catalytique. Ce site est inactif chez les ETR2 suite à des mutations dans les résidus essentiels à l'activité kinase. Cela permet une modulation fine par le biais d'interactions, c'est pourquoi ETR2 est impliqué dans des activités comme la germination qui nécessite une régulation précise. Le domaine de phosphorylation à l'extrémité C-terminale permet de moduler la réponse à l'éthylène.

L'objectif de mon stage de fin d'étude est d'étudier l'interaction entre l'éthanol et les ETRs et l'impact de l'éthanol sur la fixation de l'éthylène qui n'ont jusqu'alors jamais été étudiés.

L'objectif premier de cette étude est donc de modéliser la partie transmembranaire du récepteur VvETR2 de *Vitis vinifera*, car c'est l'ETR le plus exprimé chez la vigne au moment du développement de la baie (Chen et al., 2018). Or les baies de raisin sont soumises à des conditions hypoxiques au cours de leur développement (Xiao et al 2018) et à une augmentation des teneurs en éthanol. Alice Diot en doctorat avec Christian Chervin a choisi de faire des expériences d'expression hétérologue de ce VvETR2 (figure 2), après avoir réalisé des expériences de RNA-seq concernant les réponses de *Vitis vinifera* à de faibles concentrations en éthanol, proches de celles mesurées dans la baie de raisin (Xiao et al, 2018). Seule la partie transmembranaire est modélisée dans cette étude pour simplifier le modèle et parce que c'est là que se situe les sites de liaison putatifs à l'éthylène, identifié chez *Arabidopsis thaliana* sur les résidus 32 à 34 (Wang et al 2006) mais aussi sur les résidus 25 et 91 (Azhar et al, 2023).

Les avancées en intelligence artificielle facilitent la modélisation de protéines. Avant cela, il existait des méthodes basées sur les interactions physiques, soit sur l'histoire évolutive.

Les approches basées sur les interactions physiques permettent d'intégrer la compréhension des forces moléculaires dans les simulations thermodynamiques et cinétique de la physique des protéines. Cette approche en raison du coût computationnel très élevé des simulations moléculaires, et de la difficulté à reproduire fidèlement les interactions physiques impliquées dans le repliement des protéines, était difficile à appliquer dans la pratique même sur des petits systèmes.

Les approches bioinformatiques fondées sur l'évolution, utilisent des dérivées de l'homologie des structures connues et des corrélations évolutives entre paire de résidus (co-évolution) pour prédire la conformation des protéines. Ces méthodes reposent donc sur la quantité de structures connues, sur l'essor du séquençage génomique et les progrès en apprentissage automatique, en particulier des réseaux profonds. Cependant dans les cas où aucune structure homologue n'est connue, il est difficile d'utiliser une approche basée sur l'évolution. Malgré ces avancées, les approches actuelles basées sur la physique ou sur l'histoire évolutive, ne parviennent pas à reproduire avec précision les structures

expérimentales, dans la majorité des cas où aucun homologue proche n'a été résolu expérimentalement. Cela limite leur utilité pour de nombreuses applications biologiques.

L'arrivée de modèles fondés sur l'intelligence artificielle a renouvelé la modélisation moléculaire, rendant possible la prédiction rapide et fiable de structures protéiques même en l'absence d'un homologue proche connu. Cependant, là encore, il existe de nombreuses façons de prédire une structure tridimensionnelle basée sur l'intelligence artificielle en se reposant sur des principes et des architectures différentes.

AlphaFold2 utilise un réseau transformeur evoformer, qui exploite les alignements multiples de séquences (MSA) (Jumper et al., 2021). Son successeur, AlphaFold3 est entraîné avec une méthode de diffusion qui lui permet de retrouver une structure grâce à la suppression de bruit (Abramson et al., 2024). OmegaFold repose sur un modèle de langage pré-entraîné sur des séquences non alignées, et utilise un transformeur géométrique geoformer, pour intégrer des contraintes spatiales (Wu et al., 2022). Chai-1 adopte une approche multimodale qui combine les séquences, les données chimiques, les contraintes expérimentales et le langage naturel (Chai Discovery et al., 2024). Enfin, Boltz1, un modèle open source, fait appel à la théorie des graphes et aux réseaux de neurones graphiques contraints physiquement, intégrant également les MSA pour prédire les structures à partir d'interactions atomiques (Wohlwend et al., 2024).

Ces outils de prédiction de structure utilisent des méthodes différentes, les prédictions sont également différentes en fonction du modèle utilisé. Il sera donc nécessaire d'identifier le modèle le plus cohérent et d'apporter les modifications nécessaires afin de modéliser la partie transmembranaire du récepteur VvETR2 de *Vitis vinifera*.

L'objectif second sera de simuler l'environnement cellulaire du récepteur à l'éthylène et d'observer dans des conditions réalistes grâce à la dynamique moléculaire. La dynamique moléculaire est une méthode de simulation numérique des mouvements atomiques (Alder et Wainwright, 1959) (Frenkel & Smit, 2002). Différents systèmes sont mis en place dans cette étude afin d'étudier l'interaction de l'éthylène, de l'éthanol et des deux simultanément afin d'identifier l'interaction de l'éthanol sur le récepteur et son impact sur la fixation de l'éthylène qui est son ligand naturel. Pour cela CHARMM-gui un outil bioinformatique de préparation de systèmes (Brooks et al., 1983) et GROMACS, un logiciel de simulation de dynamique moléculaire (Abraham et al., 2015) sont utilisés pour cette étude.

Matériel et Méthodes

Modélisation de VvETR2

Identification peptide signal et du domaine transmembranaire

La séquence protéique étudiée est celle du récepteur à l'éthylène de l'organisme *Vitis vinifera* disponible sur UniProt sous le numéro d'accession F6H8H3 récupéré au format fasta. Les séquences protéiques utilisées dans cette étude sont détaillées en Annexe dans le tableau 1.

Pour détecter le peptide signal à cliver, SignalP-6.0 a été utilisé avec les paramètres par défaut (Nielsen et al, 2024).

De plus, seule la partie transmembranaire est nécessaire à cette étude. L'identification des domaines transmembranaires a été effectuée avec DeepTMHMM 1.0 (Hallgren et al, 2022). De plus, DeepTMHMM détecte aussi les séquences de peptide signal, ses résultats ont donc été comparés à ceux de SignalP-6.0.

Pour sélectionner la partie de la séquence d'intérêt, un script python 3.10 a été mis en place, prenant en entrée une séquence, une valeur de début et de fin et qui renvoie en sortie la nouvelle séquence et sa longueur. Tous les scripts python seront fournis en données supplémentaires.

Outils de prédictions tridimensionnel tout atome

La région transmembranaire prédite par DeepTMHMM a été utilisée pour modéliser la structure tridimensionnelle du domaine transmembranaire (en annexe).

La "représentation tout atome", apparue pour la première fois en 1861, permet de représenter chaque atome avec ses interactions dans le système (Loschmidt, 1861). C'est la représentation la plus précise, mais aussi la plus complexe et par conséquent la plus coûteuse en temps de calcul et en ressources.

Plusieurs outils de modélisation de structure biomoléculaire ont été utilisés pour générer le récepteur à l'éthylène : Alpha Fold2, Alpha Fold3, Chai-1, Boltz 1 et Omega Fold.

Pour créer un modèle de la protéine, chacune de ces approches a pour objectif de prédire les coordonnées de chaque atome grâce à diverses méthodes.

Afin d'optimiser la conception du récepteur à l'éthylène 2 chez la vigne (VvETR2), des modèles ont été générés avec chacun de ces outils de prédition.

Alignement multiple

Des alignements multiples ont été effectués afin de comprendre la conservation des différentes régions ETR2.

Pour cela trois jeux de données ont été faits : un premier issu d'un BlastP sur uniprot de la partie transmembranaire de ETR2, un second issu d'un BlastP de la séquence protéique entière et un dernier jeu de séquence avec une recherche par mot-clé "Ethylene Receptor" entre 700 et 900 acides aminés (Camacho et al, 2009) dans Uniprot (The UniProt Consortium, 2023). Un premier script, en python, a été utilisé pour supprimer les doublons. Afin que seules les séquences correspondant à des récepteurs à l'éthylène soient conservées dans les jeux de données, un second script python a été utilisé (figure 3).

Une fois les jeux de données filtrés obtenus (figure3), des alignements multiples ont pu être réalisés grâce à Clustal Omega (Sievers et al, 2011). Pour visualiser la conservation des résidus, WebLogo 3 a été utilisé (Crooks et al, 2004).

Les fichiers MSA générés par AF2 ont été analysés afin de les comparer aux différents jeux de données.

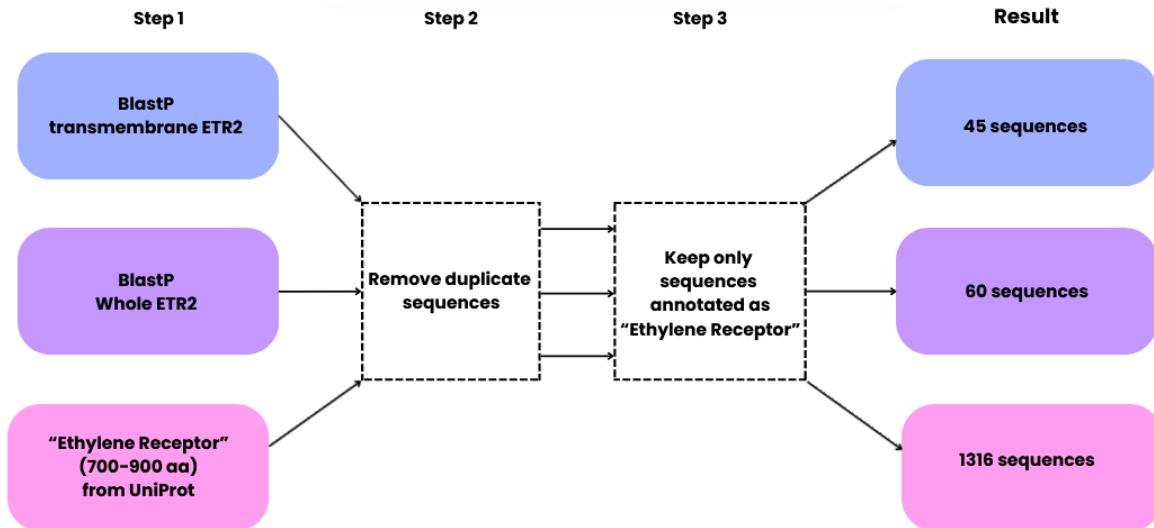


Figure 3. Pipeline de sélection des séquences de récepteurs à l'éthylène pour l'alignement multiple, depuis BLASTP et UniProt jusqu'à l'ensemble final filtré. Trois sources ont été utilisées pour récupérer les séquences candidates : une recherche BLASTP à partir de la région transmembranaire d'ETR2 (en bleu), une recherche BLASTP à partir de la séquence complète d'ETR2 (en violet) et une recherche dans UniProt pour les protéines annotées comme "Ethylene Receptor", avec une longueur comprise entre 700 et 900 acides aminés (en rose). Après suppression des doublons, seules les séquences annotées "Ethylene Receptor" ont été conservées. Le nombre final de séquences obtenues pour chaque méthode est indiqué à droite.

Intégration structurale des ponts disulfures non-prédit

Dans un premier temps, les extrémités N-terminales des deux monomères ont été modifiées à l'aide de PyMOL. Les dix premiers résidus ont été conservés. En revanche, la région située entre la seconde cystéine (résidu 10) et le début de la première hélice alpha (résidus 11 à 18) a été supprimée sur les deux monomères, afin de rapprocher spatialement les cystéines susceptibles de former un pont disulfure inter-chaîne.

Les cystéines ont ensuite été manuellement positionnées en vis-à-vis dans plusieurs configurations et à différentes distances.

Pour restaurer les segments manquants entre les résidus 10 et 18, le logiciel MoMA-FreSa (Clerc, 2024) a été utilisé. Il permet de créer différentes possibilités de conformation de la

séquence qu'on lui donne pour lier deux séquences. Dans notre cas, MoMA-FreSa a été utilisé pour relier la séquence 1–10 à 18–120, en utilisant la séquence supprimée pour qu'elle apparaisse dans le modèle final. Environ dix conformations ont été générées par monomère et par configuration.

Un script python permettant d'obtenir toutes les configurations des différentes conformations a été utilisé pour assembler toutes les possibilités (chaque monomère de la chaîne1 avec chaque monomère de la chaîne2).

Chaque configuration a ensuite été évaluée manuellement afin de déterminer si les monomères ainsi construits permettent la formation d'un pont disulfure interchaîne, tout en conservant la cohérence structurale du modèle global pour conserver le modèle le plus cohérent.

Intégration membranaire et dynamique structurale

Simulation de l'environnement de VvETR2

CHARMM-gui est un outil basé qui utilise des champs de force, qui se reposent sur des fonctions d'énergie semi-empirique, ce qui permet de faciliter la préparation de systèmes pour des simulations de dynamique moléculaire (Brooks et al., 1983). Le modèle amélioré sous format PDB a été utilisé comme entrée CHARMM-gui dans le module bilayer builder du module membrane builder. Pour l'orientation de la protéine dans la membrane, l'option Run PPM 2.0 a été sélectionnée. Les ponts disulfures ont dû être précisés dans le logiciel ainsi que la déprotonation des cystéines 68 et 188 qui lient le cuivre. Pour simplifier le modèle de la membrane lipidique un ratio de 100% en 18:1 ($\Delta 9$ -Cis) phosphocholine (DOPC) a été choisi. La taille de la boîte a été fixée à 80.0 angstroms. Le solvant utilisé est TIP3P et la concentration en KCl a été fixée à 0.100 mM parce que c'est la concentration de K dans le cytosol des plantes (Rodriguez-Navarro, 2000) et que les données sont manquantes pour le réticulum endoplasmique. Le champ de force Charmm36 a été utilisé pour la génération des fichiers de sorties. La température utilisée est de 303.15K (soit 30 °C).

Dynamique de VvETR2

Des simulations ont été effectuées sur le système protéine membrane en tout atome grâce au logiciel GROMACS dans la version 2023.03 (Abraham et al., 2015). C'est également GROMACS dans cette même version qui a été utilisée pour l'étape de minimisation et les étapes de production (figure 4).

Une étape de minimisation d'énergie a d'abord été effectuée. Cette étape a pour but de stabiliser la structure pour qu'elle soit au fond d'un minimum local du paysage énergétique. Cela signifie que les forces sur les atomes sont faibles, qu'ils ne sont pas trop proches les uns des autres évitant ainsi les contacts stériques afin que les forces de répulsion ne soient pas trop fortes au début de la simulation. Ces forces pourraient entraîner des forces infinies, et une cassure du système à cause de grands déplacements d'atome. L'algorithme de descente de gradient développé à l'origine par Cauchy est celui qui a été utilisé pour la minimisation d'énergie [Cauchy, 1847]. La procédure de minimisation se termine ici lorsque la force maximale devient inférieure à $1000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$.

Des étapes d'équilibration ont ensuite succédé celle de minimisation d'énergie.

Une première étape d'équilibration a été réalisée en conditions NVT où N signifie le nombre de molécules, V le volume et T la température, car ce sont les 3 paramètres qui ne vont pas varier. Le but d'une équilibration NVT va être de stabiliser la température du système dans un environnement thermiquement contrôlé. La température ciblé lors du NVT va être de 303.15 K à l'aide du thermostat V-rescale (velocity rescale). V-rescale est une méthode de régulation de la température qui ajuste les vitesses des particules à chaque pas de temps pour maintenir la température autour d'une valeur cible. (Bussi et al, 2007)

Les différents composants du système peuvent absorber ou libérer la chaleur différemment, par conséquent un v-rescale est appliqué indépendamment aux groupes protéine, membrane et solvant pour permettre une meilleure régulation locale de la température en évitant les déséquilibres internes.

Des contraintes positionnelles de $4000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ ont été appliquées au squelette protéique, $2000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ pour les chaînes latérales, et $1000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ aux lipides. Ces contraintes positionnelles ont pour but de conserver la structure initiale du système protéine/ solvant/ membrane.

Dans les complexes Cu²⁺–Cys et Cu²⁺–His, les longueurs de liaison typiques sont d'environ 2,1 Å pour Cu–S et entre 1,9 et 2,2 Å pour Cu–N. Ces valeurs, bien que extraites d'un exemple à géométrie tétraédrique légèrement déformée (T1), sont ici utilisées à titre indicatif, indépendamment du type de coordination du site étudié, qui ne présente pas nécessairement une géométrie de type T1 (Roger *et al.*, 2024).

La présence d'ions cuivre (Cu²⁺) dans le site actif du récepteur a nécessité des précautions particulières lors des étapes de minimisation et d'équilibration. En l'absence de modélisation explicite des liaisons de coordination dans les champs de force classiques, des contraintes souples de type flat-bottom ont été appliqués entre les atomes de coordination (soufres des cystéines et azotes des histidines) et les ions cuivre, avec des distances cibles (0.21 nm et 0.18 nm respectivement) et des constantes de force ajustées progressivement. Cette approche a permis de stabiliser la géométrie de coordination sans contraindre excessivement le système, et de prévenir les déplacements non physiques des ions métalliques pendant les premières phases d'équilibration.

De plus, des contraintes souple de type flat-bottom ont été appliquées entre les atomes de coordination du cuivre et des ions cuivres pour maintenir des distances définies. Une contrainte de type flat bottom n'applique aucune force tant que deux groupes d'atomes restent à une distance donnée ou inférieure, mais commence à exercer une force de rappel si cette distance dépasse un seuil fixé ici à 2 angstroms pour la distance avec les soufres des cystéines et 1.8 angstroms pour les azotes des histidines.

Cette phase d'équilibration a été effectuée sur 125 ps avec un pas d'intégration de 1 fs car la force des contraintes et la température ciblé permettent au système une stabilisation rapide.

Une seconde étape en conditions NVT a été réalisée afin de réduire progressivement les contraintes positionnelles appliquées sur la protéine et la membrane, tout en maintenant une température stable de 303.15 K. Cette phase intermédiaire permet au système de commencer à se réorganiser structurellement sans modification du volume, garantissant une transition douce entre l'étape NVT rigide initiale et l'équilibration sous pression (NPT) qui suit.

Les constantes de force appliquées ont été abaissées à 2000 kJ·mol⁻¹·nm⁻² pour le squelette, 1000 kJ·mol⁻¹·nm⁻² pour les chaînes latérales, et 400 kJ·mol⁻¹·nm⁻² pour les lipides. La température a été régulée via le thermostat v-rescale, appliqué indépendamment aux groupes protéine, membrane et solvant.

Cette étape, d'une durée de 125 ps avec un pas d'intégration de 1 fs, a permis une première relaxation partielle du système, essentielle pour éviter des changements brusques de structure ou de densité lors de l'équilibration sous pression.

Une dernière étape d'équilibration a été réalisée en conditions NPT (Nombre de particules, Pression, Température constants) afin de permettre au volume de la boîte simulationnelle de s'ajuster à la pression cible de 1 bar, tout en maintenant la température à 303.15 K. Le couplage à la pression a été assuré par le barostat C-rescale, en mode semi-isotrope, ce qui permet un ajustement indépendant dans le plan de la membrane et dans l'axe perpendiculaire (z). Ce couplage est particulièrement adapté aux systèmes contenant des membranes biologiques, chez qui l'anisotropie volumique est essentielle. Le temps de relaxation de la pression a été fixé à 5.0 ps et la compressibilité à 4.5×10^{-5} bar⁻¹, valeur standard pour l'eau. La température a été maintenue à 303.15 K à l'aide du thermostat V-rescale, appliqué indépendamment aux groupes protéine, membrane et solvant, afin de garantir une distribution thermique homogène.

Durant cette phase, des contraintes positionnelles allégées ont été appliquées avec des constantes de force de 1000 kJ·mol⁻¹·nm⁻² pour le squelette protéique, 500 kJ·mol⁻¹·nm⁻² pour les chaînes latérales et 400 kJ·mol⁻¹·nm⁻² pour les lipides, permettant une relaxation partielle du système tout en évitant des déformations brusques de la structure initiale.

Des contraintes de type flat-bottom ont également été conservés entre les atomes de coordination du cuivre et les groupements donneurs (soufres des cystéines et azotes des histidines), avec des distances cibles de 0.21 nm et 0.18 nm respectivement, et des constantes de rappel de 2000 kJ·mol⁻¹·nm⁻². Ces contraintes permettent de préserver une géométrie de coordination réaliste, tout en laissant une certaine flexibilité si les distances restent dans les limites tolérées.

Vient ensuite l'étape finale de production. La phase de production constitue l'étape principale de la simulation de dynamique moléculaire. Contrairement aux étapes précédentes (minimisation et équilibration), dont le but est de stabiliser le système, la production vise à observer l'évolution libre et naturelle du système dans des conditions thermodynamiques contrôlées, représentatives de son état biologique. Durant cette phase, la structure n'est plus

contrainte (à l'exception des contraintes de coordination), ce qui permet de collecter des données dynamiques sur la conformation de la protéine au cours du temps, les interactions avec la membrane, les effets des ligands (éthylène et/ou éthanol), ou encore la stabilité des sites de coordination des ions métalliques. Les coordonnées atomiques sont enregistrées régulièrement.

La simulation de production a été réalisée sur une durée totale de 1 microseconde (1 000 000 ps), avec un pas d'intégration de 2 femtosecondes. Pour se faire, l'utilisation d'un supercalculateur est requise, en l'occurrence il s'agit du calculateur national Irène du CEA à Paris, avec 256 CPUs. Les conditions de température (303.15 K) et de pression (1 bar) ont été maintenues constantes à l'aide du thermostat V-rescale et du barostat C-rescale, respectivement. Le couplage à la pression a été effectué de manière semi-isotrope, pour permettre une expansion différentielle dans le plan de la membrane et dans l'axe perpendiculaire. Aucune contrainte positionnelle n'a été appliquée durant cette phase, de manière à laisser le système évoluer librement dans des conditions thermodynamiques réalistes. Des contraintes de type flat-bottom ont été conservés entre les ions Cu²⁺ et les atomes de coordination (soufres des cystéines et azotes des histidines), avec des distances cibles de 0.21 nm et 0.18 nm et une constante de force de 5000 kJ·mol⁻¹·nm⁻², afin de maintenir une géométrie de coordination réaliste sans imposer de contrainte stricte. Les coordonnées compressées ont été sauvegardées toutes les 50 000 étapes, soit tous les 100 ps.

Ajout des ligands

La littérature actuelle suggère que l'éthanol peut exercer un effet dose-dépendant sur l'activité ou la conformation du récepteur à l'éthylène. (Chen et al, 2020).

Afin de comprendre le rôle inhibiteur de l'éthanol à l'échelle moléculaire, quatre conditions de simulation ont été définies : le récepteur seul, en présence d'éthylène seul, en présence d'éthanol seul, en présence des deux ligands simultanément. Ces conditions de simulation permettent d'évaluer l'influence de l'éthylène, de l'éthanol ainsi que des deux simultanément sur la conformation globale du récepteur.

Pour ce faire, 10 molécules d'éthylène et d'éthanol, correspondant à 23.5mM, ont été ajoutées manuellement au système à l'aide des fichiers topologiques générés via CHARMM-GUI. Ces fichiers ont été inclus dans la topologie principale et les molécules correspondantes ont été positionnées en remplacement de molécules d'eau (TIP3P) initialement présentes dans la boîte de solvatation. La topologie générale a été ajustée manuellement afin d'y inclure les définitions des nouvelles molécules d'éthanol et d'éthylène, ainsi que les déclarations du nombre de molécules ajoutées mais aussi de TIP3P remplacées. Cette étape est importante et nécessite quelques calculs car il faut que le système possède le même nombre d'atomes. Il faut donc savoir combien de molécules de TIP3P ont été supprimées pour obtenir le même nombre d'atomes. Cette modification a permis d'assurer la cohérence entre la topologie et la composition réelle du système simulé.

Les groupes d'index ont également été modifiés pour inclure les ligands dans le groupe SOLU, garantissant ainsi leur prise en compte dans le contrôle de température, le couplage de pression, et les analyses énergétiques. Cette procédure a permis une intégration cohérente et fonctionnelle des ligands dans l'environnement simulé.

Le même protocole pour les étapes de minimisation, équilibrations et production a été appliqué aux systèmes contenant les ligands (figure 4).

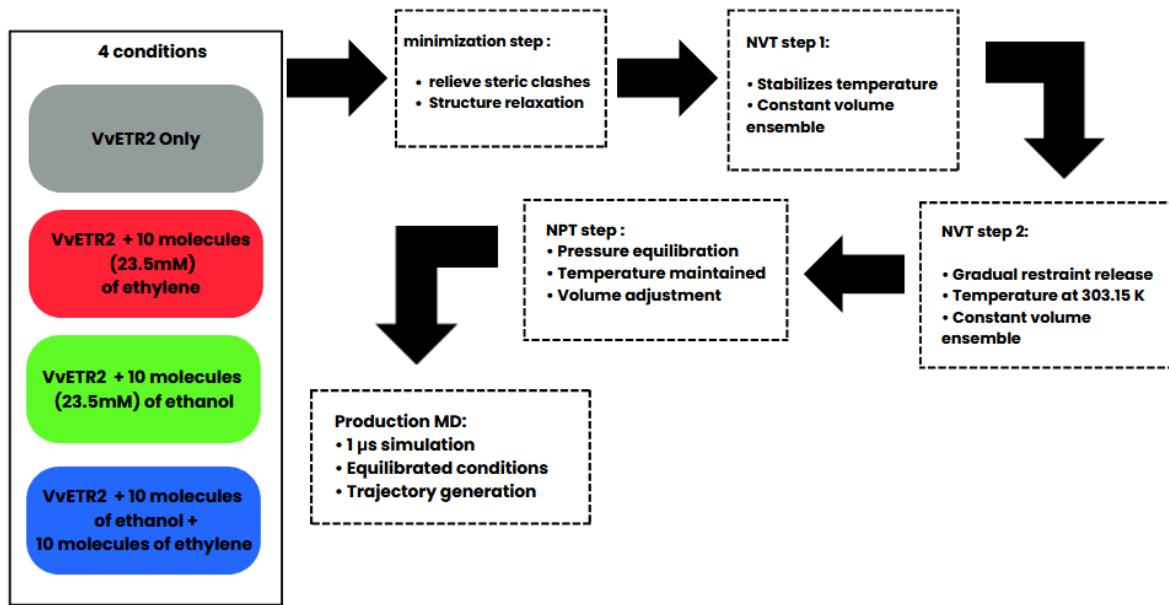


Figure 4 : Pipeline de simulation de dynamique moléculaire en tout atome du récepteur VvETR2 dans une membrane lipidique pour les 4 différentes conditions. Quatre conditions expérimentales ont été simulées : récepteur seul, avec éthanol, avec éthylène, et avec l'éthylène et l'éthanol simultanément (avec un ajout de 10 molécules à chaque fois, soit 23.5 mM). L'intégration des ligands s'est faite en remplaçant des molécules d'eau. Le pipeline comprend une minimisation d'énergie suivie de deux étapes d'équilibration en conditions NVT stabilisation thermique puis relâchement progressif des contraintes, une phase d'équilibration sous pression en NPT (ajustement du volume à 1 bar) et une phase de production de 1 μs sans contrainte positionnelle. Le protocole est identique pour toutes les conditions.

Etude des trajectoires

À l'issue des productions, les trajectoires ont été traitées afin de supprimer les conditions périodiques aux limites. Cette étape est essentielle pour reconstituer les mouvements réels des molécules.

Le traitement a été effectué en trois étapes successives à l'aide de GROMACS trjconv : suppression des sauts provoqués par le passage d'une molécule d'un bord à l'autre de la boîte périodique, recentrer de la protéine dans la boîte afin de faciliter les calculs de distances et les superpositions au cours du temps, remettre la membrane autour de la protéine .

Une fois les trajectoires corrigées, les analyses structurales ont été conduites sur les systèmes seuls et en présence de ligands (éthylène, éthanol). Les RMSD (Root Mean Square Deviation) globaux sur la protéine, sur les carbones alpha, ainsi que sur les éléments de structures secondaires ont été calculés à l'aide de la commande gmx rms.

Les fluctuations des résidus (Root Mean Square fluctuation/RMSF) ont été calculées avec gmx rmsf, en se focalisant sur les carbones alpha afin d'évaluer la rigidité du squelette protéique au cours de la simulation.

Afin de déterminer si l'un des ligands (éthylène ou éthanol) peut se fixer de manière stable aux sites de liaison hypothétiques du récepteur, une analyse a été réalisée à l'aide de scripts Python personnalisés. Pour chaque molécule de ligand simulée, la distance au centre de masse de chaque résidus a été calculée à chaque pas de temps sur l'ensemble de la trajectoire de production. L'évolution temporelle de ces distances a permis d'identifier les événements de rapprochement ou de fixation prolongée. Ces analyses ont été réalisées en utilisant les bibliothèques MDAnalysis (Michaud-Agrawal *et al*, 2011), NumPy (Harris *et al*, 2020) et Matplotlib (Hunter, 2007) sur python3. L'objectif de cette analyse est de déterminer si certaines molécules restent proches d'un site de liaison putatif, correspondant dans notre modèle aux résidus Y F S en position respective 36, 37 et 38 (Wang *et al* 2006) mais aussi sur les résidus D et K en position respective 29 et 94 chez VvETR2 (Azhar *et al*, 2023). L'identification d'une interaction prolongée sur un de ces sites suggérerait une interaction stabilisée. Cela permet de discriminer les comportements entre l'éthylène et l'éthanol, et d'évaluer si ce dernier joue le rôle potentiel d'un inhibiteur (compétitif ou non compétitif). Il peut aussi avoir une action indirecte sur les sites fonctionnels du récepteur par une action sur la membrane. En effet, des études ont mis en évidence le rôle perturbateur de l'éthanol sur les membranes biologiques, en modifiant leur fluidité et l'organisation membranaire. (Patra *et al*, 2006).

Afin de déterminer si les molécules d'éthylène ou d'éthanol s'inséraient dans la membrane ou la traversaient, une analyse de la position des centres de masse des ligands selon l'axe z, perpendiculaire au plan de la membrane, a été réalisée sur toute la durée de la simulation.

Un script Python3 a été développé pour identifier les événements d'entrée, de sortie et de traversée de la membrane et les temps d'entrée, de sortie, durée de séjour dans la membrane et le côté d'entrée et de sortie ont été enregistrés de chaque molécule. Pour y parvenir, la membrane a été définie comme une région comprise entre 4.1 nm et 7.9 nm sur l'axe Z. Ces données ont permis de quantifier les événements de traversées complètes de la membrane, mais aussi d'entrée et sortie du même côté de la membrane, d'identifier des ligands potentiellement piégés ou retenus dans la membrane.

Cette analyse permet de caractériser le comportement des ligands. Ces informations sont essentielles pour comprendre si l'éthanol peut moduler indirectement l'activité du récepteur via son interaction membranaire, indépendamment d'une fixation directe. Les résultats ont été sauvegardés sous forme de fichiers CSV et représentés graphiquement sous forme de diagrammes de présence temporelle par molécule. Cette méthode permet d'explorer l'affinité ou perméabilité de l'environnement membranaire pour chaque molécule, un paramètre qui peut être déterminant dans la modulation de l'activité du récepteur.

Une analyse complémentaire a été effectuée. Une boîte définie manuellement des régions comprises pour X de 1,9 nm à 4,7 nm et pour Y 1,9 nm à 4,6 nm pour déterminer un espace dans la membrane correspondant à l'intérieur de la protéine. Cela rend possible la détermination du temps passé par molécule d'éthylène et d'éthanol dans la protéine, dans la membrane et hors membrane. Nous allons observer quels ligands passent plus de temps dans la protéine et dans la membrane mais aussi si la présence simultanée des deux ligands empêche l'accès à la protéine et/ou à la membrane ou au contraire le favorise. Les résultats seront représentés par des histogrammes, ainsi que par un tableau avec les moyennes et les écarts type, qui contiendra aussi la vitesse de diffusion moyenne des ligands à travers la membrane.

Résultats

VvETR2 structuré selon les contraintes biologiques

Définition d'un domaine transmembranaire de 120 acides aminés

L'analyse de la séquence complète du récepteur à l'éthylène 2 de *Vitis vinifera* avec SignalP 6.0 a révélé que les 20 premiers acides aminés présentent une probabilité de 90 % d'être un peptide signal (figure 5). Les peptides signaux sont destinés à être clivés lors de la translocation dans le réticulum endoplasmique et sont donc absents dans la forme mature de la protéine. Ce résultat justifie l'exclusion des 20 premiers acides aminés pour la suite des analyses structurales.

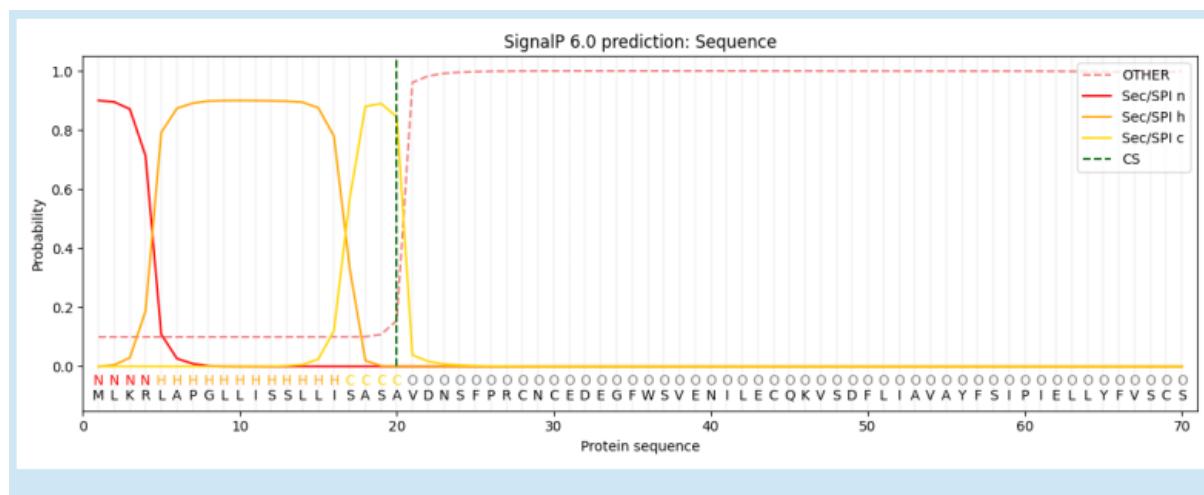


Figure 5. Prédiction du peptide signal de la protéine VvETR2 de *Vitis vinifera* à l'aide de SignalP 6.0. Le graphique montre la probabilité de classification de chaque acide aminé de la séquence protéique en différentes catégories. Sec/SPI n : Probabilité que l'acide aminé soit dans la région N-terminale du signal peptide. Sec/SPI h : Probabilité que l'acide aminé soit dans la région hélicoïdale (hydrophobe) du signal peptide. Sec/SPI c : Probabilité que l'acide aminé soit dans la région de clivage du signal peptide. OTHER : Probabilité que la séquence ne contient pas de signal peptide. CS : Site de clivage prédict, c'est-à-dire l'endroit où le signal peptide serait coupé.

De plus, l'algorithme DeepTMHMM a prédict la présence d'un peptide signal (résidus 1–20), suivi de segments extracellulaires résidus 21–46 et 98–111, intracellulaires résidus 68–76 et

133–745, et de trois segments transmembranaires (TMhelix), situés entre les résidus 47–67, 77–97, et 112–132, respectivement.

La séquence étant complète, elle comporte la partie du signal peptide, la partie N-terminale, transmembranaire, la partie à la lumière du réticulum endoplasmique ainsi que la partie C-terminale.

Un peptide signal est une chaîne peptidique qui permet d'acheminer la protéine à un compartiment cellulaire. Cette chaîne étant clivée lorsque la protéine est en place, il était nécessaire de l'identifier afin de l'écartez pour les analyses.

Le domaine transmembranaire comprenant les résidus 21–140, prédict par DeepTMHMM, a été extrait pour les prédictions structurales et simulations. Ce segment comprend les trois hélices transmembranaires, le site de coordination du cuivre, et la région de fixation de l'éthylène. Tous ces éléments étant dans la partie transmembranaire, la conservation de cette partie pour la suite de l'étude permet de maximiser la qualité des prédictions et de réduire la complexité des simulations.

DeepTMHMM	Chai1	OmegaFold	AF3	Boltz	Chai1CU	BoltzCU
1-26	1-17	1-18	1-9	1-17	1-17	1-17
27-47	18-48	19-49	10-49	18-49	18-49	18-49
48-56	49-54	50-55	50-54	50-54	50-54	50-54
57-77	55-77	56-78	55-77	55-75	55-77	55-75
78-87	78-83	79-84	78-83	76-83	78-83	76-83
88-112	84-119	85-110	84-119	84-118	84-118	84-118

Hélice Linker

Figure 6. Numérotation des régions hélicoïdales et des segments interhélices (linkers) prédites par différents outils de modélisation et DeepTMHMM. Les plages de résidus correspondant aux hélices transmembranaires (vert clair) et aux linkers (vert foncé) sont indiquées pour chaque méthode. Les versions "CU" correspondent aux modèles intégrant explicitement les ions cuivre. Le modèle AlphaFold3 a toujours des cuivres.

DeepTMHMM, basé sur l'apprentissage profond, prédit trois hélices principales aux positions 48–56, 57–77, et 88–112, tandis que les modèles tridimensionnels issus de Chai1, OmegaFold, AlphaFold3 (AF3), et Boltz confirment une topologie similaire avec de légères variations (figure 6).

Afin d'évaluer la cohérence structurale des modèles générés par différentes méthodes de prédiction tridimensionnelle, les positions des hélices transmembranaires et des segments interhélice (linkers) ont été comparées à celles prédictes par DeepTMHMM, considéré comme référence pour la topologie membranaire.

Comme le montre la Figure 6, la plupart des outils (Chai1, OmegaFold, Boltz, et leurs variantes avec cuivre) reproduisent globalement la même architecture, avec des hélices aux positions attendues et des linkers correctement espacés. Les trois segments transmembranaires observés dans DeepTMHMM (48–56, 57–77, 88–112) sont bien retrouvés par ces méthodes, avec des décalages marginaux à leurs extrémités. En revanche, le modèle issu d'AlphaFold3 (AF3) présente des variations notables : la première hélice débute très tôt (résidus 10–49), et la première zone de linker est beaucoup plus courte que chez les autres prédicteurs. Cette discordance, combinée à d'autres critères structurels analysés ultérieurement, suggère qu'AF3 pourrait être moins fiable pour modéliser ce domaine transmembranaire, et justifie son exclusion des modèles retenus pour la suite des simulations.

Plusieurs régions sont très conservées à travers l'évolution

La forte conservation de certains résidus, notamment dans les régions de liaison au cuivre et des sites putatifs de liaison de l'éthylène, confirme l'importance structurale et fonctionnelle de ces motifs dans l'activité des récepteurs à l'éthylène.

L'alignement multiple de séquences des récepteurs à l'éthylène, issus de différentes espèces végétales, a permis de générer des WEblogo de séquence illustrant la conservation des acides aminés le long du domaine transmembranaire.

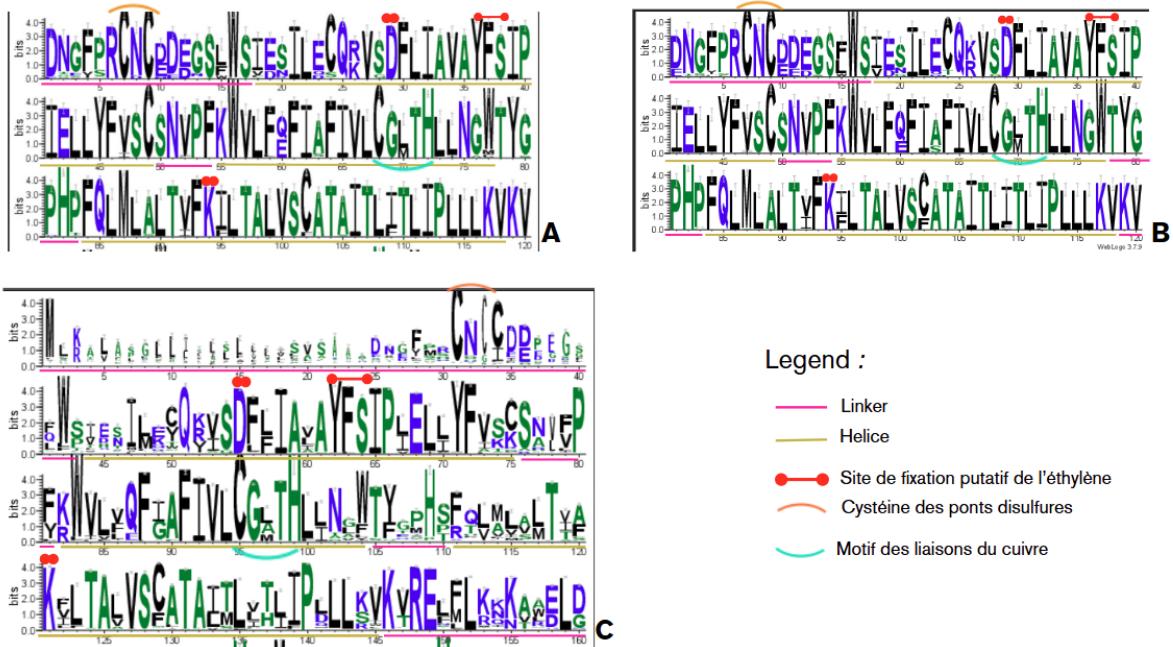


Figure 7. Logo des séquences représentant la conservation des acides aminés dans des alignements multiples de récepteurs à l'éthylène. Les logos ont été générés à partir d'alignement CLUSTAL Omega. L'axe vertical en bits indique la conservation informationnelle de chaque position de 0 à 4 bits, et la taille des lettres représente la fréquence relative des acides aminés à cette position. A : logo obtenu à partir d'un alignement de séquences récupérés avec un BLASTP de la partie transmembranaire. B : logo obtenu à partir d'un alignement de séquences récupérés avec un BLASTP de la protéine entière. C : logo obtenu à partir d'un alignement de séquences issu d'une recherche EMBL. Les annotations en couleur mettent en évidence des régions structurales ou fonctionnelles. En turquoise : motif de liaison au cuivre. En rouge : sites de fixation supposés de l'éthylène. En orange : cystéines impliquées dans des ponts disulfures. En rose : partie N-terminale et séquences inter-hélices. En jaune : les hélices transmembranaires

Il y a peu de différence entre la conservation de l'alignement des séquences issu du BLASTP de la protéine entière (figure 7 B) et de la partie transmembranaire (figure 7 A). Les parties inter-hélices, linker sont très bien conservées dans ces deux alignements. La partie transmembranaire à elle seule suffit à capturer les principales caractéristiques de conservation observées pour la protéine entière.

Certains résultats dans l'alignement multiple avec tous les ETRs témoignent de domaines extrêmement bien conservés (figure 7) y compris dans l'alignement comportant 1316 séquences issus d'une recherche sur Uniprot (figure 7C). Les positions correspondant aux cystéines et histidines du motif CxxH, impliquées dans la coordination du cuivre (Roger *et al.*, 2024), ont une conservation quasi-totale, comme les hélices transmembranaires, ainsi que le site du fixation du ligand.

Ces cuivres interviennent dans la fixation du ligand et sont proches des sites de fixation. Chaque monomère est associé à un cuivre, il semble donc que les liaisons d'un cuivre se font avec le soufre de la cystéine 68 et l'azote de l'histidine 72 du même monomère. Cependant, il est possible qu'un des rôles secondaires du cuivre soit de stabiliser le dimère. Il est donc nécessaire de s'interroger sur l'impact du cuivre dans les prédictions de modèle. Ces résultats démontrent l'importance des mécanismes de reconnaissance du ligand et de coordination métallique.

Les sites putatifs de liaison à l'éthylène sont très conservés également dans cet alignement (figure C) traduisant une forte pression de sélection. Cela démontre leur importance fonctionnelle probable, une attention particulière a naturellement été portée sur ses acides aminés lors des études d'interactions ligands/protéines.

Les ions cuivres stabilisent le modèle

Des matrices predicted alignment error (PAE) générés par AlphaFold 3 et Chai1 avec ou sans cuivre ont été comparées. Une matrice PAE est une auto-évaluation du modèle qui permet de déterminer l'erreur de prédiction attendue pour un résidu i, si un résidu j est utilisé comme position de référence. Cela est effectué pour toutes les paires de résidus existantes. En complément des matrices PAE, les prédictions ont également fourni des scores predicted TM-score (pTM) et interchain predicted TM-score (ipTM) (Jumper *et al.* 2021). Ces scores ont été utilisés pour évaluer la confiance dans le repliement de chaque chaîne mais aussi leur position relative au sein du complexe, si les chaînes sont bien positionnées l'une par rapport à l'autre.

La superposition des structures prédites avec et sans ion cuivre, pour un même outil de prédiction, révèle des conformations globales identiques. Cela suggère que l'ajout du cuivre n'entraîne pas de réarrangement majeur de la structure tertiaire.

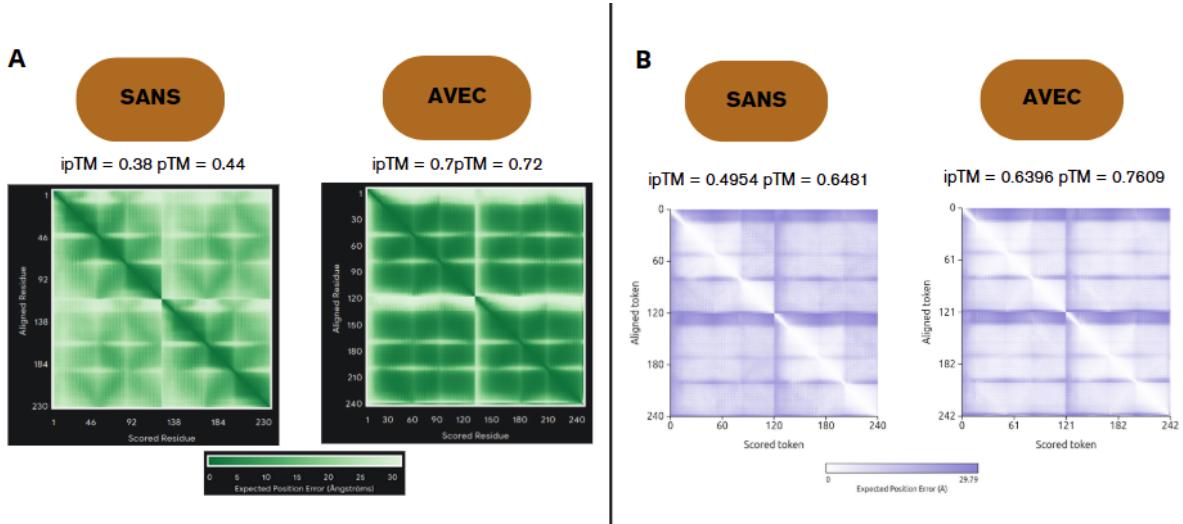


Figure 8. Comparaison des matrices Predicted Alignment Error (PAE) et des scores ipTM/pTM pour les modèles prédits avec et sans ions cuivre. A : matrices PAE et scores de confiance globale (pTM) et interchaîne (ipTM) générées par AlphaFold3. B : Matrices PAE et score pTM et ipTM, générées par Chai1. À gauche de chaque figure A et B, les modèles sans ions cuivre ; à droite, les modèles avec ions cuivre. L’ajout du cuivre améliore visiblement les scores de confiance globale (pTM) et interchaîne (ipTM), indiquant une meilleure définition des interfaces et un repliement plus cohérent.

Les matrices PAE obtenues via Chai1 et AlphaFold 3 révèlent une amélioration significative de la qualité du modèle en présence d’ions cuivre. Les scores pTM et ipTM augmentent dans les deux cas : pour Chai1, le pTM passe de 0.6481 à 0.7609 et le ipTM de 0.4954 à 0.6396 ; pour AlphaFold 3, le pTM passe de 0.44 à 0.72 et le ipTM de 0.38 à 0.70 (figure 8).

Ces améliorations suggèrent que la prise en compte des ions cuivre dans la prédiction du modèle contribue non seulement à un repliement plus fiable des chaînes individuelles, mais également à une meilleure prédiction de leur interface au sein du complexe. Cela corrobore l’hypothèse selon laquelle le cuivre stabilise le dimère et joue un rôle structurant au-delà de sa fonction de liaison à l’éthylène.

Chai1 comme modèle de référence pour VvETR2

Certains modèles prédits par les différentes méthodes de modélisation présentent une conformation similaire, tandis que d'autres montrent des différences structurelles notables. L'objectif était de voir si les différents modèles semblent construire VvETR2 de la même façon et quelles sont leurs divergences.

Les structures ont été comparées deux à deux, à l'aide de PyMOL (DeLano, 2002) et TM-align (Zhang & Skolnick, 2005). Ces outils permettent de superposer les modèles en minimisant la distance moyenne entre les atomes correspondants.

En complément de la superposition visuelle, TM-align fournit un Template Modeling Score (TM-score), qui quantifie le degré de similarité entre deux structures. Ce score permet de déterminer dans quelle mesure deux protéines partagent un repliement global similaire, indépendamment de leur taille.

Les superpositions ont également été réalisées dans PyMOL. Afin d'éviter les biais liés à des régions désordonnées, les extrémités N-terminales (résidus 1 à 19 sur chaque chaîne) ont été exclues de la commande d'alignement. Par exemple, pour aligner les modèles AlphaFold3 et Chai1 en excluant les régions N-terminales, la commande suivante a été utilisée : “align AF3 and (not (chain A and resi 1-19) and not (chain B and resi 1-19)), chai1 and (not (chain A and resi 1-19) and not (chain B and resi 1-19))”.

De plus, PyMol permet de calculer automatiquement les Root Mean Square Deviation (RMSD). Le RMSD est une mesure mathématique qui calcule la différence moyenne de position entre deux ensembles d'atomes après une superposition de deux structures.

Une superposition des structures prédites par différents outils (AlphaFold 3, Chai1, OmegaFold, Boltz1) ainsi que le modèle d'ETR1 d'*Arabidopsis thaliana* de référence issu de Schott-Verdugo *et al* (2019) et une version actualisée du modèle ETR1 issu de la prédition d'AlphaFold3 a permis d'évaluer la cohérence des prédictions. Visuellement, l'ensemble des structures présente une bonne conservation globale, en particulier entre les modèles issus de Chai1 et AlphaFold 3. En revanche, le modèle généré par OmegaFold se distingue nettement.. Une comparaison des modèles via le calcul des RMSD met en évidence une cohérence marquée entre les structures générées par Chai1, AlphaFold 3, et leurs variantes, notamment avec l'ajout du cuivre. Les prédictions issues d'OmegaFold s'écartent nettement, avec des

RMSD dépassant 8 Å par rapport aux autres outils. Cette divergence suggère que, dans ce contexte particulier, OmegaFold produit une topologie significativement différente des autres outils de prédiction.

Cela justifie que le modèle Omegafold ne soit pas retenu comme modèle de référence.

Le modèle Chai1 (figure 9) a été retenu pour la suite des manipulations. Cela s'explique car les superpositions des structures ont montré une excellente similitude entre Chai1, AlphaFold 3, et Boltz1, avec des RMSD très faibles. Cela témoigne de la convergence des modèles dans leur représentation de VvETR2. D'autre part, le modèle Alphafold3 ne peut pas être retenu en raison du N-terminale trop courte, avec une cystéine qui se situe dans une hélice alors qu'elle doit être disponible pour un pont disulfure.

De plus, Chai-1 a permis une intégration simple de l'ion cuivre, ce qui s'est traduit par une amélioration significative des scores de confiance (pTM et ipTM) et une meilleure définition de l'interface dimérique. Ces éléments ont motivé son utilisation préférentielle pour les analyses structurales approfondies.

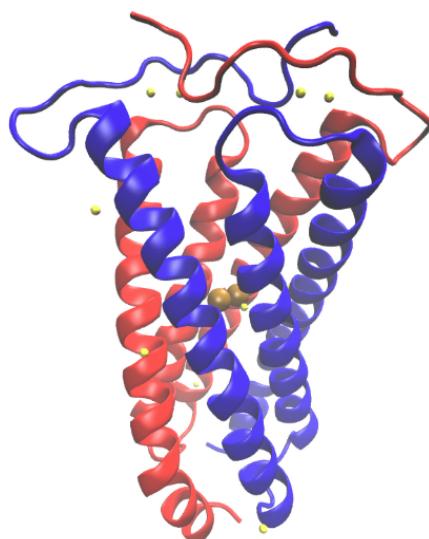


Figure 9. Modèle Chai1 visualisé avec VMD. Les structures tridimensionnelles des chaînes A et B apparaissent respectivement en bleu et en rouge en représentation New Cartoon. Les cuivres sont visibles au centre de la protéine en marron en représentation VDW. Les boules jaunes en représentation CPK représentent les soufres des cystéines qui ont servi à faire les deux ponts disulfures dans la partie N-terminale.

Croisement des chaînes N-terminale pour la création de ponts disulfures

Afin de reconstituer un dimère fonctionnel (figure 1), la formation de ponts disulfures entre les monomères a été explorée. Les extrémités N-terminales des deux monomères ont été raccourcies (résidus 11 à 18 supprimés), ce qui a permis un rapprochement manuel des cystéines (Cys8 et Cys10) via l'interface de PyMOL, dans différentes orientations spatiales. Ensuite, l'outil MoMA-FreSa a été utilisé pour reconstruire les segments manquants entre les résidus 1–10 et 18–120, en imposant la séquence native comme lien. Cette étape a généré une dizaine de conformations pour chaque monomère et configuration. Parmi toutes les combinaisons testées, un seul arrangement permettait la formation simultanée de deux ponts disulfures interchaînes, tout en maintenant une conformation stable et sans clash : Le seul modèle cohérent correspond à une création d'un premier pont disulfure avec la cystéine 8 de la chaîne A qui s'assemble avec la cystéine 10 de la chaîne B et d'un deuxième pont disulfure entre la cystéine 10 de la chaîne A qui s'assemble avec la cystéine 8 de la chaîne B (figure 10).

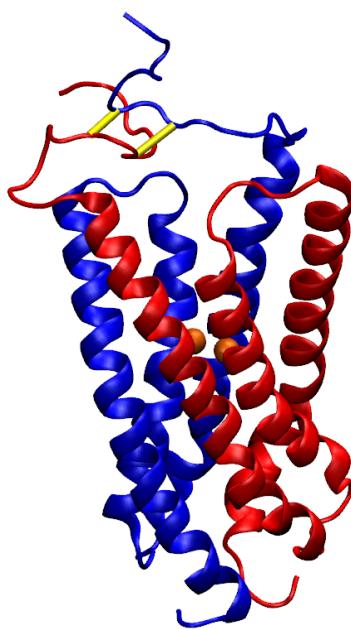


Figure 10. Modèle Chai1 avec ponts disulfures visualisé avec VMD. Les structures tridimensionnelles des chaînes A et B apparaissent respectivement en bleu et en rouge. Les cuivres sont visibles au centre de la protéine en marron en représentation VDW. Les ponts disulfures apparaissent en jaune en représentation dynamic bonds.

Analyse des trajectoires dans les différentes conditions

La stabilité de la protéine varie selon les conditions

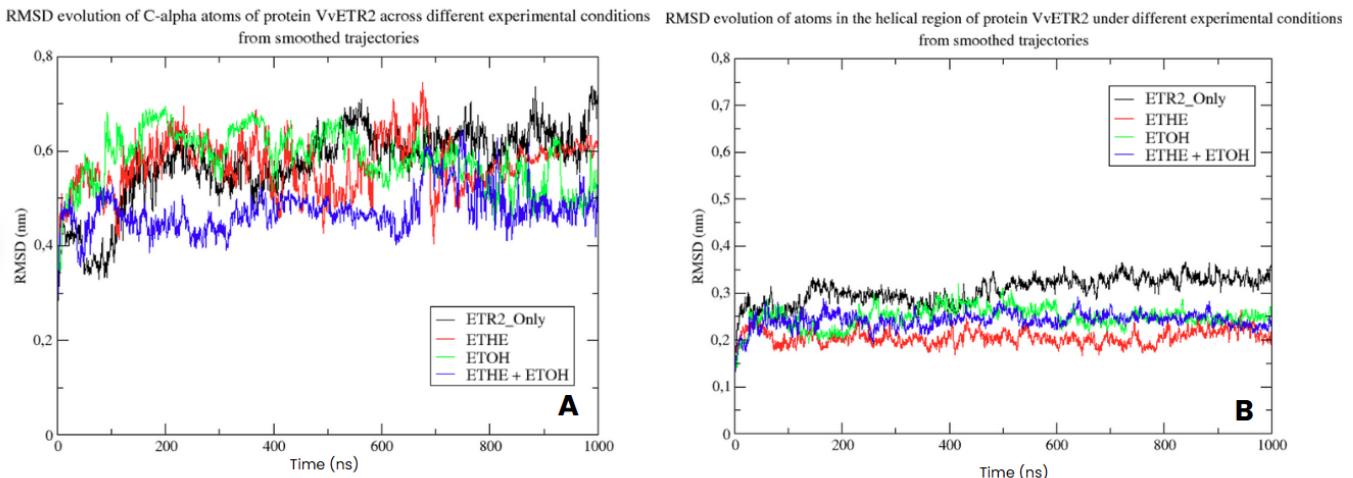


Figure 11. Evolution du Root Mean Square Deviation (RMSD) au cours du temps à partir des trajectoires lissées de chaque condition. Permet l'évaluation de la stabilité conformationnelle de la protéine dans les différentes conditions. A : le RMSD a été effectué sur les carbones alphas des résidus de la protéine VvETR2 pour chaque condition reflétant les déviations du squelette protéique. B : le RMSD a été effectué sur les atomes des résidus appartenant aux régions des hélices transmembranaires de la protéine VvETR2 pour chaque condition. En noir : le protéine seule. En rouge : avec éthylène. En vert : avec éthanol. En bleu : avec éthylène et éthanol.

Une phase de stabilisation est observable dans les 100 nanosecondes pour l'ensemble des conditions, à la fois pour le RMSD global de la protéine (Figure 11A) et pour le RMSD restreint aux hélices transmembranaires (Figure 11B).

Le RMSD effectué sur les carbones alpha atteint une valeur moyenne stabilisée entre 0,55 et 0,65 nm (figure 11A), ce qui indique une stabilité globale du squelette protéique . Le système sans ligand présente cependant des fluctuations plus importantes en fin de simulation, traduisant une plus grande variabilité conformationnelle du squelette protéique.

Le RMSD effectué sur les atomes des hélices transmembranaires (figure 11B) atteint une valeur moyenne stabilisée entre 0,2 et 0,3 nm. Les valeurs stabilisées sont significativement

plus faibles, entre que le RMSD effectué sur les C-alpha. Ce résultat confirme la rigidité des hélices transmembranaires. Il semble que les conditions contenant au moins un ligand présentent une meilleure stabilité que la condition VvETR2 seule, ce qui suggère un effet stabilisateur des ligands sur la portion transmembranaire.

L'éthylène seul (courbe rouge) semble réduire les fluctuations au niveau des hélices transmembranaires de manière plus marquée que sur l'ensemble du squelette protéique, ce qui pourrait indiquer une interaction localisée dans la région membranaire (figure 11B). À l'inverse, la présence d'éthanol seul (courbe verte) induit des variations plus importantes dans le RMSD fait sur les C-alpha (Figure 11A), ce qui suggère un effet potentiellement désorganisant sur la conformation globale du récepteur.

Dans la condition éthylène + éthanol (courbe bleue) la valeur de RMSD sur le squelette protéique est plus faible en comparaison aux autres conditions (Figure 11A). Cela peut signifier une stabilisation synergique du squelette protéique ou un effet de compensation structurale entre les deux ligands.

Le calcul du RMSF (Root Mean Square Fluctuation) des atomes C-alpha des résidus de VvETR2 permet d'évaluer la flexibilité locale des différentes régions de la protéine dans chaque condition (Figure 12). Comme attendu, les segments correspondant aux hélices transmembranaires présentent des fluctuations faibles, traduisant une stabilité structurale élevée. À l'inverse, les régions non structurées (les extrémités N- et les segments inter-hélice marqués en rose) présentent des RMSF plus élevés, indiquant une mobilité accrue.

Une légère augmentation du RMSF est observée à la fin de la dernière hélice, à partir du résidus 110, dans la chaîne B en condition éthylène et éthanol combinée (courbe bleue). Cet effet est absent dans la chaîne A. Cela suggère une possible perturbation locale ou une flexibilité accrue dans un contexte de double ligand. Cependant, cette variation n'est pas suffisante pour traduire une déstabilisation générale de la structure.

Ces résultats attestent de la robustesse de l'organisation transmembranaire de VvETR2 dans toutes les conditions testées.

RMSF of C-alpha atoms of residues of VvETR2 across different experimental conditions

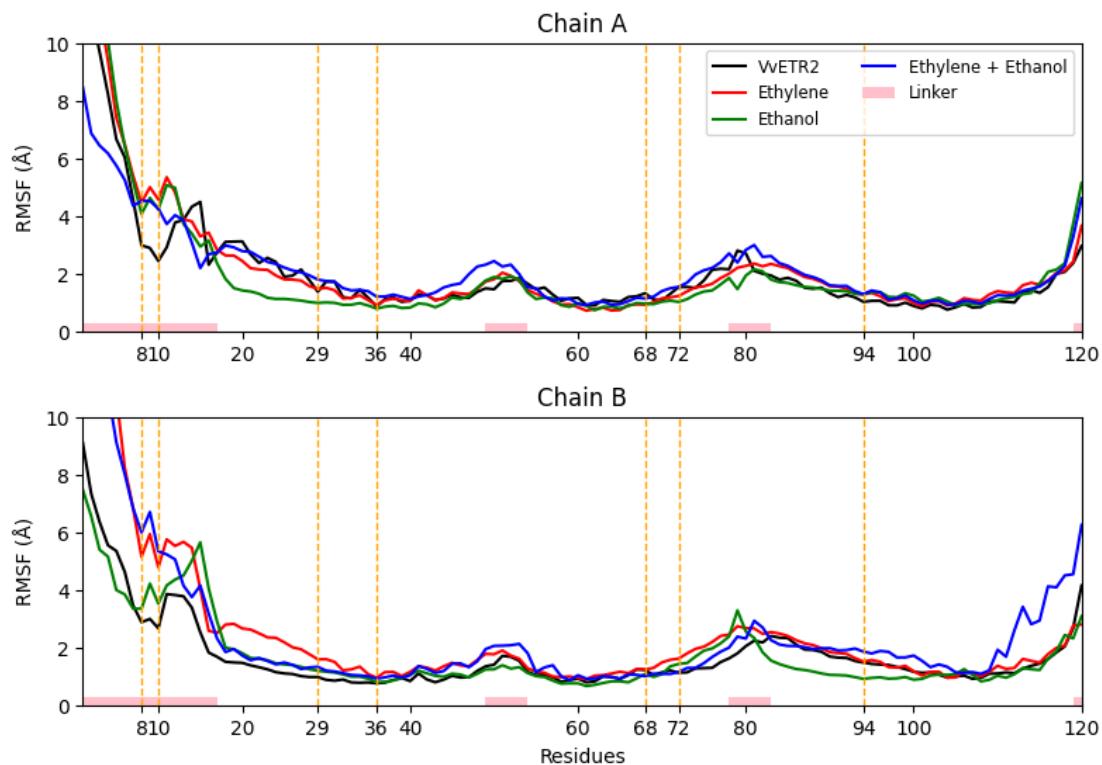


Figure 12. Fluctuation de la distance des résidus par rapport à sa position moyenne au cours du temps dans des simulations de 1 µs pour chaque chaîne de la protéine ETR2 de *vitis vinifera*. En noir : le protéine seule. En rouge : avec éthylène. En vert : avec éthanol. En bleu : avec éthylène et éthanol. Rose : Partie non-structurée de la séquence, partie N-ter et linker. En jaune : Marqueur de position correspondant aux sites putatifs de fixation (29, 36, 94), aux résidus impliqués dans la coordination du cuivre (68, 72), ou aux résidus impliqués dans les ponts disulfures (8,10).

L'analyse combinée des RMSD et RMSF confirme que la structure globale de VvETR2 est stable dans toutes les conditions simulées, en particulier au niveau des hélices transmembranaires. Les RMSD montrent une stabilité accrue en présence de ligands, notamment dans la condition combinée éthylène + éthanol, tandis que les RMSF mettent en évidence une flexibilité localisée dans certaines régions non structurées.

L'éthylène seul semble renforcer la stabilité des régions transmembranaires, alors que l'éthanol seul induit une variabilité conformationnelle plus marquée. Ces effets contrastés se retrouvent dans les deux types d'analyses, suggérant une influence structurale différenciée des deux ligands : stabilisation membranaire par l'éthylène, et perturbation plus diffuse par

l'éthanol. Le comportement observé en présence conjointe des deux ligands pourrait ainsi résulter d'un équilibre entre ces deux effets.

L'éthylène diffuse à travers la membrane

L'éthylène en condition seul et en condition éthylène + éthanol diffuse à travers la membrane lors de la simulation (figure 13) (figure 14 en annexe) en raccord avec la littérature (Dubois et al. 2018). Cependant, l'éthanol ne diffuse pas à travers la membrane, ni en condition éthanol seul, ni en condition éthylène + éthanol. (figure 15 et figure 16 en annexe). L'éthanol peut s'insérer dans la membrane mais son groupement -OH hydrophile l'empêche de diffuser librement. L'éthylène est donc capable de traverser la bicouche lipidique avec ou sans éthanol, ce qui n'est pas le cas de l'éthanol.

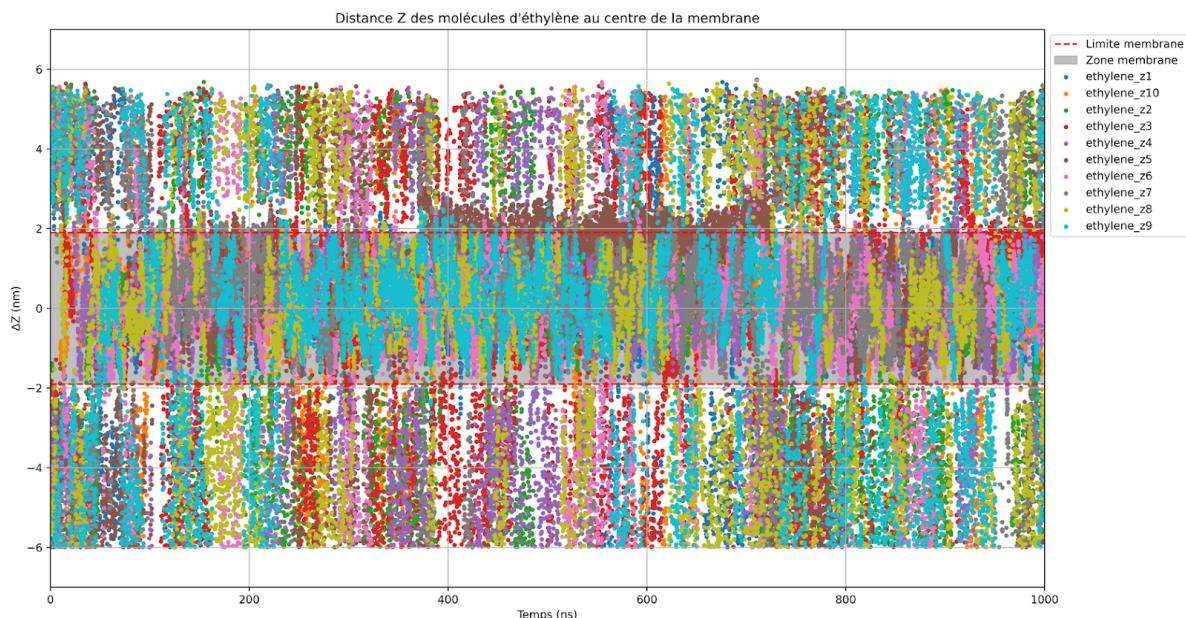


Figure 13. Distance des 10 molécules d'éthylène par rapport au centre de la membrane mesuré grâce à Delta Z en condition éthylène seul. Delta Z = position de l'éthylène sur l'axe Z - position du centre de la membrane sur l'axe Z. La zone grisée correspond à l'intérieur de la membrane. Les tirets rouges délimitent l'intérieur et l'extérieur de la membrane.

Dans la condition éthylène seul, il y a 108 diffusions qui se font en moyenne en 19.582ns +/- 16.615ns. Dans la condition éthylène couplé avec de l'éthanol il y a 123 diffusions avec un temps moyen de 21.843ns +/- 18.103ns.

Dans ce modèle, la diffusion de l'éthylène ne semble pas être impactée par la présence d'éthanol, ni en quantité, ni de vitesse de diffusion.

L'éthanol influence le comportement de l'éthylène

Afin de mieux comprendre l'impact de l'éthanol sur le récepteur et sur la fixation de l'éthylène, des analyses concernant les positions des différents ligands ont été effectuées dans les différentes conditions.

En condition éthylène seul (Figure 17A), la majorité des molécules d'éthylène sont localisées dans la membrane (barres orange) au cours de la simulation, avec une occupation modérée dans la protéine (bleu) et faible hors membrane (vert). Lorsqu'il est combiné à l'éthanol (Figure 17B), le temps d'occupation de l'éthylène dans la protéine augmente sensiblement, indiquant une possible facilitation ou stabilisation de la fixation protéique par l'éthanol. Cela est corroboré par la moyenne d'occupation dans la protéine, qui passe de $63 \text{ ns} \pm 47,6$ à $143 \text{ ns} \pm 132,0$ (figure 19).

À l'inverse, l'éthanol (Figure 18), ne pénètre presque jamais la protéine, et reste majoritairement hors de la membrane, quelle que soit la condition. En condition éthanol seul (Figure 18A), quelques événements d'insertion partielle dans la membrane sont observés, mais l'éthanol passe plus de 80% du temps de simulation hors de la membrane. La co-présence de l'éthylène (Figure 18B) ne modifie pas significativement ce comportement, et l'éthanol n'est pas attiré ou stabilisé par la protéine, contrairement à l'éthylène.

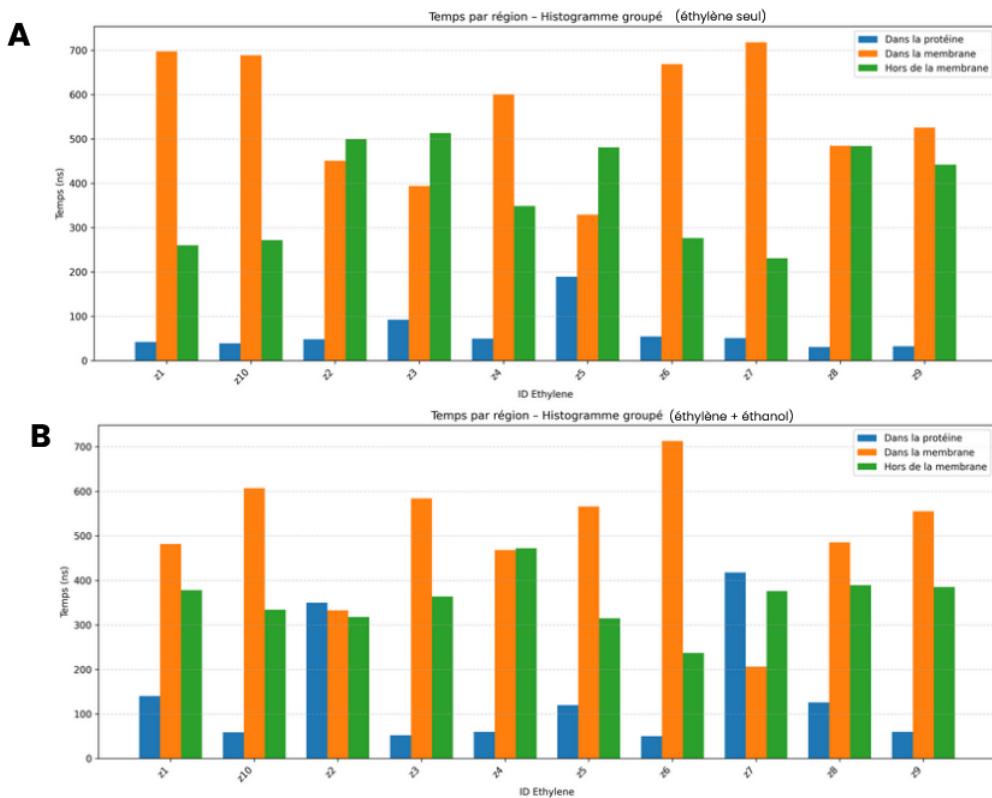


Figure 17. Temps passé par chaque éthylène dans chaque région du système. A : en condition éthylène seul. B : En condition éthylène + éthanol. Bleu : Temps passé dans la protéine. Orange : Temps passé dans la membrane. Vert : temps passé hors membrane et hors protéine.

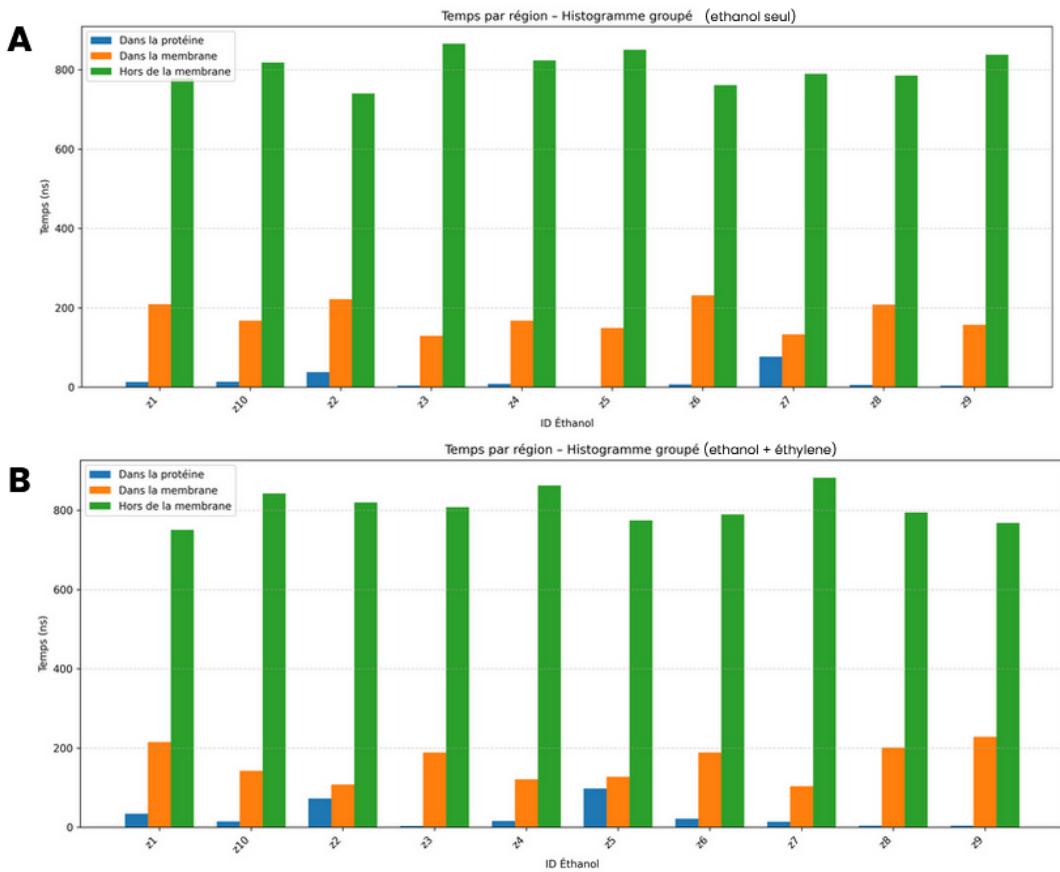


Figure 18. Temps passé par chaque éthanol dans chaque région du système. A : en condition éthanol seul. B : En condition éthylène + éthanol. Bleu : Temps passé dans la protéine. Orange : Temps passé dans la membrane. Vert : temps passé hors membrane et hors protéine.

Une forte variabilité intermoléculaire a été observée dans les temps de résidence des ligands, en particulier dans le temps passé dans la protéine (figure 19). Les écarts-types élevés par rapport à la moyenne reflètent des comportements hétérogènes entre les différentes molécules simulées, certaines interagissant plus fréquemment et plus durablement avec la protéine, tandis que d'autres restent principalement dans la membrane ou à l'extérieur du système.

Ces résultats suggèrent que l'éthanol n'agit pas comme un ligand du récepteur VvETR2, mais peut en revanche moduler la fixation de l'éthylène, possiblement en influençant l'environnement lipidique ou la conformation locale de la protéine. L'effet semble indirect mais significatif, au vu du temps de résidence accru de l'éthylène dans la protéine en présence d'éthanol.

Average time \pm standard deviation of ligand residence in the protein or membrane under different conditions

Condition :	Ethylene Only	Ethanol Only	Ethylene + Ethanol	
			Ethylene	Ethanol
Protein	63.050 \pm 47.559 (ns)	17.270 \pm 23.436 (ns)	143.370 \pm 131.996 (ns)	28.280 \pm 31.932 (ns)
Membrane	555.930 \pm 138.703 (ns)	177.57 \pm 37.186 (ns)	499.880 \pm 144.213 (ns)	162.570 \pm 46.785 (ns)

Figure 19. Temps moyen (en nanosecondes) \pm écart-type de résidence des ligands dans la protéine ou dans la membrane lipidique pour les différentes conditions de simulation.

Les valeurs correspondent au temps cumulé moyen passé par chaque molécule d'éthylène ou d'éthanol dans la protéine (zone de contact directe avec VvETR2) ou dans la membrane (région lipidique) au cours des 1000 ns de simulation. Les conditions incluent : éthylène seul, éthanol seul, et co-présence d'éthylène et d'éthanol.

L'éthylène et l'éthanol interagissent avec le récepteur

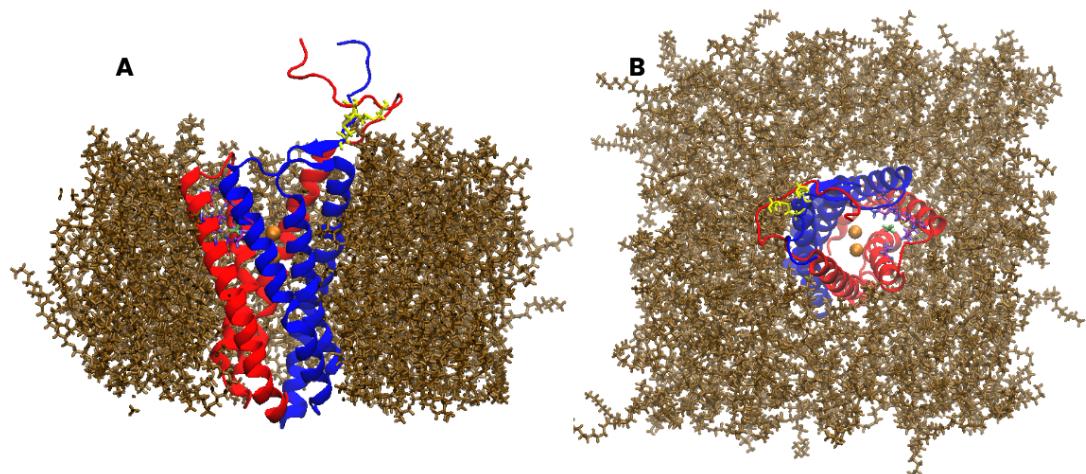


Figure 20. Images du système complet. A : vu de coupe. B : Vu de haut (côté N-ter). En rouge : Chaîne A de VvETR2. En bleu : Chaîne B VvETR2. Jaune : Ponts disulfures. Orange : Ions Cu²⁺. Marron : Membrane lipidique.

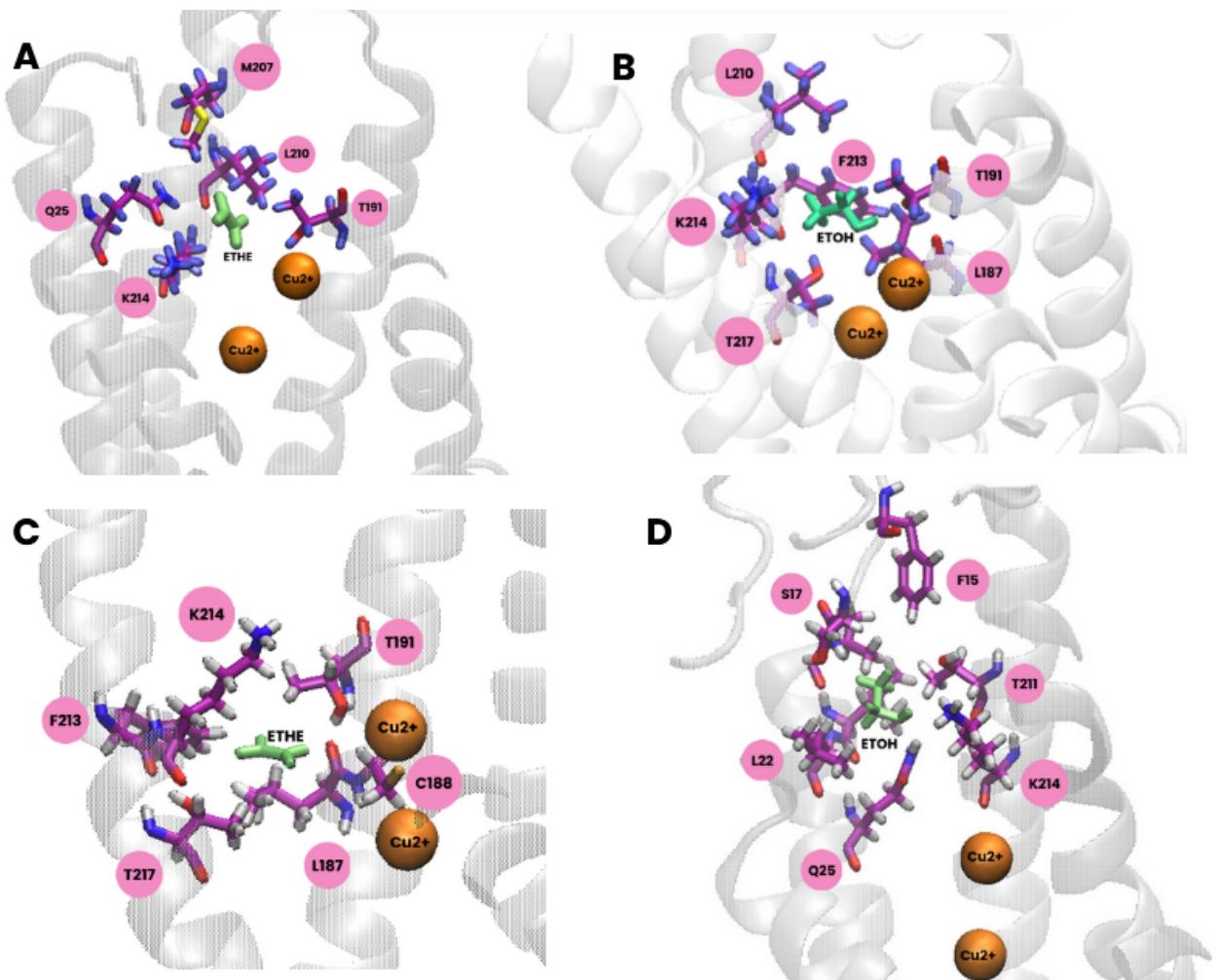


Figure 21. Image d'une interaction entre une molécule d'éthylène et le récepteur ETR2 de *Vitis Vinifera* en condition éthylène seul visualisé sur VMD. En orange : Les ions Cu²⁺ en représentation VDW. En blanc : la structure tridimensionnelle de VvETR2 représenté en new cartoon. En violet : les résidus à moins de 4 angstroms de la molécule d'éthylène ou d'éthanol. En rose : les labels avec le nom de l'acide aminé et sa position dans la séquence. ETHE : molécule d'éthylène, en vert, en représentation licorice. ETOH : molécule d'éthanol, en vert, en représentation licorice. A : molécule d'éthylène interagissant avec K214 en condition éthylène seul. B : molécule d'éthanol interagissant avec K214 en condition éthanol seul. C : molécule d'éthylène interagissant avec K214 et C188 (cystéine impliquée dans la liaison de l'ion cuivre) en condition éthylène + éthanol. D : molécule d'éthanol interagissant avec K214 en condition éthylène + éthanol.

La Figure 20 illustre la configuration globale du système simulé. La protéine VvETR2 est insérée dans une bicouche lipidique représentée en marron, dans une orientation perpendiculaire au plan membranaire. Les deux chaînes (A en rouge, B en bleu) forment une structure stable. Les ions Cu²⁺ (sphères orange) sont positionnés entre les hélices des deux chaînes. Les ponts disulfures, représentés en jaune, sont positionnés dans le lumen du réticulum endoplasmique, contribuant à la stabilité du dimère. La vue en coupe (Figure 20A) et la vue de dessus (Figure 20B) permettent de visualiser à la fois l'épaisseur membranaire, l'insertion du récepteur dans la membrane, et la répartition spatiale des éléments principaux.

Dans chaque condition, des molécules d'éthylène et d'éthanol ont été aperçues comme interagissant avec le récepteur ETR2 de *Vitis vinifera*. Pour notre étude, seules les ligands situés à moins de 5 angstroms durant au moins 20ns avec un résidu ou un groupe de résidu ont été considérées comme des interactions afin d'exclure les interactions transitoires ou aléatoires. Diverses interactions ont été trouvées, à différente position.

Le résidu correspondant à K94 (ou K214 dans la chaîne B) dans VvETR2, trouvé comme étant un site de fixation à l'éthylène par Azhar *et al* (2023) est capable d'interagir avec les éthylènes et les éthanols dans différentes conditions (figure21). De plus ce résidu spatialement proche de Y36 (Y156 dans la chaîne B) et de H72 (192 dans la chaîne B) rend possible les interactions avec plusieurs résidus identifiés comme primordiaux dans la fixation de l'éthylène et le fonctionnement de VvETR2.

D'autres interactions ont pu être identifiées, cependant aucune avec Y36. Il semble que cette partie du récepteur dans nos simulations ne soit pas accessible. De plus de longues interactions (< 300ns) entre les résidus S48 et K44 et deux éthylènes (l'un après l'autre) en condition éthylène couplé à l'éthanol ont été constatés. Cette observation suggère la possibilité d'un site alternatif ou transitoire de fixation favorisé par la présence conjointe des deux ligands. Toutefois, ce site étant localisé sur la face cytoplasmique du récepteur, il est possible qu'il ne reflète pas un site de fixation canonique, mais plutôt une zone d'accumulation transitoire favorisée par la dynamique membranaire, ou un état préliminaire à l'entrée/sortie du ligand.

Discussion

Une limite de cette étude réside dans la courte durée des phases d'équilibration (125 ps pour les étapes NVT et NPT). Bien que suffisante pour stabiliser globalement le système, une durée plus longue aurait pu permettre une meilleure adaptation de la membrane ainsi qu'une relaxation plus fine des sites de coordination des ions Cu²⁺. Dans de futures simulations, prolonger ces étapes pourrait aider à mieux dissocier les effets spécifiques des ligands des fluctuations internes du système.

Une autre limite serait que le modèle actuel ne prend pas en compte la partie du récepteur se situant dans le cytosol. L'absence de cette région limite notamment l'interprétation de certaines interactions observées à la sortie de la membrane, telles que celles impliquant les résidus K44 et S48, qui pourraient être influencées ou masquées par la présence du domaine cytosolique complet. Intégrer ce domaine dans de futures modélisations permettrait de simuler de façon plus réaliste les événements de signalisation post-liaison.

L'ensemble des analyses menées autour du récepteur à l'éthylène 2 de *Vitis vinifera* a permis de construire un modèle tridimensionnel robuste en cohérence avec les expérimentations biologiques que l'on retrouve dans la littérature.

La force de ce modèle est que sa construction repose sur des choix faits avec différents outils et la confrontation de ses outils entre eux.

Tout d'abord, l'exclusion du peptide signal, identifiée via SignalP 6.0 et DeepTMHMM, a été déterminante pour la pertinence des prédictions structurales. La forte probabilité de clivage avant le résidu 20, associée à la prédition de trois segments transmembranaires entre les résidus 47 et 132, a justifié la sélection d'un segment contenant toutes les structures fonctionnelles allant du résidus 21 à 140 pour la modélisation. Cette approche a permis d'optimiser à la fois la qualité de la prédition et la faisabilité des simulations tout-atome.

Le choix de Chai-1 comme outil de prédition principal s'est imposé à l'issue d'une évaluation comparative rigoureuse : ce modèle a montré une excellente cohérence avec les structures expérimentales. De plus, Chai1 a permis l'intégration des ions cuivre et possède une auto-évaluation de ses prédictions, ce qui a permis d'observer une amélioration significative des scores pTM et ipTM et renforcé la stabilité de l'interface dimérique.

La présence du cuivre, bien que n'entraînant pas de réarrangement de la structure, améliore la qualité prédictive. Les matrices PAE et les scores de confiance démontrent une meilleure définition des interfaces et une cohérence accrue des monomères l'un par rapport à l'autre en présence des ions cuivres. Cela suggère un rôle secondaire mais structurant du cuivre dans la stabilisation du dimère.

La modélisation manuelle des ponts disulfures via MoMA-FreSa et PyMOL a permis de reconstruire une conformation dimérique réaliste, avec un arrangement symétrique entre les cystéines Cys8 et Cys10 des deux chaînes. L'unique configuration valide obtenue renforce la pertinence du modèle retenu.

Les valeurs de RMSD montrent que le récepteur reste globalement stable au cours des 1000ns de simulation, avec des écarts modérés entre les conditions. Les hélices transmembranaires, en particulier, présentent des RMSD faibles, traduisant leur rigidité structurale. En revanche, le squelette protéique global est plus sensible à la présence ou à l'absence de ligand : l'absence de ligand s'accompagne de fluctuations plus marquées, alors que la co-présence d'éthylène et d'éthanol semble stabiliser la conformation globale du récepteur. Les profils de RMSF révèlent une flexibilité localisée dans les régions non structurées, notamment aux extrémités N-terminales ainsi que dans les segments inter-hélice. Ces fluctuations sont attendues, mais on note une augmentation modérée de la flexibilité en présence conjointe des deux ligands dans certaines zones spécifiques, comme à l'extrémité de la dernière hélice (chaîne B), ce qui pourrait indiquer une réorganisation locale ou une sensibilité accrue dans un contexte de double liaison.

Ensemble, ces données suggèrent que l'éthylène seul stabilise surtout les régions transmembranaires, tandis que l'éthanol seul semble perturber légèrement la conformation globale. En revanche, la condition avec l'éthylène et l'éthanol combiné semble préserver la stabilité du squelette protéique. Cela pourrait refléter une compensation structurale entre les effets distincts des deux ligands, ouvrant la voie à l'hypothèse d'une modulation fine de l'état conformationnel du récepteur en fonction de l'environnement membranaire.

La diffusion de l'éthylène à travers la membrane est un processus relativement indépendant de la présence d'éthanol, le nombre d'événements de diffusion et leur cinétique n'étant pas significativement modifiés par la présence d'éthanol. Cela peut s'expliquer par la nature

apolaire de l'éthylène, qui traverse librement le cœur hydrophobe de la membrane, sans nécessiter d'interaction spécifique ou facilitée. Cependant, le groupement -OH hydrophile de l'éthanol limite son passage à travers la bicouche lipidique. L'éthanol peut s'insérer de manière superficielle dans la membrane, possiblement à l'interface des têtes polaires, mais ne parvient pas à traverser la région centrale hydrophobe.

Ces observations sont en accord avec les études antérieures montrant que les petites molécules apolaires diffusent plus facilement que les molécules amphiphiles ou polaires à travers les membranes biologiques (Dubois *et al.*, 2018). Dans un environnement lipidique non perturbé, l'éthanol n'agit pas comme un modulateur direct du transport passif de l'éthylène.

De plus, l'absence d'insertion stable de l'éthanol dans la protéine, combinée à sa préférence pour les régions hors membrane, renforce l'idée qu'il n'interagit pas spécifiquement avec le site de fixation de l'éthylène, mais pourrait néanmoins influencer indirectement son comportement, par exemple en perturbant l'ordre lipidique local.

Les observations structurelles confirment l'existence d'interactions stables entre le récepteur VvETR2 et les molécules d'éthylène ou d'éthanol, en particulier avec le résidu K214 (équivalent de K94 dans la chaîne A), identifié comme un site de fixation de l'éthylène dans des travaux antérieurs (Azhar *et al.*, 2023) et qui se situe à proximité d'autres résidus clés identifiés dans la perception de l'éthylène.

Nos résultats montrent que l'éthanol peut également interagir avec K214, bien que ces interactions soient plus transitoires que celles de l'éthylène. En condition mixte, des contacts simultanés entre éthylène, K214 et C188 ont été observés, ce qui pourrait suggérer une stabilisation coopérative de la poche de fixation.

En condition éthylène + éthanol de longues interactions (> 300 ns) entre S48/K44 et des molécules d'éthylène ont été observées sur la face cytoplasmique du récepteur. Bien que seule la portion insérée dans la membrane ait été simulée, cette observation suggère l'existence possible d'un site transitoire à la sortie de la région transmembranaire, susceptible de jouer un rôle dans le recrutement, l'orientation ou la libération du ligand. Ce site ne correspond pas à une poche de fixation classique, mais sa récurrence en présence conjointe des deux ligands

pourrait indiquer une modulation conformationnelle locale, ou refléter l'effet indirect de l'éthanol sur la dynamique de diffusion de l'éthylène. Des simulations étendues à la région cytoplasmique complète seraient nécessaires pour déterminer si cette zone est fonctionnellement impliquée dans la signalisation ou le transport.

Ainsi, la combinaison des prédictions structurales, des analyses de conservation évolutive, des superpositions structurales et des simulations de dynamique moléculaire a permis de proposer un modèle robuste du domaine transmembranaire du récepteur VvETR2.

L'objectif de cette étude était d'explorer le rôle potentiel de l'éthanol dans les changements conformationnels de VvETR2, afin de mieux comprendre l'effet modulateur dose-dépendant observé expérimentalement. Grâce à la construction d'un modèle tridimensionnel réaliste, il a été possible de caractériser les effets de l'éthylène et de l'éthanol sur la conformation et la stabilité du récepteur.

Ces données permettent donc de répondre à la question initiale : l'éthanol ne mime pas l'éthylène et ne peut être considéré comme un agoniste de ce dernier. Cependant, il pourrait moduler l'effet de l'éthylène en influençant la dynamique structurale du récepteur, ce qui pourrait expliquer les observations expérimentales de modulation dose-dépendante.

Ce travail constitue une première étape dans l'exploration structurale et dynamique du récepteur ETR2 de *Vitis vinifera*. Les perspectives immédiates incluent la réalisation de simulations plus longues grâce à des modèles gros grains permettant d'explorer des échelles de temps plus longues et des phénomènes collectifs difficiles à capturer en tout-atome.

Il est également prévu de tester différentes compositions membranaires, de façon à ce que le modèle simplifié de la membrane du réticulum endoplasmique deviennent de plus en plus complexe et ressemblant à la véritable composition pour évaluer l'impact de l'environnement lipidique.

Enfin, pour approfondir les connaissances sur le comportement dose-dépendant de l'éthanol, des simulations supplémentaires seront conduites en faisant varier les concentrations d'éthanol, ce qui permettra de mieux cerner ses effets structuraux et dynamiques en fonction du contexte.

Références

- Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**: 19–25
- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A et al (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**: 493–500
- Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. I. General method. *J Chem Phys* **31**: 459–466. doi:10.1063/1.1730376
- Azhar BJ, Abbas S, Aman S, Yamburenko MV, Chen W, Müller L et al (2023) Basis for high-affinity ethylene binding by the ethylene receptor ETR1 of Arabidopsis. *Proc Natl Acad Sci USA* **120**: e2215195120
- Binder BM (2020) Ethylene signaling in plants. *J Biol Chem* **295**: 7710–7725
- Bleecker AB (1999) Ethylene perception and signaling: an evolutionary perspective. *Trends Plant Sci* **4**: 269–274
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* **4**: 187–217
- Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* **126**: 014101. doi:10.1063/1.2408420
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421
- Chai Discovery, Boitreaud J, Dent J, McPartlon M, Meier J, Reis V et al (2024) Chai-1: Decoding the molecular interactions of life. *bioRxiv*. doi:10.1101/2024.10.10.615955

Chang C (2016) Q&A: How do plants respond to ethylene and what is its importance? *BMC Biol* **14**: 7

Chang C, Kwok SF, Bleecker AB, Meyerowitz EM (1993) Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. *Science* **262**: 539–544.
doi:10.1126/science.8211181

Chen Y, Althiab Almasaud R, Carrie E, Desbrosses G, Binder BM, Chervin C (2020) Ethanol, at physiological concentrations, affects ethylene sensing in tomato germinating seeds and seedlings. *Plant Sci* **291**: 110368

Chervin C, Truett JK, Speirs J (1999) Alcohol dehydrogenase expression and alcohol production during pear ripening. *J Am Soc Hortic Sci* **124**: 71–75

Clerc Ilinka. A stochastic conformational sampling method for complex protein architectures involving disordered regions. Networking and Internet Architecture [cs.NI]. Université de Toulouse, 2024.

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190

DeLano WL (2002) The PyMOL Molecular Graphics System. *Delano Scientific*, San Carlos
Diot A, Groth G, Blanchet S, Chervin C. Responses of animals and plants to physiological doses of ethanol: a molecular messenger of hypoxia? *FEBS J*. 2024 Mar;291(6):1102-1110.
doi: 10.1111/febs.17056. Epub 2024 Jan 17. PMID: 38232057.

Dubois M, Van Den Broeck L, Inzé D (2018) The pivotal role of ethylene in plant growth. *Trends Plant Sci* **23**: 311–323

Frenkel D, Smit B (2002) *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, San Diego

Gane R (1934) Production of ethylene by some ripening fruits. *Nature* **134**: 1008

Giovannoni JJ (2007) Fruit ripening mutants yield insights into ripening control. *Curr Opin Plant Biol* **10**: 283–289

Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H et al (2022) DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* doi:10.1101/2022.04.08.487609

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D et al (2020) Array programming with NumPy. *Nature* **585**: 357–362

Hunter JD (2007) Matplotlib: A 2D graphics environment. *Comput Sci Eng* **9**: 90–95

Iqbal N, Khan NA, Ferrante A, Trivellini A, Francini A, Khan MIR (2017) Ethylene role in plant growth, development and senescence: interaction with other phytohormones. *Front Plant Sci* **8**: 475

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O et al (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589

Khan NA, Khan MIR, Ferrante A, Poor P (2017) Editorial: Ethylene: a key regulatory molecule in plants. *Front Plant Sci* **8**: 1782

Khan S, Alvi AF, Saify S, Iqbal N, Khan NA (2024) The ethylene biosynthetic enzymes, 1-aminocyclopropane-1-carboxylate (ACC) synthase (ACS) and ACC oxidase (ACO): the less explored players in abiotic stress tolerance. *Biomolecules* **14**: 90

Kimmerer TW, MacDonald RC (1987) Acetaldehyde and ethanol biosynthesis in leaves of plants. *Plant Physiol* **84**: 1204–1209

Loschmidt J (1861) *Chemische Studien I*. Wien: Carl Gerold's Sohn.
<https://books.google.fr/books?id=ksw5AAAAcAAJ>

Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**: 2319–2327

Nielsen H, Teufel F, Brunak S, von Heijne G (2024) SignalP: the evolution of a web server. *Methods Mol Biol* **2836**: 331–367. doi:10.1007/978-1-0716-4007-4_17

Patra M, Salonen E, Terama E, Vattulainen I, Faller R, Lee BW et al (2006) Under the influence of alcohol: the effect of ethanol and methanol on lipid bilayers. *Biophys J* **90**: 1121–1135

Qu X, Hall BP, Gao Z, Schaller GE (2007) A strong constitutive ethylene-response phenotype conferred on Arabidopsis plants containing null mutations in the ethylene receptors ETR1 and ERS1. *BMC Plant Biol* **7**: 3

Rodríguez-Navarro A. Potassium transport in fungi and plants. *Biochim Biophys Acta*. 2000 Mar 10;1469(1):1-30. doi: 10.1016/s0304-4157(99)00013-1. PMID: 10692635.

Roger M, Leone P, Blackburn NJ, Horrell S, Chicano TM, Biaso F et al (2024) Beyond the coupled distortion model: structural analysis of the single domain cupredoxin AcoP, a green mononuclear copper centre with original features. *Dalton Trans* **53**: 1794–1808

Schott-Verdugo S, Müller L, Classen E, Gohlke H, Groth G (2019) Structural model of the ETR1 ethylene receptor transmembrane sensor domain. *Sci Rep* **9**: 8869

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539

The UniProt Consortium (2023) UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res* **51**: D523–D531. doi:10.1093/nar/gkac1052

Wang W, Esch JJ, Shiu S-H, Agula H, Binder BM, Chang C et al (2007) Identification of important regions for ethylene binding and signaling in the transmembrane domain of the ETR1 ethylene receptor of Arabidopsis. *Plant Cell* **18**: 3429–3442

Wohlwend J, Corso G, Passaro S, Getz N, Reveiz M, Leidal K et al (2024) Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv* doi:10.1101/2024.11.19.624167

Wu R, Ding F, Wang R, Shen R, Zhang X, Luo S et al (2022) High-resolution de novo structure prediction from primary sequence. *bioRxiv* doi:10.1101/2022.07.21.500999

Xiao Z, Liao S, Rogiers SY, Sadras VO, Tyerman SD (2018) Effect of water stress and elevated temperature on hypoxia and cell death in the mesocarp of Shiraz berries. *Aust J Grape Wine Res* **24**: 487–497

Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**: 2302–2309. doi:10.1093/nar/gki524

Annexe

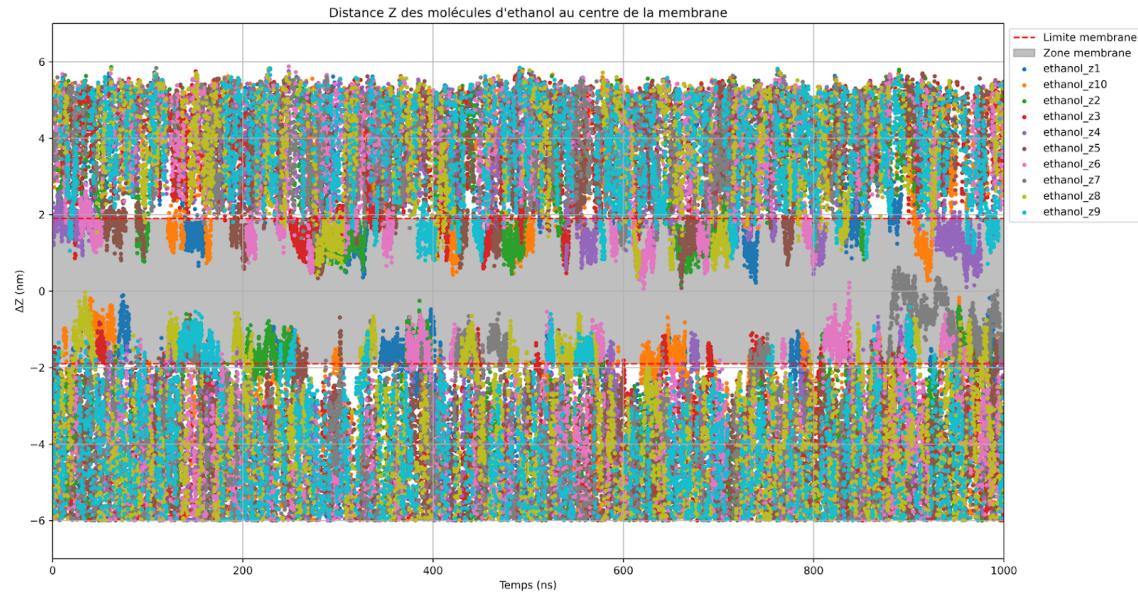


Figure 14. Distance des 10 molécules d'éthanol par rapport au centre de la membrane mesuré grâce à Delta Z en condition éthanol seul. Delta Z = position de l'éthanol sur l'axe Z - position du centre de la membrane sur l'axe Z. La zone grisée correspond à l'intérieur de la membrane. Les tirets rouges délimitent l'intérieur et l'extérieur de la membrane.

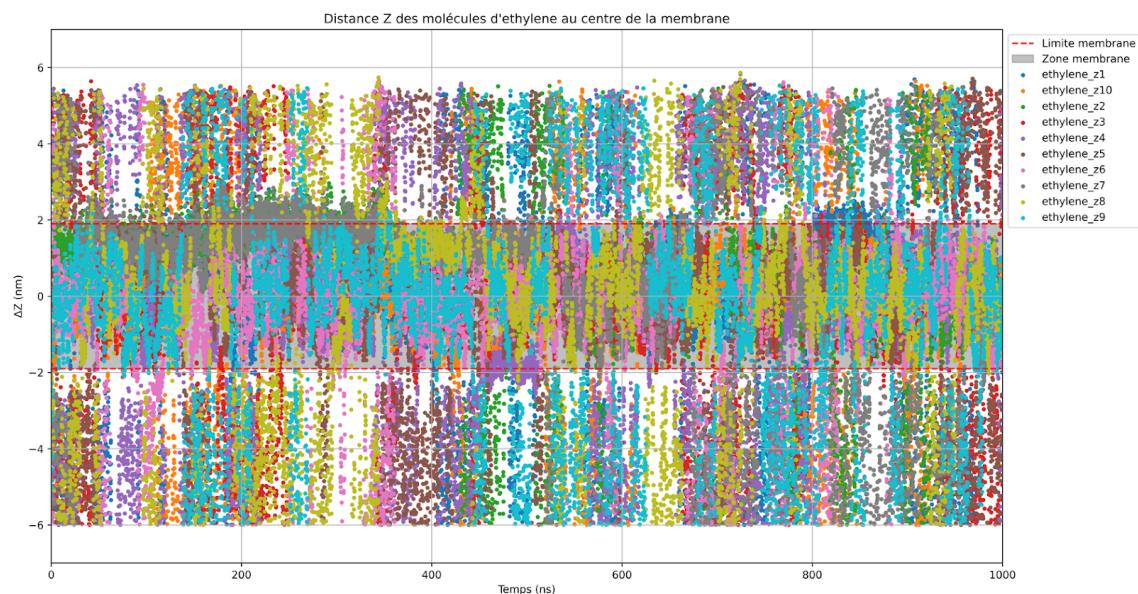


Figure 15. Distance des 10 molécules d'éthylène par rapport au centre de la membrane mesuré grâce à Delta Z en condition éthylène couplé à l'éthanol. Delta Z = position de l'éthylène sur l'axe Z - position du centre de la membrane sur l'axe Z. La zone grisée correspond à l'intérieur de la membrane. Les tirets rouges délimitent l'intérieur et l'extérieur de la membrane.

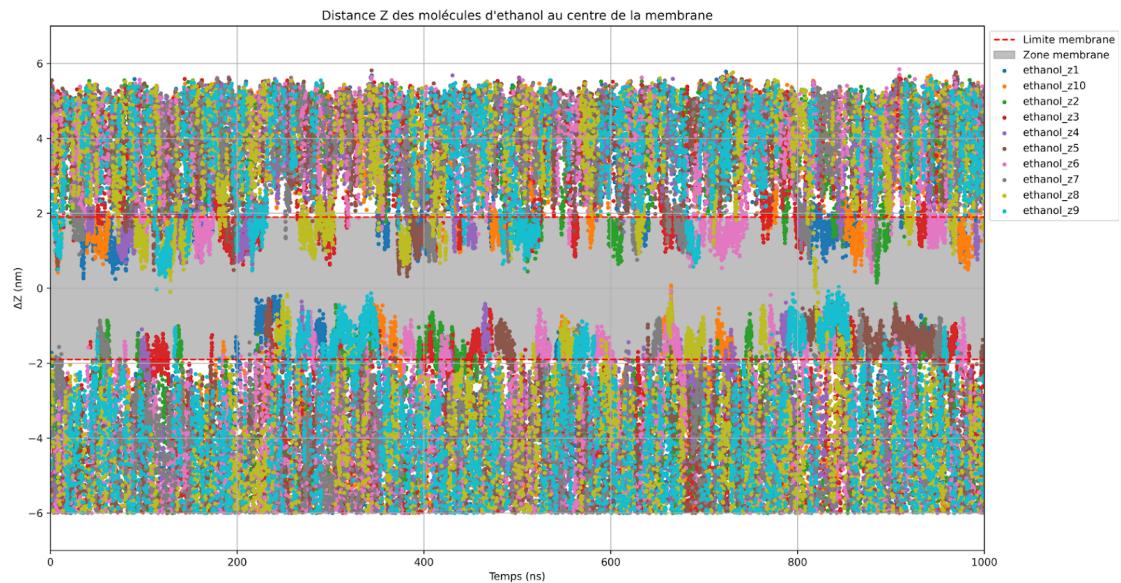


Figure 16. Distance des 10 molécules d'éthanol par rapport au centre de la membrane mesuré grâce à Delta Z en condition éthylène couplé à l'éthanol. Delta Z = position de l'éthanol sur l'axe Z - position du centre de la membrane sur l'axe Z. La zone grisée correspond à l'intérieur de la membrane. Les tirets rouges délimitent l'intérieur et l'extérieur de la membrane.

Lien git contenant les scripts personnalisés utilisés pour cette étude :

https://github.com/Strausss23/StageM2_scripts_Maeva

Tableau avec les séquences ETR2 *vitis vinifera* :

Séquence entière ETR2 <i>vitis vinifera</i> (UniProt)	MLKRLAPGLLISSLLISASAVDNSFPRCNCEDEGFWSVENILE CQKVSDFLIAVAYFSIPIELLYFVSCSNVPFKWVLFQFIAFIVLC GLTHLLNGWTYGPHPFQLMLALTIFKFLTALVSCATAITLITLIP LLLKVKVREFMLKKKTWDLGREVGIIKKKEAGLHVRMLTH EIRKSLSRHTILHTTLVELSNTLDLQNCAVWMPNENKTEMNL THELKGRNFYNFSIPINDPVVAMIKRSDEVHTLRTDSALATAS SGTSGEPGPVAAIRMPMLRVSNFKGGTPELVQACYSILVLVLK SGQARSWTSQELDIVKVVADQVAVAVSHAAVLEESQLMRDQL AEQNRALQQAKRNAMMASQARNSFQKVMSDGMRRPMHSIS GLLSMMQDEKLNEQRLLIIDAMAKTSNVLSTLINDVMEISTK DTGRFQLDVRSLHSMIKEAACLAKECLCVYRGFGFAIEVEK SLPDHVIGEERRVFQVILHMVGNLLNGTNGGSVTFRVLSET GSQGRHDQRWATWKSNSSDGYVYIKFEIGINNADQSEGSIST GSISTVQLGGGRQTSDQIDEGLSFTMCRRLAQLMQGNIWLVP NPQGFAKSMALVLRQLQPSIGINISEPGESEHHSNSLFRGL QVLLADDDDTNRAVTRKLLEKLGCVSVVSSGFECLGALGPA ASSFQIVLLDLHMPPELDGFEVAMRIRKFRSRSWPLIALTASA DEDVWERCLEIGMNGIIRKPVLLDGIAEELRRVLLQANNVV
Séquence ETR2 <i>vitis vinifera</i> sans peptide signal. Utilisée pour le BlastP de la protéine entière.	VDNSFPRCNCEDEGFWSVENILECQKVSDFLIAVAYFSIPIELL YFVSCSNVPFKWVLFQFIAFIVLCGLTHLLNGWTYGPHPFQL MLALTIFKFLTALVSCATAITLITLIPLLLKVKVREFMLKKKTW DLGREVGIIKKKEAGLHVRMLTHEIRKSLSRHTILHTTLVEL SNTLDLQNCAVWMPNENKTEMNLTHELKGRNFYNFSIPINDP VVAMIKRSDEVHTLRTDSALATASSGTSGEPGPVAAIRMPML RVSNFKGGTPELVQACYSILVLVLKSGQARSWTSQELDIVKV VADQVAVAVSHAAVLEESQLMRDQLAEQNRALQQAKRNAM MASQARNSFQKVMSDGMRRPMHSISGLLSMMQDEKLNESEQ RLIIDAMAKTSNVLSTLINDVMEISTKDTGRFQLDVRSLHS

	MIKEAACLA KCLCVYRGFC AIEVEKSLPDH VIGEERRVFQVI LHMVG NLLNGTNGGSV TRVLSETGSQGR HDQRWATWKS NSSDGYVYIKF EIGINNADQQT SDQIDEGLSFT MCRRLAQLM QGNIWLVPNPQ GFAKSMALVRF QLQPSIGNISEP GESSEHPH SNSLF RGLQVLLADD DDTNRAVTRKL LEKLGCI VVSSGFE CLGALGP AASSFQIVLL DLHMPELDG FEVAMRIRK FRSRSWPL IVALTASA DEDVWERC LEIGMNGI IRKPVLLD GIAEELRV QANNVV
Séquence ETR2 <i>vitis vinifera</i> transmembranaire. Utilisée pour le BlastP avec partie transmembranaire uniquement. Utilisée pour toutes les prédictions de modèle ainsi que pour la génération du modèle final avec Chai1.	VDNSFPRCNCEDEGFWS VENILECQKVSDFLIA VAYFSIPIELL YFVSCSNVPFKWVLFQ FIAFIVLCGLTHLLNG WTYGPHPFQL MLALTIFKFLTA LVSCATAITLITLI PLLLKVKV