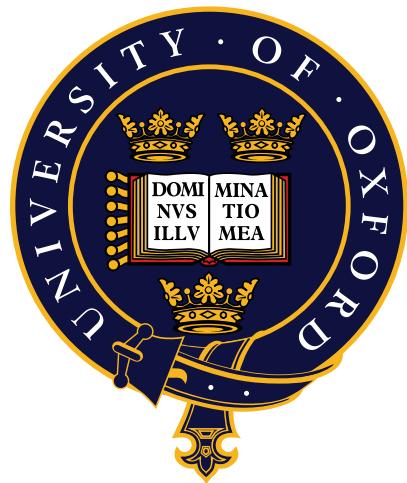


Structural and Functional Characterisation of NodD from *Rhizobium leguminosarum*

A Dissertation submitted for the Honour School of Biology

Part II 2025



Kapilan Sivanesan

Wadham College

Department of Biology

Oxford, 16 May 2025

Contents

1. Abstract.....	1
2. Background	3
2.1 An introduction to nitrogen in agriculture.....	3
2.1.1 <i>The global nitrogen cycle</i>	3
2.1.2 <i>Consequences of synthetic nitrogen fertilisers</i>	3
2.2 Biological nitrogen fixation and the rhizobium-legume symbiosis	4
2.2.1 <i>Historical and ecological context</i>	4
2.2.2 <i>Mechanisms of rhizobium-legume symbiosis</i>	4
2.3 NodD: A LysR family transcriptional regulator	5
2.3.1 <i>Role in regulation and host specificity</i>	5
2.3.2 <i>Oligomerisation and DNA binding</i>	6
2.3.3 <i>Insights from mutagenesis and truncation studies</i>	7
2.3.4 <i>Challenges in studying NodD</i>	7
2.4 Objectives.....	8
3. Materials and Methods.....	11
3.1 Experimental methods.....	11
3.1.1 <i>Plasmids</i>	11
3.1.2 <i>Bacterial culture</i>	11
3.1.3 <i>Protein expression and cell lysis</i>	12
3.1.4 <i>Protein purification</i>	12
3.1.5 <i>Native mass spectrometry (MS)</i>	13
3.2 Computational methods.....	14
3.2.1 <i>Structure prediction (Chai-1)</i>	14

3.2.2 Binding site visualisation (<i>Chimera</i> , <i>LigPlot+</i>)	14
3.2.3 Site-directed molecular docking (<i>AutoDock Vina</i>)	14
4. Results	17
4.1 Experimental results.....	17
4.1.1 Expression and purification of <i>NodD</i> constructs.....	17
4.1.2 Native mass spectrometry of <i>NodD</i> -CTD	19
4.1.3 Native mass spectrometry of <i>NodD</i> -CTD with flavonoids.....	19
4.2 Computational results.....	21
4.2.1 Predicted structures from <i>Chai-1</i>	21
4.2.2 Conserved residues concentrated in the DNA-binding domain and ligand-binding pocket.....	22
4.2.3 Ligand interaction analysis highlights conserved binding residues.....	24
4.2.4 Site-directed docking predicts strong binding of all four flavonoids to <i>NodD</i> -CTD	24
5. Discussion.....	27
5.1 Key findings	27
5.1.1 <i>NodD</i> C-terminal domain is dimeric in solution	27
5.1.2 Flavonoids induce tetramerisation of the <i>NodD</i> C-terminal domain independent of DNA-binding domain.....	27
5.1.3 Luteolin, hesperetin, naringenin, and apigenin are effective binding partners of <i>NodD</i> -CTD.....	28
5.1.4 Computational modelling guides prediction of key ligand-binding and interaction residues	29
5.2 Limitations of computational analysis	29
5.3 Implications for sustainable agriculture	30
5.4 Future work	31
6. Conclusion	33
7. Acknowledgements.....	35

8. References	37
9. Appendix	43
9.1 AutoDock Vina workflow	43
9.1.1 <i>Applications used</i>	43
9.1.2 <i>Molecular docking method</i>	43
9.1.3 <i>Visualising docking results</i>	49
9.2 Chai-1 Analysis	51
9.3 LigPlot+ Analysis.....	60

1. Abstract

Legume-rhizobia symbiosis is a key driver of biological nitrogen fixation, yet the molecular mechanisms governing this interaction remain poorly characterised. In *Rhizobium leguminosarum*, the LysR-type transcriptional regulator NodD activates nod genes in response to root-secreted flavonoids, triggering the synthesis of Nod factors that initiate nodule formation. Within these nodules, rhizobia convert atmospheric nitrogen into a bioavailable form for the host plant. As a key regulator of this signalling pathway, NodD plays a central role in host recognition and specificity. However, the structural basis of NodD function and its specificity for different flavonoids remain unclear. Here we demonstrate, using native mass spectrometry of NodD from *R. leguminosarum* bv. *viciae* 3841, that the isolated C-terminal domain undergoes flavonoid-induced oligomerisation independently of the DNA-binding domain. Docking simulations and sequence analyses further identified a conserved ligand-binding pocket within the C-terminal domain, highlighting residues potentially critical for flavonoid recognition and uncovering molecular determinants of specificity across rhizobial species. Combined, these findings provide a structural and mechanistic framework for targeted mutagenesis aimed at engineering NodD specificity for agricultural improvement.

2. Background

2.1 An introduction to nitrogen in agriculture

2.1.1 The global nitrogen cycle

Nitrogen is a fundamental component of amino acids, nucleic acids, and chlorophyll, making it essential for plant growth and crop productivity (Maathuis, 2009). Although nitrogen gas comprises around 78% of Earth's atmosphere, it is chemically inert and unavailable to most organisms without conversion into reactive forms such as ammonia or nitrate (Kabange et al., 2022). In natural ecosystems, this conversion, known as nitrogen fixation, is primarily carried out by certain bacteria and archaea, helping sustain soil fertility over time. However, as global food demand increased, the natural nitrogen cycle proved insufficient to meet agricultural needs, prompting the development of industrial methods to supplement soil nitrogen (Erisman et al., 2008).

2.1.2 Consequences of synthetic nitrogen fertilisers

The advent of the Haber-Bosch process in the early 20th century enabled large-scale synthesis of ammonia from atmospheric nitrogen, revolutionising global agriculture by vastly increasing crop yields (Galloway et al., 2013). Today, synthetic nitrogen fertilisers remain essential to modern food production, but their widespread use comes at a significant environmental cost. Fertiliser runoff contributes to eutrophication and dead zones in aquatic ecosystems, while soil acidification and biodiversity loss undermine long-term agricultural sustainability (Menegat et al., 2022). Moreover, nitrogen-based fertilisers are a major source of nitrogen oxide emissions, a greenhouse gas with a global warming potential far exceeding that of carbon dioxide (Aryal et al., 2022; Castellano-Hinojosa et al., 2020). These challenges have intensified efforts to develop more sustainable approaches to nitrogen management, including the improvement of biological nitrogen fixation.

2.2 Biological nitrogen fixation and the rhizobium-legume symbiosis

2.2.1 Historical and ecological context

Long before the emergence of synthetic fertilisers, farmers across diverse cultures relied on legume crops to enrich soil fertility (Crews & Peoples, 2004). Through practices such as crop rotation and intercropping, legumes were used to naturally replenish soil nitrogen, improving yields without external inputs (Mesfin et al., 2023; Neamatollahi et al., 2013). This benefit comes from the legume's unique ability to form a symbiotic relationship with nitrogen-fixing rhizobia: soil bacteria that convert atmospheric nitrogen into ammonia within specialised root nodules. These partnerships not only sustain legume growth in nitrogen-poor soils but also leave residual nitrogen available for subsequent crops. In ecological terms, biological nitrogen fixation plays a key role in supporting primary productivity and maintaining nutrient balance (Ladha et al., 2022). Understanding the molecular basis of this symbiosis could help improve its efficiency and broaden its applicability, contributing to reduced global reliance on synthetic fertilisers.

2.2.2 Mechanisms of rhizobium-legume symbiosis

The establishment of rhizobium-legume symbiosis begins with a molecular dialogue between the host plant and its bacterial partner (**Figure 1**). In response to nitrogen-limiting conditions, legume roots secrete flavonoids: secondary metabolites that act as signalling molecules to attract compatible rhizobia (Oldroyd, 2013). These compounds are detected by the rhizobia, triggering the expression of nodulation (*nod*) genes responsible for synthesising Nod factors: lipochitooligosaccharides that initiate symbiotic development. Nod factors are recognised by specific receptors on the plant root, leading to root hair curling, infection thread formation, and ultimately the development of nitrogen-fixing nodules. Within these nodules, rhizobia differentiate into bacteroids and use the nitrogenase enzyme complex to convert atmospheric nitrogen into ammonia, which is taken up by the plant (Poole et al., 2018). This process ensures host specificity and coordination between both partners, allowing efficient and sustained exchange of nitrogen.

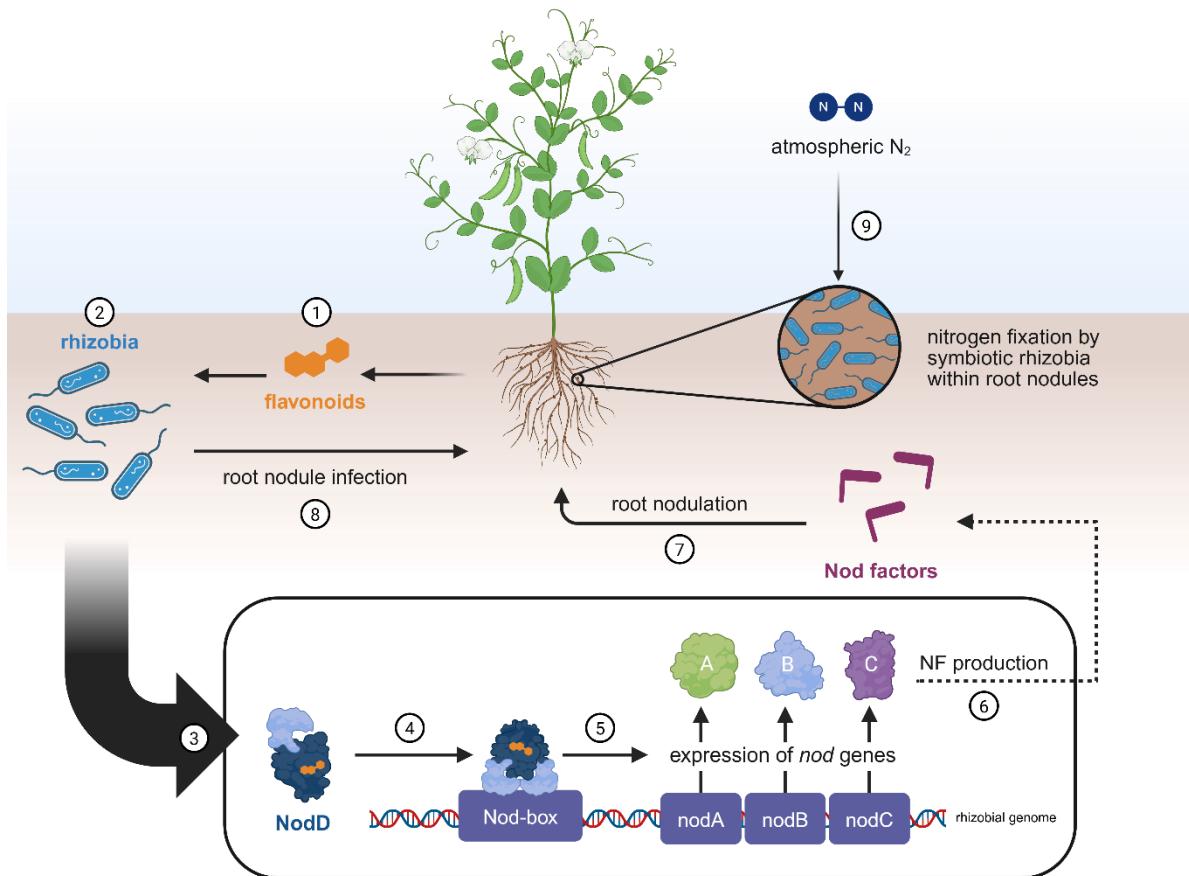


Figure 1. Schematic overview of rhizobium-legume symbiosis. (1) Legume roots secrete flavonoids, (2) which are detected by compatible rhizobia. (3) Flavonoid binding activates NodD which (4) binds the nod-box as an oligomer to activate *nod* gene transcription. (5) Expression of *nodABC* leads to (6) Nod factor (NF) production. (7) NFs trigger nodule formation in the host root, (8) allowing rhizobia to infect through infection threads and differentiate into nitrogen-fixing bacteroids. (9) Within the root nodules, the symbiotic bacteria fix atmospheric nitrogen into ammonia, which can be taken up by the plant. Figure created with BioRender.com.

2.3 NodD: A LysR family transcriptional regulator

2.3.1 Role in regulation and host specificity

Central to the transcriptional activation of *nod* genes is NodD, a LysR-type transcriptional regulator (LTTR) that directly mediates the bacterial response to flavonoid signals. Like most LTTRs, NodD consists of two primary domains: an N-terminal DNA-binding domain and a C-terminal flavonoid-binding domain. These two domains are joined by a linker helix (Haskett et al., 2025). Upon flavonoid binding, NodD undergoes a conformational change that facilitates its interaction with the nod-box, a conserved promoter sequence upstream of the *nod* gene

cluster (**Figure 1**, Feng et al., 2003). This binding event initiates the transcription of *nod* genes, ultimately leading to Nod factor biosynthesis and the establishment of symbiosis.

Early work identified NodD as a major determinant of host specificity (Bender, 1988). Sequence and structural differences between NodD proteins affect their ability to recognise particular flavonoids, ultimately influencing which legume species a given rhizobial strain can nodulate. Luteolin was the first flavonoid shown to act as a *nod* gene inducer, with the ability to activate *nod* genes in a range of rhizobial species (Liu & Murray, 2016). Peck et al. (2006) demonstrated that flavonoids luteolin, naringenin, eriodictyol, and daidzein could bind to NodD1 from *Sinorhizobium meliloti* (also known as *Ensifer meliloti*) and induce conformational changes at the *nod* gene promoter. However, only luteolin was able to stimulate transcription. Additionally, NodD variants from *Rhizobium leguminosarum* bv. *viciae* and *trifolii* responded to both luteolin and naringenin. In these strains, eriodictyol and 7-hydroxyflavone were identified as the strongest inducers, respectively. A separate study assessing the activation of *R. leguminosarum* bv. *viciae* NodD found that out of several flavonoids tested (apigenin, daidzein, genistein, hesperetin, kaempferol, luteolin, naringenin, and rutin), hesperetin and naringenin were the most effective (Begum et al., 2001).

2.3.2 Oligomerisation and DNA binding

Evidence suggests that *Rhizobium leguminosarum* NodD binds to DNA either as a V-shaped homotetramer (Feng et al., 2003; Chen et al., 2005) or a homodimer (Fisher & Long, 1993). NodD binds to two sites within the nod-box, both positioned on the same face of the DNA helix. DNase footprinting experiments have shown that NodD protects a region of approximately 55 to 60 base pairs (unusually large for a ~35 kDa protein) suggesting that it may form higher-order oligomers to span this extended sequence (Fisher & Long, 1993). This idea is supported by studies on other LTTRs such as CatR and TrpI, where inducer binding promotes increased oligomerisation. This allows the proteins to interact with larger regions of promoter DNA and may enhance transcriptional activation (Parsek et al., 1992; Chang & Crawfordt, 1990).

2.3.3 Insights from mutagenesis and truncation studies

Mutagenesis studies have identified several residues within NodD that are critical for its regulatory activity. In *Sinorhizobium meliloti* NodD1, substitutions such as L69F, S104L, D134N, and M193I resulted in reduced or abolished *nod* gene expression, while K205N produced a constitutively active variant that was repressed by luteolin (Peck et al., 2013). In *Rhizobium leguminosarum* bv. *viciae* (*Rlv*) 3841 NodD, substitutions A93V, M161I, H195Y, G228R, A274V, P278L, and D284N were shown to activate *nod* genes independently of flavonoids (Haskett et al., 2025). These findings suggest that specific residues within the ligand-binding pocket are important for sensing and transducing flavonoid signals, and may play a role in stabilising the oligomeric form of the protein.

Further insight comes from deletion experiments: removal of the regulatory domain of *Rlv* 3841 NodD yielded a truncated protein capable of driving *nod* gene expression in the absence of flavonoids, though with reduced activation strength compared to the full-length protein (Haskett et al., 2025). The authors propose that this constitutive activity could result either from the loss of an inhibitory mechanism or from changes in oligomerisation dynamics. The latter possibility could be further explored by examining the oligomerisation dynamics of the isolated regulatory domain in response to flavonoids.

2.3.4 Challenges in studying NodD

Despite its functional importance, the structural and mechanistic understanding of NodD remains limited. A major barrier is the difficulty in obtaining soluble, stable protein for biophysical analysis. NodD typically exhibits low expression yields, and maintaining its activity during purification has proven challenging (Feng et al., 2002). These issues are not unique to NodD; full-length LTTRs in general often suffer from poor solubility and a tendency to precipitate at high concentrations (Ezezika et al., 2007).

As a result, there is a lack of high-resolution structural data for NodD. While full-length structures have been resolved for some LTTRs such as CbnR and OxyR (Muraoka et al., 2003; Jo et al., 2015), most crystallographic studies have focused on the C-terminal regulatory domain (Kim et al., 2018; Tyrrell et al., 1997). A key limitation is the difficulty of crystallising

the N-terminal helix-turn-helix (HTH) DNA-binding domain, largely due to the flexibility of the “wing” region (Maddock & Oyston, 2008; Giannopoulou et al., 2021).

The molecular binding dynamics of LTTRs also remain unclear. Only in 2021 was the first full-length LTTR solved in complex with its promoter DNA (Giannopoulou et al., 2021), and obtaining such high-resolution structures continues to be challenging. Until more experimental structures become available, molecular modelling tools such as Chai-1 can provide useful predictions of NodD’s structure in complex with ligands. However, these models require validation through structural techniques such as X-ray crystallography, cryo-EM, or NMR to reliably inform our understanding of flavonoid recognition and DNA binding.

2.4 Objectives

While its functional significance in rhizobium-legume symbiosis is well established, the molecular basis for NodD ligand-binding activities remains poorly characterised. Flavonoids secreted by host plants act as inducers, binding to NodD and initiating a conformational change that promotes transcription of *nod* genes. However, the specificity and structural determinants of this interaction are not fully understood, and no crystal structure currently exists for NodD. Full-length LTTRs are also often challenging to express and purify, further limiting structural and mechanistic insight into this protein family.

The primary objective of this study is to express and purify full-length NodD and its C-terminal domain, the region involved in flavonoid binding. Recombinant expression in *E. coli*, followed by immobilised metal affinity chromatography and size-exclusion chromatography, will be used to obtain sufficient quantities of soluble protein for further analysis. We hypothesise that the C-terminal domain alone is more stable and more amenable to purification than the full-length protein, due to the absence of the flexible N-terminal DNA-binding domain, which may contribute to misfolding or aggregation during expression.

A second objective is to investigate the ligand-binding properties of NodD using native mass spectrometry. This technique will be used to assess the purity of the sample, determine oligomeric state, and detect ligand-binding through shifts in mass spectra. A number of

flavonoids identified as NodD ligands in existing literature (luteolin, naringenin, hesperetin, and apigenin) will be tested to evaluate their relative binding affinities and their effect on protein oligomerisation. We hypothesise that flavonoid binding promotes higher-order oligomerisation of NodD, facilitated by a ligand-induced conformational change. A key aim is to determine whether the C-terminal domain alone is sufficient to support oligomer formation in the absence of the DNA-binding domain.

A third objective is to explore the structural basis of flavonoid recognition in NodD using computational methods. Structure prediction with Chai-1, followed by visualisation and hydrogen bonding analysis in Chimera and LigPlot+, will be used to study potential flavonoid-interacting residues within the C-terminal domain. Conservation analysis across NodD homologues will highlight evolutionarily constrained regions and suggest residues that may be functionally important. Molecular docking simulations using AutoDock Vina will then be used to model flavonoid binding and calculate predicted binding affinities. We expect that flavonoid ligands will bind within a conserved cavity in the CTD, and that docking scores will correlate with known inducer activity. This would suggest that key ligand-recognition features are shared across active NodD inducers, and that computational docking can reflect biological relevance. Overall, we aim for this computational data to provide a structural basis for future mutagenesis studies to explore how the structure of the NodD binding pocket relates to its regulatory function.

3. Materials and Methods

3.1 Experimental methods

3.1.1 Plasmids

The full-length (FL) NodD from *Rhizobium leguminosarum* bv. *viciae* (*Rlv*) 3841 was expressed using a pre-existing plasmid (pET28a-NodD-FL). To investigate the truncated C-terminal domain (CTD), an additional plasmid was constructed (pET15b-NodD-CTD). The codon-optimised CTD sequence (residues 90-310, **Figure 2**) was cloned into a pET-15b vector via In-Fusion cloning, using gene fragments and vector digested with NdeI and NheI restriction enzymes.

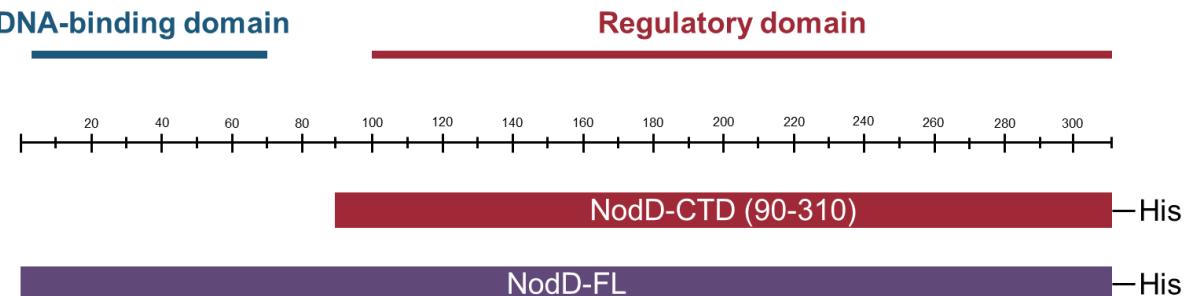


Figure 2. Representation of the NodD constructs (below), with the functional domains labelled (above).

3.1.2 Bacterial culture

StellarTM competent cells (Takara Bio) were used for plasmid propagation, and ArcticExpress (DE3) competent cells (Agilent Technologies) were used for protein expression. Cells were transformed with either pET-28a-NodD-FL or pET-15b-NodD-CTD plasmids via heat shock. Following recovery in LB medium, cells were plated on LB agar supplemented with the appropriate antibiotics for plasmid selection (see table below) and incubated overnight at 37 °C. Individual colonies were used to inoculate LB starter cultures containing the same antibiotics, and cultures were incubated at 37 °C with shaking overnight.

	pET28a-NodD-FL	pET15b-NodD-CTD
Insert	Full-length NodD	C-terminal domain (aa 90-310)
Vector backbone	pET-28a(+)	pET-15b
Cloning host	<i>E. coli</i> Stellar	<i>E. coli</i> Stellar
Expression host	<i>E. coli</i> Arctic Express	<i>E. coli</i> Arctic Express
Tag	C-terminal 6×His	C-terminal 6×His
Antibiotics	Gentamycin + Kanamycin	Gentamycin + Ampicillin

3.1.3 Protein expression and cell lysis

ArcticExpress starter cultures containing either pET-28a-NodD-FL or pET-15b-NodD-CTD were used to inoculate 6 L and 4 L of LB medium. Cultures were incubated at 30 °C with shaking for 3 hours, then induced with 1 mM IPTG at an optical density (OD_{600}) of ~0.6. Following induction, cultures were transferred to 10 °C and incubated overnight to promote protein expression.

Cells were harvested by centrifugation at 5,000 x g for 10 minutes at 4 °C, washed with lysis buffer (10% glycerol, 100 mM Tris-HCl pH 8.0, 100 mM NaCl), and stored at -80 °C. Before lysis, cells were thawed on ice and resuspended in lysis buffer supplemented with lysozyme (30 mg) and one tablet of EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by sonication on ice (40% amplitude, 10 minutes total time; 3 s on, 6 s off cycles). Lysates were clarified by centrifugation at 20,000 x g for 20 minutes at 4 °C. The supernatant containing the soluble protein fraction was filtered through a 0.45 µm membrane prior to purification.

3.1.4 Protein purification

Immobilised metal affinity chromatography was used for protein purification. Lysates were loaded onto a gravity-flow Ni-NTA affinity column (5mL resin; Qiagen) equilibrated with 25 mL of Buffer A (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole, 10% glycerol) at 4 °C. After sample loading, the column was washed with Buffer A and Buffer B (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 80 mM imidazole, 10% glycerol). Bound protein was eluted

with Buffer C (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 350 mM imidazole, 10% glycerol). Aliquots from each purification step were collected for SDS-PAGE analysis.

Eluted protein was dialysed overnight at 4 °C in 2.5 L of dialysis buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl) using a Slide-A-Lyzer™ dialysis cassette (10 kDa MWCO; Thermo Fisher).

The dialysed protein was collected and stored at -80 °C. Protein was thawed and concentrated using 30 kDa MWCO centrifugal concentrators at 4000 rpm, 4 °C. Samples of both concentrated protein and flow-through were collected for SDS-PAGE analysis.

For additional purification, NodD-FL was subjected to size-exclusion chromatography (SEC). Eluted and concentrated protein was loaded onto a Superdex 200 Increase 10/300 GL column equilibrated with SEC buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10% glycerol). Chromatography was performed at 4 °C using an ÄKTA system, and elution was monitored by absorbance at 280 nm.

3.1.5 Native mass spectrometry (MS)

Native MS was performed to analyse the oligomeric state and ligand-binding properties of purified NodD-CTD. Samples were buffer-exchanged into ammonium acetate (pH 7.4) using Micro Bio-Spin 6 columns (Bio-Rad). Samples were introduced directly into the mass spectrometer using gold-coated capillary needles (prepared in-house). Data were collected on a Q-Exactive UHMR mass spectrometer (Thermo Fisher Scientific). The instrument parameters were as follows: capillary voltage 1.1 kV, S-lens RF 100%, quadrupole selection from 1,000 to 20,000 m/z range, collisional activation in the HCD cell 100-200 V, trapping gas pressure setting 7.5, temperature 200 °C, and resolution of the instrument 12,500. The noise level was set at 3 rather than the default value of 4.64. No in-source dissociation was applied. All experiments were repeated three times with similar outcomes. For comparison, a previously purified laboratory stock of full-length NodD protein was also analysed alongside NodD-CTD. NodD-CTD was additionally analysed with 100 µM of each flavonoid (luteolin, naringenin, apigenin, hesperetin). Spectral data were analysed using Xcalibur and UniDec.

3.2 Computational methods

3.2.1 Structure prediction (Chai-1)

Predicted structures of NodD from five *Rhizobium* species (*Rlv* 3841, *Rhizobium leguminosarum* bv. *trifolii*, *Rhizobium tropici*, *Sinorhizobium meliloti*, and *Rhizobium galegae*) were generated using Chai-1 (Chai Discovery et al., 2024), with multiple sequence alignments (MSAs) constructed using JACKHMMER. Sequences were obtained from UniProt (The UniProt Consortium, 2025). Chai-1 was chosen over tools like AlphaFold and RoseTTAFold for its support of custom ligand binding. Each model was docked with one of four flavonoids (luteolin, hesperetin, naringenin, apigenin), resulting in 20 predicted complexes. ConSurf was used to map evolutionary conservation onto each structure to identify residues that may be functionally or structurally important (Ashkenazy et al., 2010, 2016; Celniker et al., 2013; Yariv et al., 2023).

3.2.2 Binding site visualisation (Chimera, LigPlot+)

Potential binding residues within each predicted protein-ligand complex were initially screened and visualised using UCSF Chimera (Pettersen et al., 2004). Protein-ligand interactions were assessed using the *Find Clashes/Contacts* and *Find H-Bonds* tools. Contacts were identified with a van der Waals overlap threshold of $\geq -0.4 \text{ \AA}$, and potential hydrogen bonds were detected using relaxed constraints ($+0.4 \text{ \AA}$). These interactions were visualised in 3D to compare binding pockets across complexes. LigPlot+ was used in parallel to generate 2D ligand interaction maps (Laskowski & Swindells, 2011). Both these methods provided a general overview of potential binding residues and allowed binding pocket features to be compared across species. This data also helped define the search space for molecular docking.

3.2.3 Site-directed molecular docking (AutoDock Vina)

Initial docking was performed using CBDock2, which integrates AutoDock Vina and automatically predicts potential binding pockets through blind docking. However, to allow for site-specific control and ensure consistency across runs, further docking was carried out

directly in AutoDock Vina (full protocol in 9.1 AutoDock Vina). Ligands were prepared in AutoDockTools by adding hydrogens, assigning Gasteiger charges, setting rotatable bonds, and defining the torsion tree root to allow the molecule to flex, before saving the files in PDBQT format. Protein structures were similarly prepared by removing water molecules, adding polar hydrogens, and converting to PDBQT format. Grid boxes were centred on residues identified from Chimera and LigPlot+ analyses and were defined to be as tight as possible around the predicted binding site, restricting docking to the identified cavity.

Docking was performed in AutoDock Vina with an exhaustiveness setting of 32 to ensure adequate sampling of conformational space. To assess reproducibility, ten independent docking runs were carried out for each protein-ligand complex using different random seeds. AutoDock Vina uses a stochastic search algorithm, so results can vary slightly between runs due to random starting conditions. As a result, some variability in docking poses and predicted binding affinities was expected. Consistency between independent runs was evaluated to confirm convergence on the predicted binding site. A batch script was written to automate the docking runs and organise the output files (see **9.1.2 Molecular docking method**). In addition to the four primary flavonoids, molecules lacking experimental evidence of NodD binding, such as kaempferol, galangin, nobiletin, and rutin, were also docked as negative controls. Binding affinity scores were automatically extracted from output files into a dataset using code (see **9.1.3 Visualising docking results**). Final docked poses were visualised in PyMol to assess convergence around the predicted binding pocket.

4. Results

4.1 Experimental results

4.1.1 Expression and purification of NodD constructs

Recombinant NodD-FL and NodD-CTD were expressed in *E. coli* with C-terminal His-tags to enable purification via immobilised metal affinity chromatography. In the initial expression attempt, NodD-FL was purified successfully; however, following concentration, the protein visibly precipitated, with white aggregates forming inside the concentrator. This precipitation may have been due to repeated freeze-thaw cycles and instability associated with the full-length protein. In a second expression attempt of NodD-FL, we used the same buffers but minimised freeze-thaw cycles, which improved initial solubility, with no visible precipitation during concentration. NodD-CTD was also expressed in the event that a truncated construct would be more stable.

SDS-PAGE analysis of lysates and buffer wash fractions confirmed successful expression and purification of both proteins at their expected molecular weights (~46 kDa for NodD-FL, ~25 kDa for NodD-CTD, **Figure 3**). Additional bands were observed at ~92 kDa and ~50 kDa,

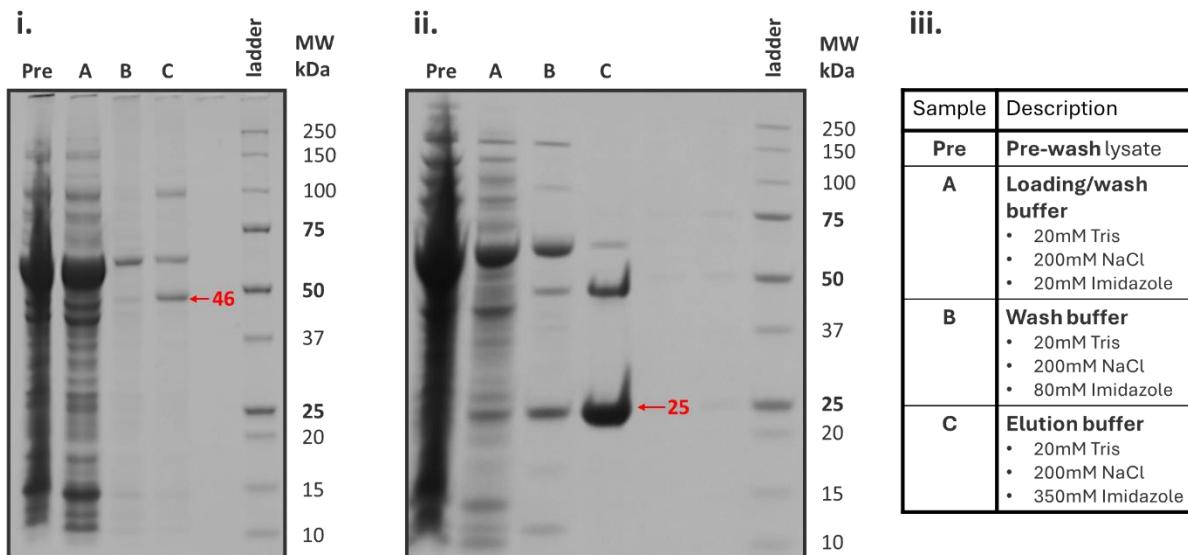


Figure 3. (i) Full-length NodD (NodD-FL) and (ii) C-terminal domain (NodD-CTD) samples were analysed at different stages of purification. The desired monomeric band is labelled in red. (iii) Table summarising the buffer used at each stage of the purification.

likely corresponding to the dimeric forms of NodD-FL and NodD-CTD, respectively. A consistent ~60 kDa band was also present in both samples, which is presumed to be the Cpn60 (GroEL) chaperone.

Following IMAC, SEC was performed to assess the purity of NodD-FL (**Figure 4**). The first SEC run (Day 1) showed a prominent peak at the void volume (8 mL, 30 mAU), indicating aggregation, with a secondary peak around 20 mL (10 mAU). The second and third SEC runs were performed using protein from the second expression attempt. Day 2 showed a smaller void volume peak (8 mL, 5 mAU) and a minor peak at 16.5 mL (4 mAU). For Day 3, a larger sample volume (600 μ L) was used, and the SEC column was washed overnight with 0.5 M NaOH. These changes led to an improved profile, with a similar void volume peak (8 mL, 26 mAU) and a more prominent peak at 16.5 mL (15 mAU). Despite this, the protein was still prone to precipitation upon concentration. Based on these observations, we concluded that our NodD-FL sample was not suitable for further analysis, and we opted to proceed with the C-terminal domain of NodD for further studies.

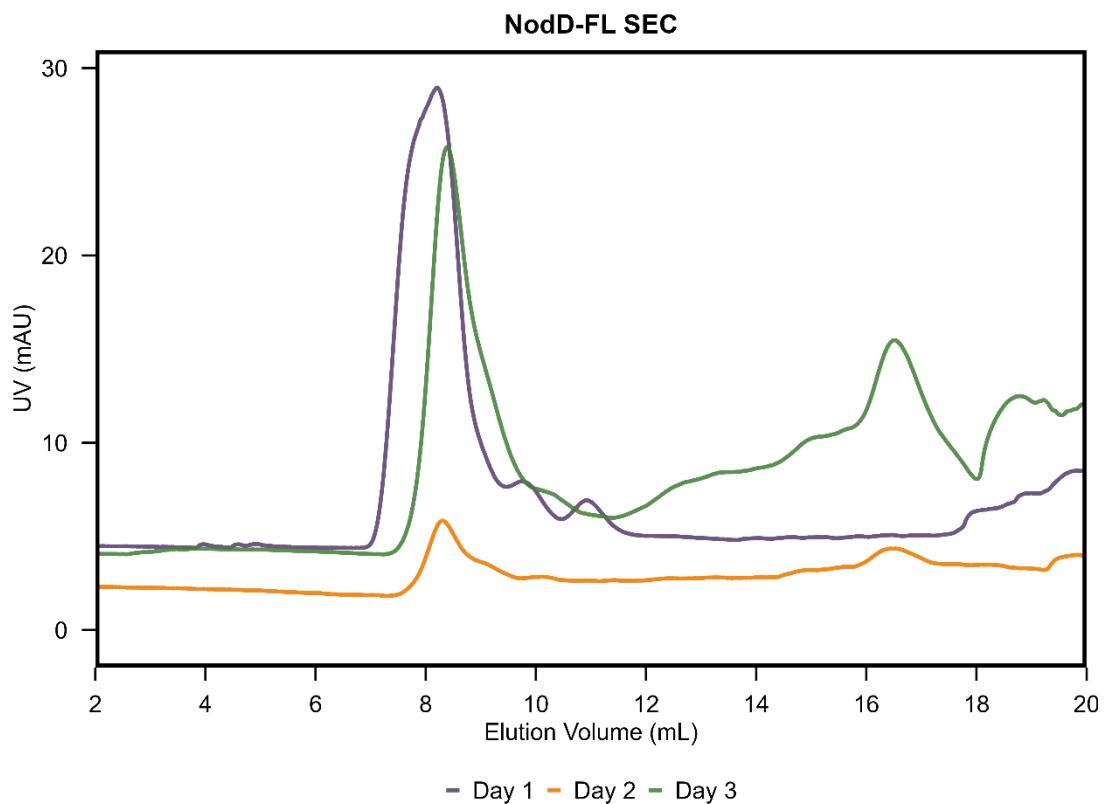


Figure 4: Size-exclusion chromatography profiles of NodD-FL across three purification attempts.

4.1.2 Native mass spectrometry of NodD-CTD

Native mass spectrometry was performed to assess the purity and oligomeric state of NodD-CTD (**Figure 5**). The resultant spectrum revealed charge states that correspond to a molecular weight of approximately 52.5 kDa, consistent with the expected dimeric form. The intensity of these peaks indicates the sample exists predominantly as a dimer under native conditions.

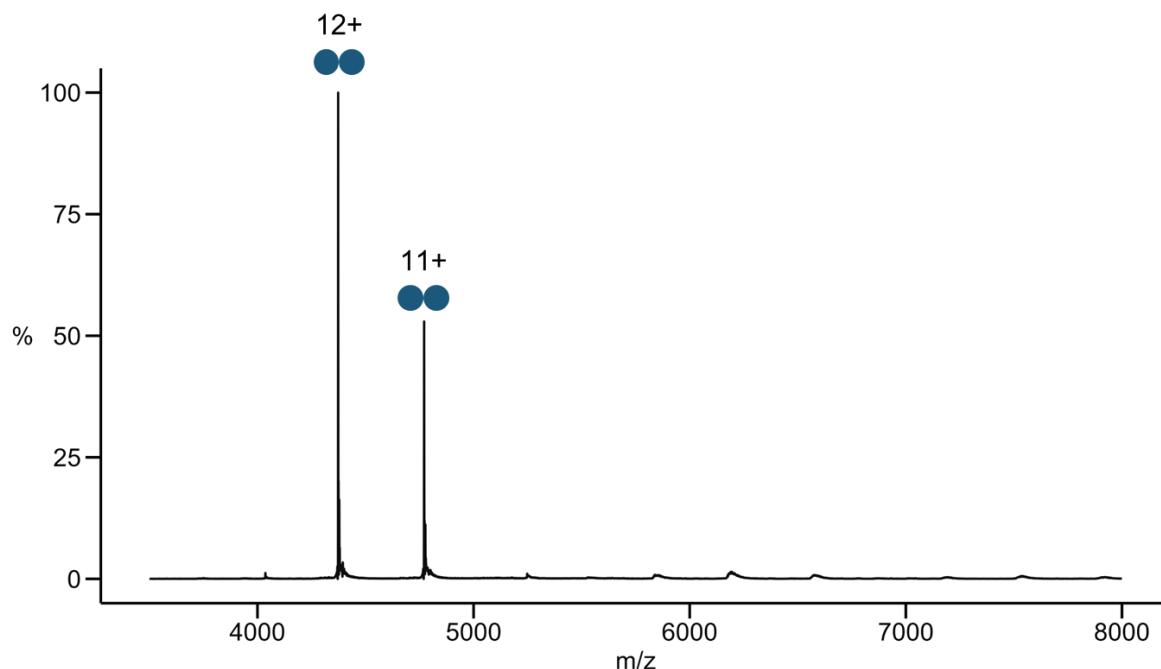


Figure 5. Native mass spectrum of NodD-CTD, showing dominant peaks corresponding to charge states 11+ and 12+, consistent with a dimeric species of ~52.5 kDa. Minor peaks at higher m/z values (between 5,500-8,000) suggest the presence of a low-abundance tetrameric form.

In addition to the major dimeric species, several low-intensity peaks were observed at m/z 5248, 5844, 6189, and 6591. These are consistent with charge states of a potential tetramer (~105 kDa).

4.1.3 Native mass spectrometry of NodD-CTD with flavonoids

To investigate the effects of ligand binding on NodD-CTD oligomerisation, native mass spectrometry was performed in the presence of four flavonoids: luteolin, naringenin, hesperetin, and apigenin. Spectra were compared to the CTD alone to assess changes in oligomeric state and detect mass shifts indicative of flavonoid binding.

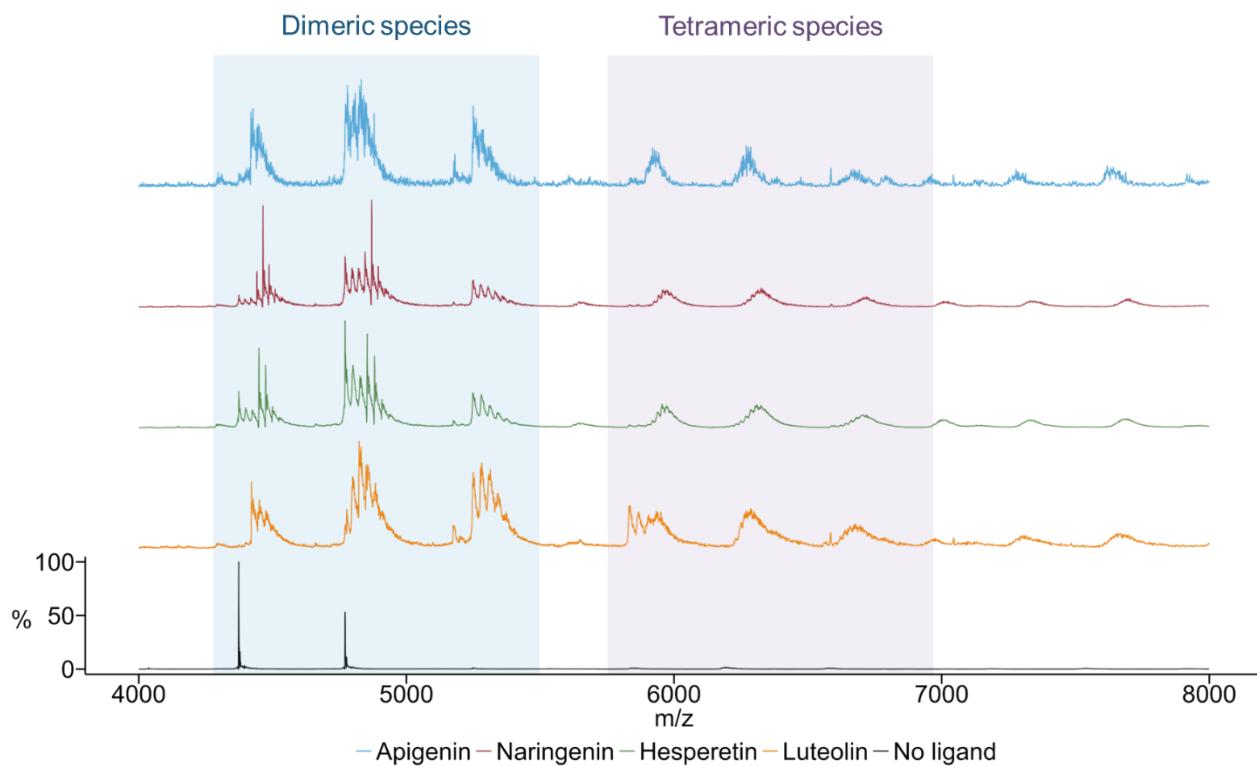


Figure 6: Overlay of native mass spectra showing oligomerisation patterns under each condition. Peaks representing dimeric and tetrameric species are labelled.

In all four conditions, flavonoid addition led to increased oligomerisation (**Figure 6**). Compared to the spectrum of NodD-CTD in the absence of ligand, which was dominated by a single species consistent with a dimer (~52.5 kDa), each flavonoid condition showed additional peaks at higher m/z values, consistent with low-abundance tetrameric species (~105 kDa). The extent of tetramer formation varied slightly between ligands, with luteolin and apigenin producing the most pronounced tetramer peaks.

Mass shifts consistent with flavonoid binding were also observed across multiple charge states (**Figure 7**). In the luteolin-bound sample, the 11+ series showed a progression of peaks at m/z 4777.1, 4798.7, 4823.1, 4849.1, and 4883.9, corresponding to the CTD dimer with 0 to 4 bound luteolin molecules. Each successive peak reflects the addition of a single luteolin (~286 Da).

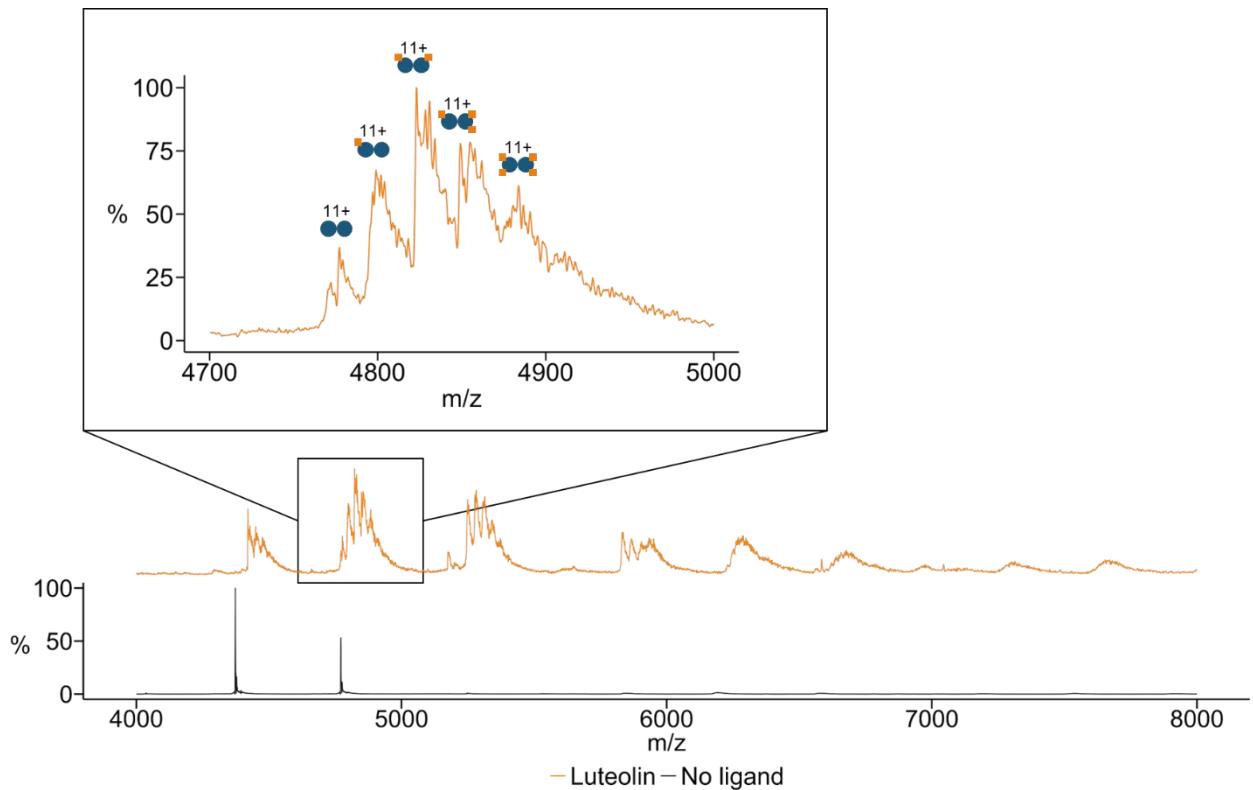


Figure 7. Zoomed-in view of the 11+ charge state series for NodD-CTD in the presence of luteolin. A progression of peaks is observed, consistent with stepwise binding of up to four luteolin molecules per CTD dimer.

4.2 Computational results

4.2.1 Predicted structures from Chai-1

Structural models of NodD from five *Rhizobium* species (*Rlv 3841*, *R. leguminosarum* bv. *trifolii*, *R. tropici*, *S. meliloti*, and *R. galegae*) were generated using Chai-1. Despite some variation in sequence identity (ranging from ~69% to ~80%), the overall folds appeared relatively conserved.

A model of the NodD dimer bound to the nod-box was also generated (Figure 8). Chai-1 predicted NodD to form a cyclically symmetric homodimer, consistent with known structures of other LTTRs. The DNA-binding N-terminal domains are also located adjacent to the DNA as expected.

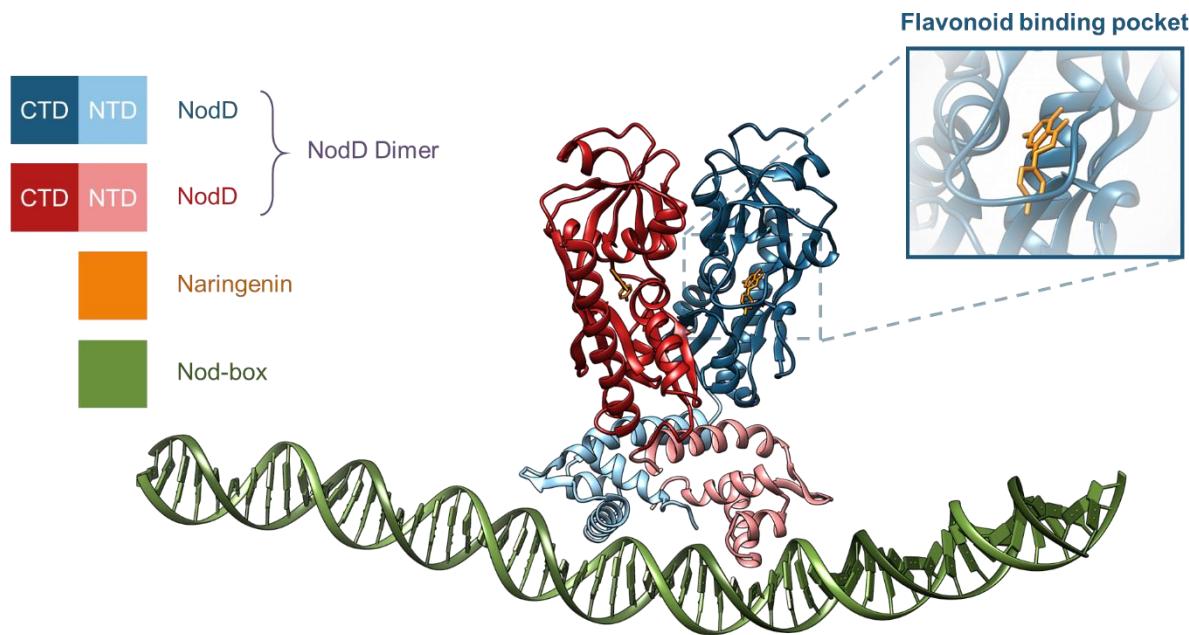


Figure 8: Dimeric structure of NodD bound to DNA, predicted by Chai-1 and visualised in Chimera. The model shows a cyclically symmetric homodimer with DNA-binding domains positioned adjacent to the DNA and naringenin located within the flavonoid-binding pocket.

4.2.2 Conserved residues concentrated in the DNA-binding domain and ligand-binding pocket

To assess which regions of NodD may be functionally or structurally important, ConSurf analysis was also performed on *Rlv* 3841 NodD (**Figure 9**). Conservation scores were mapped onto the predicted structure to highlight residues under evolutionary constraint. As expected, the helix-turn-helix domain exhibited strong conservation. Within the flavonoid-binding domain, specific residues were identified as potential functional residues (highlighted in red).

A phylogenetic tree was constructed using NodD homologues from a selection of rhizobial species (**Figure 10**). Flavonoids specific to each species were identified from literature (Ayala-García et al., 2022; Banfalvi et al., 1988; Begum et al., 2001; Hungria et al., 1992; Liu & Murray, 2016; Martínez-Romero et al., 1991; Peck et al., 2006). The tree shows that *Rhizobium leguminosarum* bv. *viciae* clusters closely with *Sinorhizobium meliloti* and *Neorhizobium galegae*, while *R. phaseoli* and *R. tropici* form a distinct clade. Notably, more closely related rhizobia tend to respond to similar flavonoids, suggesting a link between NodD sequence similarity and flavonoid specificity.

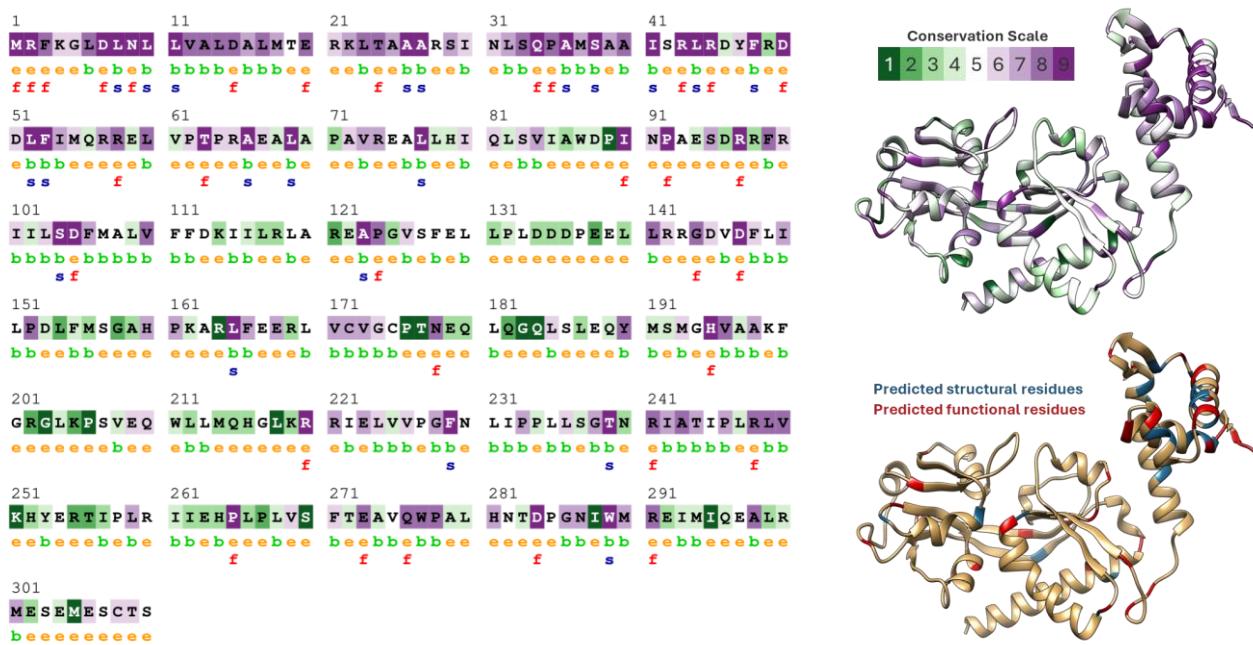


Figure 9. (A) Multiple sequence alignment of *Rlv* 3841 NodD colour-coded by ConSurf conservation scores (1-9). (B) Conservation mapped onto the AlphaFold-predicted structure of NodD. (C) Predicted structural (blue) and functional (red) residues highlighted based on residue conservation and accessibility.

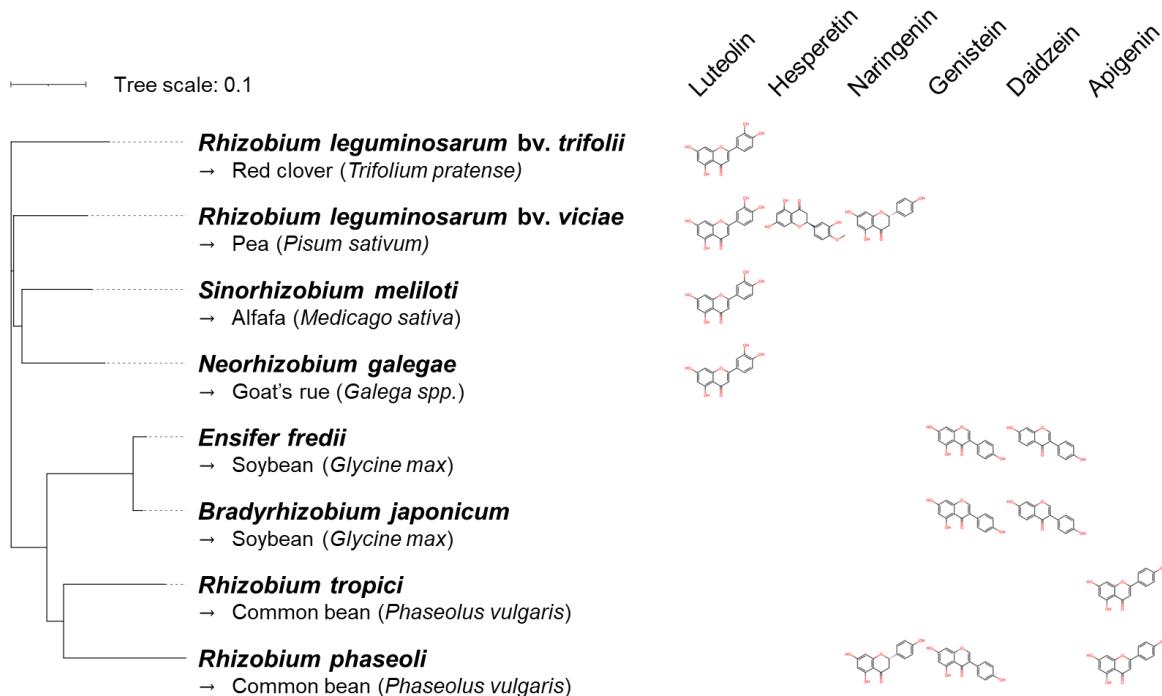


Figure 10. Phylogenetic tree of NodD homologues from selected rhizobial species. Host plant species are labelled beneath their corresponding rhizobial symbionts, and known flavonoid inducers are indicated to the right of each branch.

4.2.3 Ligand interaction analysis highlights conserved binding residues

In parallel, UCSF Chimera was used to analyse potential interactions between the predicted structures and their ligands (see **9.2 Chai-1 Analysis** for full details). Notably, S104 (numbering based on *Rlv* 3841) was predicted to form hydrogen bonds with several ligands across multiple species, and was also highly conserved according to ConSurf, suggesting a dual structural and functional role. Additional residues, such as E273 and V208, were commonly involved in ligand contacts but did not exhibit strong conservation, raising the possibility that they contribute to species-specific ligand recognition rather than core binding functionality.

4.2.4 Site-directed docking predicts strong binding of all four flavonoids to NodD-CTD

Docking simulations were performed using AutoDock Vina to refine predictions of flavonoid binding within the identified ligand-binding cavity (**Figure 11**).

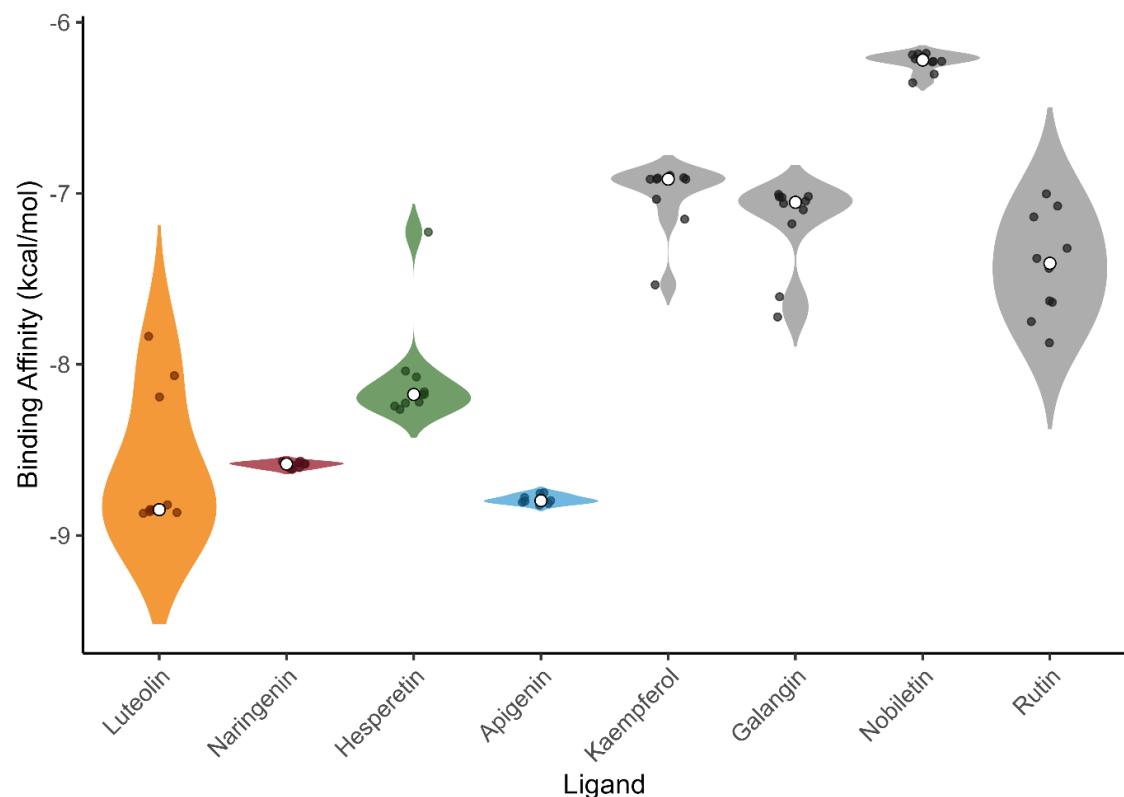


Figure 11. Predicted binding affinities of flavonoids to the NodD-CTD of *Rlv* 3841. Violin plots show range of docking results from AutoDock Vina for four known NodD-activating flavonoids (luteolin, naringenin, hesperetin, and apigenin) and four non-specific flavonoids tested for comparison. Median represented by white point. All four inducers exhibited median binding affinities below -8 kcal/mol.

Grid boxes were centred around residues highlighted by Chimera and LigPlot+ analyses, allowing site-specific docking across all species models. Each protein-ligand complex was subjected to ten independent docking runs to check pose consistency, and binding affinities were visualised on a violin plot.

All 4 flavonoids (luteolin, hesperetin, naringenin, apigenin) showed predicted binding affinities of below -8 kcal/mol, suggesting they are all effective binding partners. Out of the 4, luteolin showed the greatest median affinity (-8.85 kcal/mol), followed by apigenin (-8.80 kcal/mol), naringenin (-8.58 kcal/mol), and hesperetin (-8.18 kcal/mol), although differences were minor. Luteolin docking runs also exhibited more variability between runs, potentially indicating less convergence on a single dominant pose. In contrast, the negative controls (kaempferol, galangin, nobiletin, and rutin) had weaker predicted affinities.

5. Discussion

5.1 Key findings

5.1.1 NodD C-terminal domain is dimeric in solution

This study investigated the structure and flavonoid-responsive behaviour of NodD, a transcriptional regulator involved in rhizobium-legume symbiosis. The C-terminal domain of NodD was successfully expressed and purified, while the full-length construct proved less stable, showing a tendency to aggregate during concentration. Native mass spectrometry revealed that NodD-CTD exists primarily as a dimer under native conditions.

5.1.2 Flavonoids induce tetramerisation of the NodD C-terminal domain independent of DNA-binding domain

Upon addition of flavonoids luteolin, naringenin, hesperetin, and apigenin, increased formation of tetramers was observed. These findings indicate that ligand-induced oligomerisation is an intrinsic property of the C-terminal domain and does not require the presence of the DNA-binding domain. This behaviour is consistent with what has been observed for other LTTRs, where oligomerisation interfaces often reside within the effector-binding domain. For example, in *Shewanella oneidensis* OxyR, the C-terminal α -helix of the regulatory domain is essential for tetramer formation and transcriptional activity (Sun et al., 2022). Similarly, structural studies of LTTRs such as BenM, CbnR, and DntR have identified the tetramerisation interface within the C-terminal effector-binding region (Ezezika et al., 2007). Our findings suggest that the ability of NodD-CTD to tetramerise in response to flavonoids likely reflects a conserved structural mechanism across the LTTR family.

5.1.3 Luteolin, hesperetin, naringenin, and apigenin are effective binding partners of NodD-CTD

Native mass spectrometry revealed mass shifts consistent with flavonoid binding, with spectra showing up to four ligand molecules bound per NodD-CTD dimer. This is likely due to the high concentrations of flavonoid used, which can result in interactions outside of known binding sites. All four flavonoids tested bound effectively to NodD-CTD and induced higher-order oligomerisation. This is consistent with previous studies identifying these compounds as active *nod* gene inducers in *R. leguminosarum*, albeit with hesperetin and naringenin reportedly being the most effective (Begum et al., 2001; Liu & Murray, 2016).

Although all four flavonoids bound effectively to NodD-CTD and promoted tetramerisation, binding alone may not be sufficient to activate NodD function in vivo. Native mass spectrometry reveals ligand-induced structural changes but does not indicate whether these lead to DNA binding or *nod* gene expression. For example, Peck et al. (2006) showed that naringenin, eriodictyol, and daidzein can increase the DNA-binding affinity of NodD1 in *S. meliloti*, yet fail to induce *nod* gene expression like luteolin does. This suggests that neither ligand binding, tetramer formation, nor even DNA interaction necessarily results in transcriptional activation. Hence, while all four flavonoids appear to be effective binding partners, this does not necessarily imply that they are functional inducers or specific to *Rlv* 3841.

In nature, NodD is likely exposed to complex mixtures of flavonoids secreted by host plants. Studies have shown that combinations of flavonoids can lead to stronger *nod* promoter induction than individual flavonoids alone. For example, Begum et al. (2001) found that a mixture of hesperetin and naringenin induced higher levels of *nod* gene expression in *R. leguminosarum* than either compound alone, suggesting a cooperative effect. This may reflect the natural strategy of legumes, which secrete a diversity of flavonoids that together engage multiple NodD variants or increase activation efficiency through combined signalling. While our study focused on individual ligands, further investigation into how flavonoid mixtures influence NodD activity may be important for fully understanding its specificity and function (see **5.4 Future Work**).

5.1.4 Computational modelling guides prediction of key ligand-binding and interaction residues

Computational docking provided a useful framework for assessing the binding potential of different flavonoids to NodD-CTD. Using AutoDock Vina, binding affinities were predicted for both experimentally tested ligands and structurally related negative controls. All four active flavonoids were predicted to bind with high affinity, with mean scores below -8 kcal/mol. In contrast, flavonoids not known to induce *nod* genes in *R. leguminosarum*, such as kaempferol and nobiletin, showed weaker binding scores, supporting the idea that docking can help discriminate likely binding partners from non-functional analogues. However, no substantial differences in predicted affinity were observed among the four active ligands, suggesting that docking alone may not be sufficient to resolve the finer aspects of ligand specificity.

In parallel, structural prediction using Chai-1 identified a well-defined flavonoid-binding pocket within the C-terminal domain. This pocket was supported by conservation analysis, which revealed several highly conserved residues likely involved in ligand interaction. Hydrogen bonding models further helped pinpoint specific contacts between the ligands and surrounding residues. One such residue, S104, was not only highly conserved but also consistently predicted to form hydrogen bonds with flavonoids across most NodD variants. This residue has previously been identified as functionally important in *S. meliloti*, where the S104L substitution prevented binding to the nod gene promoter, potentially due to disrupted ligand interaction (Peck et al., 2013). Other predicted functionally important residues, such as K205 (identified in *Rlv* 3841, *S. meliloti*, and *N. galegae*) and D134 (in *Rlv* 3841), are also supported by previous mutagenesis data (Peck et al., 2013). These findings highlight the potential of computational modelling not only to identify probable ligands but also to estimate residues of functional importance. This information may guide future mutagenesis studies aimed at engineering NodD specificity or increasing response to flavonoids.

5.2 Limitations of computational analysis

While computational docking provided useful insights into potential ligand interactions, its limitations must be recognised. Tools such as AutoDock Vina estimate binding affinities based on static snapshots of protein structures and assume a mostly rigid receptor model. This means

they do not account for flexibility in either the protein or the ligand, which can play a significant role in real binding dynamics. Another limitation lies in the use of predicted structures. In this study, docking was performed on models generated by Chai-1, which are approximations and may not fully capture the true structure of the ligand-binding pocket. Small inaccuracies in side-chain orientation or backbone geometry can affect predicted binding modes and scores.

Docking algorithms like AutoDock Vina also involve a degree of stochasticity, meaning repeated runs can yield slightly different results depending on initial conditions. While averaging scores can help account for this, it still makes precise comparisons between similar ligands challenging. In our case, all four active flavonoids produced high but closely clustered binding affinities, limiting the ability to distinguish subtle differences in specificity.

Despite these caveats, computational methods can be valuable as an initial screening tool. They allow rapid testing of large numbers of candidate molecules, saving time and resources before committing to experimental validation. Ultimately, docking is best used to identify potential binding partners and prioritise targets for further study – not as a replacement for experimental approaches when it comes to understanding ligand specificity or protein function.

5.3 Implications for sustainable agriculture

A clearer understanding of NodD structure and function provides opportunities for targeted protein engineering, where specific amino acid residues could be mutated to alter ligand affinity or improve regulatory function. Previous studies have shown that single point mutations within the regulatory domain can produce constitutively active NodD variants, likely by disrupting its inhibitory control over the DNA-binding region (Peck et al., 2006). More recently, Haskett et al. (2025) showed that removing the regulatory domain from NodD in *Rlv* 3841 also uncouples its activity from flavonoid signalling, resulting in sustained *nod* gene expression even in the absence of flavonoids. These studies demonstrate the feasibility of reprogramming NodD to operate independently of host-derived cues.

Engineering more robust or flexible NodD variants could have valuable applications in agricultural contexts. Many native NodD proteins respond to a narrow set of flavonoids, and

the efficiency of rhizobial-legume symbiosis is sensitive to environmental variables such as pH, temperature and soil moisture (Yeremko et al., 2025). By modifying NodD to function effectively across a wider range of host-derived signals or in suboptimal environments, it may be possible to improve nodulation efficiency in legume crops under diverse field conditions. This could prove especially important in regions where soil quality, microbial diversity, or climate stress hinder the natural establishment of symbiosis.

As global agriculture moves toward more sustainable practices, enhancing biological nitrogen fixation is an important strategy to reduce dependence on synthetic fertilisers. Legume-rhizobium symbioses already contribute significantly to nitrogen inputs in cropping systems, but their potential remains underutilised, particularly in settings where native rhizobia fail to nodulate efficiently or adapt to new crop varieties. Engineering rhizobia with improved NodD variants could help overcome these barriers, offering a practical route to support sustainable food production in diverse environments.

5.4 Future work

Further investigation is needed to explore how NodD responds to more complex signalling environments. While this study focused on individual flavonoids, it is known that plant root exudates contain diverse combinations of inducers, which may act additively or even competitively (Peck et al., 2006). Testing the effects of flavonoid mixtures on NodD oligomerisation and gene activation could provide a more accurate representation of its function in living systems. This could be explored using native mass spectrometry or DNA-binding assays after incubating NodD with defined flavonoid mixtures, and further validated using assays to measure transcriptional activation in rhizobia.

Another valuable direction would be to mutate the key residues identified through docking and conservation analysis to test their role in flavonoid recognition. Site-directed mutagenesis of the predicted ligand-binding residues, followed by native mass spectrometry, binding assays, or EMSAs, could show whether these residues are important for ligand binding and transcriptional activation. *In silico* mutagenesis could also be used to assess how specific amino

acid substitutions might affect ligand binding, allowing for the computational screening of multiple variants before selecting those with promising results for experimental follow-up.

Characterising the full-length structure of NodD remains a key challenge. Native mass spectrometry results suggested that the protein tends to aggregate during purification, but this might be overcome by testing alternative expression systems or optimising buffer conditions to improve solubility. Structural techniques such as X-ray crystallography or cryo-EM could then be used to investigate the full-length protein or its oligomeric forms, particularly in the presence of flavonoids or DNA. Unlike structure prediction tools, which rely on computational models, crystallographic methods offer higher-resolution insight into protein conformation and ligand interactions. However, crystallising full-length NodD may be especially difficult due to the flexibility of regions within its N-terminal DNA-binding domain. As a workaround, future efforts could focus on crystallising the more stable C-terminal domain alone to gain reliable structural information without the complications introduced by flexible regions.

6. Conclusion

This study investigated the structural and functional properties of the LysR-type transcriptional regulator NodD from *Rhizobium leguminosarum*, with a focus on its C-terminal ligand-binding domain (CTD). Native mass spectrometry revealed that the isolated CTD forms dimers under native conditions and undergoes flavonoid-induced tetramerisation in the absence of the DNA-binding domain, suggesting that oligomerisation is an intrinsic property of the CTD and may represent a key step in NodD activation. Computational modelling supported the presence of a conserved ligand-binding pocket within the CTD, with ConSurf and docking analyses identifying several residues potentially involved in flavonoid recognition. Phylogenetic analysis of NodD homologues across diverse rhizobia indicated a trend toward conserved flavonoid recognition among closely related strains. This supports the idea of a structurally conserved mechanism of activation. Docking simulations confirmed that all four tested ligands (luteolin, naringenin, apigenin, and hesperetin) bind with high predicted affinities. However, no substantial differences in binding strength or pose were observed between them, limiting the ability of computational approaches alone to explain flavonoid specificity in *R. leguminosarum*. Building on these findings, further mutagenesis studies could help clarify the molecular mechanisms underlying NodD activation and support efforts to optimise symbiosis in the context of sustainable agriculture.

7. Acknowledgements

I am deeply grateful to Jani for welcoming me into his lab and for his unwavering support throughout this project. His encouragement and guidance have been invaluable at every stage of my research.

I would also like to thank Tom for guiding me through the protein expression and purification steps and for always taking the time to answer my questions.

An additional thanks to Francesco and Joe for their support as college tutors, and for their helpful feedback and advice throughout my degree.

Finally, my sincere thanks go to everyone in the Bolla lab for their assistance with all aspects of lab work and for creating such a welcoming and friendly environment. It has been a privilege to work with and learn from you all!

8. References

- Aryal, B., Gurung, R., Camargo, A. F., Fongaro, G., Treichel, H., Mainali, B., Angove, M. J., Ngo, H. H., Guo, W., & Paudel, S. R. (2022). Nitrous oxide emission in altered nitrogen cycle and implications for climate change. *Environmental Pollution*, 314, 120272. <https://doi.org/10.1016/j.envpol.2022.120272>
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, 44(W1), W344–W350. <https://doi.org/10.1093/nar/gkw408>
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T., & Ben-Tal, N. (2010). ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, 38(suppl_2), W529–W533. <https://doi.org/10.1093/nar/gkq399>
- Ayala-García, P., Jiménez-Guerrero, I., Jacott, C. N., López-Baena, F. J., Ollero, F. J., del Cerro, P., & Pérez-Montaño, F. (2022). The Rhizobium tropici CIAT 899 NodD2 protein promotes symbiosis and extends rhizobial nodulation range by constitutive nodulation factor synthesis. *Journal of Experimental Botany*, 73(19), 6931–6941. <https://doi.org/10.1093/jxb/erac325>
- Banfalvi, Z., Nieuwkoop, A., Schell, M., Besl, L., & Stacey, G. (1988). Regulation of nod gene expression in *Bradyrhizobium japonicum*. *Molecular and General Genetics MGG*, 214(3), 420–424. <https://doi.org/10.1007/BF00330475>
- Begum, A. A., Leibovitch, S., Migner, P., & Zhang, F. (2001). Specific flavonoids induced nod gene expression and pre-activated nod genes of *Rhizobium leguminosarum* increased pea (*Pisum sativum* L.) and lentil (*Lens culinaris* L.) nodulation in controlled growth chamber environments. *Journal of Experimental Botany*, 52(360), 1537–1543. <https://doi.org/10.1093/jexbot/52.360.1537>
- Castellano-Hinojosa, A., Correa-Galeote, D., González-López, J., & Bedmar, E. J. (2020). Effect of nitrogen fertilisers on nitrous oxide emission, nitrifier and denitrifier abundance and bacterial diversity in closed ecological systems. *Applied Soil Ecology*, 145, 103380. <https://doi.org/10.1016/j.apsoil.2019.103380>

- Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., & Ben-Tal, N. (2013). ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Israel Journal of Chemistry*, 53(3–4), 199–206. <https://doi.org/10.1002/ijch.201200096>
- Chai Discovery, Boitreaud, J., Dent, J., McPartlon, M., Meier, J., Reis, V., Rogozhnikov, A., & Wu, K. (2024). Chai-1: Decoding the molecular interactions of life. *bioRxiv*, 2024.10.10.615955. <https://doi.org/10.1101/2024.10.10.615955>
- Chang, M., & Crawfordt, I. P. (1990). The roles of indoleglycerol phosphate and the TrpI protein in the expression of trpBA from *Pseudomonas aeruginosa*. *Nucleic Acids Research*, 18(4), 979–988. <https://doi.org/10.1093/nar/18.4.979>
- Chen, X.-C., Feng, J., Hou, B.-H., Li, F.-Q., Li, Q., & Hong, G.-F. (2005). Modulating DNA bending affects NodD-mediated transcriptional control in *Rhizobium leguminosarum*. *Nucleic Acids Research*, 33(8), 2540–2548. <https://doi.org/10.1093/nar/gki537>
- Crews, T. E., & Peoples, M. B. (2004). Legume versus fertilizer sources of nitrogen: Ecological tradeoffs and human needs. *Agriculture, Ecosystems & Environment*, 102(3), 279–297. <https://doi.org/10.1016/j.agee.2003.09.018>
- Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z., & Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. *Nature Geoscience*, 1(10), 636–639. <https://doi.org/10.1038/ngeo325>
- Ezezika, O. C., Haddad, S., Neidle, E. L., & Momany, C. (2007). Oligomerization of BenM, a LysR-type transcriptional regulator: Structural basis for the aggregation of proteins in this family. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 63(5), 361–368. <https://doi.org/10.1107/S1744309107019185>
- Feng, J., Li, Q., Hu, H., Chen, X., & Hong, G. (2003). Inactivation of the nod box distal half-site allows tetrameric NodD to activate nodA transcription in an inducer-independent manner. *Nucleic Acids Research*, 31(12), 3143–3156. <https://doi.org/10.1093/nar/gkg411>
- Feng, J., Qing Li, F., Li, Q., Liang Hu, H., & Fan Hong, G. (2002). Expression and purification of *Rhizobium leguminosarum* NodD. *Protein Expression and Purification*, 26(2), 321–328. [https://doi.org/10.1016/S1046-5928\(02\)00535-1](https://doi.org/10.1016/S1046-5928(02)00535-1)
- Fisher, R. F., & Long, S. R. (1993). Interactions of NodD at the *nod* Box: NodD Binds to Two Distinct Sites on the Same Face of the Helix and Induces a Bend in the DNA. *Journal of Molecular Biology*, 233(3), 336–348. <https://doi.org/10.1006/jmbi.1993.1515>

- Galloway, J. N., Leach, A. M., Bleeker, A., & Erisman, J. W. (2013). A chronology of human understanding of the nitrogen cycle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1621), 20130120. <https://doi.org/10.1098/rstb.2013.0120>
- Giannopoulou, E.-A., Senda, M., Koentjoro, M. P., Adachi, N., Ogawa, N., & Senda, T. (2021). Crystal structure of the full-length LysR-type transcription regulator CbnR in complex with promoter DNA. *The FEBS Journal*, 288(15), 4560–4575. <https://doi.org/10.1111/febs.15764>
- Haskett, T. L., Cooke, L., Green, P., & Poole, P. S. (2025). Regulation of Rhizobial Nodulation Genes by Flavonoid-Independent NodD Supports Nitrogen-Fixing Symbioses With Legumes. *Environmental Microbiology*, 27(1), e70014. <https://doi.org/10.1111/1462-2920.70014>
- Hungria, M., Johnston, A. W., & Phillips, D. A. (1992). Effects of flavonoids released naturally from bean (*Phaseolus vulgaris*) on nodD-regulated gene transcription in *Rhizobium leguminosarum* bv. *Phaseoli*. *Molecular Plant-Microbe Interactions: MPMI*, 5(3), 199–203. <https://doi.org/10.1094/mpmi-5-199>
- Jo, I., Chung, I.-Y., Bae, H.-W., Kim, J.-S., Song, S., Cho, Y.-H., & Ha, N.-C. (2015). Structural details of the OxyR peroxide-sensing mechanism. *Proceedings of the National Academy of Sciences*, 112(20), 6443–6448. <https://doi.org/10.1073/pnas.1424495112>
- Kabange, N. R., Lee, S.-M., Shin, D., Lee, J.-Y., Kwon, Y., Kang, J.-W., Cha, J.-K., Park, H., Alibu, S., & Lee, J.-H. (2022). Multiple Facets of Nitrogen: From Atmospheric Gas to Indispensable Agricultural Input. *Life*, 12(8), 1272. <https://doi.org/10.3390/life12081272>
- Kim, Y., Chhor, G., Tsai, C., Winans, J. B., Jedrzejczak, R., Joachimiak, A., & Winans, S. C. (2018). Crystal structure of the ligand-binding domain of a LysR-type transcriptional regulator: Transcriptional activation via a rotary switch. *Molecular Microbiology*, 110(4), 550–561. <https://doi.org/10.1111/mmi.14115>
- Ladha, J. K., Peoples, M. B., Reddy, P. M., Biswas, J. C., Bennett, A., Jat, M. L., & Krupnik, T. J. (2022). Biological nitrogen fixation and prospects for ecological intensification in cereal-based cropping systems. *Field Crops Research*, 283, 108541. <https://doi.org/10.1016/j.fcr.2022.108541>
- Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*, 51(10), 2778–2786. <https://doi.org/10.1021/ci200227u>

- Liu, C.-W., & Murray, J. D. (2016). The Role of Flavonoids in Nodulation Host-Range Specificity: An Update. *Plants*, 5(3), Article 3. <https://doi.org/10.3390/plants5030033>
- Maathuis, F. J. (2009). Physiological functions of mineral macronutrients. *Current Opinion in Plant Biology*, 12(3), 250–258. <https://doi.org/10.1016/j.pbi.2009.04.003>
- Maddocks, S. E., & Oyston, P. C. F. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. In *Microbiology (Reading, England)* (Vol. 154, Issue 12, pp. 3609–3623). Microbiology Society. <https://doi.org/10.1099/mic.0.2008/022772-0>
- Martínez-Romero, E., Segovia, L., Mercante, F. M., Franco, A. A., Graham, P., & Pardo, M. A. (1991). Rhizobium tropici, a Novel Species Nodulating Phaseolus vulgaris L. Beans and Leucaena sp. Trees. *International Journal of Systematic and Evolutionary Microbiology*, 41(3), 417–426. <https://doi.org/10.1099/00207713-41-3-417>
- Menegat, S., Ledo, A., & Tirado, R. (2022). Greenhouse gas emissions from global production and use of nitrogen synthetic fertilisers in agriculture. *Scientific Reports*, 12(1), 14490. <https://doi.org/10.1038/s41598-022-18773-w>
- Mesfin, S., Gebresamuel, G., Haile, M., & Zenebe, A. (2023). Potentials of legumes rotation on yield and nitrogen uptake of subsequent wheat crop in northern Ethiopia. *Heliyon*, 9(6). <https://doi.org/10.1016/j.heliyon.2023.e16126>
- Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K., & Senda, T. (2003). Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: Unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *Journal of Molecular Biology*, 328(3), 555–566. [https://doi.org/10.1016/s0022-2836\(03\)00312-7](https://doi.org/10.1016/s0022-2836(03)00312-7)
- Neamatollahi, E., Jahansuz, M. R., Mazaheri, D., & Bannayan, M. (2013). Intercropping. In E. Lichtfouse (Ed.), *Sustainable Agriculture Reviews: Volume 12* (pp. 119–142). Springer Netherlands. https://doi.org/10.1007/978-94-007-5961-9_4
- Oldroyd, G. E. D. (2013). Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology*, 11(4), 252–263. <https://doi.org/10.1038/nrmicro2990>
- Parsek, M. R., Shinabarger, D. L., Rothmel, R. K., & Chakrabarty, A. M. (1992). Roles of CatR and cis,cis-muconate in activation of the catBC operon, which is involved in benzoate degradation in *Pseudomonas putida*. *Journal of Bacteriology*, 174(23), 7798–7806. <https://doi.org/10.1128/jb.174.23.7798-7806.1992>

- Peck, M. C., Fisher, R. F., Bliss, R., & Long, S. R. (2013). Isolation and characterization of mutant *Sinorhizobium meliloti* NodD1 proteins with altered responses to luteolin. *Journal of Bacteriology*, 195(16), 3714–3723. <https://doi.org/10.1128/JB.00309-13>
- Peck, M. C., Fisher, R. F., & Long, S. R. (2006). Diverse Flavonoids Stimulate NodD1 Binding to nod Gene Promoters in *Sinorhizobium meliloti*. *Journal of Bacteriology*, 188(15), 5417–5427. <https://doi.org/10.1128/jb.00376-06>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Poole, P., Ramachandran, V., & Terpolilli, J. (2018). Rhizobia: From saprophytes to endosymbionts. *Nature Reviews Microbiology*, 16(5), 291–303. <https://doi.org/10.1038/nrmicro.2017.171>
- Sun, W., Fan, Y., Wan, F., Tao, Y. J., & Gao, H. (2022). Functional Irreplaceability of *Escherichia coli* and *Shewanella oneidensis* OxyRs Is Critically Determined by Intrinsic Differences in Oligomerization. *mBio*, 13(1), e03497-21. <https://doi.org/10.1128/mbio.03497-21>
- The UniProt Consortium. (2025). UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, 53(D1), D609–D617. <https://doi.org/10.1093/nar/gkae1010>
- Tyrrell, R., Verschueren, K. H., Dodson, E. J., Murshudov, G. N., Addy, C., & Wilkinson, A. J. (1997). The structure of the cofactor-binding fragment of the LysR family member, CysB: A familiar fold with a surprising subunit arrangement. *Structure*, 5(8), 1017–1032. [https://doi.org/10.1016/S0969-2126\(97\)00254-2](https://doi.org/10.1016/S0969-2126(97)00254-2)
- Yariv, B., Yariv, E., Kessel, A., Masrati, G., Chorin, A. B., Martz, E., Mayrose, I., Pupko, T., & Ben-Tal, N. (2023). Using evolutionary data to make sense of macromolecules with a “face-lifted” ConSurf. *Protein Science*, 32(3), e4582. <https://doi.org/10.1002/pro.4582>
- Yeremko, L., Czopek, K., Staniak, M., Marenycz, M., & Hanhur, V. (2025). Role of Environmental Factors in Legume-Rhizobium Symbiosis: A Review. *Biomolecules*, 15(1), Article 1. <https://doi.org/10.3390/biom15010118>

9. Appendix

9.1 AutoDock Vina workflow

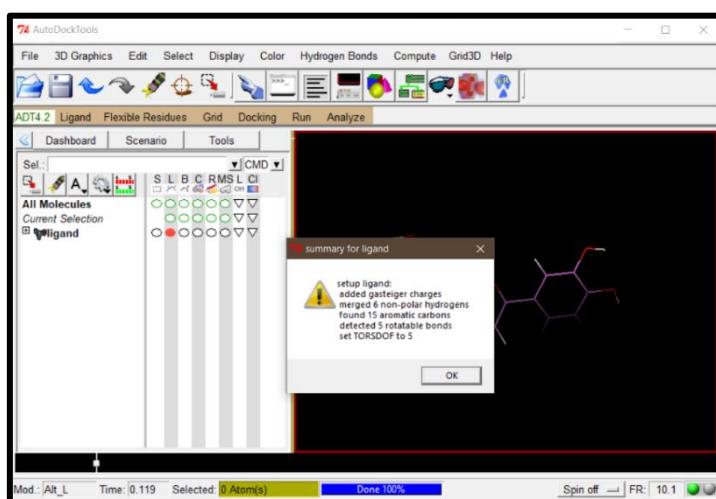
9.1.1 Applications used

- PyMol
- AutoDockTools 1.5.7
- AutoDock Vina 1.2.7

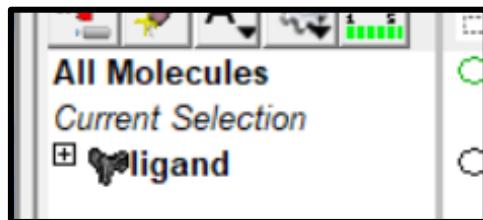
9.1.2 Molecular docking method

A) Preparing protein and ligands for docking

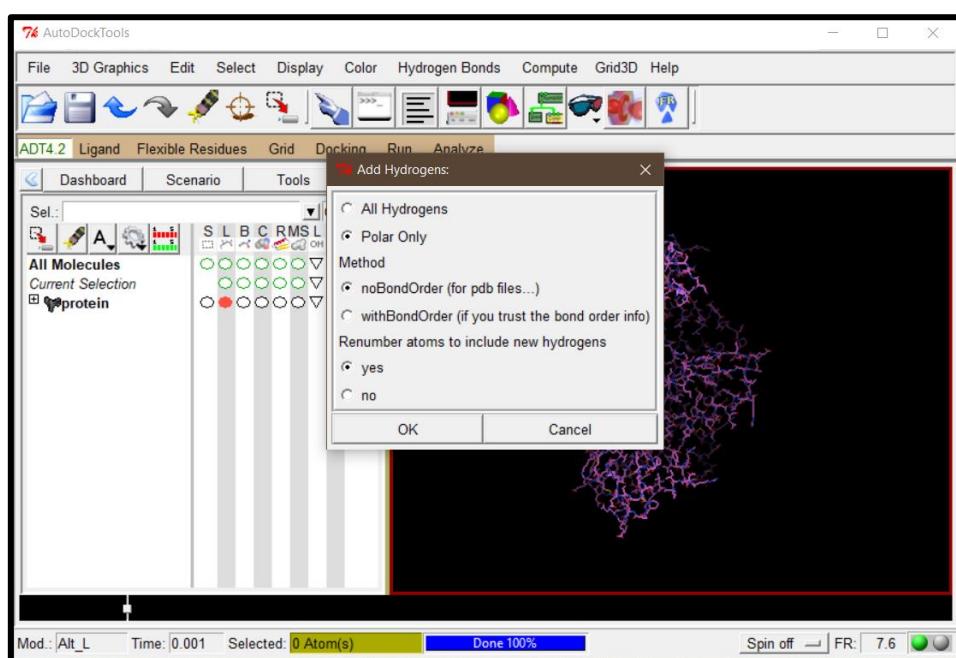
- Move protein file, ligand files, and AutoDock Vina .exe file to a single folder (e.g. “Desktop\Docking”).
 - Protein and ligand files must be in **.pdb format**. Ligands can be downloaded from PubChem as .sdf file then exported as .pdb in PyMol.
- Open AutoDockTools.
- Prepare ligand file (e.g. “ligand.pdb”).
 - Open ligand.pdb using the “Ligand” tab (Ligand > Input > Open...).



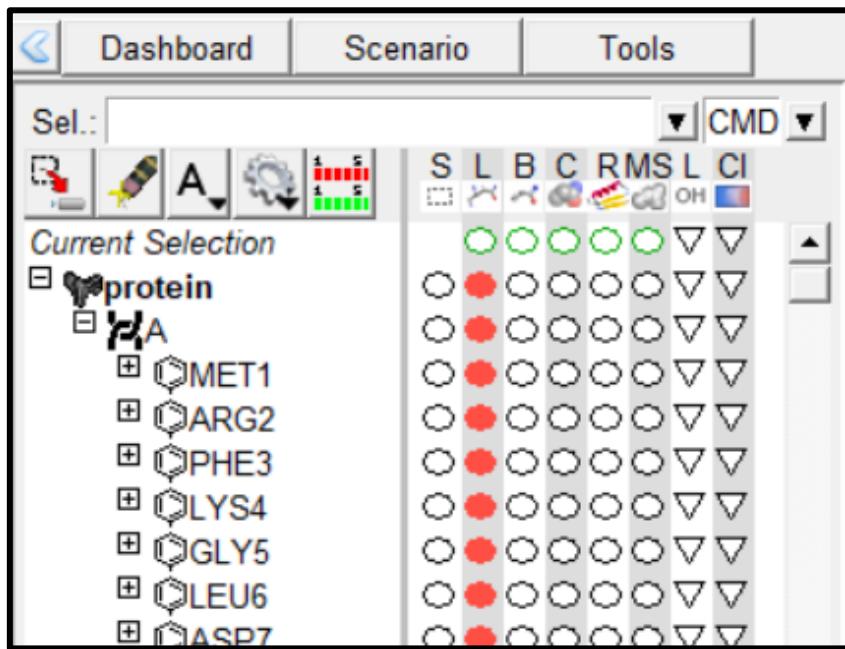
- Detect root (Ligand > Torsion Tree > Detect Root...).
- Export as .pbddqt file (Ligand > Output > Save as PDBQT...). Make sure all files are saved in your working directory.



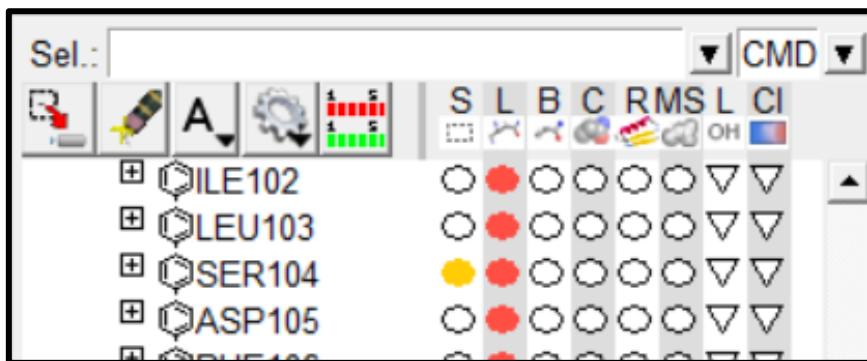
- Right click on “ligand” in the molecules panel to delete it.
- Repeat process with any other ligands to be docked.
- Prepare protein file (e.g. “protein.pdb”)
 - Open protein.pdb (File > Read Molecule).
 - Delete water molecules (Edit > Delete Water).
 - Add hydrogens (Edit > Hydrogens > Add). Select “Polar Only”, then OK.



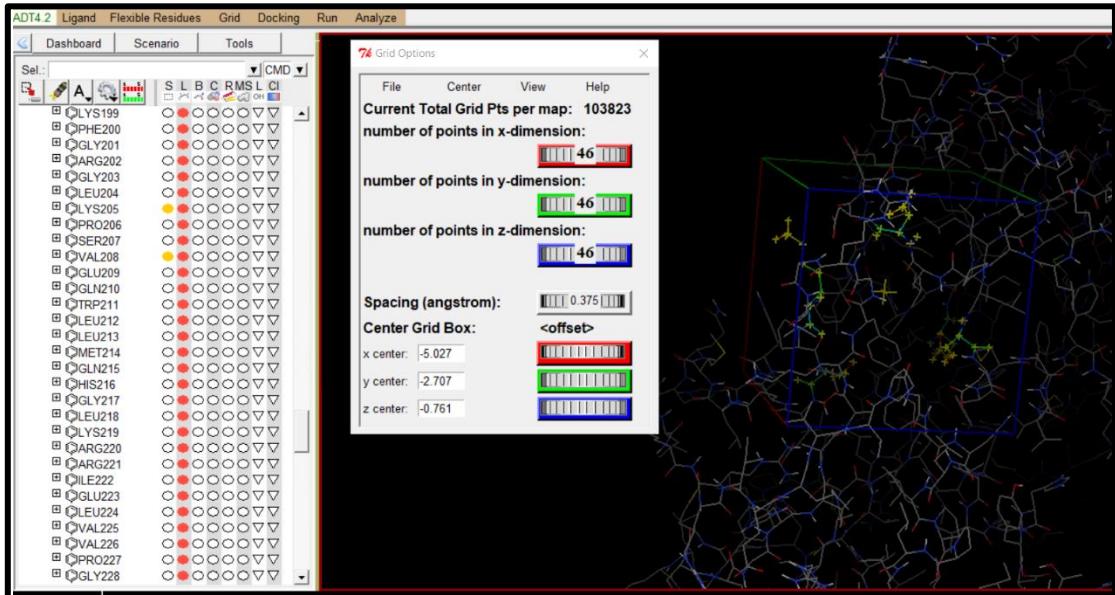
- Export protein (Grid > Macromolecule > Choose). Select your protein then save as .pdbqt file in the working directory.
- Define the grid-box.
 - To help define the region in which the ligand is docked, known binding residues can be highlighted so that the grid-box can be defined around them. To do this, click the dropdown “+” button next to the protein to reveal all residues.



- Scroll through and select a binding residue (e.g. “SER104”) by clicking the left-most “○”, then the right-most “▽” and “By rainbow” to colour it. This residue should now be highlighted within the protein. Repeat with all other binding residues.



- Add a grid-box (Grid > Grid Box). Show as lines (View > Show box as lines). Adjust x/y/z center values to define the desired search space. The default grid-box size is a 40pt cube, however this can be made bigger/smaller depending on the size of the binding cavity.



- Export grid dimensions file (File > Output grid dimensions file). Save in working directory as .txt file. The Grid window can be closed (File > Close saving current).

B) Molecular docking with AutoDock Vina

- Create a batch script by pasting the template below into NotePad and saving as a .bat file (e.g. “run_vina.bat”). Edit the following where necessary:
 - Protein and ligand file names (e.g. “protein.pdbqt” and “ligand.pdbqt”)
 - Grid center coords (e.g. “-5.027”, “-2.707”, “-0.761”)
 - Grid size (e.g. “46”, “46”, “46”)

```
@echo off
setlocal enabledelayedexpansion

:: === USER SETTINGS ===
set "receptor=protein.pdbqt"    :: change to your receptor file
set "ligand=ligand.pdbqt"        :: change to your ligand file
```

```

set runs=1 :: how many repeats

for %%A in ("%receptor%") do set "receptorName=%~nA"
for %%B in ("%ligand%") do set "ligandName=%~nB"

for /L %%i in (1,1,%runs%) do (
    echo Running docking iteration %%i...
    vina_1.2.7_win.exe ^
        --receptor "!receptor!" ^
        --ligand "!ligand!" ^
        --center_x -5.027 ^
        --center_y -2.707 ^
        --center_z -0.761 ^
        --size_x 46 ^
        --size_y 46 ^
        --size_z 46 ^
        --exhaustiveness 8 ^
        --seed %%i ^
        --out "!receptorName!_!ligandName!_%i.pdbqt"
)

echo All runs completed.
pause

```

- *Explanation of code below:*

```

@echo off
setlocal enabledelayedexpansion

:: === USER SETTINGS ===
set "receptor=protein.pdbqt"      :: change to your receptor file
set "ligand=ligand.pdbqt"          :: change to your ligand file
set runs=1                         :: how many repeats

for %%A in ("%receptor%") do set "receptorName=%~nA"
for %%B in ("%ligand%") do set "ligandName=%~nB"

for /L %%i in (1,1,%runs%) do (
    echo Running docking iteration %%i...

    vina_1.2.7_win.exe ^      # make sure this matches the name of your installed vina .exe file!

        --receptor "!receptor!" ^
        --ligand "!ligand!" ^
        --center_x -5.027 ^
        --center_y -2.707 ^
        --center_z -0.761 ^
        --size_x 46 ^
        --size_y 46 ^
        --size_z 46 ^
        --exhaustiveness 8 ^
        --seed %%i ^
        --out "!receptorName!_!ligandName!_%i.pdbqt"
)

echo All runs completed.
pause

```



- All the highlighted regions can be edited: file names in yellow, grid-box data in green, and optional values in blue.
 - `set runs` can run the docking multiple times to see if results vary. AutoDock vina uses a stochastic search algorithm (random starting points) so results can vary slightly between runs.
 - exhaustiveness is by default set to 8, however if the defined search space is large (greater than 30x30x30 Angstrom) it is recommended to increase this (<https://autodock-vina.readthedocs.io/en/latest/faq.html>).
 - `--seed %%i` sets the seed each run for reproducibility (1, 2, 3...).
 - `--out "!receptorName!!ligandName!%%i.pdbqt"` automatically names the output files by combining the names of the protein and ligand files used, plus the seed (e.g. “protein_ligand_1”, “protein_ligand_2”...).

4. Ensure batch script is saved in the working directory. When ready to dock, double click the batch script to run. The Cmd terminal should open as below.

```
C:\WINDOWS\system32\cmd.exe
Running docking iteration 1...
AutoDock Vina v1.2.3
#####
# If you used AutoDock Vina in your work, please cite: #
# J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli #
# AutoDock Vina 1.2.0: New Docking Methods, Expanded Force #
# Field, and Python Bindings, J. Chem. Inf. Model. (2021) #
# DOI 10.1021/acs.jcim.1c00203 #
# O. Trott, A. J. Olson, #
# AutoDock Vina: improving the speed and accuracy of docking #
# with a new scoring function, efficient optimization and #
# multithreading, J. Comp. Chem. (2010) #
# DOI 10.1002/jcc.21334 #
# Please see https://github.com/ccsb-scripps/AutoDock-Vina for #
# more information. #
#####

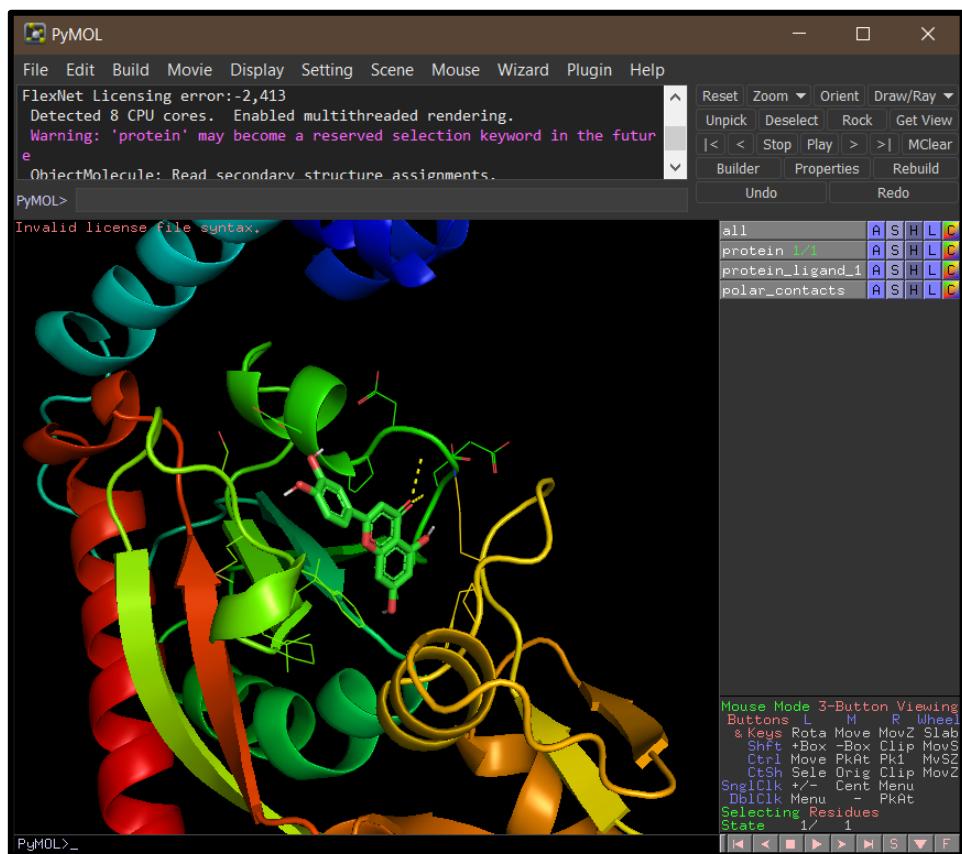
Scoring function : vina
Rigid receptor: protein.pdbqt
Ligand: ligand.pdbqt
Grid center: X -5.027 Y -2.707 Z -0.761
Grid size : X 46 Y 46 Z 46
Grid space : 0.375
Exhaustiveness: 8
CPU: 0
Verbosity: 1

WARNING: Search space volume is greater than 27000 Angstrom^3 (See FAQ)
Computing Vina grid ... done.
Performing docking (random seed: 1) ...
0% 10 20 30 40 50 60 70 80 90 100%
|---|---|---|---|---|---|---|---|---|---|
*****  
mode | affinity | dist from best mode
| (kcal/mol) | rmsd l.b. | rmsd u.b.
-----+-----+-----+
 1      -7.206      0      0
 2      -7.096     1.888    2.466
 3      -6.961     2.299    7.046
```

9.1.3 Visualising docking results

A) Visualising binding result in PyMol

- Docking results can be visualised in PyMol.
 - Open the protein .pdb file (e.g. “protein.pdb”). Then open the output .pdbqt file (e.g. “protein_ligand_1.pdbqt”).
 - View as cartoon (All > Action > Preset > Ligand sites > Cartoon).



B) Visualising binding affinities between multiple ligands

- Binding affinities between ligands can also be visualised in R.
 - Instead of manually entering the binding affinity scores into a dataset, the code below was written to automatically extract the scores from the output files. To do this, the R project working directory needs to be set to the folder containing the output files.

- The template below just requires the `protein_name` value to be changed to match the name of protein used. In this example, “vici” was used (for *R. leguminosarum* bv. *viciae*).

```
# required packages: "tidyverse"

# 1) Adjust to match protein prefix (e.g. "protein", or "vici")
protein_name <- "vici"

# 2) Find all Vina outputs with "protein_ligand_seed.pdbqt" style name
all_files <- list.files(
  pattern = paste0("^", protein_name, "_[^_]+_\\d+\\.pdbqt$"),
  full.names = TRUE
)

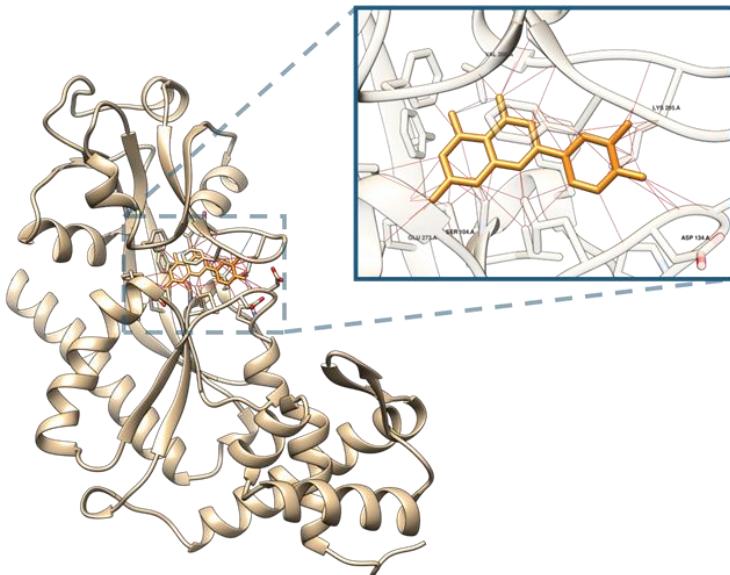
# 3) Extract numeric score from the first REMARK line of the .pdbqt file
get_score <- function(f) {
  lines <- readLines(f, warn = FALSE)
  line <- grep("^REMARK VINA RESULT", lines, value = TRUE)[1]
  if (!is.na(line)) as.numeric(strsplit(line, "\\s+")[[1]][4]) else NA
}

# 4) Build a data frame of ligands and respective score
results <- tibble(
  file    = all_files,
  Ligand = sub(
    paste0("^", protein_name, "_([^\_]+)_\\d+\\.pdbqt$"),
    "\\\1",
    basename(all_files)
  ),
  Score   = map_dbl(all_files, get_score)
)
```

- This will extract the binding results of each output file to the data frame `results`. These values can be used to plot graphs for visualisation of binding affinities, or to perform statistical tests.

9.2 Chai-1 Analysis

Chai-1: NodD *R. leguminosarum* (Target) + Luteolin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.5957, pTM = 0.7952

Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O3_1	2.819
VAL 208.A N	LIG2 1.B O1_1	2.811
LIG2 1.B O3_1	GLU 273.A OE2	2.779
LIG2 1.B O5_1	LYS 205.A O	3.105
LIG2 1.B O6_1	ASP 134.A O	2.389

Chimera: Identified residues

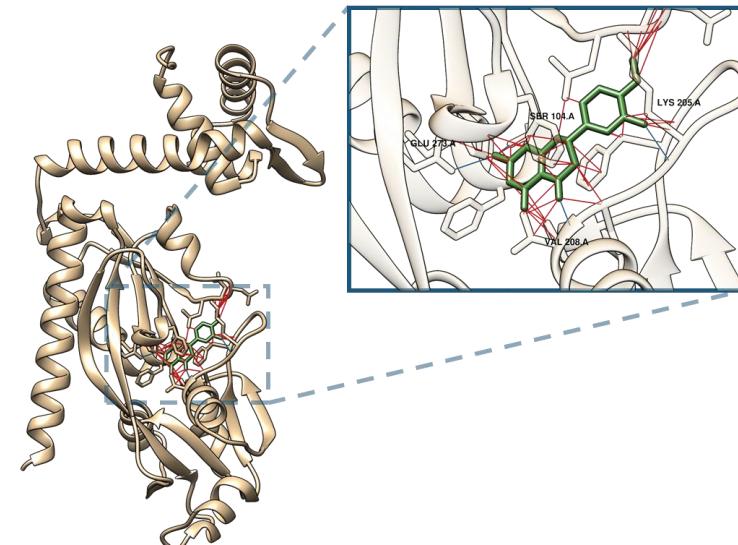
H-bonds = 5

Other Contacts = 66

CB-Dock2: Vina = -8.8, Cavity volume = 486 Å³

SER104 PHE106 MET107 LEU133 ASP134 ASP135 ASP136 PRO137
GLU138 ILE150 LEU151 PRO152 LEU154 PHE155 MET156 SER157
PHE200 GLY201 LYS205 PRO206 SER207 VAL208 GLU209 TRP211
GLN215 PHE229 PRO246 PHE271 GLU273

Chai-1: NodD *R. leguminosarum* (Target) + Hesperetin



Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O4_1	2.494
VAL 208.A N	LIG2 1.B O2_1	2.823
LIG2 1.B O4_1	SER 104.A OG*	2.494
LIG2 1.B O4_1	GLU 273.A OE2	3.099
LIG2 1.B O6_1	LYS 205.A O	3.047

Chimera: Identified residues

H-bonds = 4

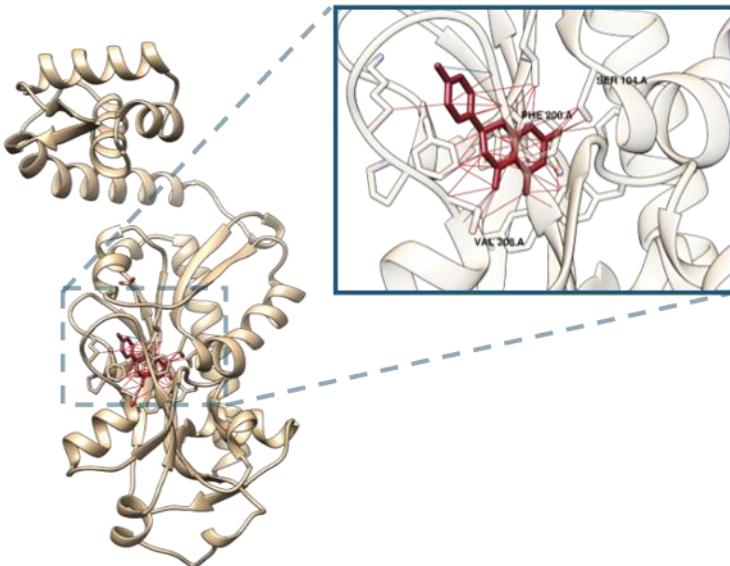
Other Contacts = 54

CB-Dock2: Vina = -8.3, Cavity volume = 486 Å³

SER104 PHE106 MET107 LEU133 ASP134 ASP135 ASP136 PRO137
GLU138 ILE150 LEU151 PRO152 LEU154 PHE155 MET156 SER157
LEU170 PHE200 GLY201 ARG202 LYS205 PRO206 SER207 VAL208
GLU209 TRP211 PHE229 PRO246 PHE271 GLU273

Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.5541, pTM = 0.7971

Chai-1: NodD *R. leguminosarum* (Target) + Naringenin



Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O4_1	2.471
VAL 208.A N	LIG2 1.B O2_1	2.862
LIG2 1.B O5_1	PHE 200.A O	3.415

Chimera: Identified residues

H-bonds = 3

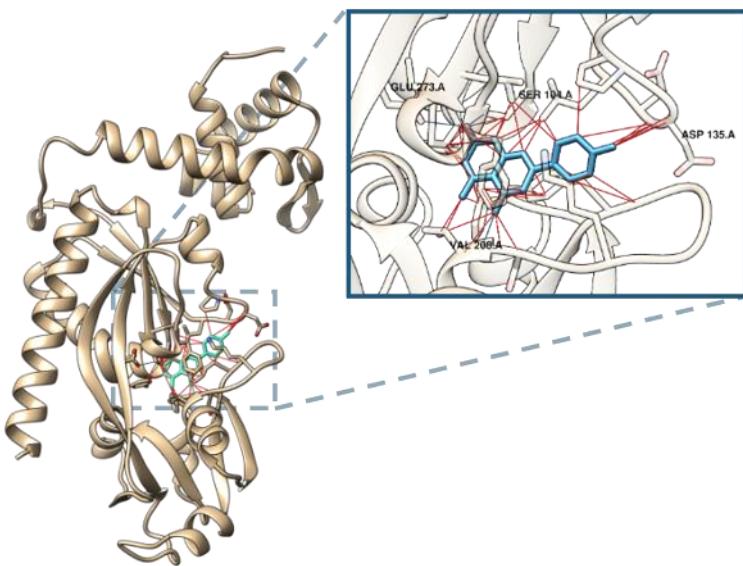
Other Contacts = 46

CB-Dock2: Vina = -8.9, Cavity volume = 486 Å³

LEU103 SER104 PHE106 MET107 LEU133 ASP134 ASP135 ASP136 PRO137 GLU138 GLU139 ILE150 LEU151 PRO152 LEU154 PHE155 MET156 LEU170 PHE200 GLY201 LYS205 PRO206 SER207 VAL208 GLU209 TRP211 PHE229 PRO246 PHE271 GLU273

Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.5542, pTM = 0.7941

Chai-1: NodD *R. leguminosarum* (Target) + Apigenin



Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O3_1	2.604
VAL 208.A N	LIG2 1.B O1_1	2.592
LIG2 1.B O3_1	SER 104.A OG*	2.604
LIG2 1.B O3_1	GLU 273.A OE2	2.976
LIG2 1.B O5_1	ASP 134.A O	1.976

Chimera: Identified residues

H-bonds = 4

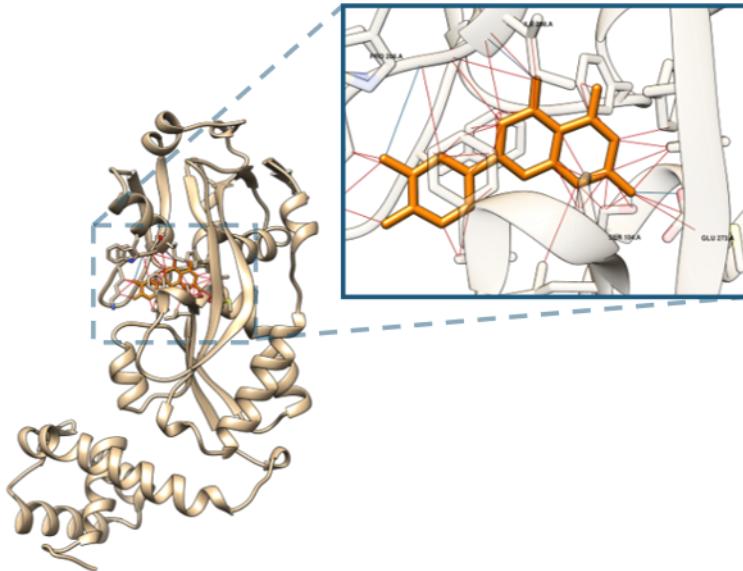
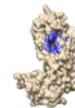
Other Contacts = 58

CB-Dock2: Vina = -8.8, Cavity volume = 486 Å³

SER104 PHE106 MET107 LEU133 ASP134 ASP135 ASP136 PRO137 GLU138 GLU139 LEU149 ILE150 LEU151 PRO152 LEU154 PHE155 MET156 LEU170 PHE200 GLY201 LYS205 PRO206 SER207 VAL208 GLU209 TRP211 PHE229 PRO246 PHE271 GLU273

Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.6182, pTM = 0.7961

Chai-1: NodD *R. leguminosarum* bv. *trifolii* + Luteolin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.7343, pTM = 0.8297

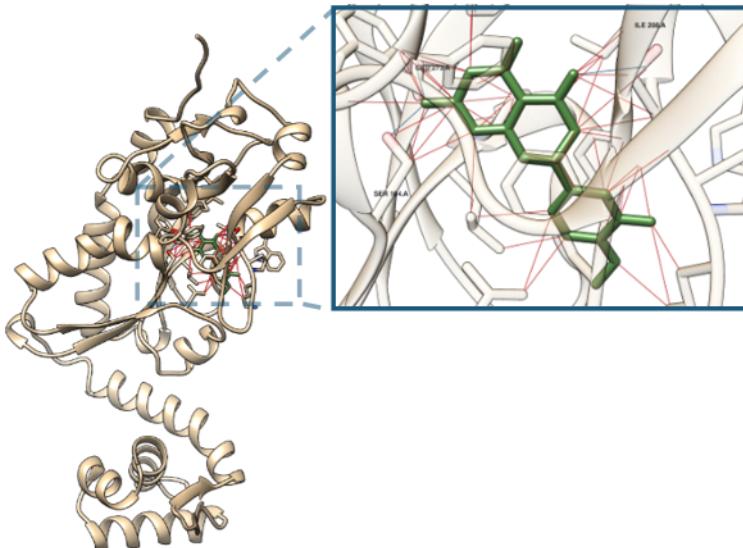
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O3_1	2.645
ILE 208.A N	LIG2 1.B O1_1	2.836
LIG2 1.B O3_1	GLU 273.A OE2	2.690
LIG2 1.B O5_1	PRO 206.A O	3.386

Chimera: Identified residues
H-bonds = 4
Other Contacts = 50

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -6.9, Cavity volume = 272 Å³
VAL133 ASN134 ASP135 ASP136 PRO137 ASP138 GLU139 ARG142
LEU149 LEU151 PRO152 ASP153 GLN154 PHE155 MET156 SER157
PHE200 LYS205 PRO206 SER207 ILE208 TRP211 GLU215 GLN276

Chai-1: NodD *R. leguminosarum* bv. *trifolii* + Hesperetin



Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O4_1	2.760
ILE 208.A N	LIG2 1.B O2_1	2.793
LIG2 1.B O4_1	GLU 273.A OE2	2.529

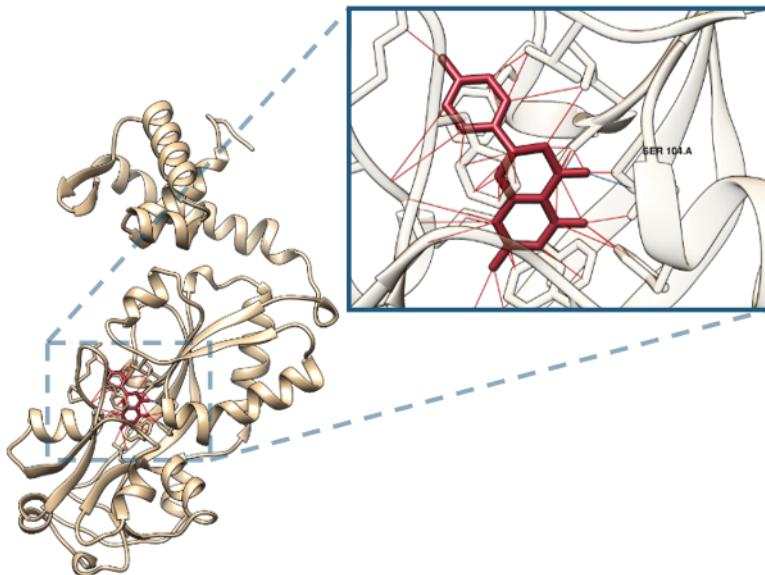
Chimera: Identified residues
H-bonds = 3
Other Contacts = 58

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -6.9, Cavity volume = 272 Å³
VAL133 ASN134 ASP135 ASP136 PRO137 ASP138 LEU149 LEU151
PRO152 GLN154 PHE155 MET156 SER157 ALA158 THR159 PHE200
LYS205 PRO206 SER207 TRP211 LEU214 GLU215

Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.7383, pTM = 0.8326

Chai-1: NodD *R. leguminosarum* bv. trifolii + Naringenin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.4381, pTM = 0.8067

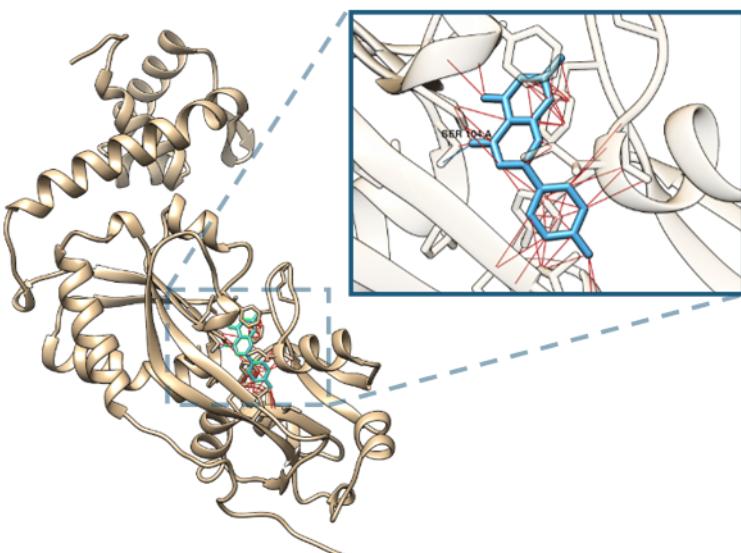
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O2_1	3.305

Chimera: Identified residues
H-bonds = 1
Other Contacts = 40

1 residue also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -6.8, Cavity volume = 272 Å³
VAL133 ASN134 ASP135 ASP136 PRO137 ASP138 LEU149 LEU151
PRO152 GLN154 PHE155 MET156 SER157 ALA158 PHE200 LYS205
PRO206 SER207 ILE208 TRP211 LEU214 GLU215 GLN276

Chai-1: NodD *R. leguminosarum* bv. trifolii + Apigenin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.4112, pTM = 0.8185

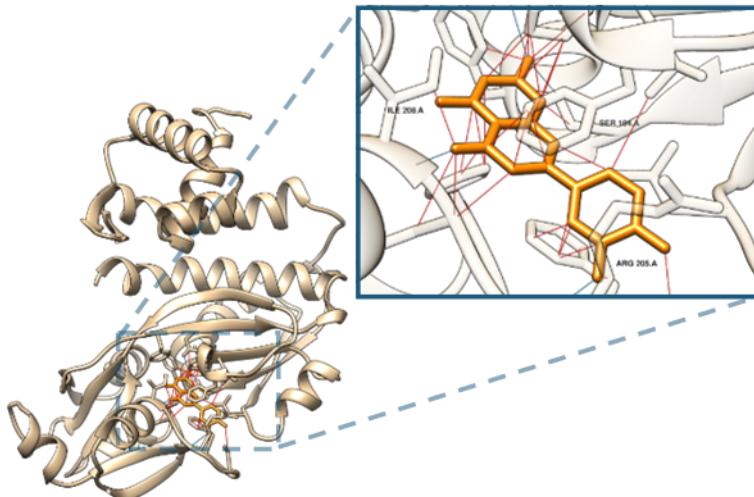
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O1_1	3.383

Chimera: Identified residues
H-bonds = 1
Other Contacts = 50

1 residue also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -7.0, Cavity volume = 272 Å³
SER104 VAL133 ASN134 ASP135 ASP136 PRO137 ASP138 GLU139
LEU149 LEU151 PRO152 GLN154 PHE155 MET156 SER157 PHE200
THR203 LYS205 PRO206 SER207 ILE208 TRP211 PHE229

Chai-1: NodD1 *R. tropici* + Luteolin



Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O3_1	2.941
ILE 208.A N	LIG2 1.B O1_1	3.104
LIG2 1.B O3_1	SER 104.A OG*	2.941
LIG2 1.B O3_1	GLU 273.A OE2	2.825
LIG2 1.B O5_1	ARG 205.A O	3.103

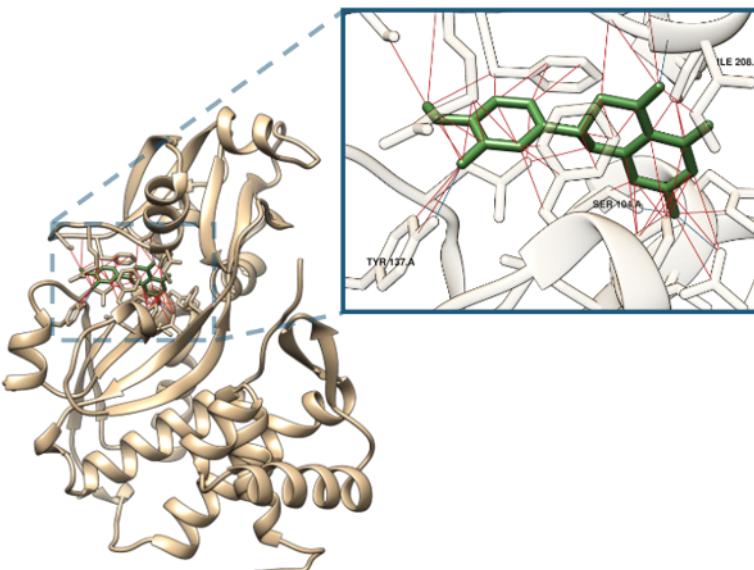
Chimera: Identified residues
H-bonds = 4
Other Contacts = 35

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -8.1, Cavity volume = 368 Å³
LEU103 SER104 PHE106 LEU107 LEU133 ASP135 TYR137 ASP138
ILE150 LEU151 PRO152 VAL154 PHE155 MET156 PRO157 LEU170
PHE200 SER203 ARG204 ARG205 PRO206 SER207 ILE208 GLU209
TRP211 LEU214 GLU215 PHE229 PRO246 PHE271 GLU273

Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.7314, pTM = 0.7940

Chai-1: NodD1 *R. tropici* + Hesperetin



Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O4_1	2.984
TYR 137.A OH	LIG2 1.B O6_1	2.769
ILE 208.A N	LIG2 1.B O2_1	3.044
LIG2 1.B O4_1	SER 104.A OG*	2.984
LIG2 1.B O4_1	GLU 273.A OE2	2.776

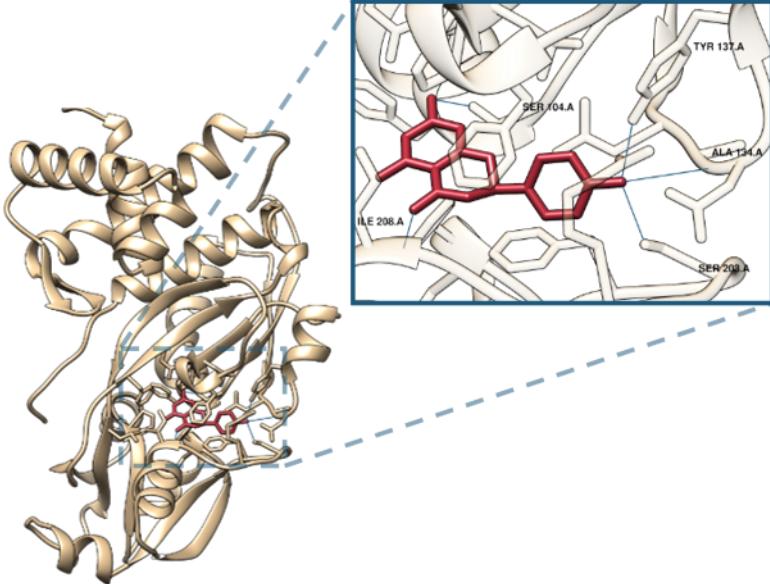
Chimera: Identified residues
H-bonds = 4
Other Contacts = 52

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -7.9, Cavity volume = 368 Å³
LEU103 SER104 PHE106 LEU107 LEU133 ASP135 TYR137 ASP138
ILE150 LEU151 PRO152 VAL154 PHE155 MET156 PRO157 THR158
LEU170 PHE200 SER203 ARG204 ARG205 PRO206 SER207 ILE208
GLU209 TRP211 LEU214 GLU215 PHE229 PRO246 PHE271 GLU273

Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.7100, pTM = 0.7895

Chai-1: NodD1 *R. tropici* + Naringenin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.7168, pTM = 0.7836

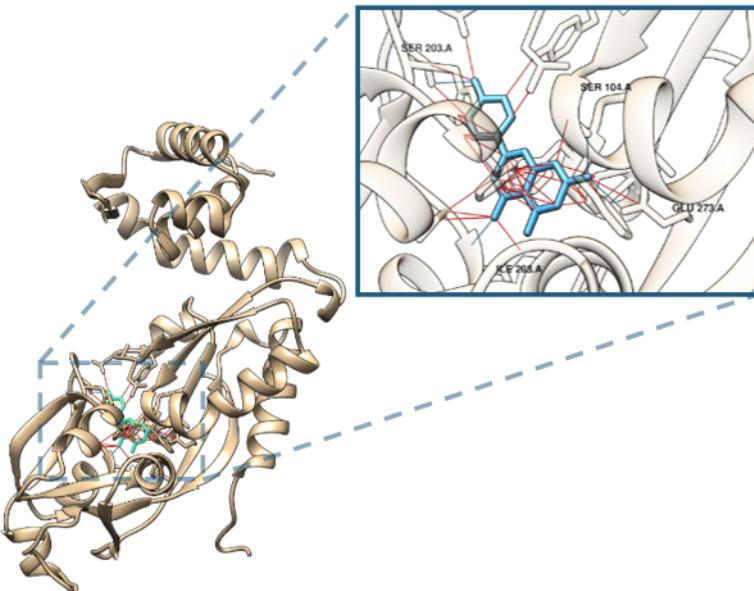
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O4_1	2.891
TYR 137.A OH	LIG2 1.B O5_1	3.356
SER 203.A OG	LIG2 1.B O5_1	2.746
ILE 208.A N	LIG2 1.B O2_1	3.004
LIG2 1.B O4_1	SER 104.A OG*	2.891
LIG2 1.B O5_1	ALA 134.A O	3.240
LIG2 1.B O5_1	TYR 137.A OH*	3.356
LIG2 1.B O5_1	SER 203.A OG*	2.746

Chimera: Identified residues
H-bonds = 5
Other Contacts = 54

1 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -8.0, Cavity volume = 368 Å³
LEU103 SER104 PHE106 LEU107 LEU133 ASP135 TYR137 ASP138
ILE150 LEU151 PRO152 VAL154 PHE155 MET156 PRO157 THR158
LEU170 PHE200 SER203 ARG204 ARG205 PRO206 SER207 ILE208
GLU209 TRP211 LEU214 GLU215 PHE229 PRO246 PHE271 GLU273

Chai-1: NodD1 *R. tropici* + Apigenin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.7171, pTM = 0.7699

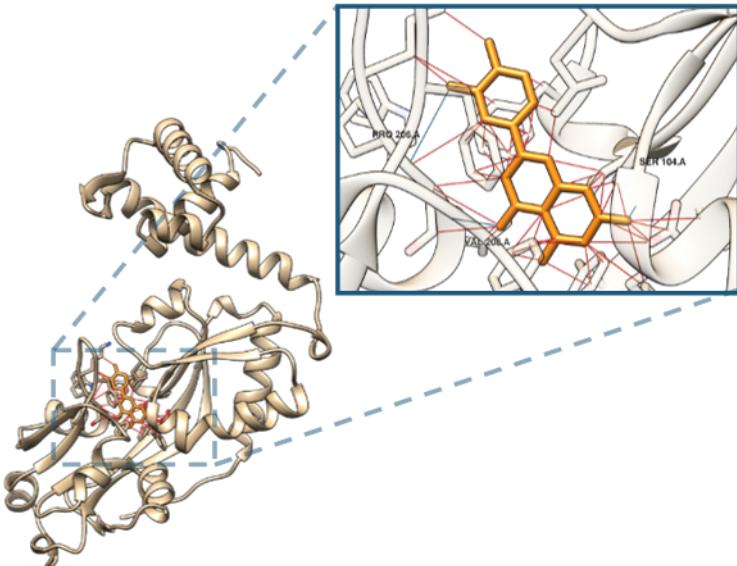
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O3_1	2.996
SER 203.A OG	LIG2 1.B O5_1	2.787
ILE 208.A N	LIG2 1.B O1_1	3.063
LIG2 1.B O3_1	SER 104.A OG*	2.996
LIG2 1.B O3_1	GLU 273.A OE2	2.812
LIG2 1.B O5_1	SER 203.A OG*	2.787

Chimera: Identified residues
H-bonds = 4
Other Contacts = 44

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -6.3, Cavity volume = 368 Å³
SER104 LEU133 TYR137 ASP138 LEU151 PRO152 VAL154 PHE155
MET156 PRO157 THR158 PHE200 SER203 ARG204 ARG205 PRO206
SER207 ILE208 TRP211 GLU215 PHE229

Chai-1: NodD1 *E. meliloti* + Luteolin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.7118, pTM = 0.8197

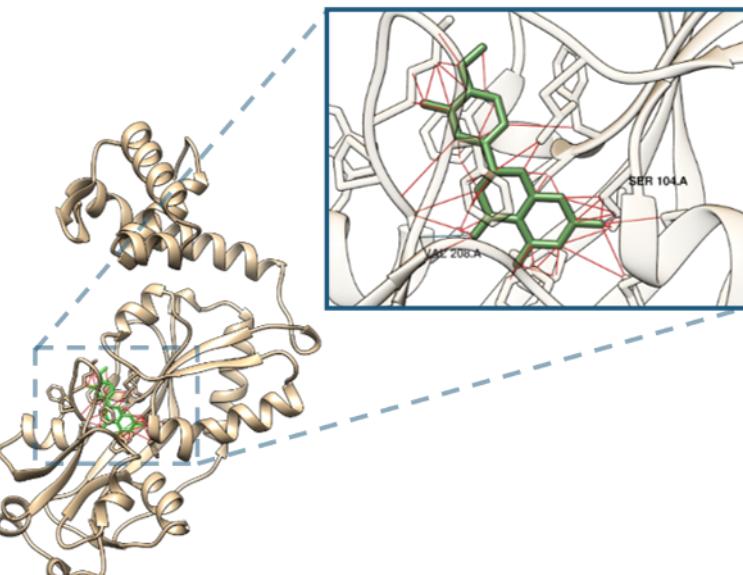
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O3_1	2.657
VAL 208.A N	LIG2 1.B O1_1	2.818
LIG2 1.B O3_1	SER 104.A OG*	2.657
LIG2 1.B O5_1	PRO 206.A O	3.070

Chimera: Identified residues
H-bonds = 3
Other Contacts = 58

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -7.6, Cavity volume = 308 Å³
LEU103 SER104 PHE106 LEU133 ASP134 ASP135 ASP136 PRO137 HIS138 GLU139 ILE150 PHE151 PRO152 VAL154 PHE155 SER157 PHE200 LYS205 PRO206 SER207 VAL208 TRP211 GLU215 PHE229 GLU273

Chai-1: NodD1 *E. meliloti* + Hesperetin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.7108, pTM = 0.8245

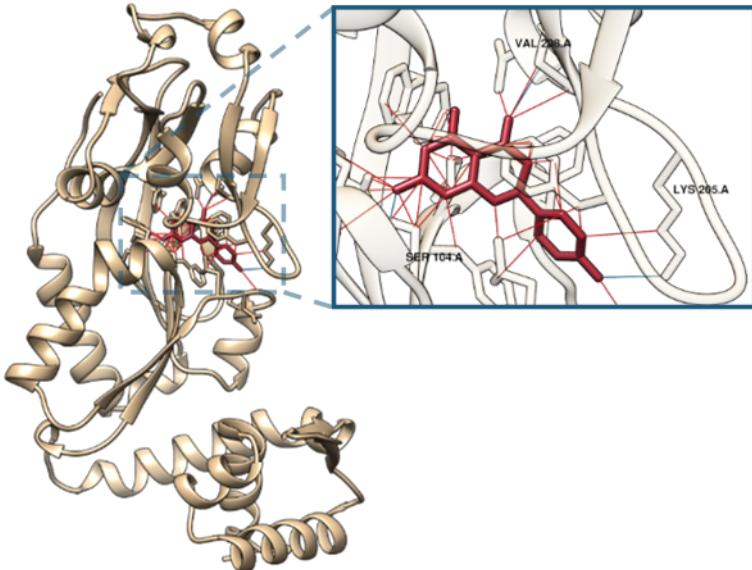
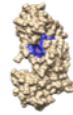
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O4_1	2.493
VAL 208.A N	LIG2 1.B O2_1	2.762
LIG2 1.B O4_1	SER 104.A OG*	2.493

Chimera: Identified residues
H-bonds = 2
Other Contacts = 60

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -8.2, Cavity volume = 308 Å³
LEU103 SER104 PHE106 MET107 LEU133 ASP134 ASP135 ASP136 PRO137 HIS138 ILE150 PHE151 PRO152 VAL154 PHE155 LEU170 PHE200 GLY201 LYS205 PRO206 SER207 VAL208 GLU209 TRP211 LEU214 GLU215 PHE229 PRO246 PHE271 GLU273

Chai-1: NodD1 *E. meliloti* + Naringenin



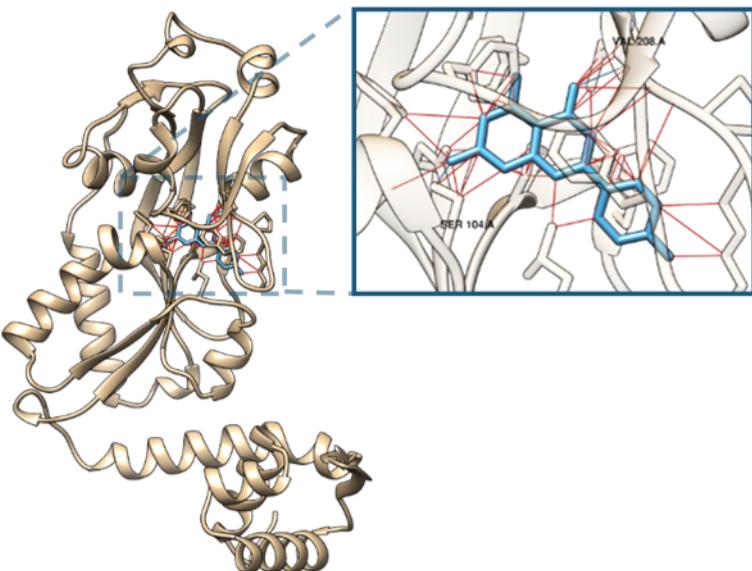
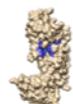
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O4_1	2.602
LYS 205.A NZ	LIG2 1.B O5_1	3.377
VAL 208.A N	LIG2 1.B O2_1	3.043
LIG2 1.B O4_1	SER 104.A OG*	2.602

Chimera: Identified residues
H-bonds = 3
Other Contacts = 45

CB-Dock2: Vina = -7.3, Cavity volume = 308 Å³
SER104 LEU133 ASP134 ASP135 ASP136 PRO137 HIS138 GLU139
PHE151 PRO152 VAL154 PHE155 PHE200 LYS205 PRO206 SER207
VAL208 TRP211 LEU214 GLU215 PHE229 GLU273

Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.6290, pTM = 0.8088

Chai-1: NodD1 *E. meliloti* + Apigenin



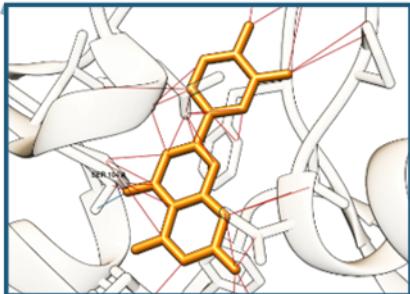
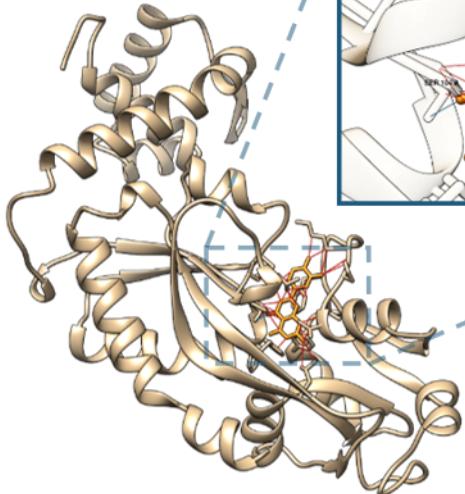
Donor	Acceptor	Distance
SER 104.A OG	LIG2.1.B O3_1	2.535
VAL 208.A N	LIG2.1.B O1_1	2.701

Chimera: Identified residues
H-bonds = 2
Other Contacts = 60

CB-Dock2: Vina = -7.8, Cavity volume = 308 Å³
LEU103 SER104 PHE106 MET107 LEU133 ASP134 ASP135 ASP136
PRO137 HIS138 ILE150 PHE151 PRO152 VAL154 PHE155 SER157
PHE200 LYS205 PRO206 SER207 VAL208 TRP211 GLU215 PHE229
PHE271 GLU273

Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.6921, pTM = 0.8176

Chai-1: NodD1 *N. galegae* + Luteolin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.4589, pTM = 0.7898

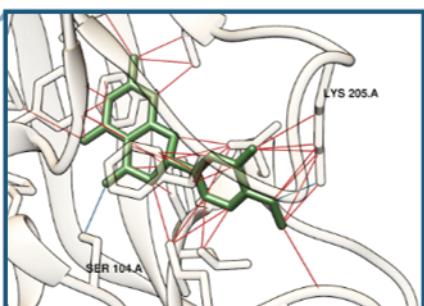
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O1_1	3.347

Chimera: Identified residues
H-bonds = 1
Other Contacts = 32

1 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -6.4, Cavity volume = 611 Å³
ILE102 LEU133 ASP134 ASP135 ASP136 PRO137 ASP138 LEU151
PRO152 LEU154 LEU155 MET156 PRO157 PHE200 LYS205 PRO206
SER207 VAL208 TRP211 GLU215

Chai-1: NodD1 *N. galegae* + Hesperetin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.4497, pTM = 0.7905

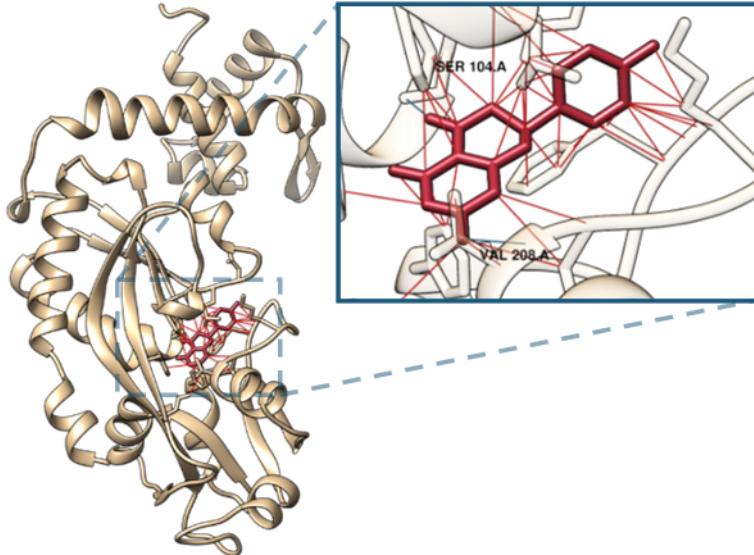
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O2_1	3.188
LYS 205.A NZ	LIG2 1.B O1_1	2.797

Chimera: Identified residues
H-bonds = 2
Other Contacts = 42

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -6.4, Cavity volume = 611 Å³
ILE102 LEU133 ASP134 ASP135 ASP136 PRO137 ASP138 LEU151
PRO152 LEU154 LEU155 MET156 PRO157 ILE159 PHE200 LYS205
PRO206 SER207 VAL208 TRP211 LEU214 GLU215

Chai-1: NodD1 *N. galegae* + Naringenin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.4402, pTM = 0.7929

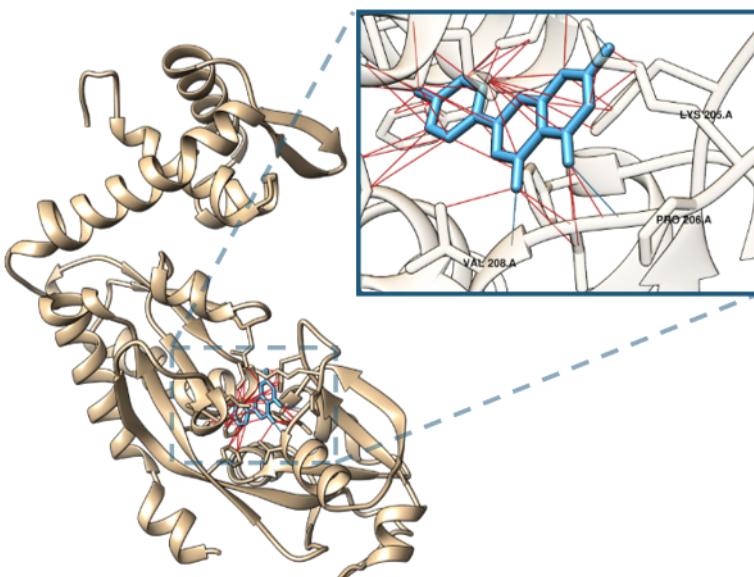
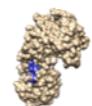
Donor	Acceptor	Distance
SER 104.A OG	LIG2 1.B O2_1	3.428
VAL 208.A N	LIG2 1.B O4_1	2.974

Chimera: Identified residues
H-bonds = 2
Other Contacts = 47

2 residues also involved in *R. leg* NodD (target)

CB-Dock2: Vina = -7.5, Cavity volume = 611 Å³
ILE102 LEU103 SER104 PHE106 MET107 LEU133 ASP134 ASP135
ASP136 PRO137 ASP138 GLU139 ILE150 LEU151 PRO152 LEU154
LEU155 PRO157 GLU168 LEU170 PHE200 LYS205 PRO206 SER207
VAL208 GLU209 TRP211 PHE229 PRO246 PHE271 LEU273

Chai-1: NodD1 *N. galegae* + Apigenin



Predicted with Chai-1 (MSAs enabled: JACKHMMER)
iPTM = 0.4408, pTM = 0.7896

Donor	Acceptor	Distance
LYS 205.A NZ	LIG2 1.B O3_1	3.557
VAL 208.A N	LIG2 1.B O1_1	2.953
LIG2 1.B O4_1	PRO 206.A O	3.116

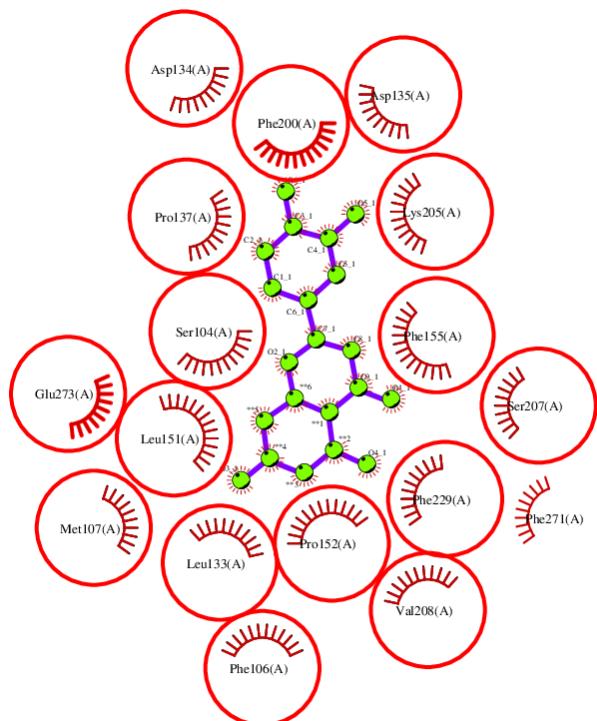
Chimera: Identified residues
H-bonds = 3
Other Contacts = 48

1 residues also involved in *R. leg* NodD (target)

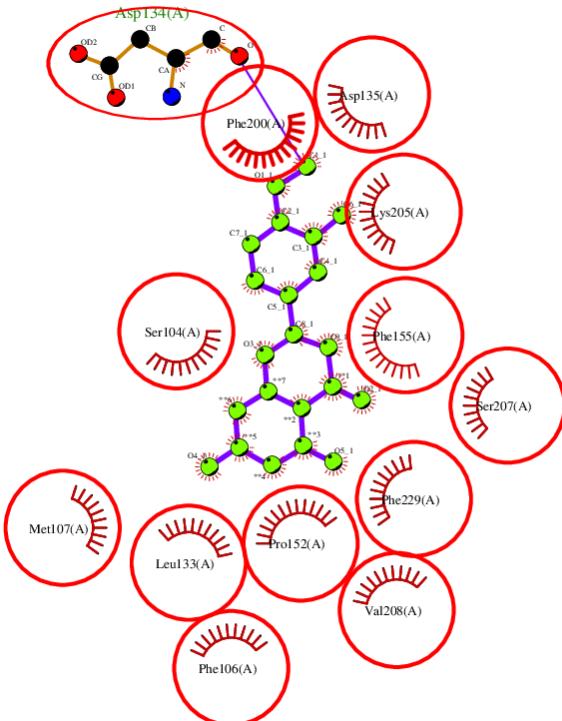
CB-Dock2: Vina = -6.4, Cavity volume = 611 Å³
ILE102 SER104 PHE106 MET107 LEU133 ASP134 ASP135 ASP136
PRO137 ASP138 GLU139 ILE150 LEU151 PRO152 LEU154 LEU155
MET156 PRO157 PHE200 LEU203 MET204 LYS205 PRO206 SER207
VAL208 TRP211 GLU215 PHE229 LEU273

9.3 LigPlot+ Analysis

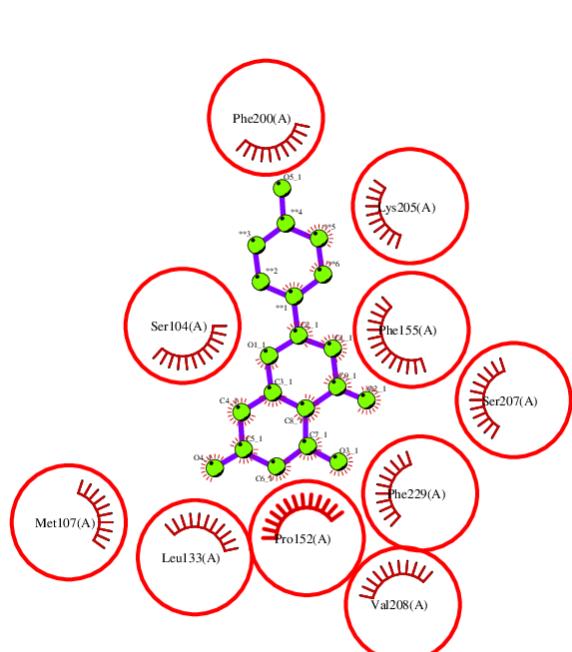
NodD from *Rhizobium leguminosarum* bv. *viciae* 3841



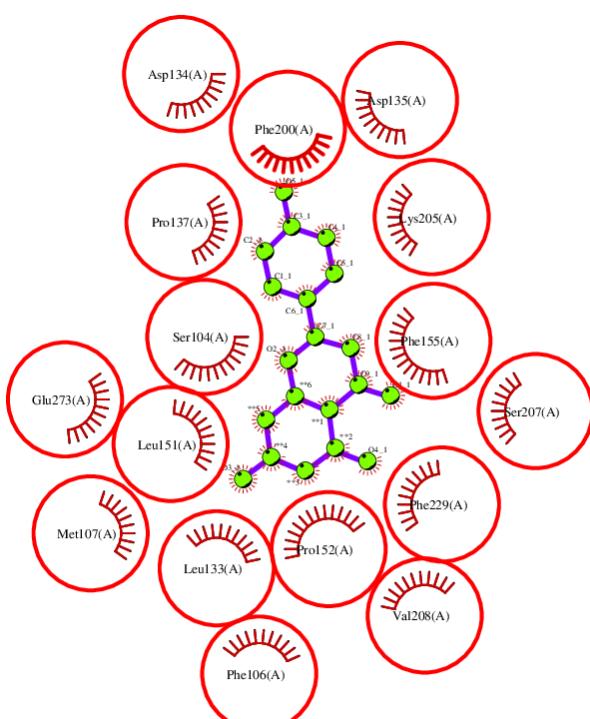
Luteolin



Hesperetin

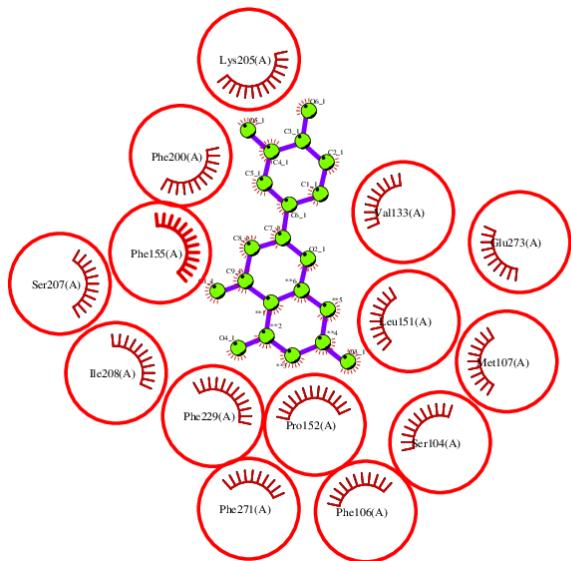


Naringenin

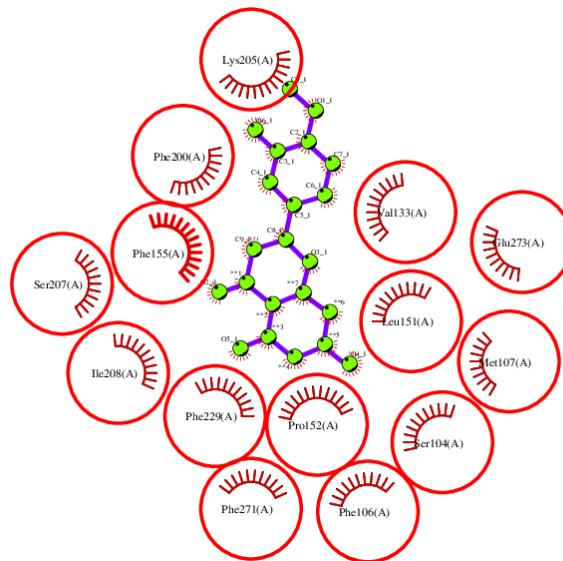


Apigenin

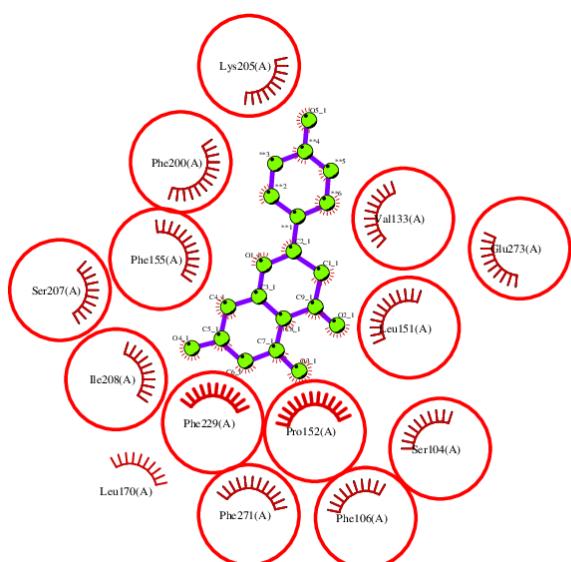
NodD from *Rhizobium leguminosarum* bv. *trifolii*



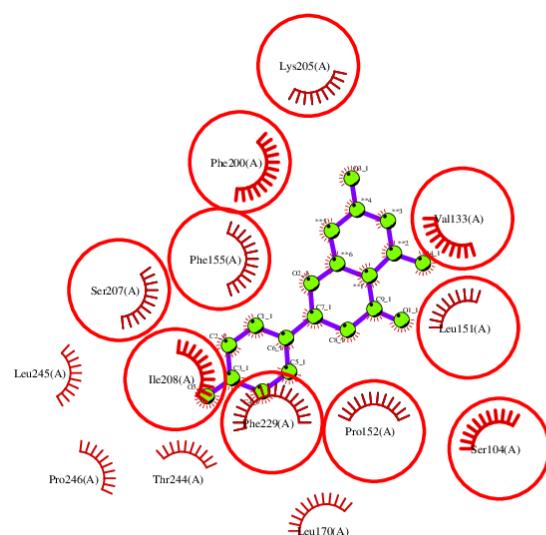
Luteolin



Hesperetin

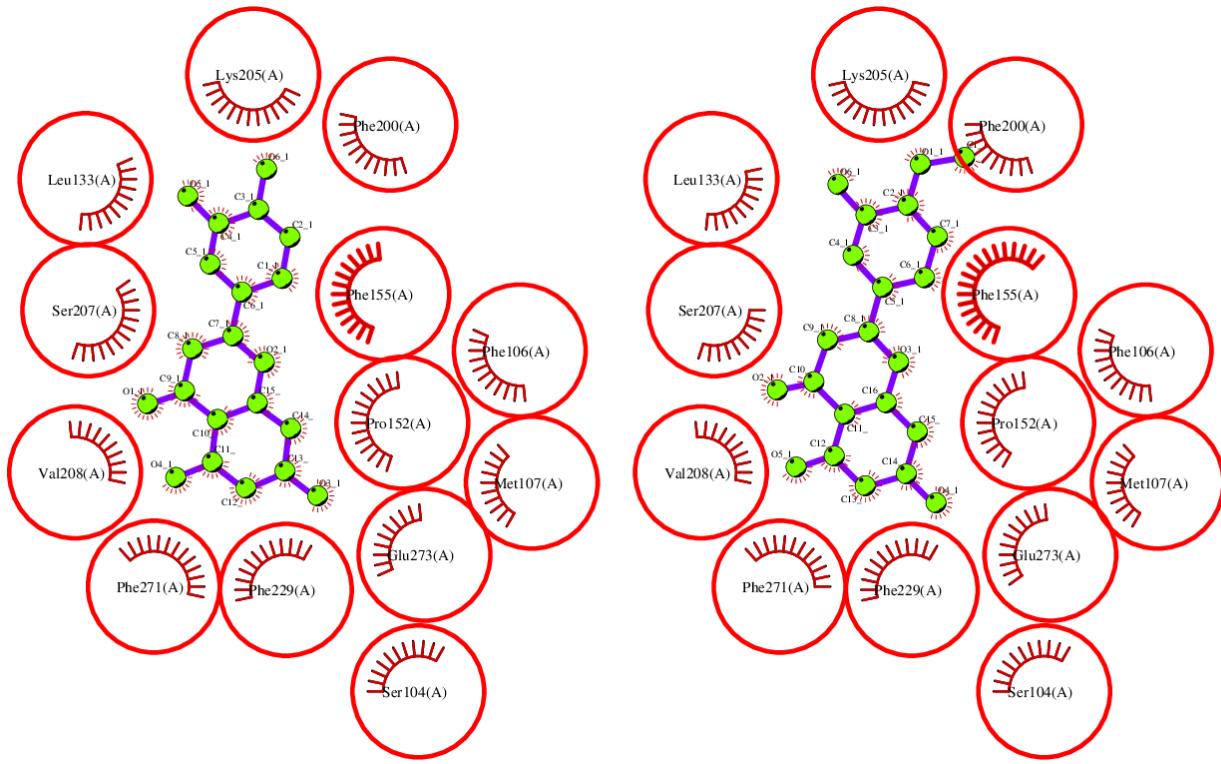


Naringenin



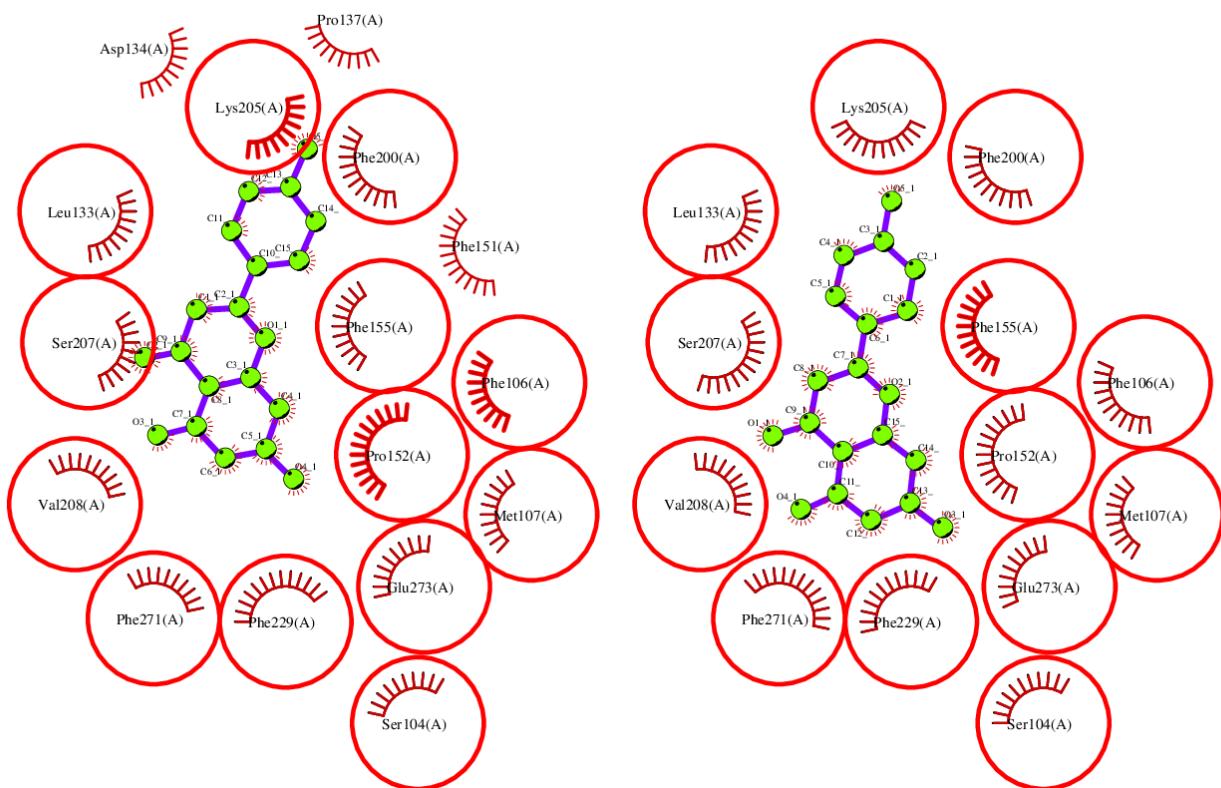
Apigenin

NodD1 from *Sinorhizobium meliloti*



Luteolin

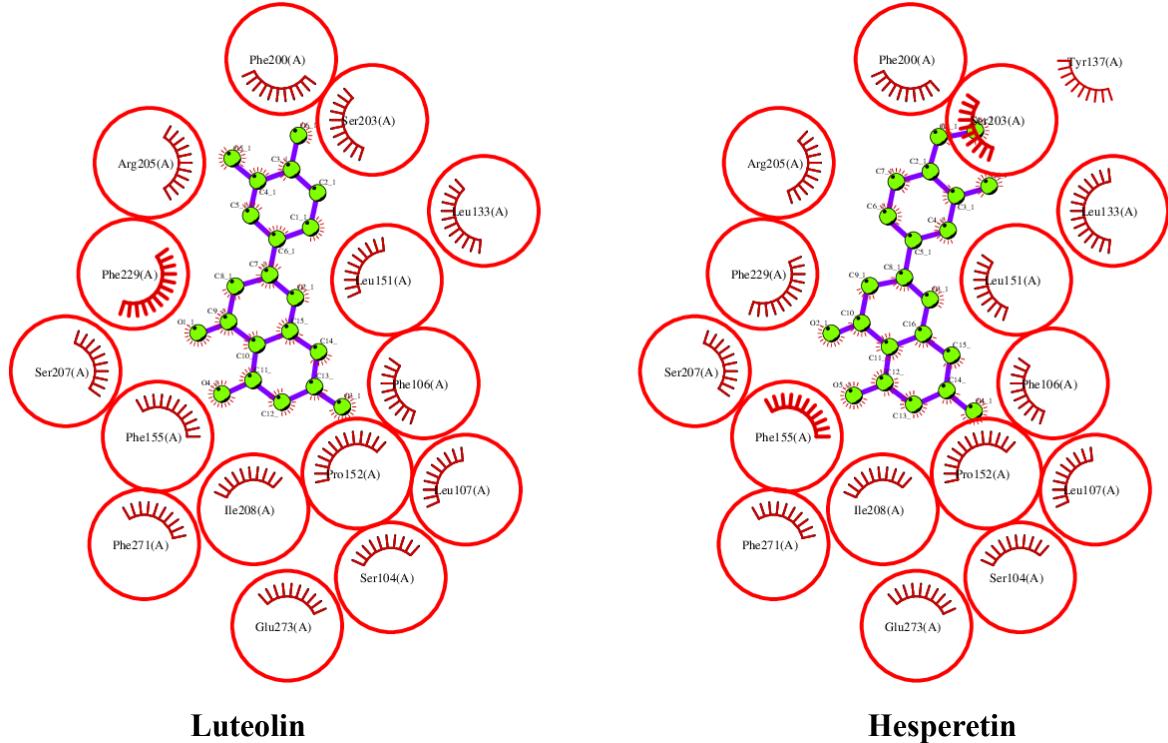
Hesperetin



Naringenin

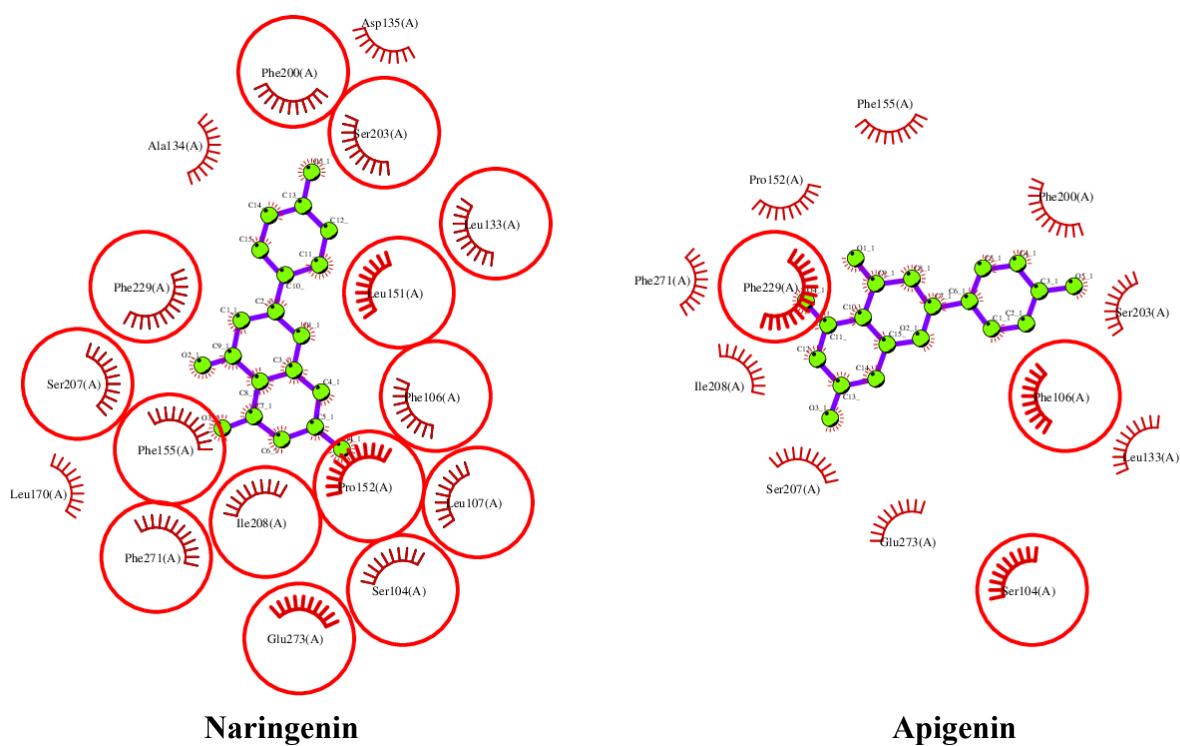
Apigenin

NodD1 from *Rhizobium tropici*



Luteolin

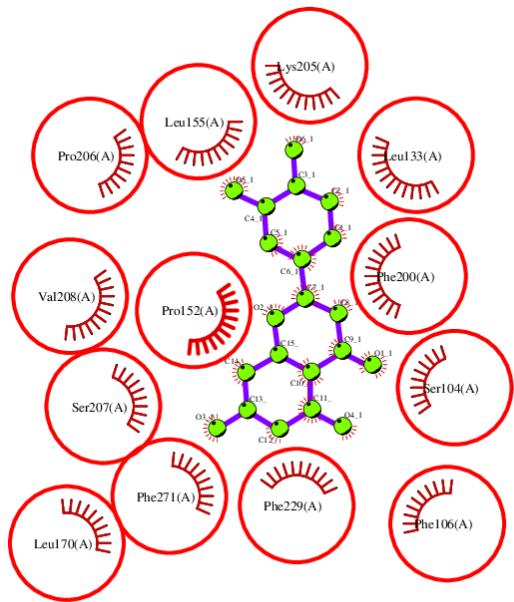
Hesperetin



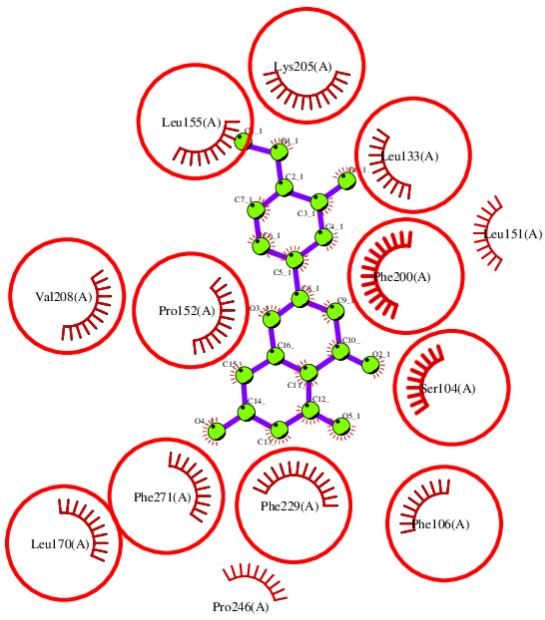
Naringenin

Apigenin

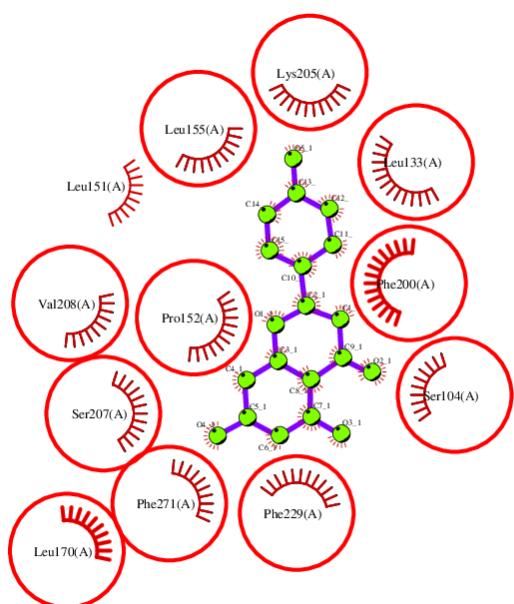
NodD1 from *Neorhizobium galegae*



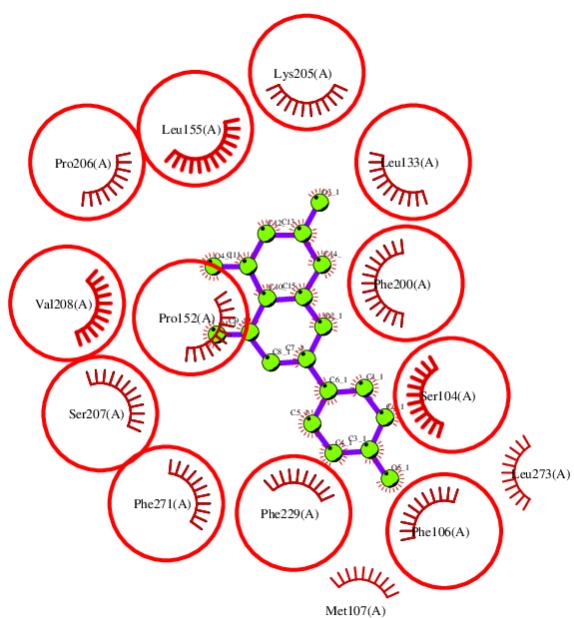
Luteolin



Hesperetin



Naringenin



Apigenin