Computational analyses of the components of Sinorhizobium meliloti ExoR-ExoS/ChvI pathway: the ExoR and ExoS proteins

Eliza M. Wiech
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds
Part of the Bioinformatics Commons, and the Biology Commons

Recommended Citation

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
COMPUTATIONAL ANALYSES OF THE COMPONENTS OF

Sinorhizobium meliloti ExoR-ExoS/ChvI PATHWAY:

THE ExoR AND ExoS PROTEINS

by

ELIZA M. WIECH

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2014
This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. SHANEEN M. SINGH

4/4/2014
Date
Chair of Examining Committee

Dr. LAUREL A. ECKHARDT

4/22/2014
Date
Executive Officer

Dr. JUERGEN POLLE
Dr. THEODORE MUTH
Dr. HAI-PING CHENG
Dr. PETER KAHN

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract


by

Eliza M. Wiech

Adviser: Professor Shaneen M. Singh

The Sinorhizobium meliloti periplasmic ExoR protein and the ExoS/ChvI two-component system form a regulatory mechanism that directly controls the transformation of free-living to host-invading cells. In the absence of crystal structures, understanding the molecular mechanism of interaction between ExoR and the ExoS sensor, which is thought to drive the key regulatory step in the invasion process, remains a major challenge. In this study, we present theoretical structural models of the active form of ExoR protein, ExoR\textsubscript{m}, as well as of the sensing domain of ExoS, ExoS\textsubscript{p}, generated using computational methods. Our model suggests that ExoR\textsubscript{m} possesses a super-helical fold comprising twelve \(\alpha\)-helices forming six Sel1-like repeats, including two that were unidentified in previous studies. The structural model of ExoS\textsubscript{p} suggests that ExoS\textsubscript{p} is a single Per-ARNT-Sim (PAS) domain. Docking analysis was used to suggest models for ExoS\textsubscript{p}-ExoS\textsubscript{p} and ExoS\textsubscript{p}-ExoR\textsubscript{m} protein interactions and interfaces. Our studies reveal three novel insights: (a) a possible extended conformation of the ExoR third Sel1-like repeat that might be important for ExoR regulatory function (b) a buried proteolytic site that implies a unique mechanism of proteolysis, central to controlling ExoR function and (c) an elongated structure of helix H4 that is unique to ExoS\textsubscript{p} and might be crucial for the association with ExoR\textsubscript{m}. This
study provides new and interesting insights into the structure of the *S. meliloti* ExoRm and ExoSp proteins, lays the groundwork for elaborating the molecular mechanism of ExoRm cleavage, ExoRm-ExoS interactions, and studies of ExoR homologs in other bacterial host interactions.
Table of Contents

Chapter I: 1-4  
Introduction

Chapter II: 5-11  
Material and methods

Chapter III: 12-32  
Molecular modeling and computational analyses of the Sinorhizobium meliloti Rm1021 ExoR protein

Chapter IV: 33-45  
Analyses of ExoR mutant proteins

Chapter V: 46-64  
Molecular modeling and computational analyses of the Sinorhizobium meliloti Rm1021 periplasmic domain of the ExoS protein

Bibliography 65-82
Lists of Tables

**Table III.1.** 13
Sell-like repeats detected within the ExoRm amino acid sequence.

**Table III.2.** 21
RMSD values for the superposition of ExoR repeats and their sequence identity (%).

**Table III.3.** 25
*In silico* prediction of protein-protein interfaces and hot spots in ExoR.

**Table IV.1.** 38
Comparison of changes in solvent-accessible surface area (SASA) of residues identified as candidate interface residues in ExoR wild type and ExoR mutants.
Lists of Figures

Figure III.1. 14
Secondary structure prediction of the ExoR protein.

Figure III.2. 15-16
Alignment of repeats found in the ExoR protein of
S. meliloti Rm1021.

Figure III.3. 19
Evaluation plots of the putative three-dimensional model
of ExoRm

Figure III.4. 20
Ribbon representation of the putative structure of the
S. meliloti Rm1021 ExoR mature protein (ExoRm, residues 31-268).

Figure III.5. 22
Secondary structure prediction of the ExoR3 repeat.

Figure III.6. 23-24
Putative functional sites mapped to the surface of
ExoRm protein.

Figure III.7. 27
Electrostatic potentials mapped to the molecular surface of
the ExoRm protein.

Figure IV.1. 33-35
Evaluation plots of the three-dimensional models of the ExoR
mutants.

Figure IV.2. 40
Electrostatic potentials mapped to the molecular surface
of the ExoR wild-type protein and the ExoR mutant proteins.
Figure IV.3.  
Orientation of Dipole moment vectors in the ExoR wild type and the ExoR mutant proteins.

Figure V.1.  
Schematic representation of the domains present in the ExoS protein.

Figure V.2.  
Evaluation plots of the three-dimensional model of the periplasmic domain of ExoS.

Figure V.3.  
Ribbon representation of the putative structure of the S. meliloti Rm1021 periplasmic domain of the ExoS protein.

Figure V.4.  
Structure-based multiple sequence alignment of the ExoS periplasmic domain and other PAS sensors.

Figure V.5.  
Orientation of the dipole moment vector mapped to the putative structural model of the sensing domain of ExoS.

Fig. V.6.  
Electrostatic potentials mapped to the molecular surface of the sensing domain of ExoS.

Figure V.7.  
Location of the putative protein-protein interaction sites within the sensing domain of ExoS.

Figure V.8.  
Ribbon diagram of the ExoS$_p$ dimer showing packing between helices H1, H3, and H6 of opposite monomers.

Figure V.9.  
ExoS$_p$-ExoR complex organization.
Chapter I: Introduction

Most bacteria rely on two-component signal transduction systems for detecting and adapting to changes in their environment [1, 2, 3, 4]. Parasitic and symbiotic bacteria rely on these two-component systems to sense the presence of host organisms and turn on appropriate signaling cascades [5, 6, 7, 8]. Two-component systems consist of a soluble cytoplasmic or membrane-bound sensor histidine kinase and a cytoplasmic response regulator [9,10,11].

A typical membrane-bound histidine kinase consists of an N-terminal extracellular sensing domain flanked by two transmembrane helices and a cytoplasmic region. The cytoplasmic region is composed of a HAMP domain (important in signal transduction), a dimerization domain, and a C-terminal catalytic (ATP/ADP-binding phosphotransfer) domain [9, 10, 12, 13, 14]. Environmental stimuli are detected either directly or indirectly by the N-terminal sensing domain. Upon transfer of the signal to the cytoplasmic domain, transphosphorylation of conserved histidine residues on the dimerization domains takes place. Then, the phosphoryl group is transferred to an aspartate residue in the regulatory domain of the response regulator. This in turn, controls the expression of target genes through its effector domain, usually a DNA-binding domain or an enzyme [11, 14].

A group of Gram-negative bacteria rely on the ExoR-ExoS/ChvI (RSI) signal transduction pathway to transit from a free-living to host-invading form [15]. ExoS and ChvI form a typical two-component system that consists of a membrane-integral histidine kinase, ExoS, and an associated cytoplasmic response regulator, ChvI [8, 16]. The ExoS/ChvI system is regulated via ExoR, a periplasmic regulatory protein [17, 18, 19, 20].

In Sinorhizobium meliloti, a nitrogen-fixing bacterium, it has been established that the ExoS/ChvI two-component regulatory system controls the transformation of free-living cells to host-invading cells which differentiate into nitrogen-fixing bacteroids inside the host cells [15, 17,
The ExoS/ChvI system remains inactive in free-living cells and allows them to produce flagella for motility. The ExoS/ChvI system becomes active in host invading cells and turns on the production of succinoglycan, which is required for invading the host through the production of the infection threads inside root hairs [15, 21]. In addition, the ExoS/ChvI system also controls the expression of various other genes related to adaptations to alternative living forms e.g. biofilms [17, 22].

The activities of the ExoS/ChvI system are thought to be regulated by ExoR through a direct interaction of ExoR and ExoS, the sensor of the ExoS/ChvI two-component system. Together, the three proteins, ExoR, ExoS, and ChvI form the RSI regulatory pathway [17, 18]. As proposed by Lu et al. [20], the current model for the RSI pathway suggests that ExoS/ChvI system is turned off when the periplasmic domain of ExoS is in a protein complex with the mature periplasmic form of ExoR, ExoR_m. In the ExoR_m-ExoS complex, ExoS acts as a phosphatase, and keeps ChvI, the response regulator, dephosphorylated and inactive. When the ExoR_m-ExoS interaction is disrupted through the proteolytic cleavage of ExoR_m, ExoS becomes an active kinase, and phosphorylates ChvI directly [8, 20]. Phosphorylated ChvI upregulates the expression of succinoglycan biosynthesis genes while repressing the flagellum biosynthesis genes allowing the cells to switch from free-living to host invading form [15, 20].

Since ExoR modulates the signaling state of ExoS [17, 18, 19], the level of the ExoR protein needs to be regulated carefully [20]. Experimental data suggests that the putative physical association between ExoR_m and ExoS proteins not only suppresses ExoS activity, but also stabilizes ExoR_m [18]. Although the trigger(s) for the disruption of stable ExoR_m-ExoS interaction is still unknown, the disruption of the complex releases ExoR_m and exposes it to proteolysis [18, 20]. Recent findings suggest that the site of proteolysis in ExoR is located between A80 and L81 of ExoR with the possibility of further
digestion that involves residues 84, 85, 86, and 87. As a consequence of the proteolysis, ExoR exists in three forms: the full-length precursor cytoplasmic ExoRcyt, the mature and active periplasmic ExoRm, and the inactive periplasmic ExoR\textsubscript{C20} [20]. The expression of the exoR gene is also under the control of the ExoS/ChvI system. Therefore, the loss of ExoR\textsubscript{m} protein in the periplasm leads to upregulation of exoR expression to replenish the ExoR\textsubscript{m} protein in the periplasm [19]. The combination of biosynthesis and proteolysis allows S. meliloti cells to maintain ExoR\textsubscript{m} levels at a fine and highly dynamic equilibrium to regulate the RSI pathway activity. This mechanism makes it possible for cells to respond to the presence of host or environmental signals which would reduce ExoR\textsubscript{m} levels and then quickly return to baseline by restoring the normal level of ExoR\textsubscript{m} in the absence of hosts [19, 20].

Phenotypes of the ExoR loss-of-function mutant ExoRL81A [20] and the ExoR reduced-function mutants ExoRG76C and ExoRS156Y [18] further strengthen this model of ExoR mediated regulation of the RSI pathway. The change of a highly conserved leucine (L81) to alanine at the proposed proteolytic site in ExoRL81A results in a dramatic reduction of both the levels and the activity of ExoR\textsubscript{m} and activation of ExoS, supporting the hypothesis that proteolysis is the key regulator of ExoR activity [20]. Similar loss of ExoR based suppression of ExoS occurs when the ExoR\textsubscript{m}-ExoS interaction is disrupted in the ExoRG76C and S156Y mutants leading to increased ExoS/ChvI activity and a phenotype of strong overproduction of succinoglycan [18].

ExoR belongs to a family of solenoid proteins with Sel1-like repeats [23] highly conserved in Rhizobia [18, 20, 24]. Even though ExoR has been established as a key regulator of the RSI pathway [17,18,19,20], the tertiary structure including structural details and associated functional aspects of this important protein remain unknown. Delineating the structure-function correlations of the ExoR protein is critical to understanding of the cleavage
mechanism of ExoR$_m$, ExoR$_m$-ExoS interactions and other aspects of its regulatory role in the RSI pathway [18, 20].

In this study, we present a robust theoretical model of the active form of ExoR protein, ExoR$_m$, and a structural model of the periplasmic domain of ExoS, ExoS$_p$, generated by means of the well-established approach of template-based modeling combined with 	extit{ab initio} modeling. Comparative modeling methods have been successfully applied in revealing key structural details, describing structure-function relationships, and modeling interactions with putative binding partners. These studies include those that belong to the family of helical repeat proteins such as modeling the repeats of proteasome-binding protein PA200 [25] and the repeats of TAL (transcriptional activator-like) effectors [26] confirmed by subsequent crystal structures of TAL effector-DNA complexes [27, 28]. We propose that the ExoR protein has an alpha-alpha super-helical fold that is known to be conducive to protein-protein interactions and has inherent flexibility built into it [29, 30, 31, 32, 33, 34]. We also propose that the periplasmic domain of ExoS takes on alpha-beta PAS fold. In the absence of structural representations of the ExoR and ExoS proteins in the structural database, our study provides a first glance into the proposed structural folds of ExoR and ExoS$_p$ and the implications of the identified folds for ExoR and ExoS functions.
Chapter II: Material and methods

ExoR protein sequence analyses

The amino acid sequence of the ExoR protein from S. meliloti Rm1021 was retrieved from the NCBI Protein database [35] (GenBank: AAA26260.1). The mature ExoR protein (238 amino acid residues long), ExoRm, i.e. without its 30-residue signal sequence [17], was used for modeling the protein as well as sequence and structure analyses.

SMART [36] was used to identify and validate boundaries of the four previously predicted Sel1-like repeats of ExoR [23]. In addition, the ExoRm sequence was analyzed with Pfam 25.0 [37], CD-Search [38] using Conserved Domain Database v3.04 [39, 40], and Prosite 20.76 [41]. To probe for any additional Sel1-like repeats previously unidentified within ExoRm, a profile-based repeat detection method, TPRpred was used [42], along with de novo repeat identification methods: a hidden Markov model (HMM) profile based HHrepID [43] and REPRO [44]. The boundaries of the identified repeats were manually assigned based on the data from the repeat/domain detection methods and the consensus of the output of a number of secondary structure prediction programs (detailed below) to ensure that each of the ExoR repeats encompasses the two α-helices characteristic of Sel1-like repeats [23, 31, 32].

The secondary structure prediction on the ExoR primary sequence was carried out using PROF v1.0 [45], PSIPRED v.3.0 [46], SAM-T08 [47], SSpro 4.0 [48], the PredictProtein server (PROFsec) [49], PSSpred v2 [50], YASPIN [51], a version of Jpred3 [52, 53, 54] that incorporates Jnet v2.2 [55], and YASSPP v1.0 [56] which was accessed through the MONSTER server [57].

The structure based sequence alignment of ExoR Sel1-like repeats was carried out using the TM-align algorithm [58]. The alignment of loop regions was manually refined by anchoring key conserved structural residues found in the Sel1-like repeat consensus sequence [23, 32]. To validate the presence of the last repeat and to determine its unique consensus sequence, sequences of
putative ExoR orthologs were retrieved using NCBI-BLASTP against the non-redundant protein sequence database [59, 60]. Redundant sequences and sequences with large inserts in this region were removed from the dataset. The sequences were aligned using the multiple sequence alignment program, T-COFFEE v. 9.01 [61].

**ExoS protein sequence analyses**

The amino acid sequence of the *S. meliloti* Rm1021 ExoS protein (595 residues) was retrieved from the *S. meliloti* 1021 database [https://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgi] (Acc CAC41430.1). The TMHMM server v2.0 [62,63] was used to identify transmembrane regions and the boundaries of the periplasmic sensing domain of ExoS (residues 68-278). The sequence of the excised periplasmic sensing domain of ExoS (211 residues) was used for modeling. The secondary structure prediction of the primary sequence of the ExoS periplasmic domain was carried out using PSIPRED v.3.3 [46], the PredictProtein server (PROFsec) [49], and PSSpred v2 [50]. The structure-based sequence alignment of the ExoS sensing domain and PAS domains of histidine kinases was carried out using Promals3D [64]. Single- and double-PAS sensor structures and sequences were retrieved from the Protein Data Bank (PDB) [65].

**Modeling methodology**

The three-dimensional models of ExoR<sub>m</sub>, ExoR mutant proteins and ExoS<sub>p</sub> were generated using comparative modeling methodology, which is widely used and the current state of the art in protein structure prediction [66, 67, 68, 69].

**a. Modeling methodology of ExoR<sub>m</sub>**

The fold of the ExoR protein was identified using the sequence
similarity based algorithm, BLAST [59, 60] limited to searches against the structural database (PDB) [65], as well as the sequence-structure threading methods, SPARKS-X [70], 3D-PSSM v.2.6.0 [71], HHpred [72, 73], LOOPP [74, 75], pGenTHREADER [76], and FUGUE [77].

To generate high-quality representations of the three-dimensional structure of ExoRm, a large number of theoretical models were generated based on various alternative alignments for each structural template and also multi-template modeling regimes. MODELLER v9.8 and v9.9 [78] was used to build models with alternative alignments generated by other programs: (1) for pairwise alignments: T-COFFEE version_9.02 [61], MAFFT v6.864 [79], and FUGUE [77] (2) for multi-sequence alignments, HHpred [72, 73]. For automatic alignment and model building of the ExoRm structure, automated homology modeling servers were also explored: 3D-JIGSAW [80], SWISS MODEL [81], (PS)²v2 [82], PHYRE [83], LOOPP [74, 75], I-TASSER [84, 85], and Robetta [86]. Some of the template based modeling algorithms, I-TASSER [84, 85] and Robetta [86], include ab initio techniques as needed to model the proteins’ structures to atomic detail. The prediction of side-chain conformations was implemented through the program SCWRL4 [87, 88, 89] via the AS2TS system [90].

b. Modeling methodology of the ExoR mutant proteins

We introduced point mutations in the ExoRm wild type sequence by manual replacement of the target residues. The best ExoRm structural model was generated with the I-Tasser server [84, 85]; therefore, to stay consistent in methodology and build comparable three-dimensional models of the ExoR mutants, their structures were generated with the I-Tasser server [84, 85] as well.
c. Modeling methodology of ExoS

The fold of the ExoS protein was identified using the sequence-structure threading methods, FFAS [91], FUGUE [77], HHpred [72, 73], SPARKS-X [70], and pGenTHREADER [76]. Several structural representations of the ExoS sensing domain were generated based on various alternative alignments. MODELLER v9.9 [78] was used to build models with alternative alignments generated by FUGUE [77] and HHpred [72, 73]. In addition, LOMETS, the Local Meta-Threading Server [92], generated full-length models of the ExoS periplasmic domain using MODELLER [78] based on alignments from a number of different threading programs (FUGUE, HHsearch, MUSTER, PROSPECT2, PPA-I, SAM, SP3, SPARKS, FFAS, and PRC). Automated homology modeling servers such as (PS)²v2 [82] and I-TASSER [84, 85] were also employed. The prediction of side-chain conformations was implemented through the program SCWRL4 [87, 88, 89] via the AS2TS system [90].

The generated models of ExoR, the ExoR mutants, and ExoS were evaluated using ProSA-web [93, 94] and Verify3D [95, 96]. ProSA-web determines the reliability of a structural model based on the model’s energy profiles and Verify3D evaluates the model by comparing the environment of residues found in the modeled structure with their observed propensities for being in those environments. To select final models of ExoR and ExoS, the best-evaluating models were checked for anomalies in stereochemical properties with WHAT_CHECK [97].

Analyses of three-dimensional models of ExoR, ExoR mutants, and ExoS

The visualization and analyses of the shape, structural alignments, solvent-accessible molecular surfaces, putative protein-protein interaction sites and hot-spots residues, and the surface electrostatic profile of the generated models were performed by using the surface property analyses tools
in PyMOL [98]. All cartoons and diagrams were constructed using PyMOL [98] and Prosite: MyDomains-Image Creator [99].

**a. Analysis of the structural conservation of ExoRm repeats**

To examine the conformation of loops and helices in the generated models, the individual repeats were structurally aligned using the TM-align algorithm [58]. The structural conservation of the repeats was further probed by computing helical packing angles of the ExoRm models using the PyMOL script to calculate angles between helices (method: helix orientation-hbond) [100]. The conserved structural residues were identified and mapped to the secondary structure elements of the ExoRm three-dimensional models to check if the positioning of these residues followed the expected pattern observed in other Sel1-like proteins. The STRIDE server [101, 102] was used to ascertain the location of secondary structure elements in the modeled proteins.

**b. Solvent accessibility analysis of the cleavage site in the ExoR wild type and mutants**

To predict the solvent-accessible surface area of the ExoR cleavage site based on primary structure the following programs were used: ACCpro at the SCRATCH Protein Predictor server [48], SABLE [103, 104, 105], NETASA [106], and RVP-NET [107]. To determine the solvent accessibility of these residues in the generated theoretical models, the GETAREA [108] server was used.

**c. Electrostatics**

The distribution of surface electrostatic potential for the ExoRm model was calculated using the Poisson-Boltzmann solver, DelPhi v.4 release 1.1 [109, 110, 111, 112]. The net charge and the dipole moment of the modeled ExoRm protein were computed using the automated system Protein Dipole Moments
d. Identification of protein-protein interaction sites in ExoR<sub>m</sub> and ExoS<sub>p</sub>

To identify putative protein-protein interaction sites on the surface of the modeled ExoR<sub>m</sub> protein and ExoS periplasmic domain, a number of automatic computational prediction methods were used: PIER [114], SPPIDER version 2 [115], and cons-PPISP [116, 117]. The PIER method depends solely on structural data that include physicochemical properties of atoms at the protein surface [114]. Only residues identified with a prediction value above 30 were considered for further analysis. The prediction method used in SPPIDER [115] incorporates solvent accessibility of surface residues, whereas the cons-PPISP method [116, 117] employs sequence conservation in addition to the accessible surface area. The detection of interaction residues with the SPPIDER server [115] was done with a tradeoff between sensitivity and specificity set to 1.0 for best precision. The accuracy of the cons-PPISP prediction [116, 117] is given as neural network scores, and residues that were reported and examined belong to cluster 1 (confidence of 35). In addition, interaction hot spot residues (residues that are essential for recognition or binding) were detected in the ExoR sequence using the alignment-dependent ISIS method [118, 119]. To determine changes in the solvent accessibility of candidate interface residues in the generated theoretical models of the ExoR mutant proteins, the GETAREA server [108] was used.

**Protein docking**

Molecular docking allows determining the orientation of two molecules relative to each other. There are several protein docking algorithms designed to model protein complexes from the protein subunits in their unbound state. Approaches include (i) using complementarity in shapes; and (ii) simulation
of binding based on interaction energies and a search for optimal subunit association that involves a transformation of one of the subunits [120]. Most of the available programs or servers are rigid-body docking algorithms and direct the transformations relying either on local shape matching, e.g. PatchDock [120] or searching the entire three-dimensional space of one of the subunits using Fast Fourier Transform (FFT) to simulate binding [121, 120], e.g. ClusPro [122, 123], GRAMM-X [124], ZDOCK [125], and HexServer [121]. The ClusPro server [122, 123] was evaluated at the 2013 Critical Assessment of Predicted Interactions meeting as the best docking algorithm in the automated docking server category, outperforming all other automated docking algorithms. Therefore, the ClusPro server [122, 123] was used to generate putative homodimers of ExoS, heterodimers of ExoS_p-ExoR_m, and complexes of ExoSR:ExoSR’ and ExoSS’:ExoR. For each group of complexes, more than 100 putative docking orientations were produced. Based on known previously identified protein interfaces of histidine kinases as well as of solenoid proteins, we selected the most probable representations of the ExoS_p-ExoS_p’ and ExoS_p-ExoR complexes. To specifically identify residues involved in the selected protein-protein interfaces, the protein complexes were analyzed with the SPIDDER server [115].
III.1. The *Sinorhizobium meliloti* Rm1021 ExoR protein is composed of six Sell-like repeats

Analysis of the ExoR sequence, using traditional domain architecture programs including the SMART database [36], suggests that the ExoR protein contains four alpha-alpha repeats categorized as Sell-like repeats [23]. Further scrutiny by the latest repeat detection methods reveal two additional repeats, repeat 5 and 6, at the C terminus of the ExoR protein (*Table III.1*). Repeat 5 (ExoR5) was predicted with similar confidence as the previously identified repeats (TPRpred, P-value 2.8E-04; HHrepID, P-value 9.0E-09; REPROM, alignment score: 31), whereas repeat 6 (ExoR6) was identified with lower confidence than the preceding repeats (TPRpred, P-value 2.2E-02; HHrepID, P-value 7.8E-02; REPROM, alignment score 24) (*Table III.1*). Secondary structure prediction analysis of ExoR (*Fig. III.1*) also suggests that the most of the ExoR protein folds into helices including the region that has the newly identified repeats, further supporting the presence of two additional Sell-like repeats. The length of each of the six repeats ranges from 32 to 40 residues and consists of two α-helices; this repeat structure is in agreement with the expected configuration of Sell-like repeats [23, 31, 32].
Table III.1. Sel-like repeats detected within the ExoR<sub>p</sub> amino acid sequence.
The repeat detection methods are sensitive but not specific in determining repeats’ boundaries. Therefore, the information from the repeat detection methods together with the data on the secondary structure elements determined from the ExoR sequence was used to determine the final boundaries of the repeats (shown in bold). The significance of the predictions is reported as <sup>a</sup> – E-values, <sup>b</sup> – P-values, and <sup>c</sup> – alignment scores. Repeats detected with lower significance shown in gray. Repeats identified as TPR repeat denoted with star (*). The numbering corresponds to the residues of the full length ExoR, ExoR<sub>p</sub>.

<table>
<thead>
<tr>
<th>Method</th>
<th>ExoR1</th>
<th>ExoR2</th>
<th>ExoR3</th>
<th>ExoR4</th>
<th>ExoR5</th>
<th>ExoR6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMART&lt;sup&gt;a&lt;/sup&gt;</td>
<td>43-74</td>
<td>75-110</td>
<td>121-159</td>
<td>161-196</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>TPRpred&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.27E+02</td>
<td>1.84E-04</td>
<td>1.12E+01</td>
<td>1.56E-07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HHrepID&lt;sup&gt;b&lt;/sup&gt;</td>
<td>43-76*</td>
<td>76-111</td>
<td>125-160</td>
<td>162-197</td>
<td>197-230*</td>
<td>235-268*</td>
</tr>
<tr>
<td>REPRO&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.4E-07</td>
<td>7.7E-11</td>
<td>2.9E-04</td>
<td>2.2E-15</td>
<td>2.8E-04</td>
<td>2.2E-02</td>
</tr>
<tr>
<td>Consensus</td>
<td>49 – 84</td>
<td>54 – 125</td>
<td>131 – 208</td>
<td>159-185</td>
<td>203-238</td>
<td>237-267</td>
</tr>
</tbody>
</table>

Consensus     43-74  75-110  121-160  161-196  197-228  229-265
**Figure III.1. Secondary structure prediction of the ExoR protein.**

(H, α-helix; “-”, random coil; “?”, low confidence α-helix prediction)

The alignment of all six ExoR repeats reveals that the ExoR6 repeat is a highly divergent repeat and its sequence does not fit the consensus sequence for the other repeats well (**Figure III.2a**). The consensus sequence (based on conservation of 80% or more) derived from an alignment of the first five ExoR repeats reveals the presence of conserved alanine and glycine residues at positions 3, 8, 14, 24, 32, 39, and 43; this consensus matches the Sell-like repeat consensus sequence for the most part [23, 32] (**Figure III.2a**). A sequence alignment of the C-terminal region from 28 putative ExoR orthologs that corresponds to the ExoR6 repeat shows the following unique consensus: sxDpsWhpshbcpAFshAscscRphAhshxxphhx (s-small, b-big, c-charged, p-polar, h-hydrophobic, x-any amino acid residue; upper case letters refer to the standard amino acid one-letter code (**Figure III.2b**). In addition, the S.
meliloti ExoR6 contains an Asp₃ tag that is also found in S. medicae WSM419 but otherwise absent from other species where the following pattern of three amino acid residues is found: xpx (x-any residue, p-polar) (figure 1b). Even though ExoR6 shows a unique sequence pattern, it still conserves three important structural residues (Ala14, Ala24, and Ala32) (Figure III.2b) [32].
Figure III.2. (a) Alignment of repeats found in the ExoR protein of S. meliloti Rm1021. The amino acid residues located within α-helices are shaded in gray. The experimentally determined proteolysis site (A80) is indicated in green. Proline residues found in the ExoR3 loop 1 and in the helix A of ExoR6 are shown in blue. (b) Amino acid sequence alignment of the ExoR C-terminal regions from 28 ExoR orthologs. The alignment is based on ExoR6 from S. meliloti Rm1021 and corresponds to residues 229-268 of the full length ExoR (ExoRp) that includes ExoR6 (black) and the ExoR Asp3 tag (gray). The following species are shown: Sinorhizobium meliloti Rm1021 (S.m.), Sinorhizobium medicae WSM419 (S.m.w.), Sinorhizobium fredii NGR234 (S.f.), Rhizobium leguminosarum bv. trifolii WSM1325 (R.l.1325), Rhizobium leguminosarum bv. trifolii WSM2304 (R.l.2304), Rhizobium etli CIAT 652 (R.e.c.), Rhizobium etli Kim 5 (R.e.k.), Agrobacterium tumefaciens str. C58 (A.t.), Agrobacterium vitis S4 (A.v.), Agrobacterium radiobacter K84 (A.r.), Bartonella grahamii as4aup (B.g.), Bartonella henselae str. Houston-1 (B.h.), Bartonella tribocorum CIP 105476 (B.t.), Bartonella schoenbuchensis R1 (B.s.), Ochrobactrum anthropi ATCC 49188 (O.a.), Ochrobactrum intermedium LMG 3301 (O.i.), Brucella abortus str. 2308 A (B.a.), Hoeflea phototrophica DFL-43 (H.p.), Chelativorans sp. BNC1 (C.sp.), Starkeya novella DSM 506 (S.n.), Beijerinckia indica subsp. indica ATCC 9039 (B.i.), Rhodopseudomonas palustris CGA009 (R.p.), Bradyrhizobiaceae bacterium SG-6C (B.b.), Bradyrhizobium japonicum USDA 110 (B.j.), Bradyrhizobium sp. BTA11 (B.sp.), Mesorhizobium ciceri biovar biserrulae WSM1271 (M.c.), Mesorhizobium loti MAFF303099 (M.l.), Mesorhizobium opportunistum WSM2075 (M.o.). The consensus sequences above the alignments reflect conservation of at least 80% of the first five repeats (ExoR1-ExoR5) (a) and of the 28 ExoR orthologs (b) and numbering indicated above the consensus sequences follows the convention of Mittl and Schneider-Brachert [23]. The numbering at the beginning and end of each repeat corresponds to the residue number of the full length ExoR, ExoRp. The conserved structural residues common to the SLR consensus sequence [23, 32] are indicated as red letters within the sequences of the S. meliloti ExoR protein and ExoR orthologs. Residues are abbreviated as follows: A, alanine; L, leucine; F, phenylalanine; G, glycine; V, valine; H, histidine; D, aspartic acid; W, tryptophan; R, arginine; s, small; b, big; h, hydrophobic; p, polar; c, charged; x, any residue.
III.2. Comparative modeling of the *S. meliloti* ExoR<sub>n</sub> protein

In BLASTP searches [59, 60] against non-redundant protein sequences, the *S. meliloti* ExoR sequence shows high sequence identity (45-98%) to other exopolysaccharide biosynthesis regulatory proteins (ExoR); however, BLASTP searches [59, 60] against Protein Data Bank proteins [65] showed that none of the identified ExoR proteins have a known structure that can be used to model the ExoR protein. Therefore, fold recognition algorithms were used to detect possible structural templates with similar fold to the ExoR protein. The closest structural match to the *S. meliloti* ExoR protein was identified as the Sell-like repeat *H. pylori* cysteine rich protein C (HcpC, PDB ID: 1OUV) [32] by several fold recognition algorithms: SPARKS-X (z-score 19.14) [70], 3D-PSSM v.2.6.0 (E-value 1.93E-05) [71], HHpred (E-value 2.3E-23) [72, 73], and LOOPP (E-value 5e-21) [74, 75]. Another Sell-like repeat protein corresponding to locus C5321 of *E. coli* strain Cft073 (PDB ID: 4BWR) [126] was identified as the closest structural match to the *S. meliloti* ExoR protein by two other fold recognition algorithms pGenTHREADER [76] and FUGUE [77] (p-value: 5E-08 and z-score: 31.98 respectively). Other distantly related structural templates identified (data not shown) also comprise solved structures of Sell-like and teratricopeptide repeat proteins confirming a super-helical fold for the ExoR protein.

Even though there is low sequence identity (below 30%) between ExoR and other solenoid proteins with experimentally solved structures, we were able to generate models that evaluated favorably. Sequence similarity between the template and the target is an accepted indicator of the accuracy and possible applications of the comparative models [67, 68, 69], however this correlation is not absolute [127]. There are examples of reliable models with root-mean-square deviation (RMSD) values below 3.0Å built from alignments with sequence identity less than 12.5% [105]. Furthermore, models constructed based on alignments with sequence identity less than 40% do not always show
significant decrease in correct modeling of the surface accessibility of residues important in analyses of point mutations and ligand docking [127] and may be suitable to provide information unique to the target [128].

Out of approximately 200 ExoR models built, the best structural ExoR$_m$ model was selected from a group of top ranked models, based on the evaluation criteria detailed in the methods section and correlated with a fine visual inspection of biophysical properties of the models. The model generated by I-Tasser, which builds the models based on threading alignments and ab initio modeling of loops and tails [84, 85] evaluated as the best three-dimensional model of the ExoR$_m$ protein. Energy profiles calculated using ProSA-web [93, 94] for the selected structural model of the ExoR$_m$ protein showed low energies (below zero) and a z-score of -6.99 comparable to solved NMR and X-ray crystal structures (Figure III.3a). The verification of the three-dimensional model of the ExoR$_m$ protein via Verify3D [95, 96] also shows a high 1D-3D profile score (above 0.2) for almost the entire length of the model (Figure III.3b). In addition, the model passed the checks of bond lengths (Z-score 0.704), bond angles (Z-score 0.901), chirality/dihedrals (Ramachandran Z-score 0.678), and chi-1/chi-2 angles (chi-1/chi-2 correlation Z-score 5.125) implemented in WHAT_CHECK [97] with values within the expected range for well-refined structures.
Figure III.3. Evaluation plots of the putative three-dimensional model of ExoR\textsubscript{m} (a) ProSA-web evaluation plots [93, 94]. The z-score of the ExoR model of -6.99 (highlighted as black dot) and the energy plot obtained with ProSA-web are shown. (b) Verify 3D evaluation plot [95, 96]. The 3D-1D profile is shown. The 3D-1D averaged scores plotted on the vertical axis (window size 21) are above 0 for the entire length of the model except for some residues at the ExoR C-terminal end.

III.3. The tertiary model of \textit{S. meliloti} ExoR\textsubscript{m} protein reveals a super-helical fold

The ExoR\textsubscript{m} three-dimensional model suggests a super-helical fold for the ExoR protein. The modeled ExoR\textsubscript{m} protein consists of twelve \(\alpha\)-helices forming six Sell-like repeats preceded by a 3 residues long 3\textsubscript{10} helix at its N-terminus. Each repeat is formed by two antiparallel helices and the N- and C-
terminal helices of the repeats are referred to as the A- and B-helices respectively. Helix A is located at the inner (concave) and helix B at the outer (convex) surfaces of the super-helix (Figure III.4a).

Figure III.4. (a) Ribbon representation of the putative structure of the S. meliloti Rm1021 ExoR mature protein (ExoR<sub>m</sub>, residues 31-268). The repeats are numbered from 1 to 6 and helix A and helix B of each repeat are colored in blue and red respectively. The N-terminal 3<sub>10</sub> helix is colored in green. The location of ExoR cleavage site, helix A of ExoR2 is marked with an arrow. (b) Backbone trace of the first three repeats of ExoR (ExoR<sub>1</sub>-ExoR<sub>3</sub>). Side-chains of conserved structural residues responsible for tight packing of helices are shown in red (3,8,32, 39 and 40) and those responsible for sharp turns in loop regions are indicated in blue (14, 24, and 43). The numbers represent the positions of the amino acid residues in repeats. (c) Schematic representation of the three forms of the ExoR protein (ExoR<sub>p</sub>, ExoR<sub>m</sub>, and ExoR<sub>c20</sub>). The cartoon shows the location of signal peptide (SP), six putative Sell-like repeats (ExoR<sub>1</sub>-ExoR<sub>6</sub>), and the cleavage site of ExoR, A80 (flag-tagged).
Individual repeats are structurally very similar and superpose with RMSD values below 1 Å, the only exception being ExoR3. The higher RMSD values (0.9–1.3 Å) for the structural alignment of ExoR3 with other ExoR repeats can be ascribed to its elongated loop region that contains two proline residues absent from the other repeats (Table III.2 and Figure III.2a). Despite the fact that ExoR6 possesses a proline residue in the middle of helix A and a reduced number of the conserved structural residues, it is modeled structurally as a typical repeat (RMSD values range from 0.6 to 0.7 Å except for the superposition with ExoR3) (Table III.2 and Figure III.2a). The ExoR3 repeat in the ExoRₘ model not only shows the extended loop conformation predicted from the ExoR sequence but its helix A is also unusual in its length of 20 amino acid residues, where the range of helix length for all other helices is 12 to 15 residues. This extended helix is predicted by 4 of the 9 secondary structure prediction programs (Fig. III.5).

### Table III.2. RMSD values for the superposition of ExoR repeats and their sequence identity (%). The RMSD values (in bold) were calculated using TM-align server [58]. The numbering corresponds to the residues of the full length ExoR, ExoRₘ.

<table>
<thead>
<tr>
<th></th>
<th>ExoR1 43-74</th>
<th>ExoR2 75-110</th>
<th>ExoR3 114-160</th>
<th>ExoR4 161-196</th>
<th>ExoR5 197-228</th>
<th>ExoR6 229-265</th>
</tr>
</thead>
<tbody>
<tr>
<td>ExoR1</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>ExoR2</td>
<td>0.7</td>
<td>0.7</td>
<td>1.3</td>
<td>0.6</td>
<td>0.7</td>
<td>-</td>
</tr>
<tr>
<td>ExoR3</td>
<td>0.9</td>
<td>1.1</td>
<td>-</td>
<td>22</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>ExoR4</td>
<td>0.5</td>
<td>0.6</td>
<td>1.1</td>
<td>-</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>ExoR5</td>
<td>0.5</td>
<td>0.7</td>
<td>1.0</td>
<td>0.5</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>ExoR6</td>
<td>-</td>
<td>16</td>
<td>22</td>
<td>19</td>
<td>16</td>
<td>7</td>
</tr>
</tbody>
</table>
Figure III.5. Secondary structure prediction of the ExoR3 repeat. The boundary of the repeat determined from the repeat detection methods is shaded in gray (region 121-160 of full length ExoR, ExoRₚ). The prediction of the secondary structure elements with the STRIDE server [101, 102] from ExoR atomic coordinates is shown in (A). (H, α-helix; “-”, random coil; “?”, low confidence α-helix prediction)

The low RMSD values for the superposition of the majority of the ExoR repeats indicate that the helical structure of the ExoR repeats follows the expected pattern for the positioning of known conserved structural residues characteristic of the Sel1-like proteins: residues 14, 24, and 43 dictate sharp turns and are present in loop regions while residues at positions 3, 8, 32, 39, and 40 are located within the α-helices and are responsible for the tight packing of the repeats [23, 32] (Fig. III.2a and III.4b). The packing of the ExoR repeats also shows high conservation: the average inter-repeat helix-packing angle found in the modeled ExoR protein is 42(±2.6)° and the average intra-repeat helix-packing angle is 18(±2.9)°.

III.4. Identification of the ExoRₚ protein-protein interaction sites

Since the canonical function of Sel1-like repeat proteins is typically mediated via protein-protein interactions [reviewed in 23] and the physical association between ExoS and ExoRₚ is known to stabilize ExoRₚ [18], the ExoR sequence and its modeled three-dimensional structure were examined for the presence of protein-protein interaction sites. Three putative non-overlapping
binding sites were identified: A, B, and C. Site A is located at the inner face of the N-terminal end of the protein, site B is found at the center of the inner face of the protein and extends to the C-terminus, and site C is positioned at the outer face of the C- terminal end of the ExoR protein (Fig. III.6a-c).
**Figure III.6.** (a) Location of the putative protein-protein interaction sites within ExoR repeats. The sites A, B, and C are colored red, blue, and green respectively. (b, c) Putative functional sites mapped to the surface of ExoR\textsubscript{m} protein. Site A (residues identified by more than one prediction program in red; residues predicted using just one program in light red) and site C (green) in (b); and site B (residues identified by more than one prediction program in blue; residues predicted using just one program in light blue) in (c). Residues identified as hot spots are shown in black. (d) Superposition of the crystal contact II (orange) of HcpC protein [32] and site A (red) of ExoR. The superposition is based on residues 28-116 from HcpC and residues 31-143 from ExoR. The side-chains of Asn66 (HcpC), Asn83 (ExoR), and Asn122 (ExoR) are shown. (e) Superposition of the crystal contact I (orange) of HcpC protein [32] and site B (blue) of ExoR. The superposition is based on residues 225-292 from HcpC and residues 162-268 from ExoR.

Site A encompasses the ExoR proteolytic site and is composed of 36 residues found in the N-terminal 3\textsubscript{10} helix, helices A of ExoR\textsubscript{1}, ExoR\textsubscript{2}, and ExoR\textsubscript{3}, intra-repeat loop of ExoR\textsubscript{2} and ExoR\textsubscript{3}, the inter-repeat loop between ExoR\textsubscript{2} and ExoR\textsubscript{3}, and helices B of ExoR\textsubscript{1} and ExoR\textsubscript{2}. Within site A, six residues were identified as protein-protein interaction hot spots: Lys55 in helix A of ExoR\textsubscript{1}, Glu115 in helix A of ExoR\textsubscript{3}, Arg132 and Arg133 in helix A of ExoR\textsubscript{3}, and Asp137 and Thr138 in intra-repeat loop of ExoR\textsubscript{3} (Table III.3 and Figure III.6). Site B is formed by 21 residues located on the surface of helices A of ExoR\textsubscript{4}, ExoR\textsubscript{5}, and ExoR\textsubscript{6}, helix B of ExoR\textsubscript{4}, and intra-repeat loops of ExoR\textsubscript{4} and ExoR\textsubscript{6}. Four residues of this binding patch were identified as hot spots: Arg169 (helix A of ExoR\textsubscript{4}), Glu175 (intra-repeat loop of ExoR\textsubscript{4}), Gln209 and Glu210 (helix A of ExoR\textsubscript{5}) (Table III.3 and Figure III.6). The putative interaction site C is formed by 8 residues located on the surface of the helix B of ExoR\textsubscript{5}, helices A and B of ExoR\textsubscript{6}, and intra-repeat loop of ExoR\textsubscript{6}. None of the residues that form site C were identified as hot spots (Table III.3 and Figure III.6).
**Table III.3.** In silico prediction of protein-protein interfaces and hot spots in ExoR. The identified interface residues were grouped into three clusters: site A, site B, and site C. Residues identified as hot spots are shown in red. The numbering corresponds to the residues of the full length ExoR, ExoR_p (Res = residue and its position; Acc = accuracy of the prediction).

<table>
<thead>
<tr>
<th>PIER</th>
<th>SPPIDER</th>
<th>cons-PPISP</th>
<th>ISIS</th>
<th>PIER</th>
<th>SPPIDER</th>
<th>ISIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Res</td>
<td>Acc</td>
<td>Res</td>
<td>Acc</td>
<td>Res</td>
<td>Acc</td>
<td>Res</td>
</tr>
<tr>
<td>Site A</td>
<td></td>
<td>Site B</td>
<td></td>
<td></td>
<td>Site C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F33</td>
<td>V37</td>
<td>F43</td>
<td>0.9</td>
<td>F43</td>
<td>0.9</td>
<td>R169</td>
</tr>
<tr>
<td>F47</td>
<td>55</td>
<td>F44</td>
<td>0.8</td>
<td>F44</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>F48</td>
<td>40</td>
<td>F47</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>K48</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F49</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>K55</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>L188</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G201</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>V202</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G204</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N205</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>V206</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>V209</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Q209</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E210</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P33</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>V37</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F51</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F52</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S114</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S114</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E115</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E115</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D116</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T117</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G118</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y119</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N122</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y130</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R132</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R133</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D137</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T138</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T221</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>L224</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G225</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H232</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>L235</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M257</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A214</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A246</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
III.5. Analyses of the electrostatic features of ExoRₘ

The surface electrostatic profile of the modeled ExoRₘ protein was examined to further analyze the biophysical features of ExoR that may be driving interactions with its binding partners since electrostatic interactions are important for protein recognition and binding [129]. The net charge of -3 at pH 7 and a dipole moment of 584 Debye oriented from the inside (concave face) of the super-helix towards the outside (convex face) was calculated from the ExoR atomic coordinates using the Protein Dipole Moments Server [113].

The putative protein-protein interaction site A, located in the N-terminal inner (concave) face of the super-helix, displays a mild acidic surface electrostatic profile due to the side chains of Asp89, Glu115, and Asp137 (Fig.III.7a). There is also a negatively charged patch at the inner side of the super-helix at its C-terminus that includes several acidic residues found in the helix A (Asp231 and Glu240) and the helix B (Asp250) of ExoR6, the intra-repeat loop of ExoR6 (Glu248 and Asp249), and Asp₃ tag (Asp266, Asp267, and Asp268) (Fig.III.7b). The protein-protein interaction site B includes a positively charged patch in the center of the protein interspersed with surface hydrophobic residues (Fig.III.7b). Another negatively charged patch is formed on the outer surface of the super-helix due to residues located on the surface of the helices B of ExoR1 (Asp60, Glu61, Glu64, Glu71) and ExoR2 (Asp95, Glu97, Glu104); and in the intra-repeat loops of ExoR2 (Asp89 and Glu93), ExoR3 (Asp137 and Asp143), and ExoR4 (Glu175) (Fig.III.7b). The outer C-terminal end of the protein shows a hydrophobic cluster and corresponds to the binding patch identified as the site C (Fig.III.7a). Adjacent to site C, there is a positively charged patch created by the side chains of Lys185, Lys186, Arg190, Arg192, Lys193 (ExoR4, helix B), Arg148 (ExoR3, helix B), and Arg215 (ExoR5, helix B) (Fig.III.7c). The electrostatic profile of the modeled ExoRₘ protein shows that the inner
concave surface of the protein is more negatively charged relative to the outer convex surface of the super-helix.

**Figure III.7.** Electrostatic potentials mapped to the molecular surface of the ExoR<sub>m</sub> protein. The ExoR protein is oriented to visualize the interfaces of the functional sites A and C in (a), site B in (b), and the basic surface patch at the outer surface of the super-helix in (c). In (a) and (b) the protein is oriented as in figure III.5. In all shown structures the N-terminal ends are placed on the left and the C-terminal ends on the right. The surface potentials are color-graded from −4 kT/e (red) to +4 kT/e (blue).

### III.6. Discussion

The ExoR protein is a key player in regulating the ExoS/ChvI two-component system responsible for the successful establishment of the symbiotic relationship between *S. meliloti* and its legume host plant, alfalfa [15, 17, 18, 19, 20]. Homologs of ExoR, ExoS, and ChvI are essential for host invasion by animal and plant pathogens [130, 131]. Yet, a major gap in knowledge exists in understanding the basis of ExoR functionality since no solved ExoR structure exists. This study attempts to bridge this gap and provides a robust theoretical structural model of the *S. meliloti* Rm1021 ExoR<sub>m</sub> protein. Here we present a first look into the structural fold of this protein and its implications in ExoS/ChvI two-component signaling.

**The ExoR<sub>m</sub> structural model reveals an alpha-alpha super-helical fold**

The proposed structure of ExoR<sub>m</sub> reveals the same overall super-helical fold as the crystal structure of the HcpC protein with PDB ID of 1OUV, an
accepted structural template for the modeling of other Sell-like repeat proteins [32]. Both of the proteins are composed of α/α repeats: six repeats consisting of 32 to 40 amino acid residues in ExoR, and seven repeats consisting of 36 amino acid residues in HcpC [32]. In contrast to HcpC repeats, the Sell-like repeats in ExoR are not interlinked with disulphide-bridges characteristic of the HcpC family of proteins [23, 32].

High structural similarity (RMSD of 0.79 Å) of ExoR to HcpC and the presence of conserved structural residues characteristic of Sell-like repeats [23, 32] in ExoR imply that the helix packing angles in HcpC and ExoR are conserved. However, the intra-repeat and inter-repeat helix packing angles are $43(\pm 3.5)^{\circ}$ and $19(\pm 3.5)^{\circ}$ in HcpC [32], compared to $18(\pm 2.9)^{\circ}$ and $42(\pm 2.6)^{\circ}$ in ExoR. This discrepancy stems from alternate definitions of Sell-like repeats [23]. The structure of the ExoR repeats is in agreement with the Sell-like repeats definition derived from SMART database [23, 36] and shown in figure III.2a with loop 1 as the flexible intra-repeat loop and loop 2 as the fixed length inter-repeat loop. In the structure of HcpC on the other hand, loop 1 (flexible) is treated as an inter-repeat loop and loop 2 (fixed length) as an intra-repeat loop [23, 32].

The structure of multi-repeat proteins is known to be a suitable scaffold for protein-protein interactions [29, 30, 31, 32]. The comparison of the ExoR and HcpC (PDB ID: 1OUV) [32] reveals that the one of the two potential peptide-binding sites identified as crystal contact II at the concave face of N-terminal end of the HcpC protein [32] matches the putative protein-protein interaction site A recognized at the concave face of the N-terminal end of the ExoR protein. The putative site B identified in ExoR partially overlaps the identified crystal contact I at the C-terminal end of HcpC [32]. Most of the crystal contact I interface residues in HcpC are found in the last helix (7B) and the extended helix after the last repeat [32]. In ExoR, most of the site B interface residues were identified in helices 4A,
5A, and 6A. There are a few interface residues identified in the last helix of ExoR (6B) but predicted with low significance and not included in our analysis (Arg251, cons-PPISP score 0.007; Ile255, cons-PPISP score 0.079; Ser258, cons-PPISP score 0.074). Site C does not have a counterpart in the HcpC protein. It is formed of just eight interface residues and possibly forms a novel interface site characteristic of the ExoR protein family. Since HcpC and ExoR belong to the same structural family and share similar structural features, we anticipate that the mode of protein-protein interactions in ExoR will be similar to that of HcpC [32].

The analyses of the electrostatic profile of the modeled ExoRₘ protein shows an asymmetry in the charge distribution on the surface of the ExoR protein with the inside of the protein being more negatively charged relative to the outer surface of the super-helix. Since all the predicted sites of protein interaction have distinct electrostatic profiles and surface hydrophobicity, electrostatic and/or hydrophobic interactions may play a key role in driving these interactions [129, 132, 133].

The C-terminus of ExoR houses two previously unidentified Sel1-like repeats

Previous studies have reported that there are four Sel1-like repeats within the S. meliloti Rm1021 ExoR protein [23]. Our analyses of the ExoR sequence suggest that there are two additional Sel1-like repeats, ExoR5 and ExoR6, present in the C-terminal part of the protein. The sequence signature of the Sel1-like repeats of the S. meliloti Rm1021 ExoR protein follows that of other Sel1-like repeat proteins [23, 32] except for the unique sequence pattern of its last repeat (ExoR6) which can be classified as a non-traditional Sel1-like repeat. The secondary structure prediction of the ExoR sequence also corresponds to the six α/α repeats and supports the helical repeats that take on modular architecture in the three-dimensional model of the ExoRₘ protein. This data suggest that the newly identified repeats are an
integral and functional part of the complete ExoR structural fold and protein.

Sell-like repeats with low sequence conservation have been found to play important roles in other solenoid proteins. For example, the C-terminal end of HcpC (PDB ID: 1OUV) which includes the last Sell-like repeat that deviates from the SLR consensus plays a key role in protein–protein interactions [32]. Similar to this divergent but functionally important terminal Sell-like repeat in the HcpC protein, we suggest that the last repeat of the ExoR protein is important for its function in vivo. The importance of ExoR6 is supported by the phenotype of the S. meliloti exoR95 mutant that exhibits a strong overproduction of succinoglycan [35, 134]. In this mutant, the helix B of ExoR6 is replaced by a sequence of 9 amino acids (ADSYTQVAS), rendering the ExoR95 protein nonfunctional [20, 134].

**Helix A of the third repeat can assume an extended conformation**

Although the modeled ExoR₉₉ protein follows the structural fold of the HcpC protein (PDB ID of 1OUV) [32] faithfully in most regions, it is unique in showing the presence of a 20 residue long extended helix A of the ExoR3 repeat. In the modeled ExoR₉₉, this third repeat is seven residues (SEDTGYF) longer than that predicted by the repeat detection algorithms. The ambiguous prediction of the length of helix A of the ExoR3 repeat in our analysis may be a reflection of pliability of the functional protein in that region with respect to secondary structure. Although speculative, this raises the possibility that a helix to random coil transformation or melting [reviewed in 135] occurs which may be important for ExoR function and/or regulation. A similar unusual extended conformation of a repeat has been observed in the crystal structure of a third repeat of the PEX5 protein from Trypanosoma brucei [136]. The PEX5 protein is a helical multi-repeat protein that contains tetratricopeptide repeats (TPRs) [136, 137] that are similar in
structure to Sel1-like repeats: two antiparallel α-helices per repeat [29, reviewed in 30] differing only in helix packing [23]. The first two repeats of the PEX5 protein assume a standard TPR fold but the third repeat is 29 residues long [136]. On the other hand, the crystallized human PEX5 does not show this unusual elongated conformation of its repeats [137]. Based on the similarity in sequence and function of T. brucei and human PEX5, it has been suggested that the third TPR in T. brucei PEX5 can adopt the extended form as well as the standard conformation characteristic for TPR motifs [136]. We propose a similar scenario for the ExoR3 repeat: although it is possible that the helix A of the ExoR3 repeat is a regular length α-helix (as suggested by some of our alternate models of ExoRm), it is equally probable that it is an extended helix that can transition to a regular length helix and vice-versa. Among the seven additional residues (SEDTGYF) found at the N-terminal end of the third repeat Ser114, Glu115, Asp116, Thr117, Gly118, and Tyr119 are predicted to be a part of protein-protein interaction site A. Assuming that this site is an actual interaction site, the strategic placement of this helix and possibility of helix length pliability for helix A suggest that this repeat may be important for ExoR function and/or its regulation.

**Conclusion**

In conclusion, we present the first attempt to generate a three-dimensional model of *S. meliloti* Rm1021 ExoR protein in the absence of its crystal structure. Our proposed structural model of ExoRm suggests the presence of six Sel1-like repeats that form a structural fold conducive to protein-protein interactions. Ongoing studies targeted towards the modeling and characterization of the ExoS protein with docking will allow us to identify the binding sites involved in ExoRm-ExoS interactions. Homologs of *S. meliloti* ExoR protein along with ExoS and ChvI have been found in many host
interacting bacteria. A systematic analysis of these homologs is currently underway to determine the origin and distribution of these three proteins. While most of the homologs have been identified based on sequence similarity alone, the *Agrobacterium tumefaciens* ExoR, ChvG (ExoS), and ChvI, and *Brucella abortus* BvrS (ExoS) and BvrR (ChvI) have been identified independently based on their crucial roles in mediating the invasion of many plants by *A. tumefaciens* and animals by *B. abortus* [5, 6, 7, 24, 131, 138, 139, 140]. The findings of our structural analyses make it possible to study the molecular mechanism of ExoR cleavage and ExoR-ExoS interactions using rational hypothesis driven approaches, which will facilitate the studies of pathogenicities of animal and plant pathogens in general.
Chapter IV. Analyses of ExoR mutant proteins

IV.1. Comparative modeling of the ExoR mutants

In order to understand the structure-function relationships of the ExoR protein, we undertook the modeling and structural analyses for two known and experimentally characterized ExoR reduced-function mutants, ExoRG76C and ExoRS156Y [18], and one loss-of-function mutant, ExoRL81A [20]. The generated models were evaluated using ProSA-web [93, 94] and Verify3D [95, 96] and the best three-dimensional representation of each mutant protein was selected. For all three mutants, energy profiles calculated using ProSA-web [93, 94] showed low energies (below zero) and z-scores of -7.26, -6.01, -6.75 for ExoRG76C, ExoRS156Y, and ExoRL81A, respectively comparable to the z-score calculated for the ExoR wild type, -6.99. The verification of the selected models of the mutant proteins via Verify3D [95, 96] also produced 3D-1D profile scores similar to wild-type ExoR, above 0.00 for almost the entire length of the models (Fig. IV.1).

I. ExoRG76C

(a)

(b)
Figure IV.1. Evaluation plots of the three-dimensional models of the ExoR mutants: I. ExoRG76C, II. ExoRS156Y, III. ExoRL81A (a) ProSA-web [93, 94] generated models’ energy profile and their z-scores (-7.26 for ExoRG76C, -6.01 for ExoRS156Y, and -6.75 for ExoRL81A). (b) Verify3D [95,96] 3D-1D models’ profiles. The 3D-1D averaged scores plotted on the vertical axis (window size 21) are above 0 for the entire length of the generated models except for some residues at their C-termini.

IV.2. Structural comparison between the ExoR wild-type protein and ExoR mutants

To detect conformational differences between the modeled wild-type ExoR and the mutant proteins, structural alignments were calculated. The overall structures of the mutant proteins were found to be comparable to the structure of the ExoR wild type. However, closer examination of the modeled single-point mutants revealed changes in the helix packing angles. The superposition of the wild type and the mutant proteins shows RMSD values of approximately 1 Å suggesting that these structures are similar and adopt the same super-helical fold. Even though there is high structural similarity between the modeled ExoR wild type and mutant proteins, the helix packing of the mutant proteins show minor changes compared to the wild-type ExoR: 18°(±2.9), 16°(±4.9), 17°(±6.3), and 16°(±5.6) for the average inter-repeat; and 42°(±2.6), 42°(±1.2), 40°(±2.9), and 45°(±3.0) for the average intra-repeat helix-packing angles of the wild type, ExoRG76C, ExoRL81A, and ExoRS156Y respectively. The effect of these changes in the helix-helix interactions on ExoR function and regulation are described and discussed in the following sections.


Since the stabilizing interactions between ExoRm and ExoS are disrupted
in the ExoRG76C and ExoRS156Y mutant proteins [18] and the ExoRL81A mutant protein undergoes higher rates of proteolysis than the wild-type protein [20], residues involved in protein-protein interactions were examined to identify possible changes that could affect ExoRExoS stabilizing interactions in these mutant proteins [18].

The analysis of the protein-protein interaction sites in the mutant proteins revealed changes in the surface accessibility of some of the residues identified as interaction hot spot residues. The mutations in the ExoR protein altered the burial category of the following interaction hot spot residues: Arg133 (site A) from exposed to partially buried in ExoRL81A; Asp137 (site A) from exposed to partially buried in the ExoRS156Y mutant; Thr138 (site A) from partially buried to exposed in the ExoRS156Y mutant; and Glu175 (site B) from exposed to partially buried in ExoRG76C and in ExoRL81A (Table IV.1).

The examination of the putative protein-protein interaction site A revealed alterations in the burial categories of Val37 from partially buried to exposed in ExoRL81A and ExoRS156Y; Phe47 from partially buried to buried in ExoRS156Y; Ser52 from partially buried to exposed in ExoRL81A; Trp79 from buried to partially buried in ExoRL81A; Asn83 from partially buried to buried in ExoRG76C and ExoRS156Y; Glu93 from partially buried to exposed in all analyzed mutants; Asp116 in ExoRL81A; Thr117 from partially buried to buried in all analyzed mutants; Gly118 from partially buried to buried in ExoRL81A and ExoRS156Y; Tyr119 from partially buried to exposed in ExoRG76C and ExoRL81A; Asn122 from exposed to partially buried in ExoRG76C and ExoRL81A; and Tyr130 from buried to partially buried in ExoRL81A (Table IV.1). The investigation of the putative protein-protein interaction site B showed changes in the burial state of Glu175 from exposed to partially buried in ExoRG76C and ExoRL81A; Gly176 from partially buried to buried in ExoRG76C and to exposed in ExoRL81A; Asn205 from partially buried to exposed in ExoRL81A;
Trp234 from exposed to partially buried in all three mutants; Ala 237 from exposed to partially buried in ExoRG76C and ExoRL81A; Ala242 from partially buried to exposed in ExoRL81A; Phe243 from exposed to partially buried in ExoRS156Y (Table IV.1). The inspection of the putative protein-protein interaction site C showed changes in the burial categories of Thr221 from exposed to partially buried in ExoRG76C; Leu224 from buried to partially buried in ExoRG76C; Met257 from partially buried to buried in ExoRS156Y (Table IV.1).

In summary, our analysis of the modeled mutant protein, showed significant alterations in the surface accessibility of key residues projected to be involved in protein-protein interactions.
Table IV.1. Comparison of changes in solvent-accessible surface area (SASA) of residues identified as candidate interface residues grouped into three putative interfaces site A (A), site B (B), and site C (C) in ExoR wild type and ExoR mutants. Residues identified as interface hot spots residues are shown in red. The program GetArea [108] was used to calculate SASA. Residues are defined as exposed if they have SASA value larger than 50 and are considered to be buried if their SASA values are less than 20. (1=WT; 2=ExoRG76C; 3=ExoRL81A; 4=ExoRS156Y)

<table>
<thead>
<tr>
<th></th>
<th>SASA (%)</th>
<th></th>
<th>SASA (%)</th>
<th></th>
<th>SASA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P33</td>
<td>96 91 88 88</td>
<td>R169</td>
<td>45 32 38 47</td>
<td>A214</td>
<td>100 100 100 96</td>
</tr>
<tr>
<td>V37</td>
<td>47 47 51 56</td>
<td>L172</td>
<td>44 44 30 44</td>
<td>T221</td>
<td>53 38 51 62</td>
</tr>
<tr>
<td>F43</td>
<td>41 39 47 47</td>
<td>S173</td>
<td>27 30 34 39</td>
<td>L224</td>
<td>12 22 7 8</td>
</tr>
<tr>
<td>F44</td>
<td>71 78 85 59</td>
<td>E175</td>
<td>50 32 40 52</td>
<td>G225</td>
<td>63 68 81 75</td>
</tr>
<tr>
<td>F47</td>
<td>22 26 31 16</td>
<td>G176</td>
<td>31 15 50 32</td>
<td>R232</td>
<td>13 12 10 13</td>
</tr>
<tr>
<td>K48</td>
<td>78 76 62 75</td>
<td>L188</td>
<td>0 0 0 1</td>
<td>L235</td>
<td>5 1 1 3</td>
</tr>
<tr>
<td>F49</td>
<td>29 31 23 27</td>
<td>G201</td>
<td>7 11 14 1</td>
<td>M257</td>
<td>29 26 42 10</td>
</tr>
<tr>
<td>F51</td>
<td>64 69 60 62</td>
<td>V202</td>
<td>38 36 29 34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S52</td>
<td>43 28 54 41</td>
<td>G204</td>
<td>0 0 1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y54</td>
<td>47 48 35 37</td>
<td>N205</td>
<td>34 38 51 48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K55</td>
<td>84 88 77 82</td>
<td>V206</td>
<td>9 9 10 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y66</td>
<td>5 6 5 4</td>
<td>F208</td>
<td>34 27 23 38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W79</td>
<td>16 16 23 18</td>
<td>Q209</td>
<td>76 65 69 63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A80</td>
<td>16 18 12 10</td>
<td>E210</td>
<td>37 26 46 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N83</td>
<td>31 17 23 19</td>
<td>D231</td>
<td>0 1 2 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M84</td>
<td>6 15 17 12</td>
<td>P233</td>
<td>9 7 7 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y87</td>
<td>58 57 55 57</td>
<td>W234</td>
<td>59 38 46 36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V91</td>
<td>19 16 10 12</td>
<td>A237</td>
<td>64 49 42 73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E93</td>
<td>48 51 71 67</td>
<td>M238</td>
<td>5 0 10 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y102</td>
<td>0 0 2 1</td>
<td>A242</td>
<td>35 31 54 45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S114</td>
<td>100 100 88 100</td>
<td>F243</td>
<td>50 60 58 48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E115</td>
<td>88 85 88 91</td>
<td>A246</td>
<td>92 88 100 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D116</td>
<td>13 18 48 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T117</td>
<td>31 5 14 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G118</td>
<td>39 26 11 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y119</td>
<td>45 51 59 48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N122</td>
<td>57 47 44 64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A123</td>
<td>0 0 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I125</td>
<td>18 9 10 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S126</td>
<td>25 28 41 31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y130</td>
<td>15 16 34 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R132</td>
<td>32 42 22 22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R133</td>
<td>63 53 45 56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D137</td>
<td>87 78 83 49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T138</td>
<td>47 38 41 76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IV.4. Changes in the electrostatic profiles of the mutant proteins.

To further our understanding of the ExoR interactions with its binding partners, the electrostatic profiles of the modeled ExoR mutant proteins were examined.

The analysis of the mutant proteins with the Protein Dipole Moments Server [113] resulted in the same net charge of -3 at pH 7 as in the wild-type protein. However, the visual inspection of the electrostatic profile of the mutant proteins revealed variations at functional site A and the C-terminal part of the concave face of the ExoR mutants. These sites became more negatively charged compared to the electrostatic profile of these regions in the wild type protein (Fig. IV.2). These observations are consistent with the calculated changes in the dipole moment and dipole moment vector orientations for the mutant proteins. The dipole moments of 584 Debye, 606 Debye, 701 Debye, and 779 Debye were calculated from the atomic coordinates of ExoR wild type, ExoRG76C, ExoRL81A, and ExoRS156Y respectively using Protein Dipole Moments Server [113]. In the ExoR wild type as well as in the mutant proteins, the dipole is oriented from the concave face to the convex face of the modeled proteins. In case of ExoRG76C and ExoRS156Y, the dipole passes between helices A and B of ExoR3 similar to the wild-type but points more toward the N-terminal part of ExoR3. On the other hand, the dipole of ExoRL81A is located between helices 3B and 4A (Fig. IV.3).

In summary, our analyses of the electrostatic profiles of the modeled mutants suggest alterations in their electrostatic profiles and dipole moment that might interfere with specific interactions with other proteins [113].
Figure IV.2. Electrostatic potentials mapped to the molecular surface of the ExoR wild-type protein (a) and the ExoR mutant proteins ExoRG76C (b), ExoRS156Y (c), and ExoRL81A (d). In all shown structures the N-terminal ends are placed on the left where the functional site A is found. The arrow points to the cleavage site. The surface potentials are color-graded from $-2 \ kT/e$ (red) to $+2 \ kT/e$ (blue).

Figure IV.3. The ExoR mutant proteins structurally aligned to show the difference in the dipole moment vectors orientations. Dipole moment of the wild type in red, ExoRG76C in blue, ExoRS156Y in magenta, and ExoRL81A in green. The dipoles were calculated using Protein Dipole Moments Server [113].
IV.5. Analysis of the accessibility of the cleavage site in the mutant proteins.

The ExoRL81A mutant displays a dramatic reduction of the active form of ExoR, ExoR_m [20]. The reduced levels of ExoR_m are most likely the result of higher rates of proteolysis of ExoRL81A [20] possibly due to the changes in the accessibility of the cleavage site. In addition, disruption of the stabilizing interactions between ExoS and ExoR in ExoRG76C and ExoRS156Y might cause the mutant proteins to be more prone to attack by periplasmic proteases [18]. Therefore, the accessibility of the cleavage site was examined in detail to probe the mutant proteins for higher susceptibility to proteolysis.

Examination of the surface accessibility of Ala80 and Leu81 in the modeled ExoR wild type supports the results obtained from the prediction of the surface accessibility of these residues based on the ExoR sequence: both residues are buried with surface accessible area of 16% for Ala80 and 0% for Leu81 in the wild-type ExoR protein. However, contrary to our hypothesis, the investigation of the proteolysis site in the mutant proteins showed no increase in the surface accessibility of this site. Residues at position 80 and 81 remained buried in the analyzed mutants: the solvent-accessible surface of Ala80 for ExoRG76C 18%, ExoRS156Y 10%, and ExoRL81A 12%; the accessible surface area of residue 81 ranges from 0 to 2% in these three mutants.

As no significant changes in the surface accessible area of the key residues at the proteolysis site (Ala80 and Leu81) was found, this site was further examined to identify structural features that possibly could lead to an increased proteolysis rate of these mutant proteins [18, 20]. Since the three-dimensional architecture of the ExoR_m protein forms a super-helix, a fold in which, function is known to be intimately associated with the flexibility of this fold [34], normal-mode analyses with ElNemo [141] was
used to probe any likely conformational fluctuations in the mutant structures compared to the wild type, in the context of the cleavage site accessibility. The open forms of the lowest frequency modes (mode 7) of the wild type and the mutant proteins were analyzed. Ala80 remained buried in the wild type (18%), ExoRL81A (15%), and ExoRS156Y (12%), but changed from buried to partially buried in the ExoRG76C mutant protein (23%). Even though the vibrational variations have not brought about any significant changes in the accessibility of the cleavage site, these conformational fluctuations take place in the region of cleavage. These fluctuations bring instability to this region raising the possibility of this site to become exposed upon binding of a protease or a ligand in order for the proteolysis to take place.

IV.6. Discussion

Structural models of the ExoR mutants reveal alterations in the packing of the helices

Our modeling studies suggest that function of the protein is impacted by the introduced missense mutations through changes in the side chain packing of residues and resultant alterations of the helix-packing angles without grossly impacting the structure. The ExoRG76C [18] is a point mutation located in helix A of ExoR2. There are two main aspects to this substitution: (1) replacement of a small hydrophobic residue with a more polar amino acid with a much bigger side chain and (2) a removal of one of the conserved structural residues responsible for the tight packing of the helices. In the ExoRL81A mutant protein, the mutation replaces a conserved hydrophobic residue at the experimentally identified cleavage site [20]. The ExoRS156Y [18] is a point mutation located in helix B of ExoR3, which replaces a small polar residue with an aromatic polar residue with a bulky side chain.
Changes in the solvent accessibility of asparagine residues might effect interactions with binding partners in mutant proteins.

Examination of potential protein-protein interaction sites in the modeled ExoR mutant proteins reveal changes in burial state of some of the binding residues. Solvent accessibility states of Glu93, Thr 117, and Trp234 residues showed changes in all three mutants. Change in the burial category of Asn83 is specific only to ExoRG76C and ExoRS156Y. Alterations in the burial category of the following residues are only found in the ExoRL81A mutant protein: Ser52, Trp79, Asp116, Tyr130, Arg133, Asn205, and Ala242.

Asparagine residues have been recognized to play an important role in peptide recognition in HcpC (Asn66) [32], in the Hsp70/Hsp90 organizing protein (Hop) [142], and in PEX5 [137]. Two asparagine residues are present in site A of ExoR: Asn83 (helix 2A) and Asn122 (helix 3A), of which Asn83 does show slight changes in solvent accessibility in ExoRG76C and ExoRS156Y, whereas analysis of solvent accessibility of Asn122 reveals variations in ExoRG76C and ExoRL81A. Based on the phenotypes of these two reduced-function mutants and the loss-of-function mutant, it is suggestive that Asn83 and Asn122 might indeed play a role in mediating protein-protein interactions in ExoR.

Mutations in ExoR affect the dipole moment of the protein

There is an asymmetric charge distribution on the surface of the ExoR wild-type protein that becomes even more pronounced in the modeled mutant proteins where the functional site A and concave face of the C terminus become more negatively charged compared to the wild-type protein. The mutations result in a bigger dipole moment and change in the orientation of the dipole in the modeled mutant proteins. Previous studies have shown that mutations that involve polar amino acids and/or changes in the size of the side chain, especially if the site of mutation is located at interface site,
can affect electrostatic properties of the protein such as the net charge or the dipole moment of the protein and influence steric interactions with other proteins [129]. The dipole may play an important role in directing the protein toward the correct binding site and partner [113], and delocalized electrostatic interactions allow for the interacting protein partners to stay in close proximity long enough to properly orient themselves toward the interface site [132]. Moreover, it is known that minor structural changes caused by point mutations coupled with changes in electrostatic profile interfere with protein complex formation [129, 132].

**The ExoR proteolysis site stays buried in the ExoR mutant proteins**

Mapping the experimentally determined cleavage site [20] on the modeled ExoR reveals that this site is buried and therefore, should not be accessible to periplasmic proteases in this native state. The residues at the experimentally determined cleavage site [20] also remained buried in all three ExoR mutant proteins. Moreover, the normal-mode analysis with ElNemo [141] also does not show conformational changes that could explain higher rates of proteolysis observed in ExoRL81A [20] and possible higher susceptibility of ExoRG76C and ExoRS156Y to periplasmic proteases [18]. Therefore, we propose a model in which conformational changes in the structure or of individual repeats would be required for proteolysis to occur. DegP, a periplasmic serine endoprotease in *Escherichia coli* recognizes three residues of the substrate protein and cleaves after a hydrophobic residue (in most cases Val, Ala, Ile, and Thr) that is almost completely buried or solvent inaccessible in most DegP substrates [143, 144, 145]. It has been suggested that a protein has to undergo a conformational change to make the cleavage site more surface accessible for the cleavage by DegP protease to take place [143]. While the protease involved in ExoR proteolysis has not been identified, *S. meliloti* Rm1021 does have a homolog of *E.coli*
DegP [146]. Even though it remains to be determined if the ExoR protein is digested by the DegP protease homolog, the presence of a solvent inaccessible hydrophobic residue in the vicinity of the experimentally determined ExoR cleavage site and the proteolysis of ExoR beyond the proposed proteolytic site fall in line with the proposed model of DegP cleavage [20, 143, 144, 145]. Experimental studies designed to ascertain if indeed a DegP is the protease responsible for ExoR cleavage would be an important step towards understanding the mechanism of ExoRm cleavage.

**Conclusion**

Our structural analysis of the experimentally characterized ExoR mutants (ExoRG76C, ExoRS156Y and ExoRL81A) provided some insight into their loss of stabilizing interactions with the ExoS protein [18] as well as higher susceptibility to proteolysis [18, 20]. We propose that alterations in the solvent accessibility of binding residues, in particular of Asn83 and Asn122, as well as changes in the electrostatic profiles of the mutant proteins destabilize complex formation between the ExoR and the ExoS proteins. Nevertheless, the structural examination of the modeled mutant proteins to explain higher susceptibility of these mutants to proteolysis [18, 20] has not produced a clear answer since the cleavage site remained buried. Therefore, we suggest that ExoR has to undergo a conformational change to make the cleavage site more surface accessible for the cleavage by DegP protease to take place [143]. It is possible that the small changes in the accessibility of the cleavage site observed in these mutant proteins might become significant upon dimer formation or interactions with other protein(s) and/or ligand(s).
Chapter V: Molecular modeling and computational analyses of the *Sinorhizobium meliloti* Rm1021 periplasmic domain of the ExoS protein

V.1. Comparative modeling of the periplasmic portion of ExoS

ExoS is a histidine kinase sensor protein that contains an N-terminal periplasmic sensor domain, two transmembrane domains (TM1 and TM2), and a C-terminal cytoplasmic domain that houses the HAMP, HisKA and HATPase C domains [8, 62, 63, 36] (Fig. V.1). We have generated homology models of the ExoS periplasmic-sensing domain, residues 68-278 (ExoS\(_p\)). Sequence-based fold prediction of the ExoS periplasmic sensor domain by Delta-Blast [59,60] and Pfam [37] indicate a stimulus-sensing domain but no specific domain was identified by SMART [36]. The closest structural match to the *S. meliloti* ExoS periplasmic domain was identified as the extracytoplasmic domain of the *Bacillus subtilis* PhoR sensor histidine kinase (PDB ID: 3CWF) [147] by the following fold recognition algorithms: FUGUE (z-score: 12.9) [77], FFAS03 (score -24.4) [91], and HHpred (E-value 1.9E-12) [72, 73]. Other histidine kinase sensor domains, DctB of *Vibrio cholerae* (PDB ID: 3BY9) [148] and of *Sinorhizobium meliloti* (PDB ID: 3E4P) [149], were identified as the closest structural match to the *S. meliloti* ExoS periplasmic domain by two other fold recognition algorithms SPARKS-X (z-score 4.96) [70] and pGenTHREADER (p-value: 7E-03) [76], respectively. All the structural templates identified by fold recognition algorithms are Per-ARNT-Sim (PAS) fold proteins.

Figure V.1. Schematic representation of the ExoS protein showing the location of two transmembrane domains (TM1 and TM2), periplasmic PAS domain (PAS\(_p\)), HAMP, HisKA and HATPase C domains.
The I-Tasser server [84, 85] generated the best three-dimensional structural representation of ExoS\textsubscript{p} based on various evaluation criteria. Analysis of the corresponding final model using ProSA-web [93, 94] generated low energy profiles (below zero) and a z-score of -7.55 that is comparable to solved NMR and X-ray crystal structures (Fig. V.2a). The sequence-structure fit in this model was validated via Verify3D [95, 96] and showed high 3D-1D profile scores (above 0) for the majority of the length of the model (Fig. V.2b). In addition, the model passed the checks of bond lengths (Z-score 0.873), bond angles (Z-score 1.004), chirality/dihedrals (Ramachandran Z-score -1.428), and chi-1/chi-2 angles (chi-1/chi-2 correlation Z-score 5.685) (all values are within expected ranges for well-refined structures) implemented in WHAT_CHECK [97].

Figure V.2. Evaluation plots of the three-dimensional model of the periplasmic domain of ExoS. (a) ProSA-web [93, 94] evaluation plot showing a z-score of -7.55 and the model’s energy profile. (b) Verify3D [95, 96] 3D-1D profile. The 3D-1D averaged scores plotted on the vertical axis (window size 21) are above 0 for the entire length of the generated model except for the H1 to H2 loop and helix H2 (residues 35 to 48).
V.2. Overall structure of the periplasmic domain of ExoS

The periplasmic domain of ExoS is an $\alpha/\beta$-structure with a central $\beta$-sheet having five antiparallel strands (S1 to S5) flanked on both sides by $\alpha$-helices. On one side of the beta-sheet there is an N-terminal helix H1 (residues 68-99) almost parallel to helix H3 (residues 132-146) at its C terminus and to the C-terminal helix H6 (residues 267-277) at its N terminus. Helices H1 and H3 are linked through helix H2 (residues 106-113) that is nearly perpendicular to these two helices. The N-terminal helix H1 and the C-terminal helix H6 are perpendicular to the bacterial inner membrane surface, however helix H6 is bent. It is likely that these two helices are the continuation of the transmembrane regions (TM1 and TM2) in the intact ExoS protein since they are positioned right next to the transmembrane regions, TM1 and TM2. The other helices, H4 (residues 188-202) and H5 (residues 218-228) as well as the long random coil regions between strand S2 and helix H4 (residues 165-187) and helices H4 and H5 (residues 203-217), lie on the other side of the $\beta$-sheet (Fig.V.3).
Figure V.3. Ribbon representation of the putative structure of the *S.* meliloti Rm1021 periplasmic domain of the ExoS protein (residues 68-278). The α-helices are numbered from H1 to H6 and the strands of the central β-sheet are numbered from S1 to S5. The periplasmic domain is oriented perpendicular to the bacterial inner membrane.

V.3. Structural similarities between ExoS<sub>p</sub> and sensor domains of other histidine kinases

A structure-based sequence alignment was performed on the ExoS sensing domain and on a selection of single and double-PAS domains (Fig. V.4). Proteins with single-PAS periplasmic domains include PhoR from *Bacillus subtilis* (PDB ID 3CWF) [147], CitA from *Klebsiella pneumoniae* (PDB ID 2V9A) [150], DcuS from *Escherichia coli* (PDB ID 3BY8) [151], and chemoreceptor T1pB from *Helicobacter pylori* (PDB ID 3UB6) [152]. Proteins with double-PAS periplasmic domains include histidine kinase 4 from *Arabidopsis thaliana* (PDB ID 3T4T) [153], DctB from *Vibrio cholerae* (PDB ID 3BY9) [148], DctB from *Sinorhizobium meliloti* (PDB ID 3E4P) [149], KinD from *Bacillus subtilis* (PDB
ID 4DAH=4JGP) [154], LuxQ from *Vibrio cholerae* (PDB ID 3C30) [155], LuxQ from *Vibrio harveyi* (PDB ID 2HJ9) [156], chemotaxis protein from *Vibrio cholerae* (PDB ID 3C8C) [157], mmHK1S-Z3 from *Methanosarcina mazei* (PDB ID 3LIB) [158], soHK1S-Z6 from *Shewanella oneidensis* (PDB ID 3LIC) [158], and vpHK1S-Z8 from *Vibrio parahaemolyticus* (PDB ID 3LIE) [158]. Despite low sequence similarity among the PAS-like domains (below 20%), they show high conservation of their secondary structure elements: the single-PAS sensors contain six $\alpha$-helices and a single beta sheet with five strands and the double-PAS sensors contain seven $\alpha$-helices and two five-stranded beta sheets, one in the membrane-distal domain and one in the membrane-proximal domain. However, it has to be noted that helices H2 and H4 are not present in all of the analyzed sensing domains (Fig. V.4). Despite its longer length of 211 residues, the sensing domain of ExoS conforms to the structure of single-PAS sensors based on the number and type of the structural elements present in the domain. The unusual length of this single PAS-like domain (almost twice the size of other single-sized PAS domains) is due to the elongated region between helices H1 and H3 as well as between S2 and H5 (Fig. V.3). These elongated structures unique to ExoS$_p$ produce relatively high root-mean-square deviation (RMSD) values (below 3Å) when ExoS$_p$ is superposed with single-PAS sensors.
Figure V.4. Structure-based multiple sequence alignment of the ExoS periplasmic domain and of single- (3CWF, 2V9A, 3BY8, 3UB6) and double-PAS (3T4T, 3BY9, 3E4P, 4DAH, 3C30, 2HJ9, 3C8C, 3LIB, 3LIC, 3LIE) domains. The RXYF motif is marked in bold. Secondary structure elements are specified at the bottom alignment based on the PSIPRED [46] calculation implemented in the PROMALS3D web server [64]. Helix residues are shown in red and strand residues are shown in blue. Numbers at the start and the end of sequences specify residues that correspond to the sensing domains of the PAS sensors used in the alignment.
V.4. Analysis of the electrostatic features of the periplasmic domain of ExoS

The surface electrostatic profile of the periplasmic domain of ExoS was examined to determine if electrostatic interactions could be the driving force for ExoS\textsubscript{p} to form homodimers and/or complexes with other proteins. ExoS\textsubscript{p} has a net charge of \(-11\) at pH 7 and a dipole moment of 774 Debye [113]. The dipole moment is oriented from the top of the sensing domain towards the membrane and crosses the beta sheet from the ligand-binding side towards the ExoS\textsubscript{p}-ExoS\textsubscript{p}’ dimerization site (Fig. V.5).

![Figure V.5. The putative structural model of the sensing domain of ExoS showing the dipole moment vector calculated using Protein Dipole Moments Server [113].](image)

The electrostatic profile of the model of the ExoS\textsubscript{p} is overall negative with regions of mild acidic surface charge interspersed with surface hydrophobic and positively charged residues (H1: R70, R77; H3: R142, R143; H3 to S1 loop: R149; S1: R151, R153; S2 to H4 loop: R176; helix 4: R197; H4 to
H5 loop: K210; S3: Arg236; S4 to S5 loop: Arg253, Arg254; H6: K272; Arg278) (Fig. V.6).

\[ \text{Fig. V.6. Electrostatic potentials mapped to the molecular surface of the} \]
\[ \text{sensing domain of ExoS. The ExoS}_p \text{ protein is oriented to visualize the ExoS-} \]
\[ \text{ExoS'} \text{ interface in (a) and the putative ExoS}_p\text{-ExoR interface in (b). The} \]
\[ \text{surface potentials are color-graded from } -4 \text{ kT/e (red) to } +4 \text{ kT/e (blue).} \]

V.6. Identification of the protein-protein interaction sites in the
periplasmic domain of ExoS

Since ExoS\textsubscript{p} is known to form a complex with ExoR\textsubscript{m} [18] and the
functional units of many of the two-component sensor kinases are dimeric
[156, 159, 160], the periplasmic domain of ExoS sequence and its modeled
three-dimensional structure were examined for the presence of protein-protein
interaction sites. The candidate interface residues were found on the surface
of helices and loops on both sides of the sheet implying two non-overlapping
interaction sites (Fig. V.7). Helices 1 and 4 show highest number of
candidate interface residues, 19 and 14 respectively. None of these residues
within helix 1 were identified as hot spot residues. Within helix 4, three
residues were identified as protein-protein interaction hot spots Glu192,
Asn198, and Arg199. Several protein-protein interaction hot spots were identified within the loop between H1 and H2 (Thr100, Asn101, and Ser102), helix 2 (Glu108), the loop between H2 and H3 (Glu117, Asp125), helix 3 (Arg143), the loop between H3 and S1 (Arg149), strand 1 (Arg151), the loop between S2 and H4 (Ser170, Leu175, Phe177, Pro184, Glu185), the loop between H4 and H5 (Pro203, Tyr209, Lys210, Glu211), the loop between S3 and S4 (Glu239, Lys240, Gly241), and the loop between S4 and S5 (Arg253, Phe254, Arg255).

**Figure V.7.** Location of the putative protein-protein interaction sites within the sensing domain of ExoS. (a) Putative ExoS<sub>p</sub>-ExoS<sub>p′</sub>dimer interface (blue). (b) Putative ExoSp-ExoR′ complex interface (red). The figures are oriented in the same orientation as in figure V.6. Residues identified as interface hot spot residues are shown in black.

**V.7. Docking analysis of the periplasmic domain of ExoS**

We generated ExoS<sub>p</sub>-ExoS<sub>p′</sub> homodimers using ClusPro [122,123] and scrutinized the differences in the molecular architecture of the predicted homodimers. Previous studies [148, 150, 158] have demonstrated that the most likely functional homodimer interfaces of the periplasmic domains of histidine kinases are interfaces involving N-terminal helices. Therefore,
this class of predicted homodimers was chosen for further analysis. One of the ExoS\textsubscript{p}-ExoS\textsubscript{p}' homodimers was selected as the best representation based on the fact that the homodimer is the most symmetric showing similar change in the surface area of the interface for both chains (1363Å\textsuperscript{2} and 1317Å\textsuperscript{2}). The N-terminal helices from both monomers are perpendicular to the putative membrane, and most of the identified interface residues were recognized as candidate interface residues or protein-protein interaction hot spots on the ExoS monomer. In this interface, each of the two helices H1 on both sides of the interface is joined by helix H3 at their C terminus and by helix H6 at their N terminus. In addition, residues found in the loop regions of H1 to H2, H3 to S1, and S2 to H4 are part of the homodimer interface (Fig. V.8).

**Figure V.8.** Ribbon diagram of the ExoS\textsubscript{p} dimer showing packing between helices H1, H3, and H6 of opposite monomers. Individual monomers are colored blue and green with interface residues shown in ball-and-stick representation. The two monomers are oriented perpendicular to the bacterial inner membrane.
V.8. Overall structure of the ExoS\textsubscript{p}-ExoR complex

To visualize the interaction between ExoR and ExoS\textsubscript{p}, we modeled the structure of this heterodimer. Analysis of the generated ExoS\textsubscript{p}-ExoR complexes revealed several distinct candidate ExoS\textsubscript{p}R interfaces. In the first type, referred to as interface A, (e.g. between ExoS\textsubscript{p} H4 and ExoR site A), the interface is mainly built by the helix H4, and the loops flanking helix 4: H3 to H4 and H4 to H5 of ExoS\textsubscript{p} and the putative protein-protein interaction site A of ExoR. Additional interactions in the H2 to H3 loop of ExoS\textsubscript{p} stabilize the C-terminal region of ExoR (Fig.V.9a). This complex interface buries 1437 Å\textsuperscript{2} and 1401 Å\textsuperscript{2} of accessible surface of ExoS\textsubscript{p} and ExoR, respectively.

Within the next putative ExoS\textsubscript{p}-ExoR interface, referred to as interface B, (e.g. between the ExoS\textsubscript{p} H5 to S3 loop and ExoR site A), there is one main contact region located between helix H1, strands S3 and S4, and loops H5 to S3 and S4 to S5 of ExoS\textsubscript{p} and the putative protein-protein interaction site A of ExoR. Additional interactions between residues located within the H2 to H3 loop of ExoS\textsubscript{p} and Asp265 of ExoR stabilize the C-terminus of ExoR (Fig.V.9b). This ExoS\textsubscript{p}-ExoR interface buries 1415 Å\textsuperscript{2} and 1374 Å\textsuperscript{2} of accessible surface of ExoS\textsubscript{p} and ExoR, respectively.

The dominant feature of the third heterodimer interface, referred to as interface C, (e.g. between ExoS\textsubscript{p} H4 and C-terminus of ExoR) is the association of helix H4, its flanking loop regions, and the H2 to H3 loop of ExoS\textsubscript{p} with the C-terminus of ExoR and its putative protein-protein interaction site B. In addition, ExoR2 and ExoR3 intra-repeat loops are stabilized through interactions with the S4 to S5 loop as well as with helix H5 and the S1 to S2 loop of ExoS\textsubscript{p}, respectively. The N-terminal end of ExoR is not stabilized by ExoS\textsubscript{p} interactions (Fig.V.9c). This heterodimer formation decreases the total accessible surface of ExoS\textsubscript{p} by 1970 Å\textsuperscript{2} and of ExoR by 1891 Å\textsuperscript{2}.

The fourth heterodimer interface (referred to as interface D), the electrostatic-favored heterodimer interface (e.g. between ExoS\textsubscript{p} H4 and the
positive patch on the ExoR convex surface) involves the association of helix H4 and loops H2 to H3, H3 to H4, and H4 to H5 of ExoS\(_p\) and mainly involves residues within helices B of ExoR4, ExoR5, and ExoR6 of ExoR that include the positive patch on the convex surface of ExoR and its putative binding site C (Fig.V.9d). This complex formation decreases the accessible surface by 1180 Å\(^2\) in ExoS\(_p\) and by 1091 Å\(^2\) in ExoR.

Each of the four different ExoR-ExoS\(_p\) complexes was dimerized, and the interface between the ExoR-ExoS\(_p\): ExoS\(_p\)’-ExoR’ complex was found to involve N-terminal helices H1 and H3, and C-terminal helix H6 of ExoS. The tetramer interface interactions are the same as observed for the ExoS\(_p\)-ExoS\(_p\)’ homodimer interactions.

To investigate other alternate possibilities of interaction modes between ExoR and ExoS\(_p\), we investigated the ExoS\(_p\)-ExoS\(_p\)’:ExoR complex (interface E). The characteristic of this trimer interface is the association of the ExoR protein with both ExoS\(_p\) monomers. The putative protein-protein interaction site A of ExoR interacts with helix H4 of ExoS\(_p\). In addition, the putative site C and the part of the positively charged patch on the convex surface of ExoR is stabilized through the interactions with the H5 to S3 and S4 to S5 loops as well as with strand S3 and helix H1 of ExoS\(_p\)’ (Fig.V.9e). This trimer formation decreases the accessible surface of ExoS\(_p\) by 2106 Å\(^2\), ExoS\(_p\)’ by 2247 Å\(^2\), and of ExoR by 1737 Å\(^2\).
Figure V.9. ExoS\textsubscript{p}-ExoR complex organization (a) interface A (b) interface B (c) interface C (d) interface D (e) interface E. ExoS\textsubscript{p} is colored blue and ExoR green. The complexes are oriented perpendicular to the bacterial inner membrane. Helix H4 of ExoS\textsubscript{p} and N terminus of ExoR are labeled in each model.

V.9. Discussion

The histidine kinase, ExoS, along with its response regulator, ChvI, forms a two-component system important for sensing and adapting to environmental changes [8,16]. Here we present the three-dimensional model of the ExoS periplasmic sensing domain and the implications of the ExoS\textsubscript{p} fold for the formation of ExoS homodimers and ExoS\textsubscript{p}-ExoR complexes.

The \textit{S. meliloti} ExoS periplasmic domain reveals a PAS fold

The three-dimensional model of the \textit{S. meliloti} ExoS periplasmic domain suggests a PAS-like fold with a central $\beta$-sheet enclosed by $\alpha$-helices. The
structure-based sequence alignment of the ExoSₚ with PAS-domain proteins further supports a single-sized PAS structure for the sensing domain of ExoSₚ. This is in agreement with previous studies that most of the periplasmic domains of histidine kinases take on a single or double PAS-like fold [158, 161]. The structure-based sequence alignment as well as the superposition of ExoSₚ with PAS domain proteins reveals structural elements that are common to single-sized PAS-like domains although the size of the ExoS sensing domain is almost twice the size of single-sized PAS domains. The unusual length for single-sized PAS domain is due to structural features that are unique to ExoSₚ. One of the distinctive features is the elongated loop region that encloses helix H2. Additional differences between ExoSₚ and other PAS sensors include an elongated loop between S2 and H4 and helix H4. Helix H4 is also present in LuxQ from V. harvey (PDB ID 1ZHH) [162], DcuS from E. coli (PDB ID 3BY8) [Cheung and Hendrickson 2008], CitA from Klebsiella pneumoniae (PDB ID 2V9A) [150], and in TlpB from H. pylori (PDB ID 3UB6) [152] but is much shorter, a few residues compared to thirteen residues in ExoSₚ. The structural differences seen in ExoSₚ fall in line with previous observations that the presence of the β-sheet in PAS-like domains is strongly conserved [163], whereas the length and the number of α-helices surrounding the central β-sheet vary considerably [164].

Docking analysis of the periplasmic domain of ExoS

Previous studies have demonstrated that the functional units of many of the two-component sensor kinases consist of a dimer and that the dimerization depends on N-terminal helices [149, 150, 151, 156, 158, 159, 160, 164]. Through our docking analysis of ExoSₚ we were able to generate ExoSₚ homodimers that show a dimerization mode similar to that observed in crystal structures of the sensing domains of other histidine kinases. The ExoSₚ dimer interface involves N-terminal helix H1, helix H3 and C-terminal helix H6. In
the proposed model of the ExoS<sub>p</sub> dimer, the C-terminal helix H6 is part of the interface but not found in other dimer interfaces of periplasmic PAS domains. However, Zhang and Hendrickson [158] speculate that the C-terminal helices might be a part of the dimer interface in intact receptors since they are oriented toward the N-terminal helices.

The dimer interface between the ExoS<sub>p</sub> monomers is overall negatively charged. Our observation is in agreement with the nature of the dimer interfaces found in extracellular domains of other histidine kinases such as mmHK1S-Z2 from *Methanosarcina mazei* (PDB ID 3LIA), mmHK1S-Z3 from *Methanosarcina mazei* (PDB ID 3LIB), and soHK1S-Z6 from *Shewanella oneidensis* (PDB ID 3LIC) [158] as well as in CitA from *Klebsiella pneumoniae* (PDB ID 2V9A) [150]. It has been suggested that the hydrophilic nature of the dimer interfaces of the periplasmic domains of histidine kinases allows for a dynamic character to the interface and permits for the ligand-induced structural changes in the sensing domain to transmit the signal across the membrane [158, 150].

**Overall structure of the ExoS<sub>p</sub>-ExoR complex**

Similar to LuxQ [162], ExoS does not bind its signaling molecule (yet to be identified) directly, but forms a complex with the periplasmic binding protein, ExoR [18]. To investigate the possible interactions between ExoS<sub>p</sub> and ExoR, we generated ExoS<sub>p</sub>-ExoR heterodimers that produced several distinct candidate ExoS<sub>p</sub>-ExoR interfaces.

The first two types of the ExoS<sub>p</sub>-ExoR dimers (interfaces A and B) involve putative protein-protein interaction site A and the C-terminus of ExoR and either helix H4 or the β-sheet and helix H1 of ExoS<sub>p</sub>. The third heterodimer interface (interface C) mainly shows association of ExoS<sub>p</sub> helix H4 with the C-terminus of ExoR and its putative protein-protein interaction site B. The electrostatically favored heterodimer interface (interface D) involves
the association of helix H4 of ExoS<sub>p</sub> and the positive patch on the convex surface of ExoR and its putative binding site C. All generated dimers were computationally shown to form ExoS<sub>p</sub>-ExoR:ExoS<sub>p</sub>′-ExoR′ tetramers. Another type of interactions between ExoR and ExoS<sub>p</sub> might involve the formation of the ExoS<sub>p</sub>-ExoS<sub>p</sub>′:ExoR trimeric complex where the ExoR protein is complexed with both ExoS<sub>p</sub> proteins.

Ligand-binding sites in the periplasmic PAS domains involve the side of the central β-sheet as well as the loops between S2 and H4, H4 to H5, S3 to S4 and helices H4 and H5; for example, ligand-binding sites of malate in DcuS from <i>E.coli</i> (3BY8) [151], succinate in DctB from <i>V.cholerae</i> (3BY9) [151] and from <i>S.meliloti</i> (3E40) [149], citrate [150] and citrate and MoO<sub>3</sub> [164] in CitA from <i>Klebsiella pneumoniae</i> (2V9A and 1P0Z), bistris in mmHK1S-Z2 (3LIA), ethylene glycol in soHK1S-Z6 (3LIC), phosphate in vpHK1S-Z8 (3LIE) [158], and urea in T1pB from <i>H.pylori</i> (3UB6) [152]. The above listed PAS-domain proteins bind their ligands directly. On the other hand, LuxQ from <i>V.harveyi</i> employs another protein (LuxP) for ligand binding. The interface between LuxQ and LuxP (PDB ID 1ZHH) in addition to the central β-sheet includes loops between S1 and S2, S4 and S5, as well as the F-G loop (H5 to S3 loop) [162].

Even though the ligand-binding site is found in similar locations in PAS-domain proteins, the binding residues show high variability due to the fact that the binding site has to accommodate diverse ligands [158]. The highest variability is observed in the length and the sequence of the region between S2 and H5 [158], which in some cases also includes helix H4. On the other hand, the RXYF motif at the beginning of helix H5 has been identified as the most conserved and therefore, important motif for ligand recognition (Fig. V.4) [158]. However, this motif is absent from LuxQ sensors due to the different mode of ligand-binding (LuxQ employs periplasmic binding protein LuxP to bind its signaling molecule) [158] and from sensors with single PAS-like domains included in our analysis as well as from ExoS<sub>p</sub> (Fig. V.4). In the
case of LuxQ, it was shown that the F-G loop (H5 to S3 loop) is critical for LuxQp-LuxP interactions [162]. Association of ExoS_p with ExoR, an elongated solenoid protein, might require a different approach to recognize and bind this protein partner compared to binding of smaller ligands or the LuxP protein that has a more compact fold of two parallel α/β/α domains [165]. Therefore, the unique elongated random coil regions between S2 and H4 as well as between H4 and H5 of ExoS_p might provide flexibility, while helix H4 offers a more rigid scaffold needed to accommodate binding of this elongated protein.

Although it is not possible to determine which of the generated ExoS_p-ExoR interfaces is physiologically relevant based on theoretical data alone, we hypothesize that the first type of the analyzed heterodimer (interface A) (Fig. V.9a) is the most promising model for the ExoS_p-ExoR interactions. In this interface helix H4 and its surrounding random coil regions of ExoS_p interact with the putative protein-protein interaction site A of ExoR, whereas the H2 to H3 loop of ExoS_p stabilizes the C-terminal region of ExoR. The interactions seen in ExoR in this dimer interface are supported by the interactions observed in another Sell-like repeat protein, HcpC [32], and interactions that involve mainly helix H4 of ExoS_p are in agreement with the predicted protein-protein interaction sites within ExoS_p (Fig. V.7). It is possible that the direct interaction of helix H4, its flanking loop regions, and the loop between H2 to H3 (the elongated nature of these secondary elements is unique to ExoS_p) with ExoR might cause conformational changes necessary for signal transduction. It can be speculated that the ExoS_p-ExoR interactions might change the ExoS_p-ExoS_p′ dimerization interface similar to the changes observed in another sensing domain of the histidine kinase, DctB. In DctB the movement of the loop between S2 and H5 is caused by ligand binding and leads to rearrangement in the dimer interface that leads to the transfer of a signal across the membrane [149].
Conclusion

In conclusion, we present for the first time a three-dimensional representation of *S. meliloti* Rm1021 ExoS periplasmic domain generated through computational methods. Our model suggests the Per-ARNT-Sim (PAS)-like fold for ExoS$_p$. Even though ExoS$_p$ shares a similar overall structure as other PAS domains, it does not bind small ligands and therefore probably uses a different mechanism to recognize and bind its periplasmic binding protein, ExoR. Elongated structural elements that are unique to ExoS$_p$ might provide flexibility to this fold to accommodate binding of ExoR that is an elongated multi-repeat protein. Further studies including experimental analysis that target key proposed functional sites of ExoS$_p$ and interface of the ExoS$_p$-ExoR complexes can help validate the ExoS$_p$ structure and the essential residues employed in ExoS$_p$-ExoR interactions.
Bibliography


17. Wells DH, Chen EJ, Fisher RF, Long SR. 2007 ExoR is genetically coupled to the ExoS-ChvI two-component system and located in the periplasm of


eds. Proc. Sixth Int. Conf. on Intelligent Systems for Molecular Biology, pp 175-182. AAAI Press.


98. The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.


131. Wu CF, Lin JS, Shaw GC, Lai EM. 2012 Acid-induced type VI secretion
system is regulated by ExoR-ChvG/ChvI signaling cascade in Agrobacterium tumefaciens. PLoS Pathog. 8, e1002938. (doi:10.1371/journal.ppat.1002938)


139. Vanderlinde EM, Yost CK. 2012 Mutation of the sensor kinase chvG in Rhizobium leguminosarum negatively impacts cellular metabolism, outer
(doi:10.1128/JB.06357-11)

(doi:10.1371/journal.pone.0010216)


155. Slama B, Hendrickson W. Crystal structure of the Vibrio Cholerae LuxQ
periplasmic domain (SeMet), to be published


157. Patskovsky et al., Crystal structure of Mcp_N and cache N-terminal domains of methyl-accepting chemotaxis protein from Vibrio cholerae, to be published


165. Fox NK, Brenner SE, Chandonia JM. 2013. SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of