

# Molecular modeling and computational analyses suggests that the *Sinorhizobium meliloti* periplasmic regulator protein ExoR adopts a superhelical fold and is controlled by a unique mechanism of proteolysis

Eliza M. Wiech,<sup>1,2</sup> Hai-Ping Cheng,<sup>1,3</sup> and Shaneen M. Singh<sup>1,2\*</sup>

<sup>1</sup>Department of Biology, The Graduate Center of the City University of New York, New York, New York 10016

<sup>2</sup>Department of Biology, Brooklyn College, The City University of New York, Brooklyn, New York 11210

<sup>3</sup>Biological Sciences Department, Lehman College, The City University of New York, Bronx, New York 10468

Received 8 July 2014; Revised 26 November 2014; Accepted 1 December 2014

DOI: 10.1002/pro.2616

Published online 9 December 2014 proteinscience.org

**Abstract:** 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 believed to drive the key regulatory step in the invasion process, remains a major challenge. In this study, we present a theoretical structural model of the active form of ExoR protein, ExoR<sub>m</sub>, generated using computational methods. Our model suggests that ExoR possesses a super-helical fold comprising 12  $\alpha$ -helices forming six Sel1-like repeats, including two that were unidentified in previous studies. This fold is highly conducive to mediating protein–protein interactions and this is corroborated by the identification of putative protein binding sites on the surface of the ExoR<sub>m</sub> protein. Our studies reveal two novel insights: (a) an extended conformation of the third Sel1-like repeat that might be important for ExoR regulatory function and (b) a buried proteolytic site that implies a unique proteolytic mechanism. This study provides new and interesting insights into the structure of *S. meliloti* ExoR, lays the groundwork for elaborating the molecular mechanism of ExoR<sub>m</sub> cleavage, ExoR<sub>m</sub>–ExoS interactions, and studies of ExoR homologs in other bacterial host interactions.

**Keywords:** ExoR; *Sinorhizobium meliloti* Rm1021; sel1-like repeats; superhelical fold; molecular modeling; computational analyses; ExoS

**Abbreviations:** RMSD, root-mean-square deviation; RSI, ExoR–ExoS/ChvI signal transduction pathway; SLR, sel1-like repeat

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Institute of Health of U.S.; Grant number: SGM081147.

\*Correspondence to: Shaneen M. Singh, Department of Biology, Brooklyn College, The City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210. E-mail: ssingh@brooklyn.cuny.edu

## Introduction

Most bacteria, including parasitic and symbiotic species, rely on two-component signal transduction systems for detecting and adapting to changes in their environment.<sup>1</sup> A group of Gram-negative bacteria rely on the ExoR–ExoS/ChvI (RSI) pathway, to transit from a free-living to host-invading form.<sup>2</sup> This pathway is best understood in *Sinorhizobium meliloti*, the model organism for bacterium–plant

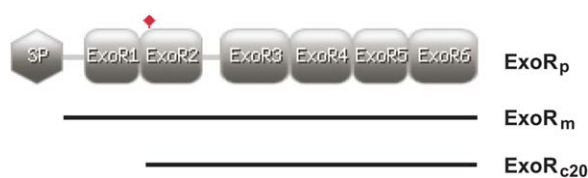
symbiosis. ExoS and ChvI form a typical two-component system that consists of a membrane-integral histidine kinase, ExoS, and an associated cytoplasmic response regulator, ChvI.<sup>3</sup> In *S. meliloti*, the activities of the ExoS/ChvI system are regulated by a periplasmic regulatory protein, ExoR, through a direct interaction with ExoS.<sup>4</sup> The current model for the ExoR-ExoS/ChvI 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<sub>m</sub>.<sup>5,6</sup> In the ExoR<sub>m</sub>-ExoS complex, ExoS acts as a phosphatase and keeps ChvI dephosphorylated and inactive. When the ExoR<sub>m</sub>-ExoS interaction is disrupted through the proteolytic cleavage of ExoR<sub>m</sub>, ExoS becomes an active kinase, and phosphorylates ChvI directly,<sup>3,6</sup> resulting in upregulation of succinoglycan biosynthesis and repression of flagellum biosynthesis, allowing the cells to switch from a free-living to host-invading form.<sup>2,4-6</sup>

Even though ExoR has been established as a key regulator of the RSI pathway,<sup>4-7</sup> its tertiary structure, including structural details and associated functions, remains unknown. Delineating the structure-function correlations of the ExoR protein is critical to understanding the cleavage mechanism of ExoR<sub>m</sub>, ExoR<sub>m</sub>-ExoS interactions, and other aspects of its regulatory role in the RSI pathway.<sup>5,6</sup> Comparative modeling methods have been successfully applied in revealing key structural details, structure-function relationships, and interactions with putative binding partners, including those of helical repeat proteins (e.g., repeats of proteasome-binding protein PA200<sup>8</sup> and repeats of transcriptional activator-like [TAL] effectors),<sup>9</sup> confirmed by subsequent crystal structures of TAL effector-DNA complexes.<sup>10</sup> In this study, we present a robust theoretical model of the active form of ExoR protein, ExoR<sub>m</sub>, generated by means of the well-established approach of template-based modeling.<sup>11</sup> We propose that the ExoR protein adopts an alpha-alpha superhelical fold, an inherently flexible fold designed for protein-protein interactions.<sup>12,13</sup> In the absence of any structural representations of the ExoR family of proteins in the structural database, our study provides a first look into the structure of ExoR and its implications to ExoR function.

## Results

### *S. meliloti* Rm1021 ExoR protein is composed of six sel1-like repeats

In addition to the four Sel1-like repeats already identified in ExoR,<sup>14</sup> our analysis of the ExoR sequence with the latest repeat detection methods reveal two additional repeats at the C terminus of ExoR, corroborated by secondary structure prediction (Supporting Information Table SI; Supporting Information Fig. S1; Fig. 1). The length of each of



**Figure 1.** 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 Sel1-like repeats (ExoR1-ExoR6), and the cleavage site of ExoR, A80 (flag-tagged).

the six repeats ranges from 32 to 40 residues and consists of two  $\alpha$ -helices; this repeat structure is in agreement with the expected configuration of Sel1-like repeats.<sup>12,14</sup> The alignment of all six repeats reveals expected conserved alanine and glycine residues based on the Sel1-like repeat consensus except for ExoR6 that maintains only three out of seven conserved structural residues (Fig. 2).<sup>12,14</sup>

## Comparative modeling of *S. meliloti*

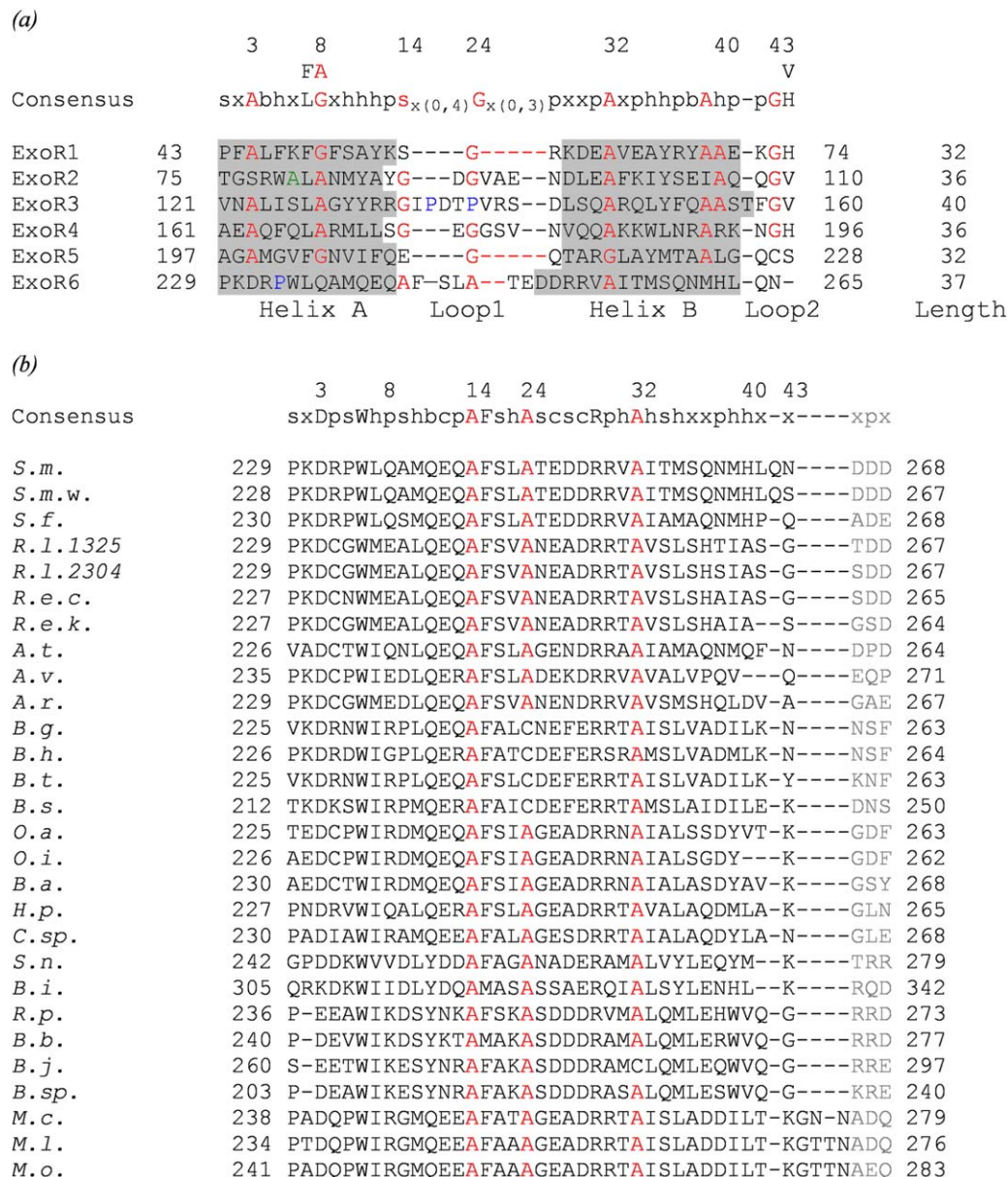
### ExoR<sub>m</sub> protein

The closest structural matches to the *S. meliloti* ExoR protein were found to be *Helicobacter pylori* cysteine-rich protein C (HcpC, PDB ID:1OUV)<sup>12</sup> and a protein corresponding to locus C5321 of *Escherichia coli* strain Cft073 (PDB ID:4BWR),<sup>15</sup> both of which are Sel1-like repeat containing proteins, using fold recognition algorithms (Supporting Information Table SII).

Out of approximately 200 ExoR models built, the structural ExoR<sub>m</sub> model generated by I-Tasser server<sup>16</sup> was selected as the best structural representation from a group of top-ranked models (Supporting Information Table SIII). Energy profiles calculated using ProSA-web<sup>17</sup> for the selected model showed low energies and a z-score of -6.99 comparable to solved structures. The evaluation via Verify3D<sup>18</sup> shows high 1D-3D profile scores for almost the entire length of the model (Supporting Information Fig. S2). In addition, the model passed the checks of various stereochemical parameters implemented in WHAT\_CHECK<sup>19</sup> (Supporting Information Table SIV) and an overall favorable Ramachandran plot<sup>20</sup> (Supporting Information Fig. S3).

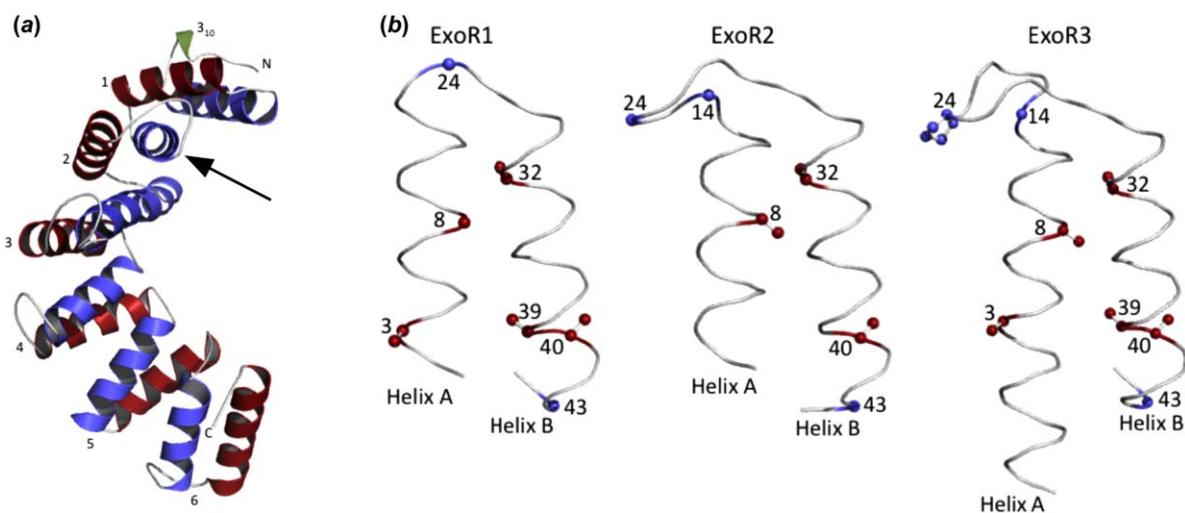
### Tertiary model of *S. meliloti* ExoR<sub>m</sub> protein shows a superhelical fold

The ExoR<sub>m</sub> sequence adopts a superhelical fold consisting of 12  $\alpha$ -helices forming six Sel1-like repeats preceded by a small  $3_{10}$  helix at its N terminus with high structural similarity to the 1OUV template<sup>12</sup> (Supporting Information Table SV). Each repeat is formed by two antiparallel helices: A (N-terminal) and B (C-terminal), located at the inner (concave) and outer (convex) surfaces of the superhelix, respectively [Fig. 3(a)]. Individual repeats are structurally similar and superpose with low root-mean-



**Figure 2.** (a) Alignment of repeats found in the ExoR protein of *Sinorhizobium meliloti* Rm1021. Residues located in  $\alpha$ -helices are shaded gray. The cleavage site (A80) is indicated in green. Proline residues in the ExoR3 loop 1 and helix A of ExoR6 are colored blue. (b) Sequence alignment of the ExoR C termini 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 that includes ExoR6 (black) and the ExoR Asp<sub>3</sub> tag (gray). The following species are shown: *S. 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. BTai1 (*B.sp.*), *Mesorhizobium ciceri* biovar *biserrulae* WSM1271 (*M.c.*), *Mesorhizobium loti* MAFF303099 (*M.l.*), and *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); numbering indicated above the consensus sequences follows the convention of Mittl and Schneider-Brachert.<sup>14</sup> The numbering at the beginning and end of each repeat corresponds to the residue number of full length ExoR. The conserved structural residues common to the SLR consensus sequence<sup>12,14</sup> are colored red. Residues are abbreviated as follows: s, small; b, big; h, hydrophobic; p, polar; c, charged; x, any residue.





**Figure 3.** (a) Ribbon representation of the *S. meliloti* Rm1021 ExoR<sub>m</sub> model (residues 31–268). The repeats are numbered from 1 to 6 and helix A and helix B of each repeat are colored blue and red, respectively. The N-terminal 3<sub>10</sub> helix is colored green. The cleavage site is indicated with an arrow. (b) Backbone trace of the first three repeats of ExoR. Conserved structural residues responsible for tight packing of helices are colored red and those responsible for sharp turns in loop regions are colored blue. The numbers represent the positions of the amino acid residues in repeats.

square deviation (RMSD) values below 1 Å except ExoR3 (Table I). The marginally higher RMSD values (0.9–1.3 Å) for ExoR3 can be ascribed to its extended helix A and elongated loop region containing two proline residues absent from the other repeats (Supporting Information Fig. S1). The helical structure of the ExoR repeats follows the expected positioning of conserved structural residues characteristic of Sel1-like proteins: residues 14, 24, and 43 that dictate sharp turns are present in loop regions, whereas residues responsible for the tight packing of the repeats at positions 3, 8, 32, 39, and 40 are located within  $\alpha$ -helices [Figs. 2(a) and 3(b)].<sup>12,14</sup> The packing of the ExoR repeats is also conserved: the average inter-repeat helix-packing angle in the ExoR model is 42 ( $\pm$  2.6) $^\circ$  and the average intra-repeat helix-packing angle is 18 ( $\pm$  2.9) $^\circ$ .

### Identification of the ExoR<sub>m</sub> protein–protein interaction sites

Analysis of the ExoR structural model predicted three non-overlapping binding sites: A, B, and C. Site A, composed of 36 residues, is located at the inner face of the N-terminal end of the protein and encompasses

the ExoR proteolytic site. Site B that is found at the center of the inner face of the protein and extends to the C terminus is formed by 21 residues. The putative interaction site C is formed by eight residues located on the C-terminal convex surface of ExoR<sub>m</sub> (Fig. 4(a–c)); Supporting Information Table SVI).

### Analysis of the electrostatic features of ExoR<sub>m</sub>

The surface electrostatic profile of the modeled ExoR<sub>m</sub> protein was examined to further analyze the biophysical features of ExoR that may drive interactions with its binding partners. The modeled ExoR<sub>m</sub> has a net charge of  $-3$  at pH 7 and a dipole moment of 584 Debye oriented from the inside toward the outside of the superhelix. The concave face of the ExoR superhelix is highly negatively charged with acidic patches at the putative protein–protein interaction site A and the C terminus. The convex face of ExoR is predominantly hydrophobic with a prominent basic patch (Fig. 5).

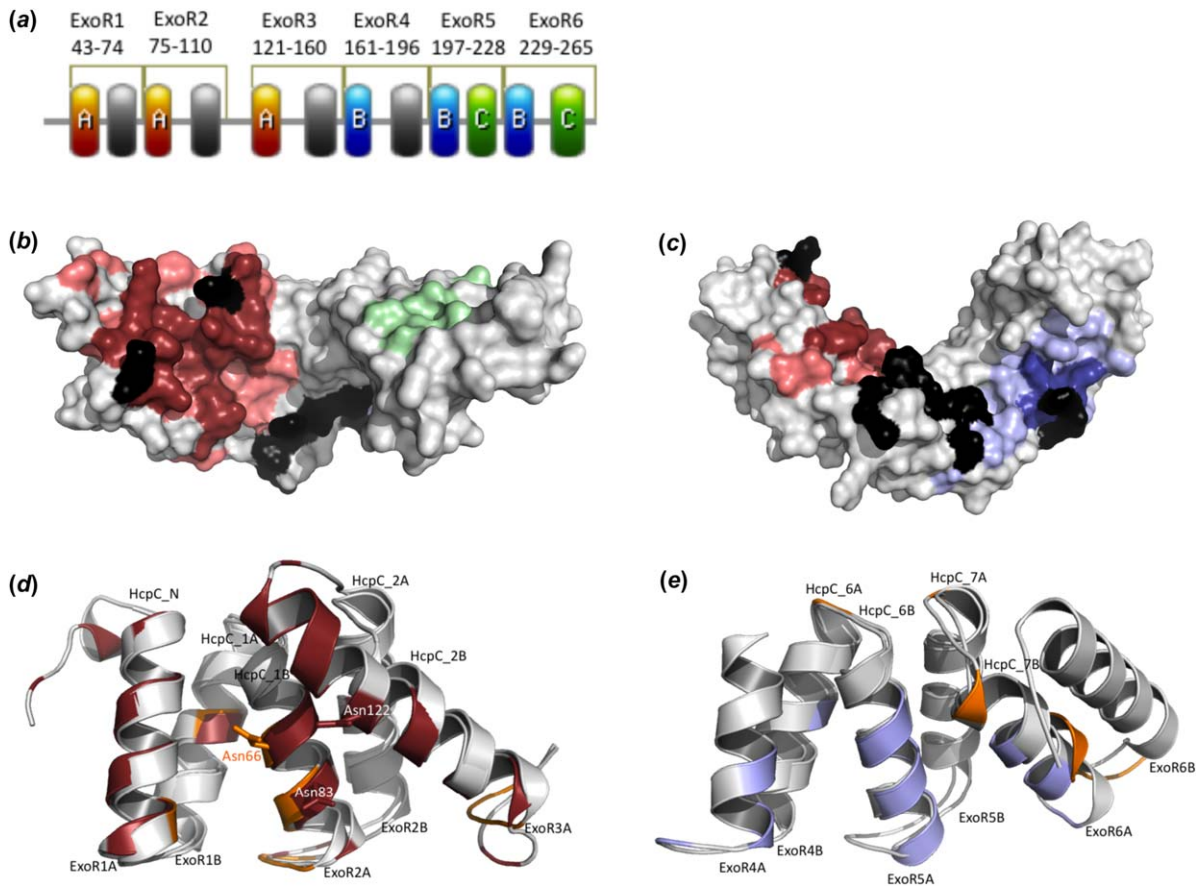
### Analysis of the accessibility of the cleavage site

Because the proteolysis of the ExoR<sub>m</sub> protein has been implicated as a key step in regulating its

**Table I.** RMSD Values for the Superposition of ExoR Repeats and Their Sequence Identity (%)

	ExoR1 43–74	ExoR2 75–110	ExoR3 114–160	ExoR4 161–196	ExoR5 197–228	ExoR6 229–265
ExoR1	—	16	22	19	16	7
ExoR2	<b>0.7</b>	—	28	25	9	17
ExoR3	<b>0.9</b>	<b>1.1</b>	—	22	6	9
ExoR4	<b>0.5</b>	<b>0.6</b>	<b>1.1</b>	—	9	9
ExoR5	<b>0.5</b>	<b>0.7</b>	<b>1.0</b>	<b>0.5</b>	—	7
ExoR6	<b>0.7</b>	<b>0.7</b>	<b>1.3</b>	<b>0.6</b>	<b>0.7</b>	—

The RMSD values (in bold) were calculated using TM-align server.<sup>36</sup> The numbering corresponds to the residues of the full length ExoR, ExoR<sub>p</sub>.

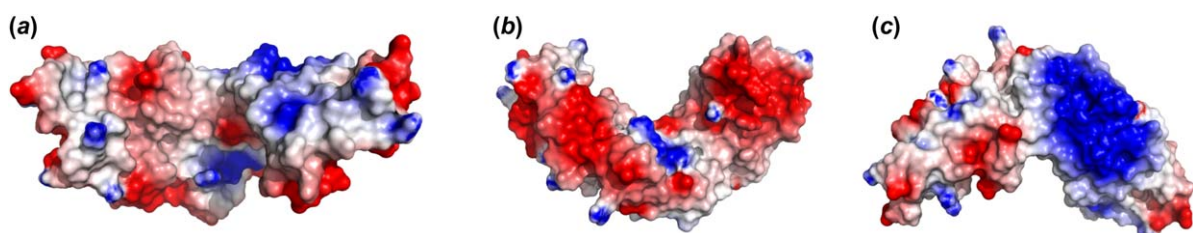


**Figure 4.** (a) Location of putative protein–protein interaction sites. The sites A, B, and C are colored red, blue, and green, respectively. (b, c) Putative functional sites mapped to the surface of ExoR<sub>m</sub> protein. Site A (residues identified by more than one prediction program in red; residues 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<sup>12</sup> 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<sup>12</sup> and site B (blue) of ExoR. The superposition is based on residues 225–292 from HcpC and residues 162–268 from ExoR.

function, the experimentally determined proteolysis site<sup>6</sup> was mapped on the modeled structure of ExoR [Fig. 3(a)]. Surface accessibility of Ala80 and Leu81 in the modeled ExoR<sub>m</sub> protein corroborates the sequence-based surface accessibility predictions and suggests that both residues are not surface exposed (16% surface accessible area for Ala80 and 0% for Leu81).

## Discussion

ExoR 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 leguminous host, alfalfa.<sup>2–7</sup> Similarly, homologs of ExoR, ExoS, and ChvI in *Agrobacterium tumefaciens* and *Brucella abortus* are essential for host invasion by these plant and animal



**Figure 5.** Electrostatic potentials mapped to the molecular surface of ExoR<sub>m</sub>. 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 4. 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).

pathogens.<sup>21,22</sup> Yet, a major gap in knowledge exists in understanding ExoR functionality because no solved ExoR structure exists. This study attempts to bridge this gap, providing a robust theoretical structural model of *S. meliloti* Rm1021 ExoR<sub>m</sub>, providing a first look into the details of the structural fold of this protein and predicting its implications to ExoS/ChvI two-component signaling.

Our analysis of the ExoR sequence suggests that it houses six Sel1-like repeats, in contrast to previous studies that report only four.<sup>14</sup> Of these, we classify ExoR6 as a nontraditional low conservation repeat. Sel1-like repeats with low sequence conservation have been found to play important roles in other solenoid proteins such as the last Sel1-like repeat of *H. pylori* HcpC that deviates from the SLR consensus but plays a key role in protein–protein interactions.<sup>12</sup> The “succinoglycan overproduction” phenotype of the *S. meliloti* *exoR95* mutant with a disrupted C terminus<sup>23</sup> supports the importance of ExoR6 and suggests that the newly identified repeats are an integral and functional part of the complete ExoR structural fold and protein.

The modeled ExoR<sub>m</sub> adopts the typical superhelical fold observed for Sel1-like repeat proteins known to be a suitable scaffold for protein–protein interactions.<sup>12</sup> A comparison of ExoR and HcpC reveals that the crystal contact II at the concave face of the HcpC N terminus<sup>12</sup> matches the putative protein–protein interaction site A at the concave face of the ExoR N terminus [Fig. 4(d)]. Because asparagine residues are recognized to be important to peptide recognition in HcpC,<sup>12</sup> Hsp70/Hsp90 organizing protein,<sup>24</sup> and PEX5,<sup>25</sup> we propose that two asparagine residues present in site A of ExoR (Asn83 and Asn122) play a role in mediating protein–protein interactions, although their role would have to be confirmed through site-directed mutagenesis. The putative site B identified in ExoR protein also partially overlaps with the identified crystal contact I at the HcpC C terminus [Fig. 4(e)].<sup>12</sup> Because 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.<sup>12</sup> Although the prediction of Site C is not as reliable as Sites A and B, it is possible that it forms a novel interface site characteristic of the ExoR protein family. Further validation of the identified interaction sites comes from our ongoing modeling studies (data not shown) of the experimentally characterized ExoR reduced-function mutants, ExoRG76C and ExoRS156Y, which show loss of stabilizing interactions with the ExoS protein.<sup>5</sup>

In addition, ExoR is predicted to form unique 20 residues long extended helix A of ExoR3 that encompasses a part of the protein–protein interaction site A. A similar unusual extended conformation of a

repeat has been observed in the third repeat of the crystallized *Trypanosoma brucei* PEX5 protein, a helical multirepeat protein.<sup>26</sup> On the other hand, the crystallized human PEX5 does not show this unusual elongated conformation.<sup>25</sup> Based on similarity in sequence and function of *T. brucei* and human PEX5, it has been suggested that the third repeat in *T. brucei* PEX5 can adopt the extended form and the standard conformation.<sup>26</sup> We speculate a similar helix to random coil pliability,<sup>27</sup> which may be important for ExoR function and/or regulation.

Electrostatic interactions are important for selective binding of interacting partners in multirepeat  $\alpha$ - $\alpha$  proteins.<sup>14,28</sup> Our analysis of the electrostatic profile of the modeled ExoR<sub>m</sub> protein shows 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 superhelix that may play a key role in driving interactions between ExoR and its binding partners.

Mapping the experimentally determined cleavage site<sup>6</sup> on the modeled ExoR<sub>m</sub> [Fig. 3(a)] reveals that it is buried and therefore would not be accessible to periplasmic proteases without undergoing conformational changes. The requirement for a similar conformational change is observed for substrates of DegP, a periplasmic serine endoprotease in *E. coli*, which recognizes three residues of the substrate protein and cleaves after a hydrophobic residue (Val, Ala, and Ile) that is almost completely buried in most DegP substrates.<sup>29</sup> Although the protease involved in ExoR proteolysis has not been identified, *S. meliloti* Rm1021 does have a homolog of *E. coli* DegP.<sup>30</sup> The presence of a solvent inaccessible hydrophobic residue in the vicinity of the experimentally determined ExoR cleavage site falls in line with the proposed model of DegP cleavage<sup>6,29</sup> and can be further investigated experimentally.

In conclusion, we present the first attempt to generate a 3D model of *S. meliloti* Rm1021 ExoR in the absence of its crystal structure revealing important insights toward understanding its function. The findings of our structural analyses make it possible to study the molecular mechanism of ExoR cleavage and ExoR<sub>m</sub>–ExoS interactions using rational hypotheses-driven approaches and facilitate studies of pathogenicities of animal and plant pathogens in general.

## Materials and Methods

### Protein sequence analyses

The amino acid sequence of the ExoR protein from *S. meliloti* Rm1021 was retrieved from the NCBI Protein database (GenBank: AAA26260.1).<sup>31</sup> The mature ExoR protein, ExoR<sub>m</sub>, that is, without its 30-residue signal sequence,<sup>4</sup> was used for modeling the



protein and sequence and structure analyses. SMART<sup>32</sup> was used to identify and validate boundaries of the four previously predicted Sell-like repeats of ExoR<sup>14</sup> and cross-checked with other programs (Supporting Information Table SI). Previously unidentified Sell-like repeats were detected with TPRpred,<sup>33</sup> HHrepID,<sup>34</sup> and REPRO<sup>35</sup> and validated for correspondence with secondary structure prediction (Supporting Information Fig. S1). Structure-based sequence alignment of ExoR repeats was performed using the TM-align algorithm<sup>36</sup> and manually refined by anchoring key conserved structural residues<sup>14</sup> (Fig. 2). The sequences of putative ExoR orthologs were retrieved using NCBI-BLASTP<sup>31</sup> against the nonredundant protein sequence database and aligned using T-COFFEE server v.9.01.<sup>37</sup>

### Modeling methodology

The fold of the ExoR protein was identified using BLAST<sup>31</sup> and threading algorithms (Supporting Information Table SII). To generate high-quality representations of the 3D structure of ExoR<sub>m</sub>, a large number of theoretical models were generated based on various alternative alignments for each template and also multitemplate modeling using comparative modeling methodology (detailed in Supporting Information Fig. S4). MODELLER versions 9.8 and 9.9<sup>38</sup> were used for model-building using alignments generated from T-COFFEE v9.02,<sup>37</sup> MAFFT v6.864,<sup>39</sup> FUGUE,<sup>40</sup> and HHpred.<sup>41</sup> I-TASSER<sup>16</sup> and other modeling servers (Supporting Information Table SII) were explored for automated model building. The final model was generated with I-TASSER<sup>16</sup> using the templates 1OUV<sup>12</sup> and 3E4B.<sup>42</sup> I-Tasser is currently ranked as the top server for automated protein structure prediction<sup>43</sup> and its superior performance stems from the use of multiple threading alignments, iterative template fragment assembly simulation and *ab initio* modeling of unaligned regions.<sup>16</sup> The prediction of side-chain conformations was implemented through the program SCWRL4<sup>44</sup> via the AS2TS system.<sup>45</sup> The generated models were evaluated using ProSA-web,<sup>17</sup> Verify3D,<sup>18</sup> WHAT\_CHECK,<sup>19</sup> and RAMPAGE.<sup>20</sup>

### Analyses of the 3D ExoR<sub>m</sub> model

The visualization and analyses of the shape, structural alignments, solvent-accessible molecular surfaces, and the electrostatic potential profile of the generated models were performed using the surface property analyses tools in PyMOL.<sup>46</sup> All cartoons and diagrams were constructed with PyMOL<sup>46</sup> and Prosite: MyDomains-Image Creator.<sup>47</sup> STRIDE server<sup>48</sup> was used to ascertain the location of secondary structure elements in the modeled ExoR<sub>m</sub> protein. Individual repeats were structurally aligned using the TM-align algorithm.<sup>36</sup> The helical packing angles of the ExoR<sub>m</sub> models were computed using

the PyMOL script to calculate angles between helices.<sup>49</sup>

### Identification of protein–protein interaction sites

The putative protein–protein interaction sites were identified with PIER,<sup>50</sup> SPPIDER v2,<sup>51</sup> and consPISP.<sup>52</sup> Interaction hot spot residues were detected using the ISIS method.<sup>53</sup>

### Electrostatics

The distribution of surface electrostatic potential for the ExoR<sub>m</sub> model was calculated using the Poisson-Boltzmann solver, DelPhi v.4 release 1.1.<sup>54</sup> The net charge and the dipole moment of the modeled ExoR<sub>m</sub> protein were computed using the Protein Dipole Moments Server.<sup>55</sup>

### Solvent accessibility analysis of the cleavage site

The solvent-accessible surface area of the ExoR cleavage site based on primary structure of ExoR was determined using several programs, including SABLE<sup>56</sup> and RVP-NET.<sup>57</sup> To determine the solvent accessibility of these residues in the generated theoretical models, the GETAREA<sup>58</sup> server was used.

### Acknowledgments

The authors thank the members of the Singh Laboratory and the Cheng Laboratory, specifically Mary Ellen Heavner, for helpful discussions and critical comments.

### References

1. Stock JB, Ninfa AJ, Stock AM (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol Rev* 53:450–490.
2. Yao SY, Luo L, Har KJ, Becker A, Ruberg S, Yu GQ, Zhu JB, Cheng HP (2004) *Sinorhizobium meliloti* ExoS and ExoS proteins regulate both succinoglycan and flagellum production. *J Bacteriol* 186:6042–6049.
3. Cheng HP, Walker GC (1998) Succinoglycan production by *Rhizobium meliloti* is regulated through the ExoS-ChvI two-component regulatory system. *J Bacteriol* 180:20–26.
4. 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 *Sinorhizobium meliloti*. *Mol Microbiol* 64:647–664.
5. Chen EJ, Sabio EA, Long SR (2008) The periplasmic regulator ExoR inhibits ExoS/ChvI two-component signaling in *Sinorhizobium meliloti*. *Mol Microbiol* 69:1290–1303.
6. Lu HY, Luo L, Yang MH, Cheng HP (2012) *Sinorhizobium meliloti* ExoR is the target of periplasmic proteolysis. *J Bacteriol* 194:4029–4040.
7. Lu HY, Cheng HP (2010) Autoregulation of *Sinorhizobium meliloti* ExoR gene expression. *Microbiology* 156:2092–2101.
8. Kajava AV, Gorbea C, Ortega J, Rechsteiner M, Steven AC (2004) New HEAT-like repeat motifs in proteins regulating proteasome structure and function. *J Struct Biol* 146:425–430.

9. Bradley P (2012) Structural modeling of TAL effector–DNA interactions. *Protein Sci* 21:471–474.
10. Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335:716–719.
11. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biom* 29:291–325.
12. Lüthy L, Grütter MG, Mittl PRE (2004) The crystal structure of *Helicobacter* Cysteine-rich protein C at 2.0 Å resolution: similar peptide-binding sites in TPR and SEL1-like repeat proteins. *J Mol Biol* 340:829–841.
13. Forwood JK, Lange A, Zachariae U, Marfori M, Preast C, Grubmüller H, Stewart M, Corbett AH, Kobe B (2010) Quantitative structural analysis of importin-beta flexibility: paradigm for solenoid protein structures. *Structure* 18:1171–1183.
14. Mittl PRE, Schneider-Brachert W (2007) Sel1-like repeat proteins in signal transduction. *Cell Signal* 19:20–31.
15. Urošev D, Ferrer-Navarro M, Pastorello I, Cartocci E, Costenaro L, Zhulenkova D, Marechal JD, Leonchik A, Reverter D, Serino L, Soriani M, Daura X (2013) Crystal structure of c5321: a protective antigen present in uropathogenic *Escherichia coli* strains displaying an SLR fold. *BMC Struct Biol* 13:19.
16. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.
17. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407–W410.
18. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356:83–85.
19. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272–272.
20. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2002) Structure validation by C $\alpha$  geometry:  $\phi/\psi$  and C $\beta$  deviation. *Proteins* 50:437–450.
21. Guzman-Verri C, Manterola L, Sola-Landa A, Parra A, Cloeckert A, Garin J, Gorvel JP, Moriyon I, Moreno E, Lopez-Goni I (2002) The two-component system BvrR/BvrS essential for *Brucella abortus* virulence regulates the expression of outer membrane proteins with counterparts in members of the Rhizobiaceae. *Proc Natl Acad Sci USA* 99:12375–12380.
22. 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.
23. Doherty D, Leigh JA, Glazebrook J, Walker GC (1988) *Rhizobium meliloti* mutants that overproduce the *R. meliloti* acidic calcofluor-binding exopolysaccharide. *J Bacteriol* 170:4249–4256.
24. Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H, Hartl FU, Moarefi I (2000) Structure of TPR domain–peptide complexes: critical elements in the assembly of the Hsp70–Hsp90 multi-chaperone machine. *Cell* 101:199–210.
25. Gatto GJ, Geisbrecht BV, Gould SJ, Berg JM (2000) Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nat Struct Biol* 7:1091–1095.
26. Kumar A, Roach C, Hirsh IS, Turley S, deWalque S, Michels PA, Hol WG (2001) An unexpected extended conformation for the third TPR motif of the peroxin PEX5 from *Trypanosoma brucei*. *J Mol Biol* 307:271–282.
27. Hite KC, Kalashnikova AA, Hansen JC (2012) Coil-to-helix transitions in intrinsically disordered methyl CpG binding protein 2 and its isolated domains. *Protein Sci* 21:531–538.
28. Cortajarena AL, Kajander T, Pan W, Cocco MJ, Regan L (2004) Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins. *Protein Eng Des Sel* 17:399–409.
29. Kolmar H, Waller PRH, Sauer RT (1996) The DegP and DegQ periplasmic endoproteases of *Escherichia coli*: specificity for cleavage sites and substrate conformation. *J Bacteriol* 178:5925–5929.
30. Galibert F, Finan TM, Long SR, Puehler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Guzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293:668–672.
31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
32. Schultz J, Milpetz F, Bork P, Ponting C (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* 95:5857–5864.
33. Karpenahalli MR, Lupas AN, Söding J (2007) TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* 8:2.
34. Biegert A, Söding J (2008) *De novo* identification of highly diverged protein repeats by probabilistic consistency. *Bioinformatics* 24:807–814.
35. George RA, Heringa J (2000) The REPRO server: finding protein internal sequence repeats through the Web. *Trends Biochem Sci* 25:515–517.
36. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302–2309.
37. Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217.
38. Šali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.
39. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res* 30:3059–3066.
40. Shi J, Blundell TL, Mizuguchi K (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* 310:243–257.
41. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–W248.
42. Keiski CL, Harwich M, Jain S, Neculai AM, Yip P, Robinson H, Whitney JC, Riley L, Burrows LL, Ohman DE, Howell PL (2010) AlgK is a TPR-containing protein and the periplasmic component of a novel exopolysaccharide secretin. *Structure* 18:265–273.



43. Huang YJ, Mao B, Aramini JM, Montelione GT (2014) Assessment of template-based protein structure predictions in CASP10. *Proteins* 82(Suppl 2):S43–S56.
44. Krivov GG, Shapovalov MV, Dunbrack RL Jr (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 77:778–795.
45. Zemla A, Zhou CE, Slezak T, Kuczmarski T, Rama D, Torres C, Sawicka D, Barsky D (2005) AS2TS system for protein structure modeling and analysis. *Nucleic Acids Res* 33:W111–W115.
46. The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.
47. Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuče BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJA (2008) The 20 years of PROSITE. *Nucleic Acids Res* 36:D245–D249.
48. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* 32:W500–W502.
49. Holder T. Angle between helices, PyMOLWiki; 2010. Available at: <http://pymolwiki.org/index.php/AngleBetweenHelices>. Retrieved March 9, 2012.
50. Kufareva I, Budagyan L, Raush E, Totrov M, Abagyan R (2007) PIER: protein interface recognition for structural proteomics. *Proteins* 67:400–417.
51. Porollo A, Meller J (2007) Prediction-based fingerprints of protein–protein interactions. *Proteins* 66:630–645.
52. Zhou H, Shan Y (2001) Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins* 44:336–343.
53. Ofra Y, Rost B (2007) ISIS: interaction sites identified from sequence. *Bioinformatics* 23:e13–e16.
54. Nicholls A, Honig B (1991) A rapid finite difference algorithm, utilizing successive over-relaxation to solve the poisson-boltzmann equation. *J Comp Chem* 12:435–445.
55. Felder CE, Prilusky J, Silman I, Sussman JL (2007) A server and database for dipole moments of proteins. *Nucleic Acids Res* 35:W512–W521.
56. Adamczak R, Porollo A, Meller J (2004) Accurate prediction of solvent accessibility using neural networks based regression. *Proteins* 56:753–767.
57. Ahmad S, Gromiha MM, Sarai A (2003) RVP-net: online prediction of real valued accessible surface area of proteins from single sequences. *Bioinformatics* 19: 1849–1851.
58. Fraczek R, Braun W (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comp Chem* 19:319–333.