

Article

## Structural Insights into a Novel Interkingdom Signaling Circuit by Cartography of the Ligand-Binding Sites of the Homologous Quorum Sensing LuxR-Family

Sonia Covaceuszach <sup>1</sup>, Giuliano Degrassi <sup>2,3</sup>, Vittorio Venturi <sup>2,\*</sup> and Dorian Lamba <sup>1,\*</sup>

<sup>1</sup> Institute of Crystallography, National Research Council, Trieste Outstation, Area Science Park-Basovizza, S.S. n° 14 Km 163.5, I-34149 Trieste, Italy; E-Mail: sonia.covaceuszach@ts.ic.cnr.it

<sup>2</sup> International Centre for Genetic Engineering and Biotechnology, Padriciano 99, I-34149 Trieste, Italy; E-Mail: degrassi@icgeb.org

<sup>3</sup> IBIIBA-CONICET-ICGEB, International Centre for Genetic Engineering and Biotechnology, Scientific and Technological Center, Godoy Cruz 2390, C1425FQD, Buenos Aires, Argentina

\* Authors to whom correspondence should be addressed; E-Mails: venturi@icgeb.org (V.V.); dorian.lamba@ts.ic.cnr.it (D.L.); Tel.: +39-40-3757319 (V.V.); +39-40-3758514 (D.L.); Fax: +39-40-226555 (V.V.); +39-40-9221126 (D.L.).

Received: 1 August 2013; in revised form: 13 September 2013 / Accepted: 1 October 2013 /

Published: 15 October 2013

---

**Abstract:** Recent studies have identified a novel interkingdom signaling circuit, via plant signaling molecules, and a bacterial sub-family of LuxR proteins, bridging eukaryotes and prokaryotes. Indeed pivotal plant-bacteria interactions are regulated by the so called Plant Associated Bacteria (PAB) LuxR solo regulators that, although closely related to the quorum sensing (QS) LuxR family, do not bind or respond to canonical quorum sensing *N*-acyl homoserine lactones (AHLs), but only to specific host plant signal molecules. The large body of structural data available for several members of the QS LuxR family complexed with different classes of ligands (AHLs and other compounds), has been exploited to dissect the cartography of their regulatory domains through structure-based multiple sequence alignments, structural superimposition and a comparative analysis of the contact residues involved in ligand binding. In the absence of experimentally determined structures of members of the PAB LuxR solos subfamily, an homology model of its prototype OryR is presented, aiming to elucidate the architecture of its ligand-binding site. The obtained model, in combination with the cartography of the regulatory domains of the

homologous QS LuxRs, provides novel insights into the 3D structure of its ligand-binding site and unveils the probable molecular determinants responsible for differences in selectivity towards specific host plant signal molecules, rather than to canonical QS compounds.

**Keywords:** quorum sensing; bacterial-plant communication; interkingdom signaling; Plant Associated Bacteria LuxR solos; ligand binding site; molecular modeling; structure-based multiple alignment

---

## 1. Introduction

Research over the last 15 years has evidenced that intercellular communication frequently occurs in bacteria regulating gene expression in a cell density-dependent signaling, referred to as “quorum sensing” (QS) [1]. Gram-negative bacteria most commonly use *N*-acyl homoserine lactones (AHLs) as signal molecules; AHLs are synthesized by LuxI-family proteins and at high concentration (*i.e.*, high cell density) they bind to a cognate LuxR-family regulatory protein, which consequently binds target gene promoters. There is a class of LuxR-family proteins having the typical modular structure of QS LuxRs, which do not possess a cognate LuxI AHL synthase; these LuxR proteins have been called orphans or solos [2,3]. A sub-family of LuxR solos of Plant Associated Bacteria (PAB) has recently been shown to be part of a novel interkingdom signaling circuit, involved in communication between plant with both pathogens and beneficial bacteria [4,5]. It is likely that this sub-family of LuxRs of this interkingdom signaling circuit shares structural and functional similarities with the archetypical LuxI/R QS systems [6]. These PAB LuxR solos bind and respond to plant signals and probably have undergone coevolution with the host plant.

Five members of this PAB LuxR solos subfamily have been studied: XccR of *Xanthomonas campestris* pv. *campestris* (Xcc), OryR of *Xanthomonas oryzae* pv. *oryzae* (Xoo), PsoR of *Pseudomonas fluorescens*, XagR of *Xanthomonas axonopodis* pv. *glycines* (Xag) and NesR of *Sinorhizobium meliloti* [2,7–11]. With the exception of NesR, all have been demonstrated to respond to as yet chemically uncharacterized low molecular weight signal molecules synthesized by the host plant, with the final outcome of regulating crucial aspects of plant-bacteria interactions. Namely, OryR of the rice vascular pathogen Xoo, is involved in virulence; it responds to plant signals since the protein is solubilized and activates the expression of the neighboring *pip* and of motility genes only in the presence of plant extracts [4,7,9]. XccR of the crucifer pathogen Xcc also responds to an as yet unidentified plant compound and regulates the neighboring *pip* gene; the presence of the plant extracts allows XccR to bind to the *pip* promoter *in vitro* [8]. XagR of the soybean pathogen Xag which causes bacterial leaf pustule on soybean (*Glycine max*) is also involved in virulence [11]. As for XccR in Xcc, XagR in Xag also activates *pip* transcription *in planta* and temporal studies have indicated that *pip* transcription increases gradually after infection, reaching its greatest activity after 72 h, before slowly decreasing. PsoR responds to plant compounds of different plant species playing a role in biocontrol by rhizospheric *Pseudomonas fluorescens* via the control of transcriptional regulation of various anti-microbial-related genes [10]. NesR of *Sinorhizobium meliloti* is important for survival under stress and utilization of various carbon sources; the response to plant compounds has not yet been addressed [2].

Nevertheless, being closely related to the QS LuxRs, this sub-family of PAB LuxR solos share the same overall protein architecture, comprising two functional domains. In particular, members of the QS LuxR family are mainly composed of an *N*-terminal ligand-binding domain (the regulatory domain) [12,13] and a *C*-terminal helix-turn-helix DNA-binding domain [14,15], joined together by a short linker region. In QS systems, a conformational change is induced upon binding of the regulatory domain to the cognate AHL, most commonly then allowing the recognition of specific promoter regions by the DNA-binding domain and leading to transcriptional activation [16,17]. Indeed binding to the AHL is responsible for stability, correct folding [16] and most commonly dimerization, which in turns stabilizes the transcription factor allowing DNA binding [18].

Surprisingly conservation of primary structure among LuxR-family proteins is quite low (18%–25%), however, multiple sequence alignments performed have identified nine highly conserved residues (Figure 1): six of these residues delineate the cavity of the ligand-binding domain (W57, Y61, D70, P71, W85 and G113, according to TraR numbering) and the remaining three are located within the DNA-binding domain (E178, L182 G188) [19–22]. On these bases in a recent review Gonzales and Venturi [4] pinpointed that members of PAB LuxR solos subfamily show substitutions in one or two of these highly conserved amino acids in the regulatory domain, namely, W57M and Y61W, thus suggesting an involvement of these residues in the different selectivity of this subfamily towards specific host plant signal molecules rather than to AHLs.

Considering that the experimental three dimensional structures of several members of the LuxR family [23–30] show a quite conserved overall folding, with the regulatory domain composed of five anti-parallel  $\beta$ -sheets flanked by three  $\alpha$ -helices on each side, our aim here is to: (i) validate and extend the previous analysis of LuxR-family, based on primary sequence alignment, in order to dissect the structural determinants involved in ligand recognition; and (ii) extend the outcomes of the detailed molecular cartography to the PAB LuxR solos subfamily in order to identify the molecular determinants responsible for the different ligand selectivity of this subfamily.

In the present study we take advantage of the large body of experimental structural data available for several members of the LuxR-family in complex not only with cognate AHLs, but also with unrelated signaling molecules [23–30], focusing on structure-based sequence alignment, structural superimposition and a comparative analysis of the contact residues involved in ligand binding; this should allow the identification of the key residues characterizing the ligand-binding sites. Moreover, in the absence of experimentally determined structures of members of the PAB LuxR solos subfamily, the homology model of its prototype, OryR, is expected to provide us with sufficient information to gain insights into the architecture of its ligand-binding site, as well as to elucidate the likely structural basis of the reported different ligand selectivity between the PAB LuxR solos subfamily and the canonical QS LuxR receptors.

## 2. Results and Discussion

Structure-based multiple sequence alignment of the regulatory domains of all the QS LuxRs whose experimental three-dimensional structural information (obtained by X-ray crystallography or by NMR spectroscopy) is available (*i.e.*, TraR from *Agrobacterium tumefaciens* and from *Sinorhizobium fredii* NGR234, LasR and QscR from *Pseudomonas aeruginosa*, CviR from *Chromobacterium violaceum*

and SdiA from *Escherichia coli*) was performed by Expresso [31]. Figure 1 shows the multiple sequence alignment, based on structural information, having a main score (the total consistency value) of 71 (100 being the full agreement between the considered alignment and its associated primary library that has been computed as a first step of the consistency-based protocol exploited by Expresso), albeit the overall level of sequence identity or homology is quite low according to the calculated consensus sequence. It is interesting to note that even if the individual scores are 74 for LasR, 74 and 77 for the *A. tumefaciens* TraR and for its homolog from *S. fredii* NGR234 respectively, 76 for QscR, 56 for SdiA and 71 for CviR, regions encompassing residues 21–132 (TraR numbering) are characterized by an even higher level of consistency that has been prompted to reflect a higher level of accuracy [31].

**Figure 1.** Structure-based multiple sequence alignment of the quorum sensing (QS) LuxRs whose structures have been solved to date. The residues directly involved in ligand binding are in bold and colored in red. The residues highlighted in yellow and in gray are engaged in hydrogen bonding and in hydrophobic interactions with the ligand, respectively. Stars highlight conserved positions; colon and period indicate respectively conservation between groups having strong or weak similarity. The unique identification PDB code (PDB\_ID) of the 16 analyzed structures is given.

2UV0_E PDBID LasR	1	---MAL-----	VDGFLELERSSGKLEWSAILQKMASDLGFSKILFGLLPKDSQ	45
3IX3_A PDBID LasR	1	---MAL-----	VDGFLELERSSGKLEWSAILQKMASDLGFSKILFGLLPKDSQ	45
3IX4_A PDBID LasR	1	---MAL-----	VDGFLELERSSGKLEWSAILQKMASDLGFSKILFGLLPKDSQ	45
3IX8_A PDBID LasR	1	---MAL-----	VDGFLELERSSGKLEWSAILQKMASDLGFSKILFGLLPKDSQ	45
3JPU_A PDBID LasR	1	---MAL-----	VDGFLELERSSGKLEWSAILQKMASDLGFSKILFGLLPKDSQ	45
1L3L_A PDBID TraR	1	-MQHWL-----	DKLTDLAAIEGDECILKTGLADIADHFGFTGYAYLHIQ----	43
1HOM_A PDBID TraR	1	-MQHWL-----	DKLTDLAAIEGDECILKTGLADIADHFGFTGYAYLHIQ----	43
2Q0O_A PDBID TraR	1	---MS-----	VNGNLRSLIDMLEAAQDGHMIKIALRSFAHSCGYDRFAYLQKDG---	46
3S2T_A PDBID QscR	1	---MHDE-----R-----	EGYLEILSRITTEEEFFSLVLEICGNYGFEFFSFGARAPFPPL	47
2AVX_A PDBID SdiA	1	---MSDK--DFFSWR-----	RTMLLRFORMETAEEVYHEIELOAQQLYDYYSLCVVRHPVVF	52
3QP1_A PDBID CviR	1	---GSHM--RPLPAGLTASQQWTLLEWIHMAGHIETENELKAFLDQVLSQAPSERLLALGRLNNQ	61	
3QP2_A PDBID CviR	1	---GSHM--RPLPAGLTASQQWTLLEWIHMAGHIETENELKAFLDQVLSQAPSERLLALGRLNNQ	61	
3QP4_A PDBID CviR	1	---GSHM--RPLPAGLTASQQWTLLEWIHMAGHIETENELKAFLDQVLSQAPSERLLALGRLNNQ	61	
3QP5_B PDBID CviR	1	---GSHM--RPLPAGLTASQQWTLLEWIHMAGHIETENELKAFLDQVLSQAPSERLLALGRLNNQ	61	
3QP6_A PDBID CviR	1	MVTSKPINARPLPAGLTASQQWTLLEWIHMAGHIETEGELKAFLDNILSQAPSDRIILVLGRLNNQ	66	
3QP8_A PDBID CviR	1	---GSHMNARPLPAGLTASQQWTLLEWIHMAGHIETEGELKAFLDNILSQAPSDRIILVLGRLNNQ	63	
cons	1		:	63
2UV0_E PDBID LasR	46	D--YENAFIV-GNYPAAWREHYDRAGYARVDPVSHCTQSVLPWFPEPSI---YQTRKQHEFFEEA	105	
3IX3_A PDBID LasR	46	D--YENAFIV-GNYPAAWREHYDRAGYARVDPVSHCTQSVLPWFPEPSI---YQTRKQHEFFEEA	105	
3IX4_A PDBID LasR	46	D--YENAFIV-GNYPAAWREHYDRAGYARVDPVSHCTQSVLPWFPEPSI---YQTRKQHEFFEEA	105	
3IX8_A PDBID LasR	46	D--YENAFIV-GNYPAAWREHYDRAGYARVDPVSHCTQSVLPWFPEPSI---YQTRKQHEFFEEA	105	
3JPU_A PDBID LasR	46	D--YENAFIV-GNYPAAWREHYDRAGYARVDPVSHCTQSVLPWFPEPSI---YQTRKQHEFFEEA	105	
1L3L_A PDBID TraR	44	---HRHITAV-TNYHRQWQSTYFDKKFEALDPVVKRARSRKHIFTWSGEHERPTLSKDERAFYDHA	105	
1HOM_A PDBID TraR	44	---HRHITAV-TNYHRQWQSTYFDKKFEALDPVVKRARSRKHIFTWSGEHERPTLSKDERAFYDHA	105	
2Q0O_A PDBID TraR	47	---TQVRIF-HSYPGFWESIYLGSDYFNIDPVLAEAKRRRDVFFWTADAWPARGSSPLRRFRDEA	107	
3S2T_A PDBID QscR	48	T--APKYHFL-SNYPGEWKSRYISEDYTSIDPIVRHGLLEYTPLIWNGED-----FQENRFWEEA	105	
2AVX_A PDBID SdiA	53	T--RPKVAFY-TNYPEAWVSYQAKNFLAIDPVLNPNENFSQGHLMWNDDL-----FSEAQPLWEAA	110	
3QP1_A PDBID CviR	62	NQIQRLERVLNVSYPDWLDQYMKENYAQHDPIILRIH-LGQGPVME-ERFNRAKGAEKRFIAEA	125	
3QP2_A PDBID CviR	62	NQIQRLERVLNVSYPDWLDQYMKENYAQHDPIILRIH-LGQGPVME-ERFNRAKGAEKRFIAEA	125	
3QP4_A PDBID CviR	62	NQIQRLERVLNVSYPDWLDQYMKENYAQHDPIILRIH-LGQGPVME-ERFNRAKGAEKRFIAEA	125	
3QP5_B PDBID CviR	62	NQIQRLERVLNVSYPDWLDQYMKENYAQHDPIILRIH-LGQGPVME-ERFNRAKGAEKRFIAEA	125	
3QP6_B PDBID CviR	67	NQIQRMKVLNVSYPDWLDQYMQENFAQHDPIIMRIH-LGQGPVME-ERFSRAKGSEKRFIAEA	130	
3QP8_B PDBID CviR	64	NQIQRMKVLNVSYPDWLDQYMQENFAQHDPIIMRIH-LGQGPVME-ERFSRAKGSEKRFIAEA	127	
cons	64	. * * * * : ** : . * . : : *	129	

Figure 1. Cont.

2UV0_E PDBID LasR	106	SAAGLVYGLT <b>MPLHG-ARGE</b> <b>LGAL</b> <b>SLS</b> <b>SVEAENRAE</b> <b>ANRF</b> <b>FMSV</b> <b>LP</b> <b>TLWMLKDYALQSGAGLA</b> ---F	167
3IX3_A PDBID LasR	106	SAAGLVYGLT <b>MPLHG-ARGE</b> <b>LGAL</b> <b>SLS</b> <b>SVEAENRAE</b> <b>ANRF</b> <b>FMSV</b> <b>LP</b> <b>TLWMLKDYALQSGAGLA</b> ---F	167
3IX4_A PDBID LasR	106	SAAGLVYGLT <b>MPLHG-ARGE</b> <b>LGAL</b> <b>SLS</b> <b>SVEAENRAE</b> <b>ANRF</b> <b>FMSV</b> <b>LP</b> <b>TLWMLKDYALQSGAGLA</b> ---F	167
3IX8_A PDBID LasR	106	SAAGLVYGLT <b>MPLHG-ARGE</b> <b>LGAL</b> <b>SLS</b> <b>SVEAENRAE</b> <b>ANRF</b> <b>FMSV</b> <b>LP</b> <b>TLWMLKDYALQSGAGLA</b> ---F	167
3JPU_A PDBID LasR	106	SAAGLVYGLT <b>MPLHG-ARGE</b> <b>LGAL</b> <b>SLS</b> <b>SVEAENRAE</b> <b>ANRF</b> <b>FMSV</b> <b>LP</b> <b>TLWMLKDYALQSGAGLA</b> ---F	167
1L3L_A PDBID TraR	106	SDFG <b>IR</b> SGIT <b>PIKT-ANGF</b> MS <b>MT</b> <b>MASDK</b> ---PVIDLDREIDAVAAAATIGQIHARISF-----L	162
1HOM_A PDBID TraR	106	SDFG <b>IR</b> SGIT <b>PIKT-ANGF</b> MS <b>MT</b> <b>MASDK</b> ---PVIDLDREIDAVAAAATIGQIHARISF-----L	162
2Q00_A PDBID TraR	108	ISHG <b>IR</b> CGVT <b>IPVEG-SYGS</b> AMML <b>T</b> <b>FASPE</b> ---RKVDISGVLDPKKAVQLLMVHYQLKIIAA---K	167
3SZT_A PDBID QscR	106	LHHG <b>IR</b> HGWS <b>IPVRG-KYGL</b> <b>ISM</b> <b>LS</b> <b>LVRSSE</b> ---SIAATEILEKESFLLWIT <b>S</b> MLQAT <b>F</b> GD <b>L</b> LAPRI	168
2AVX_A PDBID SdiA	111	RAHG <b>L</b> RRGVT <b>QY</b> LML- <b>PERAL</b> G <b>F</b> LS <b>F</b> SRCSAR-E <b>I</b> P <b>L</b> S <b>D</b> EL <b>Q</b> L <b>K</b> M <b>Q</b> LL <b>V</b> RES <b>L</b> M- <b>AL</b> MRL-----N	169
3QP1_A PDBID CviR	126	TQNG <b>M</b> SGG <b>I</b> TFSAASER <b>N</b> IG <b>S</b> <b>T</b> LS <b>I</b> AGRE---PG---RNAALVAM <b>L</b> NCL <b>T</b> PHL <b>HQ</b> AA <b>I</b> RV---A	181
3QP2_A PDBID CviR	126	TQNG <b>M</b> SGG <b>I</b> TFSAASER <b>N</b> IG <b>S</b> <b>T</b> LS <b>I</b> AGRE---PG---RNAALVAM <b>L</b> NCL <b>T</b> PHL <b>HQ</b> AA <b>I</b> RV---A	181
3QP4_A PDBID CviR	126	TQNG <b>M</b> SGG <b>I</b> TFSAASER <b>N</b> IG <b>S</b> <b>T</b> LS <b>I</b> AGRE---PG---RNAALVAM <b>L</b> NCL <b>T</b> PHL <b>HQ</b> AA <b>I</b> RV---A	181
3QP5_B PDBID CviR	126	TQNG <b>M</b> SGG <b>I</b> TFSAASER <b>N</b> IG <b>S</b> <b>T</b> LS <b>I</b> AGRE---PG---RNAALVAM <b>L</b> NCL <b>T</b> PHL <b>HQ</b> AA <b>I</b> RV---A	181
3QP6_B PDBID CviR	131	SSNG <b>M</b> SGG <b>I</b> TFSAASDR <b>N</b> VG <b>S</b> <b>T</b> LS <b>I</b> GGKE---PG---RNAALVAM <b>L</b> NCL <b>T</b> PHL <b>HQ</b> AA <b>V</b> RI---A	187
3QP8_B PDBID CviR	128	SSNG <b>M</b> SGG <b>I</b> TFSAASDR <b>N</b> VG <b>S</b> <b>T</b> LS <b>I</b> GGKE---PG---RNAALVAM <b>L</b> NCL <b>T</b> PHL <b>HQ</b> AA <b>V</b> RI---A	183
cons	130	*: * :	:::

The regulatory domains of all the QS LuxRs complexes in the PDB database [32] have been analyzed using Pymol [33] and the Protein Interfaces, Surfaces and Assemblies (PISA) interactive tool for the exploration of macromolecular—ligand interfaces at the European Bioinformatic Institute [34]. The results have been summarized in Figure 1 and will be discussed using TraR numbering as a reference.

It is interesting to note that most of the residues involved in ligand binding (see Figure 1, in bold and colored in red) are conserved in all the QS LuxRs complexes and seems to be invariant regardless of the chemical nature of the ligand (AHLs, chloro-lactones, triphenyl ligands) as observed in the LasR complexes. This finding supports the strategy to dissect the cartography of the ligand-binding sites of QS LuxRs in order to gain insight on the structural basis of PAB LuxR solos specificity.

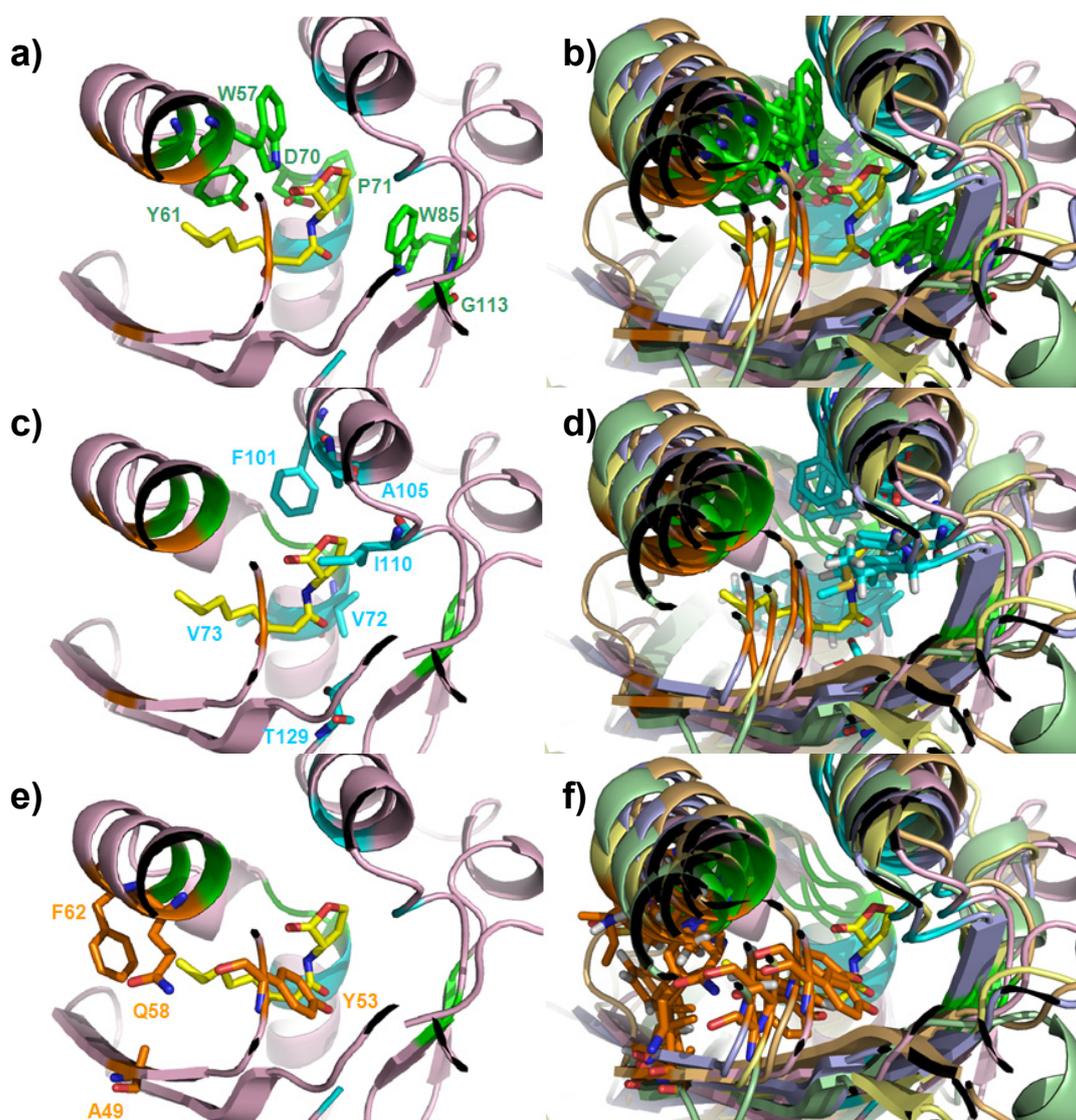
Previous studies have suggested, based on multiple sequence alignment of QS LuxR transcriptional regulators, that six conserved hydrophobic/aromatic residues of the regulatory domain *i.e.*, W57, Y61, D70, P71, W85 and G113, delineate the binding site [19–22]. The present structure-based multiple sequence alignment validates the above mentioned six residues pinpoints to an additional three conserved residues of the QS LuxR family regulatory domain, *i.e.*, Y53, A105 and G109 (identified by a star in Figure 1) and clearly indicates that residues P71, G109 and G113, although located very close to the binding site, are not directly involved in ligand binding. Furthermore, an additional four, but not fully conserved, among the 10 residues with similar physico-chemical properties (identified by a semicolon in Figure 1) directly interact with the ligands in all the analyzed complexes.

Besides the 10 residues, *i.e.*, Y53, W57, Y61, D70, W85, A105; V/L73, F/L101, I/L/M110 and T/S129, in a number of QS LuxRs complexed with AHLs—*i.e.*, TraR from *A. tumefaciens* in complex with OC8-HLS (PDB\_ID 1L3L [23], TraR from *S. fredii* NGR234 in complex with OC8-HSL (PDB\_ID 2Q00 [25]); LasR from *P. aeruginosa* in complex with OC12-HSL (PDB\_ID 2UV0 [26], PDB\_ID 3IX3), QscR from *P. aeruginosa* in complex with OC12-HSL (PDB\_ID 3SZT [28])—a water molecule is present at the ligand-binding sites, mediating protein-AHLs interactions by bridged hydrogen bonds.

From a structural perspective, it is interesting to note that not only the C<sub>α</sub> positions but also the side chains orientations of all the six conserved residues W57, Y61, D70, P71, W85 and G113 (hereafter called Cluster 1, highlighted by a star in Figure 1 and colored in green in Figure 2) superimpose in all the analyzed structures rather well (Figure 2b). To this end only W57, Y61, D70 and W85 are directly

involved in ligand binding (see Figures 1 and 2a). Residues P71 and G113 are located close to the binding site (Figure 2a) and are likely involved in the proper side chain orientation of D70 and W85, respectively. In this respect it is worth noting that in all of the analyzed structures, residues P71 and G113 adopt a *trans* and *cis* peptide conformation, respectively.

**Figure 2.** Mapping of the identified Clusters residues on the regulatory domains of the QS LuxRs. Residues, belonging to Cluster 1 (in green), Cluster 2 (in cyan) and Cluster 3 (in orange) and labeled according to the TraR numbering scheme, have been mapped in (a), (c) and (e) on the X-ray crystal structure of TraR in complex with OC8-HSL (PDB\_ID 1H0M [24]); in (b), (d) and (f) on the superimposed C $\alpha$  traces of the representative structures of the regulatory domains of QS LuxR, showing the side chains orientations, with TraR (PDB\_ID 1H0M [24]) in light pink, LasR (PDB\_ID 3IX3) in light orange, QscR (PDB\_ID 3SZT [28]) in light yellow; SdiA (PDB\_ID 2AVX [30]) in light blue, CviR (PDB\_ID 3QP1 [29]) in light green. The carbon, nitrogen and oxygen atoms of OC8-HSL, are colored in yellow, blue and red respectively. Figures produced by Pymol [33].



The present analysis reveals that the regulatory domain of the QS LuxR family includes, besides Cluster 1, an additional cluster of residues, namely V72, V73, F101, A105, I110, T129 (hereafter Cluster 2, colored in cyan in Figure 2) that is reasonably conserved and also directly involved in ligand binding (see Figures 1 and 2c). In all of the analyzed structures, the C<sub>α</sub> positions and the side chain orientations superimpose rather well, as shown in Figure 2d.

Beyond these two conserved clusters, the residues A49, Y53, Q58 and F62 (hereafter Cluster 3, colored in orange in Figure 2), represent a less conserved cluster (see Figures 1 and 2e) within the regulatory domain of the QS LuxR family. Besides Y53 that is conserved in a number of members belonging to the LuxR family (Figure 3), the residues A49, Q58 and F62 are highly substituted. Nevertheless the C<sub>α</sub> positions and the side chain orientations of these residues superimpose rather well in all of the analyzed structures (Figure 2f).

To extend this detailed molecular cartography of the regulatory domain of the QS LuxR family to the PAB LuxR solos subfamily [35], OryR, the prototype of this subfamily, has been modeled and structurally aligned, based on secondary structure prediction, using I-TASSER [36] (see Figure 3).

**Figure 3.** Structure-based multiple sequence alignment of the QS LuxRs with the prototype of the PAB LuxR solos subfamily. The residues of the QS LuxR involved in ligand binding are colored in red. The homologous residues of OryR, the prototype of PAB LuxR solos subfamily, belonging to Cluster 1 are highlighted by a star while the ones belonging to Cluster 2 and Cluster 3 (see text) are highlighted in cyan and in green respectively.

I-TASSER	OryR	1	--MFE--IL-----ASLGRDLQASQTVNSCLDRVFRDVC--ALGFQSLVYDYAPVPLSMEGAL	53
3IX3_A PDBID LasR		1	--MAL-----VDGFLFLERSGKLEWSAILQKMASDLGFSKILFGLLPKDSQD----	46
1HOM_A PDBID TraR		1	MQHWL-----DKLTDLAAIEGDECI LRTGLADIADHFQFTGYAYLHIQ-----	43
2Q00_A PDBID TraR		1	--MS-----VNGNLRSLIDMLEAAQDGHMIKIALRSFAHSCGYDRFAYLQKRDG----	46
3S2T_A PDBID QscR		1	--MHDE----R-----EGYLEILSRITTEEEFFSLVLEICNGYGFEEFFSGGARAPFPLT----	48
2AVX_A PDBID SdiA		1	--MSDKDFFSWR-----RTMLLRFORMETAEEVYHEIELQAQQLLEYDYYSLQVLRHPVPFT----	53
3QP1_A PDBID CviR		1	--GSHMRPLPAGLTASQQWTLLEWIHMAGHIETENEKRAFLDQVLSQAPSERLLALLGRLLNNQNI---	64
* * ** *				
I-TASSER	OryR	54	TPTVEHQ-RNAPGDMOHVWCEHGYQHDHPVQQRATRNRTPFVWSYRTDGCAGVEYVGGQHRQVTRYLC	121
3IX3_A PDBID LasR		47	YENAFIV-GNYPAAREHYDRAGYARVDPVSHCTQSVLPWFPEPSI-----YQTRKQHEFFEEAS	106
1HOM_A PDBID TraR		44	HRHITAV-TNYHRQWQSTYFDKRFKFEALDPVVKRARSRKHIPTWSGEH----ERPTLSKDERAFYDHAS	106
2Q00_A PDBID TraR		47	--TQVRTF-HSYPGWESIYLGSDYFNIDPVLAEAKRRRDVFFWTADA----WPARGSSPLRRFRDEAI	108
3S2T_A PDBID QscR		49	APKYHFL-SNYPGEWKSRYISEDYTSIDPIVRHGLLEYTPLIWNGED-----FQENRFWEEAL	106
2AVX_A PDBID SdiA		54	RPKVAFY-TNYPEAWVSYQAKNFLAIDPVLNPNFNSQGHLMWNDL-----FSEAQPLWEAAR	111
3QP1_A PDBID CviR		65	QRLERVLNVSYPDWLDQYMKENYAQHDPIRLRIH-LGQGPVMWE-ER-----FNRAKGAEEKRFLAEAT	126
*				
I-TASSER	OryR	122	DSGMGTGVTVPLHL-PGGAFATFSAIDA--VA-AEALRLAESQLLPFLLLAHAFQARAQELLDPOE	184
3IX3_A PDBID LasR		107	AAGLVYGLTmplhg-ARGLGALSLSVEAENRAE---ANRFMESVLPTLWMLKDYALQSGAGLA---F	167
1HOM_A PDBID TraR		107	DFGIRSGITPIKRT-ANGFMSMFTMASDK--PV---IDLREIDAVAAAATIGQIHARISF----L	162
2Q00_A PDBID TraR		109	SHGIRCGVTIPVEG-SYGSAMMLTFASPE---RK---VDISGVLDPKKAVQLMMVHYQLKIIAA-K	167
3S2T_A PDBID QscR		107	HHGIRHGWSIPVRG-KYGLISMLS LVRSSSE-SI---AATEILEKESFLLWITSMQLQATFGDLLAPRI	168
2AVX_A PDBID SdiA		112	AHGLRRGVTQYMLM-PERALGFLSFSRCSAR-EI---PILSDELQKMQLLVRESLM-ALMRL----N	169
3QP1_A PDBID CviR		127	QNGMGGGITFSAASERNIGSILSIAGRE---PG-----RNAALVAMLNCLTPHLHQAAIRV----A	181

The obtained homology model allowed to inspect the architecture of its ligand-binding site and to map the residues belonging to the three clusters, pinpointing the molecular determinants that are responsible for the observed differences in the ligand selectivity of this subfamily compared to QS LuxRs.

Mapping Cluster 2 residues on the regulatory domain of OryR shows that residues V72 and T129 are conserved, whereas residue F101 is substituted by L (like in QscR) and I110 is substituted by M (similarly to CviR). V73 and A105 instead are substituted by Q and L respectively, these residues

being rather conserved and specific for the subfamily of PAB LuxR solos (highlighted in cyan in Figures 3 and 4).

**Figure 4.** Multiple sequence alignment of the PAB LuxR solos subfamily members. The residues of the PAB LuxR solos subfamily homologous to Cluster 1 are highlighted by a star while the ones homologous to Cluster 2 and Cluster 3 (see text) are highlighted in cyan and in green, respectively.

OryR	-----MFEILASLGRDLQASQTVNSCLDRVFRDVCA LGFQSLVYDYAPVPLSMEGALITPTVFMORNP	PGDMQ*	68
XccR	-----MFDTLASLGRDLQARETLNAACLDVFRDVCA LGFQSLVYDYAPVPLSMEGALITPSVFLQRNAPDDMH		68
YP_242384.1	-----MFDTLASLGRDLQALHTLNAACLDVFRDVCA LGFQSLVYDYAPVPLSMEGALITPSVFMORNAPGDMR		68
ZP_06703918.1	-----MFDTLASLGRDLQALQTLDSCLDRVFRDVCA LGFQSLVYDYAPVPLSLEGLALITPTVFMORNAPGDMQ		68
ZP_08179515.1	-----MFDTLASLGRDLQALHTLDAACLDVFRDVCA LGFQSLVYDYAPVPLSMEGGLITPTVFLQRNAPGDMQ		68
XagR	-----MFDILASLGRDLQALRTLDSCLDRVFRDVCA LGFQSLVYDYAPVPLSLEGLALITPTVFMORNAPGDMQ		68
YP_276356.1	-----MHIRLSDFNRSR LQSASTVDQMDCA LLLASE LGFDAVITYDYS PVPVSHDGA LIT PSL LSLRNTPADWH		68
YP_004351870.1	-----MQAKLSDFNTR LLSGRS LDEQMDNT LGMVQALGF EALVYDYS PVP LDHDGALITPTV LKLRNTPRDWH		68
YP_004215311.1	-----MKLAVQESAVS LR-ETTVDGAFDQLFAQTSR LGFDALTYD YTPVPRSLEGE LIT PSL LRMCNVPGDMQ		68
YP_002541315.1	-----MLDEIGTIRRKFTAHDTLDGRI DQAF EAMQ LGF EALVYD YTPVPFDLDGE LMPVPSF LKLRN ISDDML		68
ZP_03522316.1	MRTRELPLDDIGTIRRQFTAHE TLDGRI DQAF EAMKR LGF EALTYD YTPVPYDL DGA IMI PSL LKLRN IADDMH		75
YP_765467.1	-----MLDDPILDDIGTIRRQFTAHE TLDGRI DQAF EAMQ LGF EALTYD YTPVPYDL DGA IMI PSL LKLRN ISDDMH		73
Psor	-----MQSRLAEFDTH LLSGNS LHERMTGTMTLASE LGFDALVYDYS PVPFDQAGE LIT PPSAMVC-NTPRDWL		67
NesR	-----MFDELGTIRNQFTAHD TLDGRI DQAF EAMKS IGF EALTYD YTPVPRDL DGT IMVPSL LKLRN ISEDMH		68
YP_427266.1	-----MTTASPGFDAHLDAACRQAMA LGFAAVLYDYS PVPRAHDGT LIT PTL LKLRN LPVDMA		58
ZP_05784535.1	-----MHDTEESI RNR LSE GSH LGQRVDAAFROMQ VGF DGM IYD YTPMVSGP GAD VPI PSVFEQRN LGDEMR		68
BAD15091	-----	ITPSVLQLRNTPPDWH	14
AER-0001405	-----MHIRLSDFNRSR LQSASTVDQMDCA LLLASD LGFDAVITYDYS PVPVSHDGA LIT PSL LSLRNTPADWH		68
OryR			
XccR			
YP_242384.1			
ZP_06703918.1			
ZP_08179515.1			
XagR			
YP_276356.1			
YP_004351870.1			
YP_004215311.1			
YP_002541315.1			
ZP_03522316.1			
YP_765467.1			
Psor			
NesR			
YP_427266.1			
ZP_05784535.1			
BAD15091			
AER0001405			
OryR	HVWCEHGYYQHDPVQQRATRRTT PFVWS YRTDGD CAGVEYVGGQHRQVTRY LCD SGMGTGVTVP LHLPGGAFAT		142
XccR	NVWCEHGYYQNDPVQQRATRRTT PFVWS YRTEGR TEGVEYVGGQHRQVTRY LCD SDMGTVTVPLHLPGGAFAT		142
YP_242384.1	HVWCEHGYYQHDPVQQRATRRTT PFVWS YRTDGNQEGVEYVGGQHRQVTRY LCD SGMGTGVTVP LHLPGGAFAT		142
ZP_06703918.1	HVWCEHGYYQHDPVQQRATRRTT PFVWS YRTDGS CPGEYVGGQHRQVTRY LCD SGMGTGVTVP LHLPGGAFAT		142
ZP_08179515.1	HVWCEHGYYQHDPVQQRATRRTT PFVWS YRTDGNRPGVEYVGAQHRQVTRY LCD SGMGNGVTVP LHLPGGAFAT		142
XagR	HVWCEHGYYQHDPVQQRATRRTT PFVWS YRTDGHCPGEYVGGQHRQVTRY LCD SGMGTGVTVP LHLPGGAFAT		142
YP_276356.1	ALWCSQGYQIDPVQHLAVASVSPFVWS YQPKAE TVLQTF ITDMHKPVVRY LHD SHMTCGVTVP I HMPKGGFAT		142
YP_004351870.1	SLWCEE GYYQIDPVQHLAINAIS PFVWS YRPEAD TVLQPF IQGSHAPVSY LQD SRMTCGVTVP MHLPRGGFAT		142
YP_004215311.1	RLWLENGYQVDPVQHYALE SCA PFVWS YQRPDNTSLQNR LSLDNKPVTHY MCDNNMPSGATVPLHLNPGGFVT		142
YP_002541315.1	EYWFDRGYFRIDPVQVALRTSAPFFWNYNVEAD TAINRY LGENTE PVARY LNERDMSTGVTVPVHMPRGDYAT		142
ZP_03522316.1	DYWFGRGYFRIDPVQVALRTSAPFFWNYDTPAD TLI NRFMSDTPAVTRY LSERDMSTGVTVPVHMPRGDYAT		149
YP_765467.1	DYWFNRGYFRIDPVQVALRTSAPFFWNYDTPAD TLI NRFMTEDTAPVTRY LSERDMSTGVTVPVHMPRGDYAT		147
Psor	ELWCDQGYHIDPVQVALDSSS PFVWS YKPEAE TVLQQA LGQHPVSSY LHHQMAHGMVPI HLPKGGFAS		141
NesR	DYWFDRGYFRIDPVQVALRTSAPFFWNYDTPAD TLI NRFMSDTPAVARY LSERDMSTGVTVPVHMPRGDYAT		142
YP_427266.1	EIWCCKGYQIDPVQDAALLVSAPFFVWS HEGDQSDVMARV LRAEHRPVLDY LRHRGLTRGVTVP I RCGEGLAT		132
ZP_05784535.1	QHWSEDEYI RLDPVQRVAMRSAPFCWS YDSDID SII RPLM LTPDIE PVMSY LREREVS GVTVP I HLPNGDFAT		142
BAD15091	ALWCSEGFYQIDPVQHLALS RVSPFVWS YDVKAD TPLQKI IDPCHAPVSSY LHDQQLTCGVSP I HLPKGGFAS		90
AER0001405	ALWCSEGYQIDPVQHLAVASVSPFVWS YQPKAE TVLQTF ITDMHKPVVRY LHD SHMTCGVTVP I HMPKGGFAT		142
OryR	FSAAIDAVAAEALRLAE SQLLPF LLLAHAFQARAQEL LDPQE		184
XccR	FSAAVNVTQAEAPRLAE AQLPPF LLLAHAFQARAQEL LDPQE		184
YP_242384.1	FSGAVDAVAAEAPRLAE AQLSPF LLLAHAFQARAQEL LDPQE		184
ZP_06703918.1	FSGAIDAAAQAPRLAE AQLLPF LLLAHAFQARAQEL LDPQE		184
ZP_08179515.1	LSGAVDAVAAEAPRLAE AQLSPF LLLAHAFQARAQEL LDPQE		184
XagR	FSGAIDAAAQAPRLAE AQLLPF LLLAHAFQARAQEL LDPQE		184
YP_276356.1	LTGLCSDSSDAALDARQSLAE FGLLAHAFQAEVAYPLFDQRM		184
YP_004351870.1	LTGLRADTKGSTLQDVRQTLGDF SLI SHALQEAAPFL LGKEA		183
YP_004215311.1	VTGIHTGAG-- HELDIHDVLAEL SLLALTFQECVFP LFDSSI		181
YP_002541315.1	VTGIRFGANKD FERHALRYIADFGLLAHVPHETAYALFDSSE		190
ZP_03522316.1	VTGIRFGEDSD FERHALRYIADFNLLAHVPHETAYSLFDRKA		195
YP_765467.1	VTGIRGRNKD FERHALRYIADFNLLAHVPHETAYSLFDTRA		200
Psor	LTGLRSGNATVALQDLQVQGHFTLLAHALQEHLYPQLSKCV		184
NesR	VTGVRF GGNRAFEHGHALRYIADFNLLAHVPHETAYSLFDAQA		190
YP_427266.1	FTAIAPSGAPFPPEP---RLSSLGLLGQSFHDRVYPGPPAAS		169
ZP_05784535.1	VTGVSLGRRQNWGGTSRAIADLSL IAH LFNQSAVEL LGFDH		183
BAD15091	LTGLRRTGNARNVLQDAQQLSDFGLI SHALQEAAYPLFSKEL		132
AER0001405	LTGLCSDSSDAALDARQSLAE FGLLAQAFQAEVAYPLFDQRM		184

Regarding Cluster 3, the residues A49, Q58 and F62 are highly substituted in the PAB LuxR solos subfamily (highlighted in green in Figures 3 and 4) similar to what has been found in the QS LuxRs. In contrast, Y53 is highly variable within the PAB LuxR solos subfamily members (Figure 4), while it is conserved in a number of QS LuxRs (Figure 3).

Details of the residues type and the frequencies for the residues belonging to each of the three clusters both in QS LuxRs and in PAB LuxR solos are summarized in Table 1.

**Table 1.** Clusters residue type and frequency in the ligand-binding site of the regulatory domain of QS LuxRs and of the PAB LuxR solos subfamily.

Cluster	Position	QS LuxRs: residue and frequency	QS LuxRs: ligand binding frequency	PAB LuxR solos: residue and frequency
1	57	W (6/6)	16/16	M (13/18), W (5/18)
1	61	Y (6/6)	16/16	W (18/18)
1	70	D (6/6)	2/16	D (18/18)
1	71	P (6/6)	16/16	P (18/18)
1	85	W (6/6)	0/16	W (18/18)
1	113	G (6/6)	16/16	G (18/18)
2	72	V (3/6), I (2/6), T (1/6)	16/16	V (18/18)
2	73	V (3/6), L (3/6)	16/16	Q (18/18)
2	101	F (5/6), L (1/6)	16/16	V (18/18)
2	105	A (6/6)	16/16	L (17/18), M (1/18)
2	110	I (3/6), L (2/6), M (1/6)	16/16	M (15/18), L (2/18), V (1/18)
2	129	S (4/6), T (2/6)	16/16	S (6/18), T (12/18)
3	49	F (2/6), I (1/6), V (1/6), T (1/6), A (1/6)	16/16	M (5/18), K (5/18), E (2/18), S (2/18), L (1/18), V (1/18), Q (1/18), R (1/18)
3	53	Y (6/6)	16/16	A (6/18), T (5/18), I (4/18), L (2/18), V (1/18)
3	58	E (1/6), K (1/6), Q (1/6), R (1/6), L (1/6), V (1/6)	14/16	H (8/18), Q (5/18), R (2/18), L (2/18), A (1/18)
3	62	M (1/6), F (1/6), Q (1/6), D(1/6), L (1/6), I (1/6)	14/16	C (12/18), F (4/18), S (1/18), L (1/18)

The three dimensional architecture of the boundaries of the ligand-binding site of the QS LuxRs is outlined in Table 2.

**Table 2.** Clusters three-dimensional spatial distribution in the ligand-binding site of the regulatory domain of the QS LuxRs.

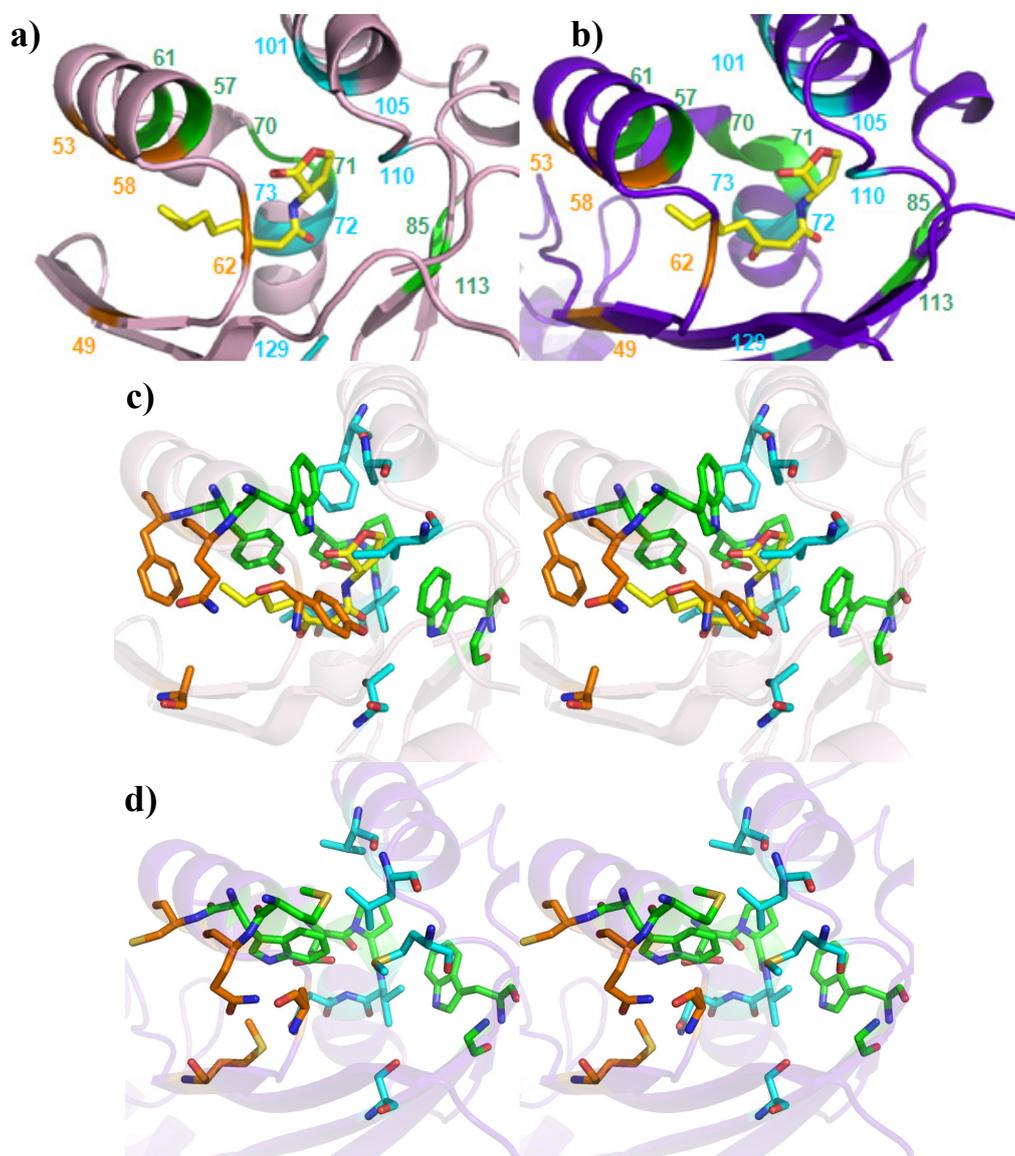
	Roof	Distal Wall	Floor	Proximal Wall
Structural determinants	Two anti-parallel $\alpha$ -helices	Two perpendicular $\alpha$ -helices fragments	Four anti-parallel $\beta$ -strands	Loops
Cluster 1	W57, Y61	D70, P71	W85, G113	-
Cluster 2	F101, A105	V72, V73	T129	I110
Cluster 3	Q58, F62	-	A49	Y53

The contribution of the three clusters to the binding site topology of QS LuxRs and in PAB LuxR solos can be seen in Figure 5.

The residues belonging to each of the above described three clusters have been mapped on TraR (PDB\_ID 1H0M) [24] and on the homology model of OryR regulatory domains (Figure 5a,b) in order

to obtain the cartography of their respective ligand-binding sites: the resulting comparison (Figure 5c,d) indicates a tripartite architecture.

**Figure 5.** Comparison of the ligand-binding sites of prototypes of QS LuxRs and PAB LuxR solos subfamily (Top view). Mapping of the three Clusters of residues in (a) and (c) on the X-ray crystal structure of TraR, in complex with OC8-HSL (PDB\_ID 1H0M [24]), showing the C<sub>α</sub> trace and stereo view of the side chains, respectively; in (b) and (d) on the homology model of OryR, the C<sub>α</sub> trace and stereo view of the side chains orientations can be seen. Residues have been labeled according to the TraR numbering scheme. The carbon, nitrogen and oxygen atoms of OC8-HSL, are colored in yellow, blue and red respectively. Figures produced by Pymol [33].



Firstly, a shared part (conserved core), delimited by the floor and the distal wall (residues 70, 71, 72, 85, 110, 113, 129), appears to be crucial for the folding mechanism of the regulatory domain in the presence of the signal molecule, regardless of its chemical nature. Then a specific part (specificity patch), mainly delimited by the roof and the nearby regions of the proximal and distal walls (residues

57, 61, 73, 101, 105), is conserved only within the QS LuxRs or within the PAB LuxR solos subfamily members respectively. It is therefore likely that the selectivity of LuxR family and of the PAB LuxR solos subfamily towards diverse ligands is modulated by these residues. In all the experimental structures analyzed, these are the ones interacting with the lactone ring of the AHL, while the PAB LuxR solos do not bind to AHLs. Finally a variable part (variability patch), delimited by the proximal wall and the nearby regions of the roof and of the floor (residues 49, 53, 58, 62), is less conserved even within the members of the QS LuxR family or of the PAB LuxR solos subfamily. It is interesting to note that in all the experimental structures analyzed, these residues interact with the fatty acyl side chain moieties of the AHLs that are found to adopt different position/orientation and conformations. Therefore they are likely to be responsible for the different selectivity towards molecules belonging to the same family of ligands or for the modulation of the degree “promiscuity” towards members of the same family of compounds.

In order to corroborate the results of this analysis, the shape and the physico-chemical properties of the ligand-binding sites were evaluated. In Figures 6 and 7, on comparing the overall shape and the electrostatic and lipophilic potentials, respectively, of the ligand-binding sites of QS LuxRs, confirms the tripartite topology previously outlined (mapping the conserved core in yellow, the specificity patch in magenta and the variability patch in orange). Furthermore the physico-chemical properties mapped on the ligand-binding site of the OryR model reveal an increased negative potential (Figure 6) and a decreased hydrophobicity (Figure 7) in comparison to QS LuxRs, which most likely accounts for the structural determinants that are responsible for differences in the selectivity of the PAB LuxR solos subfamily with respect to QS LuxRs.

**Figure 6.** Electrostatic potential of the ligand-binding sites of the QS LuxR and of the prototype of the PAB LuxR solos subfamily. (a) TraR (PDB\_ID 1H0M [24]) in light pink (showing the side chains of the conserved core in yellow, of the specificity patch in magenta and of the variable patch in orange); (b) SdiA (PDB\_ID 2AVX [30]) in light blue; (c) LasR (PDB\_ID 3IX3) in light orange; (d) CviR (PDB\_ID 3QP1 [29]) in light green; (e) QscR (PDB\_ID 3SZT [28]) in light yellow; (f) OryR model in blue purple. **Left** column: top view of the ligand binding sites (same orientation of the previous Figures); **Right** column: bottom view (obtained by 180 degrees rotation around y axis). Figures produced by Pymol [33].

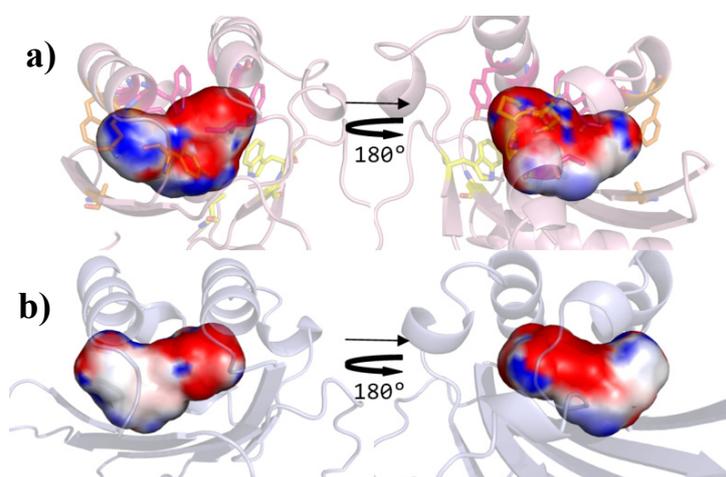
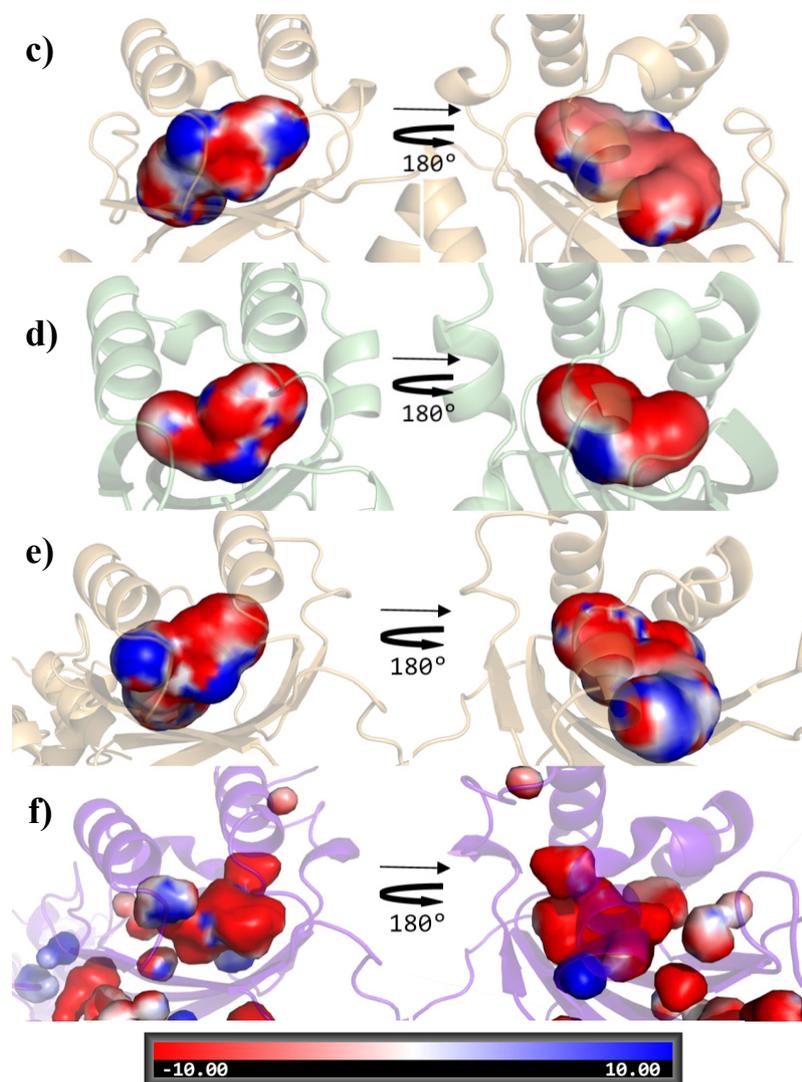


Figure 6. Cont.



**Figure 7.** Lipophilic potential of the ligand-binding sites of the QS LuxR and of the prototype of the PAB LuxR solos subfamily. (a) TraR (PDB\_ID 1H0M [24]) in light pink (showing the side chains of the conserved core in yellow, of the specificity patch in magenta and of the variable patch in orange); (b) SdiA (PDB\_ID 2AVX [30]) in light blue; (c) LasR (PDB\_ID 3IX3) in light orange; (d) CviR (PDB\_ID 3QP1 [29]) in light green; (e) QscR (PDB\_ID 3SZT [28]) in light yellow; (f) OryR model in blue purple. The hydrophobic surface is indicated by yellow and the hydrophilic surface by cyan. **Left** column: top view (same orientation of the previous Figures); **Right** column: bottom view (obtained by 180 degrees rotation around y axis). Figures produced by Pymol [33].

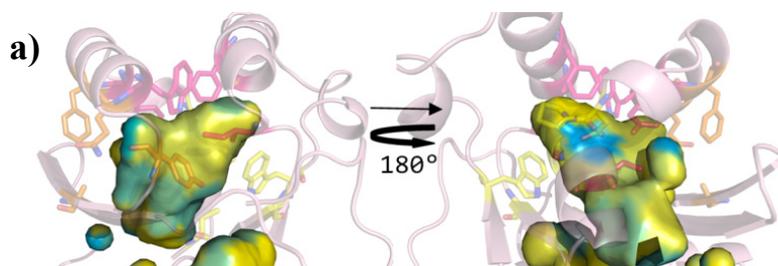
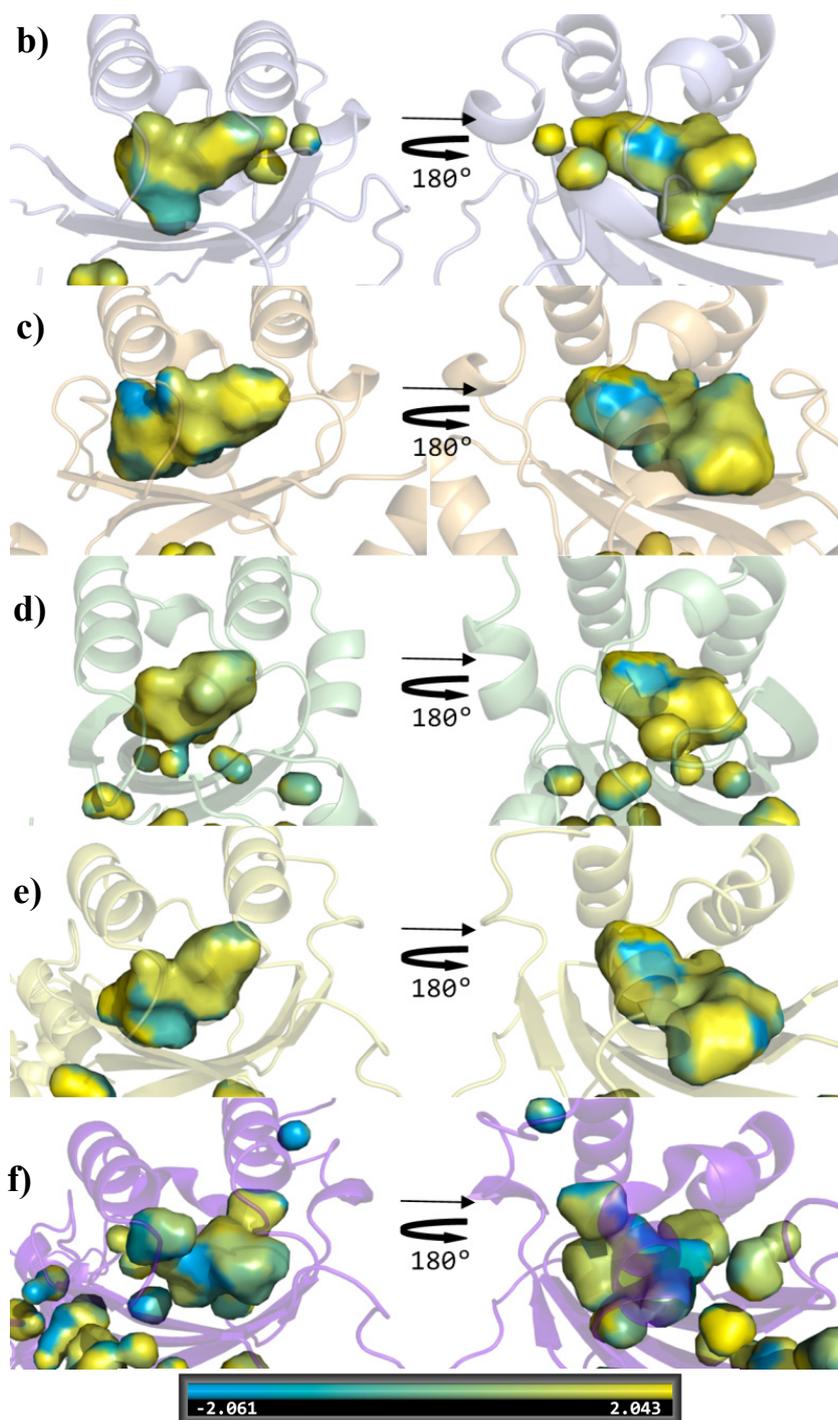
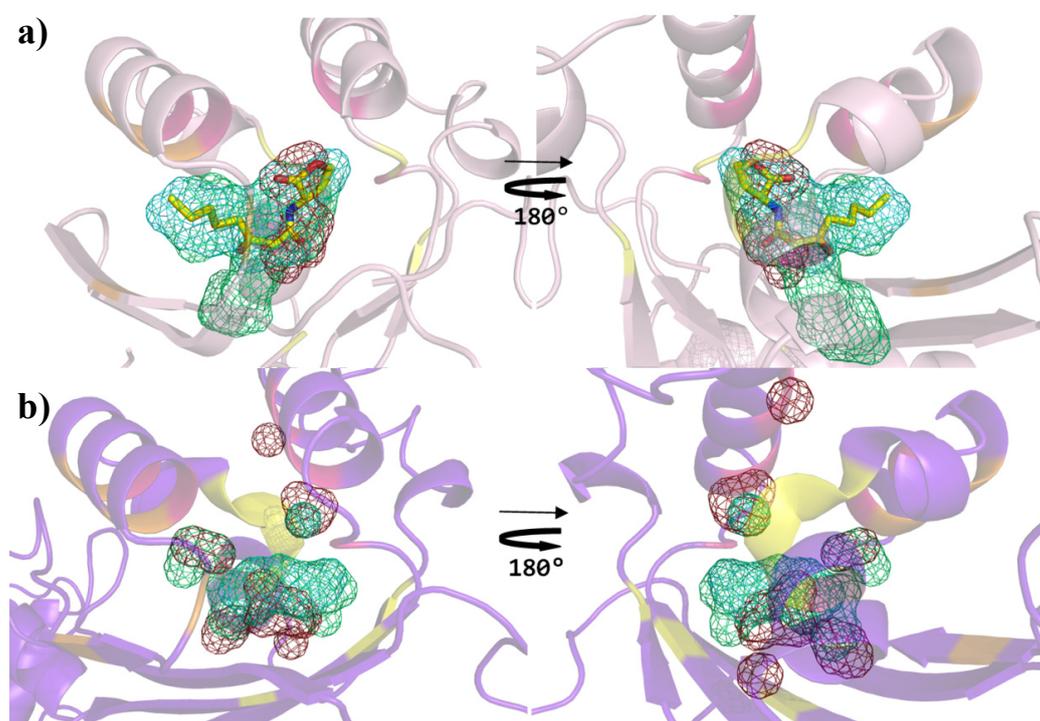


Figure 7. Cont.



An additional validation was carried out by comparing the binding sites topochemical preferences of the QS LuxRs and of the PAB LuxR solos prototypes by SITEHOUND [37] that exploits favorable interactions of three different structural probes (methyl carbon, aromatic carbon and hydroxyl oxygen). This analysis discloses clear differences between the prototypes of the two families (Figure 8). Indeed the comparison of the binding sites of the OryR and of the prototype of QS LuxRs shows the former to prefer hydroxylic groups rather than aliphatic and/or aromatic groups; therefore providing further support to the molecular determinants responsible for the differences in selectivity of the PAB LuxR solos towards specific host plant signal molecules rather than towards canonical quorum sensing ligands.

**Figure 8.** Comparison of the binding site preferences of prototypes of QS LuxRs and of PAB LuxR solos subfamily. Cartoon diagram (a) of TraR X-ray crystal structure in complex with OC8-HSL (PDB\_ID 1H0M [24]) (showing the C<sub>α</sub> trace of the conserved core in yellow, of the specificity patch in magenta and of the variable patch in orange) and (b) of the model of OryR, showing the top ranking clusters obtained by SITEHOUND [37] as mesh: methyl carbon probe clusters (green), aromatic carbon probe clusters (cyan) and hydroxyl oxygen probe cluster (red). The carbon, nitrogen and oxygen atoms of OC8-HSL are colored in yellow, blue and red respectively. Left column: top view of the ligand binding sites (same orientation as in previous figures); right column: bottom view (obtained by 180 degree rotation around y axis). Figures produced by Pymol [33].



### 3. Experimental Section

Sequence alignment was performed by Expresso [31], that exploits structural aligners algorithms like SAP [38] or TMalign [39] to generate structure-based alignments that are used as a template for realigning the original sequences.

All the complexes of QS LuxR deposited in PDB both solved by X-ray crystallography, *i.e.*, TraR from *Agrobacterium tumefaciens* in complex with OC8-HSL (PDB\_ID 1L3L [23], PDB\_ID 1H0M [24]), TraR from *Sinorhizobium fredii* NGR234 in complex with OC8-HSL (PDB\_ID 2Q0O [25]); LasR from *Pseudomonas aeruginosa* in complex with OC12-HSL (PDB\_ID 2UV0 [26], PDB\_ID 3IX3), with TP1 (PDB\_ID 3IX4 [27]), with TP3 (PDB\_ID 3IX8 [27]) and with TP4 (PDB\_ID 3JPU [27]); QscR from *Pseudomonas aeruginosa* in complex with OC12-HSL (PDB\_ID 3SZT [28]); CviR from *Chromobacterium violaceum* in complex with C6-HSL (PDB\_ID 3QP1 and PDB\_ID 3QP6 [29]), in complex with OC8-HSL (PDB\_ID 3QP2 [29]), in complex with OC10-HSL (PDB\_ID 3QP4 and PDB\_ID 3QP8 [29]) and in complex with an antagonist chlorolactone (PDB\_ID 3QP5 [29]) and by NMR, *i.e.*, SdiA from *Escherichia coli* in complex with OC8-HSL (PDB\_ID 2AVX [30]), have been

superimposed using Pymol [33] and analyzed by using the Protein Interfaces, Surfaces and Assemblies (PISA) web tool at the European Bioinformatic Institute [34].

Homology-based protein modelling has been performed on the full-length amino acidic sequence of OryR protein from *Xanthomonas oryzae* using five molecular modelling strategies based on different criteria for template selection. SWISS-MODEL [40] performs a search in a library of experimental protein structures extracted from the PDB: up to five template structures per batch are superposed using an iterative least squares algorithm generating a structural alignment after removing incompatible templates, improved by a heuristic step after the calculation of a local pair-wise alignment of the target sequence to the main template structures [41]. ModWeb [42] depends on the large scale protein structure modeling pipeline, ModPipe, which performs a search in a set of non-redundant chains extracted from structures in the PDB and establishes sequence-structure matches using multiple variations of sequence-sequence, profile-sequence, sequence-profile and profile-profile alignment methods [43–45]. M4T (Multiple Mapping Method with Multiple Templates) [46] is based on two of major modules, Multiple Templates (MT) and Multiple Mapping Method (MMM) [47], developed to produce accurate alignments and models by minimizing the errors associated with the first two steps of modeling procedure (template recognition and alignment). HHpred (Homology detection & structure prediction by HMM-HMM comparison) [48] implements pairwise comparison of profile hidden Markov models (HMMs) to generate pairwise query-template alignments or multiple alignments of the query with a set of templates selected from the search results. I-TASSER (iterative threading assembly refinement) [36] generates three-dimensional atomic models from multiple threading alignments and iterative structural assembly simulations.

The five top-scored models generated have been ranked and validated by two protein model quality predictors ProQ [49] and AIDE [50], that have different and often complementary ability to properly assess the quality of protein structures and therefore their combined use can increase the reliability in the evaluation of model quality. The resulting outputs were consistent, pinpointing the top-scored model (confidence score 0.64) produced by I-TASSER, based on the QscR template (PDB\_ID 3SZT [28]), as the most reliable candidate. Indeed the correctness of the selected model was confirmed by ProQ [49] (having Predicted LGscore of 4.299 and Predicted MaxSub of 0.437) and its overall best quality was validated by AIDE [50] (with a Predicted TM-score of 0.69 and a Predicted RMSD of 6.97).

Electrostatic potentials calculations were performed by PDB2PQR [51] and visualized by Pymol [33] on the ligand surface in the ligand-binding sites of the QS LuxRs and on the cavity surface of the ligand-binding site of OryR model. Lipophilic potential representation, based on the hydrophobicity scale derived by Black and Mould [52], was performed by Pymol [33] on cavity surface of the ligand-binding site of the QS LuxRs and of OryR model.

Binding sites preferences of QS LuxRs and of PAB LuxR solos prototypes, TraR and OryR respectively, were analyzed by SITEHOUD [37] employing three different structural probes (*i.e.*, methyl carbon, aromatic carbon and hydroxyl oxygen).

#### 4. Conclusions

The present study was aimed to dissect the cartography of the ligand-binding sites of the QS LuxRs by structure-based sequence alignment, structural superimposition and comparative mapping residues

directly involved in ligand binding. The structure-based analysis pinpointed the key residues crucial for ligand recognition and led us to identify a tripartite architecture of the ligand-binding sites that may account for differences in selectivity within the QS LuxR family. In order to extend this detailed molecular cartography of the regulatory domain to the PAB LuxR solos subfamily, an homology model of its prototype OryR has been obtained. Its comparative structural analysis allowed the identification of the key molecular determinants shaping the physico-chemical properties (electrostatic and lipophilic potentials). The resulting binding sites topochemical preferences are likely to be responsible for the difference in selectivity of the PAB LuxR solos towards specific host-plant signal molecules rather than towards canonical QS ligands.

### Acknowledgments

V.V.'s laboratory is funded by ICGEB core funding and by the Progetto AGER.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Fuqua, W.C.; Winans, S.C.; Greenberg, E.P. Quorum sensing in bacteria: The LuxR-LuxI family of cell density-responsive transcriptional regulators. *J. Bacteriol.* **1994**, *176*, 269–275.
2. Patankar, A.V.; Gonzalez, J.E. An orphan LuxR homolog of *Sinorhizobium meliloti* affects stress adaptation and competition for nodulation. *Appl. Environ. Microbiol.* **2009**, *75*, 946–955.
3. Subramoni, S.; Venturi, V. LuxR-family “solos”: Bachelor sensors/regulators of signalling molecules. *Microbiology* **2009**, *155*, 1377–1385.
4. González, J.F.; Venturi, V. A novel widespread interkingdom signalling circuit. *Trends Plant Sci.* **2013**, *18*, 167–174.
5. Venturi, V.; Fuqua, W.C. Chemical signaling between plants and plant-pathogenic bacteria. *Annu. Rev. Phytopathol.* **2013**, *51*, 17–37.
6. Fuqua, C.; Greenberg, E.P. Listening in on bacteria: Acylhomoserine lactone signalling. *Nat. Rev. Mol. Cell. Biol.* **2002**, *3*, 685–695.
7. Slock, J.; VanRiet, D.; Kolibachuk, D.; Greenberg, E.P. Critical regions of the *Vibrio fischeri* LuxR protein defined by mutational analysis. *J. Bacteriol.* **1990**, *172*, 3974–3979.
8. Ferluga, S.; Bigirimana, J.; Hofte, M.; Venturi, V. A LuxR homologue of *Xanthomonas oryzae* pv. *oryzae* is required for optimal rice virulence. *Mol. Plant. Pathol.* **2007**, *8*, 529–538.
9. Ferluga, S.; Venturi, V. OryR is a LuxR-family protein involved in interkingdom signaling between pathogenic *Xanthomonas oryzae* pv. *oryzae* and rice. *J. Bacteriol.* **2009**, *191*, 890–897.
10. Zhang, L.; Jia, Y.; Wang, L.; Fang, R. A proline iminopeptidase gene upregulated *in planta* by a LuxR homologue is essential for pathogenicity of *Xanthomonas campestris* pv. *campestris*. *Mol. Microbiol.* **2007**, *65*, 121–136.

11. Subramoni, S.; Gonzalez, J.F.; Johnson, A.; Pechy-Tarr, M.; Rochat, L.; Paulsen, I.; Loper, J.E.; Keel, C.; Venturi, V. Bacterial subfamily of LuxR regulators that respond to plant compounds. *Appl. Environ. Microbiol.* **2011**, *77*, 4579–4588.
12. Chatnaparat, T.; Prathuangwong, S.; Ionescu, M.; Lindow, S.E. XagR, a LuxR homolog, contributes to the virulence of *Xanthomonas axonopodis* pv. *glycines* to soybean. *Mol. Plant Microbe Interact.* **2012**, *25*, 1104–1117.
13. Shadel, G.S.; Young, R.; Baldwin, T.O. Use of regulated cell lysis in a lethal genetic selection in *Escherichia coli*: Identification of the autoinducer-binding region of the LuxR protein from *Vibrio fischeri* ATCC 7744. *J. Bacteriol.* **1990**, *172*, 3980–3987.
14. Choi, S.H.; Greenberg, E.P. Genetic dissection of DNA binding and luminescence gene activation by the *Vibrio fischeri* LuxR protein. *J. Bacteriol.* **1992**, *174*, 4064–4069.
15. Choi, S.H.; Greenberg, E.P. The C-terminal region of the *Vibrio fischeri* LuxR protein contains an inducer-independent lux gene activating domain. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 11115–11119.
16. Zhu, J.; Winans, S.C. The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1507–1512.
17. Welch, M.; Todd, D.E.; Whitehead, N.A.; McGowan, S.J.; Bycroft, B.W.; Salmond, G.P.C. N-acyl homoserine lactone binding to the CarR receptor determines quorum-sensing specificity in *Erwinia*. *EMBO J.* **2000**, *19*, 631–641.
18. Nasser, W.; Reverchon, S. New insights into the regulatory mechanisms of the LuxR family of quorum sensing regulators. *Anal. Bioanal. Chem.* **2007**, *387*, 381–390.
19. Whitehead, N.A.; Barnard, A.M.; Slater, H.; Simpson, N.J.; Salmond, G.P. Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol. Rev.* **2001**, *25*, 365–404.
20. Fuqua, C.; Winans, S.C.; Greenberg, E.P. Census and consensus in bacterial ecosystems: The LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annu. Rev. Microbiol.* **1996**, *50*, 727–751.
21. Soulère, L.; Frezza, M.; Queneau, Y.; Doutheau, A. Exploring the active site of acyl homoserine lactones-dependent transcriptional regulators with bacterial quorum sensing modulators using molecular mechanics and docking studies. *J. Mol. Graph. Model.* **2007**, *26*, 581–590.
22. Ahumado, M.; Díaz, A.; Vivas-Reyes, R. Theoretical and structural analysis of the active site of the transcriptional regulators LasR and TraR, using molecular docking methodology for identifying potential analogues of acyl homoserine lactones (AHLs) with anti-quorum sensing activity. *Eur. J. Med. Chem.* **2010**, *45*, 608–615.
23. Zhang, R.G.; Pappas, K.M.; Brace, J.L.; Miller, P.C.; Oulmassov, T.; Molyneaux, J.M.; Anderson, J.C.; Bashkin, J.K.; Winans, S.C.; Joachimiak, A. Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. *Nature* **2002**, *417*, 971–974.
24. Vannini, A.; Volpari, C.; Gargioli, C.; Muraglia, E.; Cortese, R.; De Francesco, R.; Neddermann, P.; Marco, S.D. The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J.* **2002**, *21*, 4393–4401.
25. Chen, G.; Jeffrey, P.D.; Fuqua, C.; Shi, Y.; Chen, L. Structural basis for antiactivation in bacterial quorum sensing. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16474–16479.

26. Bottomley, M.J.; Muraglia, E.; Bazzo, R.; Carfi, A. Molecular insights into quorum sensing in the human pathogen *Pseudomonas aeruginosa* from the structure of the virulence regulator LasR bound to its autoinducer. *J. Biol. Chem.* **2007**, *282*, 13592–13600.
27. Zou, Y.; Nair, S.K. Molecular basis for the recognition of structurally distinct autoinducer mimics by the *Pseudomonas aeruginosa* LasR quorum-sensing signaling receptor. *Chem. Biol.* **2009**, *16*, 961–970.
28. Lintz, M.J.; Oinuma, K.; Wysoczynski, C.L.; Greenberg, E.P.; Churchill, M.E. Crystal structure of QscR, a *Pseudomonas aeruginosa* quorum sensing signal receptor. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15763–15768.
29. Chen, G.; Swem, L.R.; Swem, D.L.; Stauff, D.L.; O’Loughlin, C.T.; Jeffrey, P.D.; Bassler, B.L.; Hughson, F.M. A strategy for antagonizing quorum sensing. *Mol. Cell.* **2011**, *42*, 199–209.
30. Yao, Y.; Martinez-Yamout, M.A.; Dickerson, T.J.; Brogan, A.P.; Wright, P.E.; Dyson, H.J. Structure of the *Escherichia coli* quorum sensing protein SdiA: Activation of the folding switch by acyl homoserine lactones. *J. Mol. Biol.* **2006**, *355*, 262–273.
31. Armougom, F.; Moretti, S.; Poirot, O.; Audic, S.; Dumas, P.; Schaeli, B.; Keduas, V.; Notredame, C. Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* **2006**, *34*, W604–W608.
32. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
33. *The PyMOL Molecular Graphics System*, Version 1.5.0.4, Schrödinger, LLC: Palo Alto, CA, USA, 2002.
34. Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, *372*, 774–797.
35. Patankar, A.V.; González, J.E. Orphan LuxR regulators of quorum sensing. *FEMS Microbiol. Rev.* **2009**, *33*, 739–756.
36. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinforma.* **2008**, *9*, 40–47.
37. Hernandez, M.; Ghersi, D.; Sanchez, R. SITEHOUND-web: A server for ligand binding site identification in protein structures. *Nucleic Acids Res.* **2009**, *37*, W413–W416.
38. Taylor, W.R. Protein structure comparison using SAP. *Methods Mol. Biol.* **2000**, *143*, 19–32.
39. Zhang, Y.; Skolnick, J. TM-align: A protein structure alignment algorithm based on TM-score. *Nucleic Acids Res.* **2005**, *33*, 2302–2309.
40. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22*, 195–201.
41. Huang, X.; Miller, W. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **1991**, *12*, 337–357.
42. Eswar, N.; John, B.; Mirkovic, N.; Fiser, A.; Ilyin, V.A.; Pieper, U.; Stuart, A.C.; Marti-Renom, M.A.; Madhusudhan, M.S.; Yerkovich, B.; *et al.* Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res.* **2003**, *31*, 3375–3380.
43. Martí-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.
44. Pieper, U.; Eswar, N.; Stuart, A.C.; Ilyin, V.A.; Sali, A. MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res.* **2002**, *30*, 255–259.

45. Sanchez, R.; Sali, A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13597–13602.
46. Fernandez-Fuentes, N.; Madrid-Aliste, C.J.; Rai, B.K.; Fajardo, J.E.; Fiser, A. M4T: A comparative protein structure modeling server. *Nucleic Acids Res.* **2007**, *35*, 363–368.
47. Rai, B.K.; Fiser, A. Multiple mapping method: A novel approach to the sequence-to-structure alignment problem in comparative protein structure modeling. *Proteins* **2006**, *63*, 644–661.
48. Biegert, A.; Mayer, C.; Remmert, M.; Soding, J.; Lupas, A.N. The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res.* **2006**, *34*, W335–W339.
49. Wallner, B.; Elofsson, A. Can correct protein models be identified? *Protein Sci.* **2003**, *12*, 1073–1086.
50. Mereghetti, P.; Ganadu, M.L.; Papaleo, E.; Fantucci, P.; De Gioia, L. Validation of protein models by a neural network approach. *BMC Bioinforma.* **2008**, *9*, 66–76.
51. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: An automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.
52. Black, S.D.; Mould, D.R. Amino acid scale: Hydrophobicity of physiological L-alpha amino acids. *Anal. Biochem.* **1991**, *193*, 72–82.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).