Identification of novel autoinducer-2 receptors in Clostridia reveals plasticity in the binding site of the LsrB receptor family

Inês M. Torcato[1,2], Meghann R. Kasal[3#], Patrícia H. Brito[1,4], Stephen T. Miller[3*], Karina B. Xavier[1*]

From the [1]Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156, Oeiras, Portugal, the [2]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Avenida da República, 2780-157, Oeiras, Portugal, the [3]Department of Chemistry and Biochemistry, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081, USA and the [4]Chronic Disease Research Center (CEDOC), NOVA Medical School, Universidade NOVA de Lisboa, Rua Câmara Pestana, 6, 1150-082 Lisboa, Portugal.

Running title: *AI-2 receptors in Clostridia*

[#]Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

*To whom correspondence should be addressed: Stephen Miller: Department of Chemistry and Biochemistry, Swarthmore College, 500 College Ave., Swarthmore PA 19081, USA; Telephone (+1) 6109576063; E-mail: smiller1@swarthmore.edu OR Karina Xavier: Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal; Telephone: (+351) 214464655; E-mail: kxavier@igc.gulbenkian.pt

Keywords: autoinducer-2; quorum-sensing; LsrB receptors; microbiome; *Clostridium saccharobutylicum*; *Clostridium autoethanogenum*; cell signaling; crystal structure; ITC; ABC transporters

---

ABSTRACT

Autoinducer-2 (AI-2) is unique among quorum sensing signaling molecules, as it is produced and recognized by a wide variety of bacteria and thus facilitates interspecies communication. To date, two classes of AI-2 receptors have been identified: the LuxP-type, present in the *Vibrionales,* and the LsrB-type, found in a number of phylogenetically distinct bacterial families. Recently, AI-2 was shown to affect the colonization levels of a variety of bacteria in the microbiome of the mouse gut, including members of the genus *Clostridium*, but no AI-2 receptor had been identified in this genus. Here, we identify a noncanonical, functional LsrB-type receptor in *Clostridium saccharobutylicum.* This novel LsrB-like receptor is the first one reported with variations in the binding site amino acid residues that interact with AI-2. The crystal structure of the *C. saccharobutylicum* receptor, determined at 1.35 Å, resolution revealed that it binds the same form of AI-2 as the other known LsrB-type receptors, and isothermal titration calorimetry (ITC) assays showed that binding of AI-2 occurs at a submicromolar concentration. Using phylogenetic analysis, we inferred that the newly identified noncanonical LsrB receptor shares a common ancestor with known LsrB receptors and that noncanonical receptors are present in bacteria from different phyla. This led us to identify putative AI-2 receptors in bacterial species in which no receptors were known, as in bacteria belonging to the Spirochaetes and Actinobacteria phyla. Thus, this work represents a significant step towards understanding how AI-2-mediated quorum sensing influences bacterial interactions in complex biological niches.

Bacteria are able to colonize and adapt to different environments, and the process of quorum sensing greatly enhances this ability. Quorum sensing is a mechanism of cell-to-cell communication mediated by exchange of small chemical molecules, named autoinducers, that allows bacteria to monitor their population density and regulate gene expression

accordingly (1-3). Behaviors regulated by this process include biofilm formation, virulence factor expression, antibiotic production, and bioluminescence (4,5). Autoinducer-2 (AI-2) is a signaling molecule produced and recognized by a wide variety of bacteria and thus facilitates interspecies communication (4,6-8). Since bacteria often produce species-specific signals as well, the relative proportion of AI-2 depends on the identity and number of bacterial species present in the environment; thus AI-2-sensitive bacteria can, in principle, regulate their behavior according to the species composition of the community (9).

To date, all characterized AI-2 receptors have high sequence similarity to, and complete binding site identity with, one of two receptors: LuxP or LsrB (4,6,7). Both receptors belong to the high affinity substrate-binding protein family. Importantly, the two receptors bind chemically distinct derivatives of the AI-2 precursor molecule DPD ((S)-4,5-dihydroxypentane-2,3-dione) (6,7). Whereas LuxP binds S-THMF-borate [(2S,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran-borate], LsrB binds R-THMF [(2R,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran], with discrimination between these AI-2 adducts being presumably due to differences in the amino acid composition of the binding sites of these receptors (7). Regulatory mechanisms also differ by receptor type. In response to AI-2 binding, LuxP modulates the activity of a membrane-spanning sensor histidine kinase protein, thus regulating a phosphorylation signal transduction cascade, whereas LsrB is part of an ATP binding cassette (ABC) transport system that internalizes AI-2 (10-12). LsrB was first identified in *Salmonella enterica* serovar Typhimurium but, unlike LuxP, which has only been found in the Vibrionales, it is also present in other enteric bacteria and members of the *Rhizobiaceae* and *Bacillaceae* families (7,13,14).

The LsrB-mediated transport system is comprised of two transmembrane proteins, LsrC and LsrD, and an ATPase, LsrA, that are encoded in the *lsr* (for LuxS regulated) operon together with LsrB. As the density of AI-2-producing bacteria increases, AI-2 accumulates extracellularly. When the threshold concentration is reached, LsrB-bound AI-2 is internalized via this transport system (14,15). Internalized AI-2 is subsequently phosphorylated by the kinase LsrK (16-18), and phosphorylated AI-2 relieves the repression caused by the *lsr* regulator, LsrR (19,20). This derepression leads to *lsrABCDGF* expression and thus rapid depletion of AI-2 from the extracellular medium. In most bacteria that possess the *lsr* operon, the enzymes responsible for further AI-2 processing, LsrG and LsrF, are also encoded in the *lsr* operon (see Fig. 1 for an explanatory scheme) (21,22). It has been shown that Lsr-mediated AI-2 internalization influences the expression of downstream genes involved in regulating aggregation, attachment, and biofilm formation in *Escherichia coli* (23-25). Moreover, in pathogenic enterohemorrhagic *E. coli*, extracellular AI-2 acts as a chemoattractant; this activity, mediated by the LsrB receptor, is required for cell aggregation and biofilm formation (24-28). Thus, AI-2 internalization presumably regulates these behaviors via the different responses exerted by intra and extracellular AI-2.

In previous work, we identified LsrB receptors in several species (including *Bacillus anthracis* and *Bacillus cereus,* the first Firmicutes to be shown to have functional LsrB receptors) through an approach that used bioinformatics to identify candidate receptors and biochemical and genetic studies to confirm their function (13). Based on this work, we proposed the following criteria to identify functional LsrB receptors: i) more than 60% sequence identity with the LsrB from *Salmonella* serovar Typhimurium, ii) conservation of all six-binding site amino acid residues that interact with AI-2 and iii) co-occurrence with orthologs to the other key transport proteins encoded by the *lsr* operon. With the constant increase in the number of sequenced genomes available, new strains of bacteria can now be studied. It became clear in the course of this work that, while these criteria do identify functional receptors, the identification of a more diverse range of AI-2 receptors requires the expansion of these criteria.

AI-2 receptors are relevant in complex communities such as the gut microbiota, a clinically significant community where species imbalance has been linked to bowel disease, obesity and susceptibility to pathogen colonization. However, little is known about the interspecies communication mechanisms typically employed in this environment, partly due to the inability to identify the AI-2 receptors of certain members of the microbiota (29,30). Recently, it was shown that AI-2 can influence the proportions of the major phyla in antibiotic-

treated mice gut microbiota (29). Specifically, the Firmicutes phylum was shown to respond positively to this signal. Interestingly, the majority of Firmicutes, including many Clostridia, have orthologs of the AI-2 synthase (LuxS) and are therefore putative AI-2 producers (29); however, no AI-2 receptors were known or readily identified by prior criteria in the Clostridiales. These findings, together with increasing understanding of the importance of Clostridia members as both commensals and pathogens in the gut microbiota, led us to search for AI-2 receptors in this class (31-33).

In this work we identify new LsrB-type receptors in *Clostridium saccharobutylicum* and *Clostridium autoethanogenum*, species shown to exist in the microbiota of mammals (34,35). We characterize the *C. saccharobutylicum* receptor and show that this novel receptor binds R-THMF like previously characterized LsrBs, despite the fact that two of the amino acids involved in AI-2 binding differ from those of the canonical LsrBs (13). This variation demonstrates that the plasticity of the LsrB binding site is higher than previously thought and expands our understanding of the diversity among AI-2 binding receptors. To the best of our knowledge, this is the first report of the identification and characterization of AI-2 receptors in Clostridia. Moreover, this knowledge allowed us to identify new noncanonical putative AI-2 receptors in other bacterial species belonging to the Firmicutes phylum, including bacterial species isolated from human microbiota. We also identified putative noncanonical LsrB receptors in organisms belonging to other phyla where AI-2 receptors had not previously been identified and where AI-2 mediated quorum sensing can now be studied. The identification of AI-2 receptors in these species represents a key step in understanding the mechanisms by which AI-2 regulates the levels of Firmicutes in the mouse gut and, more broadly, how AI-2-mediated quorum sensing influences the behavior of this and other complex communities.

Results

Identification of LsrB orthologs and in vitro binding to AI-2

To identify novel LsrB receptors, we searched the complete genomes deposited in the Kyoto Encyclopedia of Genes and Genomes (KEGG) for LsrB orthologs. We were seeking to identify functional LsrB receptors in the Firmicutes phylum, but we only obtained hits in members of the *Bacillus* genus using the previous established criteria (13). However, we were intrigued by the identification of LsrB homologous proteins in the Clostridia class. These hits had sequence identities lower than 60%, the minimum observed in previously identified LsrB functional receptors, but still higher than 36%, the maximum observed for the homologs unable to bind AI-2. Proteins identified from *C. saccharobutylicum* and *C. autoethanogenum* shared a sequence identity with the LsrB from *B. cereus* of 39.2 and 37.9%, respectively (Table S2). Sequences for these proteins were submitted to the Phyre2 fold-recognition server and the top two hits for both were the LsrBs from *B. anthracis* and *Salmonella* serovar Typhimurium, suggesting that these proteins share the same overall tertiary structure as previously characterized LsrB receptors. We aligned the resulting predicted structures to the LsrB from *B. anthracis* (Protein Data Bank (PDB) ID: 4PZ0) and found that the putative binding site in the Clostridia orthologs varied from canonical LsrBs in two of the six AI-2 binding residues (P223N, A225S; numbering based on that of the PDB structure of CsLsrB). We also analyzed the genome context of these proteins and found the genome location of the genes encoding the putative LsrB receptors of *C. saccharobutylicum* and *C. autoethanogenum* to be identical to that of bacteria with functionally characterized LsrB receptors. The genes are located in a putative operon with homologs for an ABC transport system (Fig. 2, A; Table S2). These findings, combined with the biological relevance of the Clostridia class, led us to test if these proteins were functional LsrB-like receptors despite their lower sequence identity and the absence of a fully conserved AI-2 binding site.

To determine the functionality of these proteins, *in vitro* AI-2 binding studies using *Vibrio harveyi* strain MM32 (a reporter strain for AI-2 activity) were conducted (10). Candidate receptors from *C. saccharobutylicum* and *C. autoethanogenum* were expressed in AI-2-producing *E. coli* strains, purified, and heat-denatured to release the bound ligand. The ligands released from the putative receptors induced light production in *V. harveyi* at levels similar to that of our positive control, ligand released from *B. anthracis* LsrB (Fig. 2, B). As a negative control, receptors were also expressed

in a LuxS$^-$ *E. coli* strain (i.e. non-AI-2-producing); as predicted these proteins elicited no response from *V. harveyi* showing that the response is AI-2-dependent (Fig 2, B). Given this demonstration of AI-2 binding ability by the candidate proteins, we concluded that they are functional AI-2 receptors. We named these receptors CsLsrB for *C. saccharobutylicum* LsrB and CaLsrB for *C. autoethanogenum* LsrB.

AI-2 internalization in C. saccharobutylicum

As *C. saccharobutylicum* and *C. autoethanogenum* have orthologs for the Lsr transporter system, we tested whether accumulation and internalization of AI-2 was similar to previously studied bacteria with the *lsr* operon (13,15,16). Due to the similarity of the putative LsrBs identified in *C. saccharobutylicum* and *C. autoethanogenum*, we focused our studies on *C. saccharobutylicum* since the strain is more amenable to manipulation. Cell-free supernatants from *C. saccharobutylicum* cultures were collected at different time points post-inoculation and the activity of extracellular AI-2 was assayed via a *V. harveyi* MM32 luminescence assay (Fig. 3, A). We observed that the levels of extracellular AI-2 increased until a certain threshold level was reached, then rapidly decreased starting at mid-exponential growth. To further support the observation that *C. saccharobutylicum* is capable of internalizing AI-2, we added synthetic AI-2 to cultures of *C. saccharobutylicum* and of *E. coli* at time zero. We then compared the concentration of extracellular AI-2 present in *C. saccharobutylicum* supernatants to that in supernatants of the *E. coli* mutant that can neither produce or internalize AI-2 (*ΔlsrKΔluxS* mutant) (Fig. 3, B). As this *E. coli* mutant is not capable of internalizing AI-2, any decrease in AI-2 activity in the supernatants from these *E. coli* cultures would be due to degradation of AI-2 in these conditions. At 6h, supernatants of *C. saccharobutylicum* only stimulated approximately 280-fold induction, whereas supernatants of *E. coli ΔlsrKΔluxS* caused a 2000-fold induction, supporting the conclusion that *C. saccharobutylicum* depletes AI-2 from the culture by internalization. Moreover, AI-2 is not degraded extracellularly, as no significant degradation was observed in the course of 6h incubation of 40 μM of AI-2 in cell-free supernatants (Fig. S1, A) collected from a late stationary phase culture of *C. saccharobutylicum* (culture shown in Fig. S1, B). Thus, we concluded that *C. saccharobutylicum* is able to internalize AI-2 as previously observed for other LsrB receptors.

C. saccharobutylicum LsrB binds R-THMF

To determine the form of AI-2 recognized by CsLsrB and CaLsrB and the identity of the AI-2 binding residues, we determined the crystal structure of CsLsrB. This receptor was expressed in a LuxS$^+$ *E. coli* strain and crystallized; the structure was solved at 1.35Å resolution by molecular replacement omitting the ligand from the molecular replacement model (Table S3). The crystal structure shows that CsLsrB has a classic periplasmic binding protein fold like the canonical LsrB receptors from *Salmonella* serovar Typhimurium (7), *Salmonella enterica* serovar Typhi (PDB ID: 5GTA), *Sinorhizobium meliloti* (36), *Yersinia pestis* (37), and *B. anthracis* (PDB ID: 4PZ0), with two α/β domains connected through a three-stranded hinge. The binding site is located near this hinge in the cleft between the two domains (Fig. 4, A). Superimposition of the CsLsrB structure with that of *B. anthracis* LsrB shows that the structures of the two proteins are very similar (RMSD 0.78 Å (Fig. 4, A)). After refinement of the structure, clear electron density showed the ligand to be R-THMF, the same form of AI-2 recognized by the canonical LsrB receptors. The crystal structure also revealed that, as predicted by Phyre2, the AI-2 binding residues in CsLsrB (and, presumably by extension, CaLsrB) differ from those of the canonical LsrBs in two positions: a proline replaced with an asparagine and an alanine substituted with a serine (P225N, A227S numbering using *B. anthracis* LsrB as reference; compare Fig. 4, B with 4, C). The structure has been deposited in the PDB under the description "LsrB from *Clostridium saccharobutylicum* in complex with AI-2" with the PDB ID 6DSP.

The LsrB binding site is more plastic than previously thought

The observation that CsLsrB binds R-THMF with two of the six canonical AI-2 binding residues altered, led us to investigate whether other amino acid residues could be altered without loss of AI-2 binding ability. We performed alanine scanning mutagenesis on each of the six AI-2-interacting residues in the binding site (K29, D110, D167, Q168, N223, S225). To assess AI-2 binding of these mutants, each alanine mutant protein was expressed,

purified, and tested for AI-2 binding ability via the *V. harveyi* MM32 assay. We observed that alanine substitutions in three out of the four amino acid residues that bind AI-2 through side chains (K29, D110 and Q168) caused loss of AI-2 binding ability and are thus necessary for binding (Fig. 5, A). The exception to this observation was D167, which is thought to make a hydrogen bond with AI-2 through the carboxylic group of the γ carbon and did not cause CsLsrB to lose its ability to bind AI-2 when substituted by an alanine.

A D167N substitution was also constructed since the asparagine mutation at this site was previously observed as one of two recurring amino acid substitutions in putative LsrB homologs that failed to bind AI-2 (13); AI-2 binding was abolished in this mutant (Fig. 5, A). The second substitution commonly observed in non-functional LsrB orthologues is a threonine at position 225. As CsLsrB has a serine in that position, we performed the single S225T substitution and observed that AI-2 binding was abolished (Fig. 5).

We also tested whether reverting the noncanonical binding site residues to the canonical ones present in previous characterized LsrBs would hinder AI-2 binding by replacing the serine at position 225 with an alanine (S225A) and the asparagine at position 223 with a proline (N223P). Both mutants were able to bind AI-2, showing that the binding site of the CsLsrB receptor is plastic enough to accept reversion to the canonical LsrB binding site composition. Altogether, these results allow us to conclude that the residues interacting with AI-2 in the LsrB binding site are not constrained to a single composition, as previously thought.

LsrB receptors bind AI-2 with high affinity

Both the canonical LsrB proteins and the noncanonical CsLsrB studied here share the fold with the proteins of the large family of substrate binding proteins. Proteins of this family usually bind their ligands with high affinity; in fact, the affinity is sufficiently strong as to allow their purification with bound substrates (38). Dissociation constants in this family of proteins and their cognate ligands typically range between 0.1 and 1 μM for sugars and around 0.1 μM for amino acids (38). Thus, we would expect that LsrB receptors bind AI-2 with a dissociation constant in the submicromolar range. Supporting this hypothesis, CsLsrB purified and crystallized with AI-2 bound (above). To quantify this interaction, we used isothermal titration calorimetry (ITC) to determine the affinity of CsLsrB for AI-2 and compared it with the affinities measured for the canonical LsrBs from *B. anthracis* and *E. coli*. As expected, we observed that all LsrB proteins tested were able to bind AI-2 with a dissociation constant (K_d) smaller than 1 μM (Fig. 6). We measured K_d of 0.81 ± 0.13 μM for CsLsrB (Fig. 6, *A*), 0.20 ± 0.04 μM for *B. anthracis* LsrB (Fig. 6, *B*) and 0.19 ± 0.03 μM for *E. coli* LsrB (Fig. 6, *C*). As a negative control, we performed the ITC experiments with a substrate binding protein from *Rhizobium etli*, that is homologous to LsrB and thought to bind rhamnose (13,39). Both the raw data and the heat of reaction curve show that, even with AI-2 at a concentration nearly 4 times higher than in the other ITC assays, there is no specific binding, though nonspecific binding seems to occur without apparent saturation of the binding site (Fig. 6, *D*). Together, these results show that the AI-2 receptors tested bind AI-2 with a submicromolar affinity.

CsLsrB shares a common ancestor with canonical LsrB receptors

The identification of a functional LsrB receptor with naturally occurring variations in the amino acid residues of the binding site led us to inquire if CsLsrB arose from the same ancestor as the canonical LsrBs. To answer this question, a phylogenetic analysis was performed by querying the protein sequence of CsLsrB against the UniProt database of Reference Proteomes. Our analysis indicates that the large majority of LsrB proteins are distributed between two unrelated phyla, the Proteobacteria, particularly in Gammaproteobacteria and Alphaproteobacteria and the Firmicutes, from the orders Clostridiales and Bacillales (Fig. 7). The LsrB sequence of *C. saccharobutylicum* clusters in close proximity with other sequences from Clostridiales and in the same evolutionary lineage as the LsrB proteins from the *B. cereus* group previously shown to have functional AI-2 receptors (Fig. 7). There is clear evidence of lateral gene transfer events in the evolution of LsrB sequences, which explains the disjoint presence of LsrB orthologs within the two unrelated phyla (Proteobacteria and Firmicutes), as well as the presence of LsrB sequences from Actinobacteria and Spirochaetes clustering within a clade largely composed of Firmicutes species.

To obtain further support for the evolutionary position of CsLsrB, we performed a similar phylogenetic analysis focusing on the ATP-binding proteins (LsrA), a protein typically encoded by a gene co-transcribed with *lsrB* as part of the *lsr* operon. Given that the function of these two proteins is usually associated, we expected their evolutionary histories to be congruent. Importantly, the ATP-binding proteins from the ABC transporters are known to evolve more slowly than the substrate-binding proteins and are therefore better genetic markers for tracing the evolution of the genes in this operon (40,41). The evolution of LsrA sequences corroborated all major conclusions obtained with LsrB (Fig. S3). Although some local differences in topology and bootstrap support between the two trees do occur, these differences are likely due to the different rates of protein evolution and do not cause conflict with the interpretation of the evolution of *C. saccharobutylicum* Lsr genes. The inferred evolution of LsrA sequences also corroborates the close relationship of the same species of Spirochaetes and Actinobacteria with the Clostridiales, which strongly suggests that events of operon lateral transfer are responsible for the observed taxonomic distribution of Lsr genes.

To determine which organisms have LsrB proteins with canonical (KDDQPA) or noncanonical (KDDQNS) predicted binding sites, we analyzed sequence conservation at the binding sites across species in Figure 7. Interestingly, some of the protein sequences in this phylogenetic tree have a serine instead of an alanine in position 225, as in CsLsrB, but retain the proline in position 223 observed for canonical receptors (KDDQPS) (Fig. 7). According to the results of our mutagenesis studies we infer that these noncanonical receptors with one substitution in the predicted AI-2 interacting residues will also be functional AI-2 receptors since the CsLsrB N223P mutant was able to bind AI-2 (Fig. 5, A).

Overall, the evolutionary history of the LsrB receptors supports the conclusion that lateral gene transfer is responsible for the appearance of canonical LsrB receptors in the Firmicutes, which subsequently evolved to noncanonical receptors in a stepwise manner accumulating one, then two, substitutions in the AI-2 binding residues. Most noncanonical receptors cluster together, but events of lateral gene transfer are again evident, as members from the Spirochaetes and Actinobacteria seem to have acquired noncanonical LsrB receptors with one or two substitutions.

Identification of additional putative noncanonical LsrB receptors

Following the phylogenetic analysis of the LsrBs, it became clear that the noncanonical LsrB receptors were present in organisms other than *C. saccharobutylicum* and *C. autoethanogenum*. For this reason, we conducted a comprehensive bioinformatic analysis of proteomes currently available in the NCBI RefSeq non-redundant protein database using the sequence of CsLsrB to further identify organisms that possess noncanonical LsrB proteins with one (KDDQPS) or two substitutions (KDDQNS) in the putative binding site.

We identified 95 noncanonical putative LsrB receptors of which 26 have one substitution and 69 have two substitutions, as CsLsrB. The organisms encoding these proteins belong to the same phyla identified in our phylogenetic analysis with the large majority belonging to the Firmicutes phylum. To characterize these hits, all 95 protein sequences were submitted to the fold recognition server Phyre2. The predicted structures for all putative LsrB hits were consistent with the class I periplasmic binding protein fold shared by the canonical LsrBs and the noncanonical CsLsrB (Table S4). We also determined whether these hits had Lsr transport proteins next to LsrB in their genomes (Table S4). From the 95 hits, only the LsrB proteins from *Treponema primitia* and *Treponema azotonutricium* did not have Lsr transport proteins next to the LsrB protein (Table S4). In these genomes the LsrB homolog is located near a hybrid sensor histidine kinase/response regulator, although homologs for Lsr proteins were identified in another location in the genome (Table S4). The presence of the enzymes involved in AI-2 metabolization was also assessed. Even though LsrF was present in 93 of the 95 hits, LsrG was only present (with a query cover higher than 75%) in 9 proteomes. Finally, we checked the proteomes for the existence of the AI-2 synthase LuxS. We did not identify LuxS homologs in all these organisms (79 LuxS homologs in 95 hits; Table S4). However, this is also true for *Sinorhizobium meliloti*, a species which carries a functional *lsr* operon but is unable to produce AI-2 (36). Altogether, these results suggest that functional,

noncanonical LsrB receptors are present in many bacterial species, particularly in the Firmicutes phylum.

Discussion

LsrB AI-2 receptors share an identical fold to substrate-binding proteins that recognize a wide variety of substrates including 5-carbon sugars similar to AI-2. Thus, the identification of LsrB receptors has been possible due to the conservation of the six amino acid residues that bind AI-2, which allows the distinction between LsrB receptors and other substrate-binding proteins. In this study, we provide the first report of a functional LsrB receptor in Clostridia; this receptor has a previously unobserved binding site variation and gives insights into the plasticity of the binding site of LsrB receptors. The crystal structure of this receptor identified in *C. saccharobutylicum*, CsLsrB, showed that it binds the R-THMF form of AI-2 with two variations in the amino acid residues that bind AI-2: an asparagine instead of a proline (P223N) and a serine instead of an alanine (A225S) (Fig. 4 B and C). The fact that the substitution of the nonpolar proline and alanine for the polar asparagine and serine do not abolish AI-2 binding was surprising both due to the dissimilar nature of the amino acids and the fact that the canonical AI-2 binding residues were thought to be essential. However, examination of both canonical and noncanonical AI-2 binding pockets reveals that the amino acids in these positions bind AI-2 through the backbone. Moreover, the crystal structure of CsLsrB shows that the amino and carboxyl groups are still positioned to form hydrogen bonds with similar geometry to canonical receptors; thus the local region must be plastic enough to accept side chain variation without significantly changing the overall structure of the protein. Still, not all amino acid residues can be accommodated in these positions as the S225T substitution caused loss of AI-2 binding. The side chain hydroxyl group of the threonine would possibly be shifted by the presence of the additional methyl group, causing a change in geometry that would likely impact the interaction with the ligand. In addition, this shift could alter the interaction between S225T and N33, potentially causing a conformational change that impacts the surrounding amino acids, altering the conformation of the binding site further (Fig. 5). As for D167N, which similarly to S225T, is a naturally occurring substitution in sugar binding proteins with homology to LsrB, the inability to bind AI-2 might be a result of the change from a presumably negatively-charged side chain to a side chain containing an amine. This could lead to unfavorable interactions with the ligand or the adjacent Q168, which also binds AI-2 through side chain interactions. Moreover, the introduction of the amine in this mutation could lead to more fundamental conformational changes in the binding site and surrounding environment than might be expected for a simple alanine mutation, explaining the loss of binding activity in the Asn but not the Ala mutant (Fig. 5). Accordingly, S225T and D167N substitutions also caused loss of AI-2 binding ability in *B. anthracis* LsrB (13), a canonical receptor, indicating that the presence of a threonine in position 225 and an asparagine in position 167 might be characteristic of sugar-binding proteins.

The two variations in AI-2 binding residues of CsLsrB might explain the slightly lower affinity of this receptor to AI-2 when compared with the canonical LsrBs tested. Nevertheless, CsLsrB bound AI-2 with high affinity (submicromolar K_d; Fig. 6). Interestingly, in a previous study by Zhu and Pei a significantly higher K_d (160 μM) was reported for the binding of the LsrB from *Salmonella* serovar Typhimurium to AI-2 (42). However, our results were obtained in the absence of boron and, as was shown in Miller *et. al*, the removal of boron is essential to avoid shifting the equilibrium of AI-2 molecules toward the S-THMF-borate form (7). Hence, the higher apparent K_d value obtained by Zhu and Pei might result from the presence of boron. Moreover, the authors reported a K_d of 0.16 μM for LuxP from *V. harveyi* (42), a value closer to the ones we obtained for the LsrBs tested here and more consistent with the values expected for substrate binding proteins towards their cognate ligands (38).

Similarly to canonical LsrB receptors, CsLsrB was able to internalize AI-2 (Fig. 3), likely through interaction with the ABC transporter encoded in the same operon as *lsrB* (Fig. 2, A). Curiously though, AI-2 uptake started at mid-exponential phase (Fig. S1, B) earlier than what is usually observed for *E. coli* K12 MG1655 that starts internalization in the transition from exponential to stationary phase (Fig. S1, C and (9,13)). Additionally, *C. saccharobutylicum* supernatants induced 10-fold less light production, indicating that it accumulates lower concentrations of extracellular AI-2. The

observed difference is likely because *C. saccharobutylicum* starts AI-2 uptake at lower extracellular AI-2 concentrations indicating that, at least under the laboratory conditions tested, *C. saccharobutylicum* has a lower concentration threshold than *E. coli.* This might indicate that the AI-2 concentration needed for induction of the *lsr* operon in *C. saccharobutylicum* is lower than in *E. coli* K-12 MG1655. As *C. saccharobutylicum* and *E. coli* LsrB receptors have high affinity to AI-2, it is unlikely that differences in affinities of the receptors cause the variation in concentrations required for the induction of the operons and, thus, in the start of internalization. Moreover, *E. coli* can accumulate extracellular AI-2 to concentrations up to 40 μM without starting to internalize AI-2. This indicates that, in *E. coli,* additional mechanisms are preventing this bacterium from starting internalization of AI-2 at lower concentrations (43). In fact, in *E. coli,* the Lsr system is regulated not only by LsrR but also by catabolite repression and the PTS system, such that the availability of PTS substrates inhibits Lsr-mediated AI-2 internalization (43,44). Thus, it is possible that these additional regulatory mechanisms are either different or not present in *C. saccharobutylicum*, allowing internalization to start at lower concentrations of AI-2. Further work on the regulation of these systems is necessary to address this hypothesis. The search for new noncanonical receptors revealed that in most organisms where putative noncanonical LsrB receptors were identified, the *lsrB* neighboring genes encoded homologs for the components of the Lsr transport system. In contrast, homologs of the AI-2 processing enzymes, LsrG and LsrF, identified in *E. coli* and *Salmonella* serovar Typhimurium*,* were not as widely conserved. For example, in *C. saccharobutylicum* no LsrF ortholog was identified and in both *C. saccharobutylicum* and *C. autoethanogenum*, the gene that encodes the LsrG orthologs is present in another region of the genome (Fig. 2, A). These results, together with previous reports of canonical LsrB receptors encoded in operons that also encode functional Lsr transport systems but have one or both LsrG and LsrF orthologs located in other regions of the genome (13,36) indicate that AI-2 metabolism might not be as conserved as the steps required for the uptake, phosphorylation and regulation of the *lsr* operon. Thus, the lack of orthologs for the proteins responsible for AI-2 metabolization (LsrG and LsrF) should not be an excluding factor when searching for new LsrB receptors. Curiously, in *T. primitia* and *T. azotonutricium* we identified putative noncanonical LsrBs that were near a sensor kinase instead of an ABC transporter in the genome. LuxP, the other type of known AI-2 receptor known, interacts with a sensor kinase (LuxQ) that facilitates signal transduction through a phosphorylation cascade (2). However, when searching for homologs of the Lsr proteins, we found hits for all the proteins, except LsrR, located in a more distant region of the genome (Table S4). Hence, further studies are needed to understand if these putative LsrBs evolved to interact with the nearby sensor kinase or if they instead use an ABC transporter located in another region of the genome.

In summary, the characterization of the amino acid variation in the AI-2 binding site of CsLsrB fulfills our initial motivation as it allowed the identification of new putative LsrB receptors in microbes present in biologically relevant niches like *Roseburia inulinivorans* (45) and *Clostridium merdae* (46), Firmicutes isolated from human microbiota that appear to have a noncanonical receptor with two substitutions (like CsLsrB). This is a significant step towards understanding the molecular mechanisms involved in interspecies communication in these niches. In particular, we expect the identification of AI-2 receptors in Firmicutes to help in the characterization of the mechanism by which their colonization is favored in the presence of AI-2 in the mammalian gut microbiome, a clinically relevant niche where interspecies interactions are highly prevalent.

Experimental procedures

Identification of LsrB orthologs

The search for LsrB orthologs was performed as described previously (13). Briefly, we searched the complete genomes in the KEGG SSDB (Sequence Similarity Database, 3478 bacterial genomes in September 2015) for amino acid sequences with similarity to LsrB from *B. cereus*. All pairwise genome comparisons were performed using SSEARCH program and best bidirectional hits with a Smith-Waterman score of at least 120 were selected. From the obtained hits with sequence identity between 30-60% we focused on the putative LsrBs of *C. saccharobutylicum* and *C. autoethanogenum.* The genome context of the CsLsrB and CaLsrB genes was assessed through analysis of the

neighboring genes. To determine the presence/absence of Lsr orthologs in *C. saccharobutylicum* and *C. autoethanogenum,* we looked for best bidirectional hits using the Lsr proteins of *B. cereus* ATCC 10987 and *B. anthracis* Sterne as query. Simultaneously, we determined their similarity (Table S2). The amino acid sequences of the putative LsrBs from *C. saccharobutylicum* and *C. autoethanogenum* were submitted to the Phyre2 fold recognition server. The presence/absence of conserved amino acid residues in AI-2 binding was determined through structural alignment of the structures predicted by Phyre2 with the LsrB from *B. anthracis* (PDB ID: 4PZ0) using COOT (47).

AI-2 synthesis

The AI-2 precursor DPD was synthesized as previously described (48). Boron-free DPD was obtained by performing the synthesis with plastic material and using boron-free water (prepared by batch incubation with Amberlite IRA743 resin for 2h at room temperature, as described before (49)).

AI-2 internalization assays

Anaerobic frozen stocks of *C. saccharobutylicum* DSM13864 were revived on modified PY+X solid medium with galactose (see Table S1 for media composition). 13 mL of broth (modified PY+X with arabinose and 100 mM MOPS buffer, pH 7.0) was inoculated with 3-6 colonies of *C. saccharobutylicum* and grown overnight at 37ºC in an anaerobic chamber (PlasLabs, USA) under a gaseous mix of 80% nitrogen, 15% carbon dioxide and 5% hydrogen to an OD_{600nm} lower than 4. A 20% (v/v) inoculum was added to fresh medium and incubated at 37ºC without agitation in a 100 mL Hungate Schott flask. Culture suspensions were collected at the stated time points for optical density measurement at 600 nm (UV-visible spectrophotometer; Helios Delta, ThermoSpectronic, USA) and for detection of AI-2. *E. coli ΔluxSΔlsrK* (*E. coli* ARO093) (29) was revived directly from the frozen stock to liquid medium and a 8% inoculum was performed. Exogenous AI-2 was supplemented to a final concentration of 40 µM in fresh media. For AI-2 activity measurements, culture suspensions were filtered using multiscreen filter plates (Millipore). The cell-free culture was frozen at -20ºC overnight and AI-2 activity was accessed in triplicate using the *V. harveyi* MM32 bioluminescence reporter assay, as described previously (10,36). Light production was measured at 7h for *C. saccharobutylicum* cell-free culture fluids and at 5h for the cultures supplemented with 40 µM AI-2. Luminescence was measured with a Glomax Explorer microplate luminometer (Promega, EUA). AI-2 activity is reported as the induction of light production compared with the background light obtained with the appropriate growth medium. Standard deviation was calculated from three technical replicates. Propagation of uncertainty was employed to calculate the standard deviation after normalization. This experiment was repeated on three different days with three independent cultures. For simplicity, a representative experiment of the three independent experiments is shown (Fig. 3). It was not possible to analyze the data from the three different experiments together due to variations of the growth curves, in particular with respect to differences in the lag phase of the cultures that varied from 30 min to 4 hours. The three experiments are shown in Fig. S2.

Protein expression and purification

The gene encoding LsrB ortholog in *C. saccharobutylicum* DSM13864 was amplified from genomic DNA (DSMZ) and cloned in the plasmid pDEST-527 (gifted by Dominic Esposito, Addgene plasmid #11518) via the pENTR/TEV/D-TOPO cloning kit (Thermo Fisher Scientific) for expression as a His_6-tagged protein. The signaling sequence for secretion, as determined by SignalP 4.1 (MKKKAVALALIGAMIFTTLVGCG), was excluded from the construct. *E. coli* BL21 (DE3) $LuxS^+$ or $LuxS^-$ cells were transformed with the construct and grown in LB with 1 µg/mL ampicillin at 37°C until the optical density at 595 nm was 0.3. At this point, temperature was decreased to 22°C. At OD_{595} = 0.9, 0.3 mM of IPTG was added and the cells were induced for 6h before being harvested by centrifugation. Cells were resuspended in 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM imidazole, 2.5 µg/mL DNase and 2.5 µg/mL leupeptin and lysed using a M-110Y microfluidizer (Microfluidics, USA). The lysate was centrifuged and the tagged protein was purified from the clarified supernatants using Nickel-nitrilotriacetic (Ni^+-NTA) acid affinity chromatography (Qiagen). The protein was eluted from the column in 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl and 250 mM imidazole and

subsequently swapped into 25 mM Tris-HCl (pH 8.0), 50 mM NaCl and 1 mM dithiothreitol (DTT) using Sephadex-G25 agarose. The His_6-tag was removed by cleavage with tobacco etch virus (TEV) protease at a proportion of 1 mg TEV protease per 250 mg of protein at 4°C, overnight. A second round of Ni^+-NTA affinity chromatography was performed to remove the His_6-tagged TEV protease, the cleaved tag and any uncut fusion protein. The protein collected from the flow-through was buffer swapped into 25 mM Tris-HCl (pH 8.0), 50 mM NaCl and 1 mM DTT via Sephadex-G25 agarose and further purified by anion exchanging chromatography (SourceQ column, GE Healthcare Life Sciences) using a NaCl gradient from 0 to 1 M. As a final purification step, the protein underwent a size exclusion chromatography on a Superdex75 (GE healthcare) column and was eluted in 25 mM Tris-HCl (pH 8.0), 150 mM NaCl and 1 mM DTT.

Crystallization studies

Crystals of *C. saccharobutylicum* LsrB expressed in *E. coli* BL21 $LuxS^+$ were grown via the sitting drop method with a well solution of 0.1 M Citric acid pH 2.75 and 26% w/v PEG 3350 and developed in approximately one week at room temperature. Crystals were frozen after a 30 second soak in a solution of 27% w/v PEG 3350 plus 15% v/v glycerol. Diffraction data was collected on beamline BL14-1 at the Stanford Synchrotron Radiation Light Source. Data was processed with CCP4 software (50). A molecular replacement solution was determined via PHENIX (51) using 1TJY as the search model. The initial model was built by PHENIX with subsequent manual building in Coot (47) and refinement in PHENIX. The resulting PDB file is used as a reference for amino acid numbering of CsLsrB throughout this manuscript. Thus based on the structure, the amino acids of the binding site are numbered K29/D110/D167/Q168/N223/S225, while the numbering based on the complete amino acid sequence of the protein is K48/D129/D186/Q187/N242/S244.

Site directed mutagenesis

The QuikChange Lightning Site Directed Mutagenesis kit (Agilent) was employed to make the single amino acid substitutions in pDEST527/*C. saccharobutylicum* LsrB constructs. Primers were designed using the QuikChange primer design program (Agilent) and sequences are given in Table S5. The mutant proteins were expressed as described above. The His_6-tagged mutants were purified through Ni^+-NTA affinity chromatography (Qiagen) as above and then buffer swapped into 25 mM Tris-HCl (pH 8.0), 150 mM NaCl and 1 mM DTT using PD-10 desalting columns (GE Healthcare). Protein was concentrated in 10 kDa Centricons (Millipore) until a concentration of at least 10 mg/mL was achieved.

In vitro AI-2 binding assay

The *in vitro* AI-2 binding ability of the studied proteins was assessed as previously described (13). Briefly, LsrB proteins at a concentration of 10 mg/mL were heated at 70ºC for 10 min. The denatured protein was pelleted via centrifugation and 10 µL of the supernatant were added to 90 µL of a 1:5000 dilution of an overnight culture of *V. harveyi* MM32, as described above. Bioluminescence after 5h of incubation at 30ºC was measured. A 1420 Victor 2 plate reader (Perkin Elmer, USA) and a 1420 Victor 3 plate reader were used for data in Fig. 2 and 5, respectively. Binding data is representative of three independent experiments and the standard deviations are derived from three technical replicates. Propagation of uncertainty was used to calculate the standard deviation after normalizing the light produced by AI-2 by the light produced by buffer alone.

Isothermal titration calorimetry

ITC measurements were performed in a MicroCal iTC200 microcalorimeter (GE Healthcare Biosciences, Northampton, MA, USA) at 25ºC. Boron-free AI-2 was diluted in boron-free buffer containing 25 mM sodium phosphate buffer pH 8.0, 150 mM NaCl 1 mM β-mercaptoethanol. Amberlite IRA743 resin was employed to remove the boron from the water used to prepare buffer as previously described (49). To avoid contamination by boron silicates present in glass, only plastic material was used. AI-2 at 800 µM was injected into 117.4 µM and 108 µM of LsrB protein from *C. saccharobutylicum* and *E. coli* K-12 MG1655, respectively. For *B. anthracis* Sterne 34F2 LsrB, a solution of 750 µM AI-2 was added to 106.9 µM of protein. As a negative control, 3 mM of AI-2 was injected into 114.7 µM of the LsrB ortholog from *Rhizobium etli* CFN42 (RHE-PE00289 in KEGG, annotated as substrate binding protein involved in the rhamnose transport system). Measurements were made

with the reference power at 5 µcal/s and a syringe stirring speed of 800 rpm. The heat of dilution for successive injections of AI-2 into buffer was included in the final analysis. The heat of reaction for each injection was calculated by integrating the area under each titration peak under the assumption of a one-site binding model using MicroCal/Origin 7.0 software (provided by the manufacturer). Injections were: *C. saccharobutylicum*: 1 x 0.5 µL + 14 x 1.5 µL + 10 x 1 µL + 3x0.5 µL; *B. anthracis*: 1 x 0.5 µL + 4 x 2 µL + 23 x 1 µL; *E. coli*: 1 x 0.5 µL + 1 x 3 µL + 8 x 1.5 µL + 18 x 1 µL; *R. etli:* 1 x 0.5 µL + 18 x 2 µL. Binding curves shown are representative of three different runs.

Phylogenetic analysis

A homology search was performed by querying the protein sequence of CsLsrB against the UniProt database of Reference Proteomes using *phmmer* (52). This database contains proteomes that have been selected either manually or algorithmically in order to provide a broad coverage of the tree of life and a balanced cross-section of the taxonomic diversity found within UniProtKB, thus removing over-representation of certain species of bacteria caused by over sampling of preferred niches. At the time of analysis (June 22, 2017) it contained a total of 10454 proteomes, 6469 of which were from Bacteria. To extract significant hits we applied different e-value thresholds of increasing stringiness starting at 10^{-4} and performed phylogenetic analyses on the aligned datasets to evaluate its evolutionary structure. The results presented here resulted from applying an e-value threshold of 10^{-30}, retrieving 167 protein sequences (Table S6). This threshold excluded distant homologue sequences while keeping all sequences in our focal group and in the sister clade identified as rhamnose-binding proteins. To this dataset we added other sequences that were previously shown to be either functional LsrB receptors (11 sequences) or rhamnose-binding proteins (2 sequences) (Table S7) (13). This dataset was aligned with Mafft (v7.310) using the L-INS-i method and default parameter values (53). The aligned dataset was analyzed with Prottest version 3.42 (54) to estimate the most likely model of protein evolution and RaxML v. 8.0.26 (55) to produce a maximum-likelihood inference of the phylogenetic history of these proteins. Nodal support in the phylogenetic analysis was estimated with nonparametric bootstrap using an automatic frequency-based criterion (autoFC option) to determine the number of replicates. We identified two truncated proteins, one from *Clostridium magnum* str. DSM 2767 and the other from *Klebsiella pneumoniae* str. ISC21, that were eliminated from final analyses. The phylogenetic analysis of the 180 protein sequences homologous to CsLsrB indicated a clear separation between *bona fide* LsrB sequences and the paralogous sequences identified as rhamnose binding proteins (Fig. S4). This result was used to define a smaller dataset focused on exclusively LsrB sequences keeping all sequences within the most inclusive and highly supported clade that includes all confirmed functional LsrB sequences and our target LsrB sequence from *C. saccharobutylicum*. This final dataset of 97 protein sequences was realigned and reanalyzed as before. All sequences included have an e-value smaller than 10^{-63} and an average sequence identity of 43%. This smaller dataset produced a better alignment and a more accurate phylogenetic inference. For the phylogenetic study of LsrA, we performed a homology search by querying the ATP-binding protein of *C. saccharobutylicum* against the UniProt database of Reference Proteomes using *phmmer* algorithm. We extracted all significant hits using an e-value threshold of 10^{-95} which retrieved 470 sequences. To this dataset we added all ATP-binding protein sequences that pertain to the same strain/operon as the functional LsrB (11 sequences), if not previously included, as well as the two ATP-biding proteins from the rhamnose-biding operon (Table S8). These sequences were analyzed as explained above for LsrB sequences to produce a phylogenetic analysis of *bona fide* LsrA sequences present in the Reference Proteome database (Fig. S3).

Identification of noncanonical LsrBs

The search for putative noncanonical LsrB receptors was accomplished using BLASTP against the NCBI RefSeq non-redundant protein database using CsLsrB as query sequence (in May 2018). RefSeq allows sequence submission and includes complete and incomplete genomes, thus having more genomes available than KEGG Genome, which is a collection of organisms with complete genome sequences or than UniProt database of Reference Proteomes, which is a balanced database to inference evolutionary history (56,57). The alignments between the amino acid sequence of CsLsrB and the hit

proteins were examined to determine the identity of the six amino acid residues involved in AI-2 binding. To simplify the analysis, the search was conducted for each one of the 35 phyla described in KEGG separately. From the functional mutations identified in this study, only putative LsrBs with N223P and N223P/S225A were found. The annotation of the proteins in the vicinity of these noncanonical hits was assessed. For the hits with nearby Lsr orthologs, the amino acid sequences of these proteins were submitted to a comparative BLASTP with the respective ortholog in *C. saccharobutylicum*. For the absent orthologs and for LuxS, a BLASTP directed to the proteome of the organisms using the matching ortholog in *C. saccharobutylicum* as query was performed. As no LsrF homolog was identified in *C. saccharobutylicum*, the amino acid sequence of LsrF from *E. coli* K12 MG1655 was employed. Additionally, very few hits for LsrG homologs were obtained using the *C. saccharobutylicum* LsrG as query, so we also checked for homologs using the biochemically characterized LsrG from *E. coli* K12 MG1655. All of the 95 putative LsrB sequences were submitted to the Phyre2 fold recognition server.

Acknowledgements: The work performed in the laboratory of K.B.X. was supported by Portuguese National funds from Fundação para a Ciência e Tecnologia (FCT) PTDC/BIA-MIC/4188/14. K.B.X. and I.M.T. acknowledge FCT for individual grants IF/00831/2015 and PD/BD/105736/2014, being the latter within the scope of the PhD program Molecular Biosciences PD/00133/2012. The laboratory of S.T.M acknowledges Swarthmore College. M.R.K. acknowledges Mayer Davidson '57 and Frances Velay Womens Science research fellowships. The authors also thank Ana Rita Oliveira and Roberto Balbontín for helpful suggestions, Osvaldo Ascenso, Vanessa Miranda and Rita Ventura for the boron-free DPD, Steven Chen and Kendall Yoshii for help with the *in vitro* AI-2 binding assay of CsLsrB and CaLsrB and Paula Chicau and Marco Domingues for technical support with ITC.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Fuqua, W. C., Winans, S. C., and Greenberg, E. P. (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *Journal of bacteriology* **176**, 269-275
2. Waters, C. M., and Bassler, B. L. (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annual review of cell and developmental biology* **21**, 319-346
3. Whiteley, M., Diggle, S. P., and Greenberg, E. P. (2017) Progress in and promise of bacterial quorum sensing research. *Nature* **551**, 313-320
4. Pereira, C. S., Thompson, J. A., and Xavier, K. B. (2013) AI-2-mediated signalling in bacteria. *FEMS microbiology reviews* **37**, 156-181
5. Papenfort, K., and Bassler, B. L. (2016) Quorum sensing signal-response systems in Gram-negative bacteria. *Nature reviews. Microbiology* **14**, 576-588
6. Chen, X., Schauder, S., Potier, N., Van Dorsselaer, A., Pelczer, I., Bassler, B. L., and Hughson, F. M. (2002) Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* **415**, 545-549
7. Miller, S. T., Xavier, K. B., Campagna, S. R., Taga, M. E., Semmelhack, M. F., Bassler, B. L., and Hughson, F. M. (2004) Salmonella typhimurium recognizes a chemically distinct form of the bacterial quorum-sensing signal AI-2. *Molecular cell* **15**, 677-687
8. Surette, M. G., Miller, M. B., and Bassler, B. L. (1999) Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: a new family of genes responsible for autoinducer production. *Proc Natl Acad Sci U S A* **96**, 1639-1644
9. Xavier, K. B., and Bassler, B. L. (2005) Interference with Al-2-mediated bacterial cell-cell communication. *Nature* **437**, 750-753
10. Bassler, B. L., Wright, M., Showalter, R. E., and Silverman, M. R. (1993) Intercellular Signaling in Vibrio-Harveyi - Sequence and Function of Genes Regulating Expression of Luminescence. *Molecular microbiology* **9**, 773-786

11. Bassler, B. L., Wright, M., and Silverman, M. R. (1994) Multiple Signaling Systems Controlling Expression of Luminescence in Vibrio-Harveyi - Sequence and Function of Genes Encoding a 2nd Sensory Pathway. *Mol Microbiol* **13**, 273-286
12. Jung, S. A., Chapman, C. A., and Ng, W. L. (2015) Quadruple quorum-sensing inputs control Vibrio cholerae virulence and maintain system robustness. *PLoS pathogens* **11**, e1004837
13. Pereira, C. S., de Regt, A. K., Brito, P. H., Miller, S. T., and Xavier, K. B. (2009) Identification of functional LsrB-like autoinducer-2 receptors. *Journal of bacteriology* **191**, 6975-6987
14. Taga, M. E., Semmelhack, J. L., and Bassler, B. L. (2001) The LuxS-dependent autoinducer Al-2 controls the expression of an ABC transporter that functions in Al-2 uptake in Salmonella typhimurium. *Molecular microbiology* **42**, 777-793
15. Xavier, K. B., and Bassler, B. L. (2005) Regulation of uptake and processing of the quorum-sensing autoinducer AI-2 in Escherichia coli. *Journal of bacteriology* **187**, 238-248
16. Taga, M. E., Miller, S. T., and Bassler, B. L. (2003) Lsr-mediated transport and processing of Al-2 in Salmonella typhimurium. *Molecular microbiology* **50**, 1411-1427
17. Xavier, K. B., Miller, S. T., Lu, W. Y., Kim, J. H., Rabinowitz, J., Pelczer, I., Semmelhack, M. F., and Bassler, B. L. (2007) Phosphorylation and processing of the quorum-sensing molecule autoinducer-2 in enteric bacteria. *Acs Chem Biol* **2**, 128-136
18. Wang, L., Li, J., March, J. C., Valdes, J. J., and Bentley, W. E. (2005) luxS-dependent gene regulation in Escherichia coli K-12 revealed by genomic expression profiling. *Journal of bacteriology* **187**, 8350-8360
19. Wu, M., Tao, Y., Liu, X., and Zang, J. (2013) Structural basis for phosphorylated autoinducer-2 modulation of the oligomerization state of the global transcription regulator LsrR from Escherichia coli. *The Journal of biological chemistry* **288**, 15878-15887
20. Xue, T., Zhao, L., Sun, H., Zhou, X., and Sun, B. (2009) LsrR-binding site recognition and regulatory characteristics in Escherichia coli AI-2 quorum sensing. *Cell research* **19**, 1258-1268
21. Marques, J. C., Lamosa, P., Russell, C., Ventura, R., Maycock, C., Semmelhack, M. F., Miller, S. T., and Xavier, K. B. (2011) Processing the interspecies quorum-sensing signal autoinducer-2 (AI-2): characterization of phospho-(S)-4,5-dihydroxy-2,3-pentanedione isomerization by LsrG protein. *The Journal of biological chemistry* **286**, 18331-18343
22. Marques, J. C., Oh, I. K., Ly, D. C., Lamosa, P., Ventura, M. R., Miller, S. T., and Xavier, K. B. (2014) LsrF, a coenzyme A-dependent thiolase, catalyzes the terminal step in processing the quorum sensing signal autoinducer-2. *P Natl Acad Sci USA* **111**, 14235-14240
23. Li, J., Attila, C., Wang, L., Wood, T. K., Valdes, J. J., and Bentley, W. E. (2007) Quorum sensing in Escherichia coli is signaled by AI-2/LsrR: Effects on small RNA and Biofilm architecture. *Journal of bacteriology* **189**, 6011-6020
24. Hegde, M., Englert, D. L., Schrock, S., Cohn, W. B., Vogt, C., Wood, T. K., Manson, M. D., and Jayaraman, A. (2011) Chemotaxis to the quorum-sensing signal AI-2 requires the Tsr chemoreceptor and the periplasmic LsrB AI-2-binding protein. *Journal of bacteriology* **193**, 768-773
25. Wood, T. K., Gonzalez Barrios, A. F., Herzberg, M., and Lee, J. (2006) Motility influences biofilm architecture in Escherichia coli. *Appl Microbiol Biotechnol* **72**, 361-367
26. Bansal, T., Jesudhasan, P., Pillai, S., Wood, T. K., and Jayaraman, A. (2008) Temporal regulation of enterohemorrhagic Escherichia coli virulence mediated by autoinducer-2. *Appl Microbiol Biot* **78**, 811-819
27. Jani, S., Seely, A. L., Peabody, G. L., Jayaraman, A., and Manson, M. D. (2017) Chemotaxis to self-generated AI-2 promotes biofilm formation in Escherichia coli. *Microbiol-Sgm* **163**, 1778-1790
28. Laganenka, L., Colin, R., and Sourjik, V. (2016) Chemotaxis towards autoinducer 2 mediates autoaggregation in Escherichia coli. *Nat Commun* **7**, 12984
29. Thompson, J. A., Oliveira, R. A., Djukovic, A., Ubeda, C., and Xavier, K. B. (2015) Manipulation of the quorum sensing signal AI-2 affects the antibiotic-treated gut microbiota. *Cell reports* **10**, 1861-1871

30. Zhang, Y. J., Li, S., Gan, R. Y., Zhou, T., Xu, D. P., and Li, H. B. (2015) Impacts of gut bacteria on human health and diseases. *International journal of molecular sciences* **16**, 7493-7519
31. Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L. L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., Littmann, E., van den Brink, M. R. M., Jenq, R. R., Taur, Y., Sander, C., Cross, J. R., Toussaint, N. C., Xavier, J. B., and Pamer, E. G. (2015) Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517**, 205-U207
32. Lee, A. S. Y., and Song, K. P. (2005) LuxS/autoinducer-2 quorum sensing molecule regulates transcriptional virulence gene expression in Clostridium difficile. *Biochem Bioph Res Co* **335**, 659-666
33. Solbach, P., Woltemate, S., Chhatwal, P., Tacconelli, E., Buhl, M., Gerhard, M., Vehreschild, M., Jazmati, N., Rupp, J., Manns, M. P., Bachmann, O., and Suerbaum, S. (2018) BaiCD gene abundance is negatively correlated with Clostridium difficile infection. *J Crohns Colitis* **12**, S558-S558
34. Abrini, J., Naveau, H., and Nyns, E. J. (1994) Clostridium Autoethanogenum, Sp-Nov, an Anaerobic Bacterium That Produces Ethanol from Carbon-Monoxide. *Arch Microbiol* **161**, 345-351
35. van Sambeek, D. M., Tran, H., Fernando, S. C., Ciobanu, D. C., Miller, P. S., and Burkey, T. E. (2016) Alteration of the pig intestinal microbiome when vaccinated against or inoculated with porcine circovirus 2 using a multivariate analysis model. *J Anim Sci* **94**, 387-390
36. Pereira, C. S., McAuley, J. R., Taga, M. E., Xavier, K. B., and Miller, S. T. (2008) Sinorhizobium meliloti, a bacterium lacking the autoinducer-2 (AI-2) synthase, responds to AI-2 supplied by other bacteria. *Molecular microbiology* **70**, 1223-1235
37. Kavanaugh, J. S., Gakhar, L., and Horswill, A. R. (2011) The structure of LsrB from Yersinia pestis complexed with autoinducer-2. *Acta Crystallogr F* **67**, 1501-1505
38. Ames, G. F. L. (1986) Bacterial Periplasmic Transport-Systems - Structure, Mechanism, and Evolution. *Annu Rev Biochem* **55**, 397-425
39. Rivers, D. M., and Oresnik, I. J. (2015) The Sugar Kinase That Is Necessary for the Catabolism of Rhamnose in Rhizobium leguminosarum Directly Interacts with the ABC Transporter Necessary for Rhamnose Transport. *Journal of bacteriology* **197**, 3812-3821
40. Saurin, W., Hofnung, M., and Dassa, E. (1999) Getting in or out: Early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evol* **48**, 22-41
41. Wang, B., Dukarevich, M., Sun, E. I., Yen, M. R., and Saier, M. H. (2009) Membrane Porters of ATP-Binding Cassette Transport Systems Are Polyphyletic. *J Membrane Biol* **231**, 1-10
42. Zhu, J. G., and Pei, D. H. (2008) A LuxP-based fluorescent sensor for bacterial autoinducer II. *Acs Chem Biol* **3**, 110-119
43. Pereira, C. S., Santos, A. J., Bejerano-Sagie, M., Correia, P. B., Marques, J. C., and Xavier, K. B. (2012) Phosphoenolpyruvate phosphotransferase system regulates detection and processing of the quorum sensing signal autoinducer-2. *Molecular microbiology* **84**, 93-104
44. Ha, J. H., Hauk, P., Cho, K., Eo, Y., Ma, X., Stephens, K., Cha, S., Jeong, M., Suh, J. Y., Sintim, H. O., Bentley, W. E., and Ryu, K. S. (2018) Evidence of link between quorum sensing and sugar metabolism in Escherichia coli revealed via cocrystal structures of LsrK and HPr. *Science advances* **4**, eaar7063
45. Duncan, S. H., Aminov, R. I., Scott, K. P., Louis, P., Stanton, T. B., and Flint, H. J. (2006) Proposal of Roseburia faecis sp nov., Roseburia hominis sp nov and Roseburia inulinivorans sp nov., based on isolates from human faeces. *Int J Syst Evol Micr* **56**, 2437-2441
46. Alou, M. T., Ndongo, S., Fregere, L., Labas, N., Andrieu, C., Richez, M., Couderc, C., Baudoin, J. P., Abrahao, J., Brah, S., Diallo, A., Sokhna, C., Cassir, N., La Scola, B., Cadoret, F., and Raoult, D. (2018) Taxonogenomic description of four new Clostridium species isolated from human gut: 'Clostridium amazonitimonense', 'Clostridium merdae', 'Clostridium massilidielmoense' and 'Clostridium nigeriense'. *New microbes and new infections* **21**, 128-139

47. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501
48. Ascenso, O. S., Marques, J. C., Santos, A. R., Xavier, K. B., Ventura, M. R., and Maycock, C. D. (2011) An efficient synthesis of the precursor of AI-2, the signalling molecule for inter-species quorum sensing. *Bioorganic & medicinal chemistry* **19**, 1236-1241
49. Darwish, N. B., Kochkodan, V., and Hilal, N. (2015) Boron removal from water with fractionized Amberlite IRA743 resin. *Desalination* **370**, 1-6
50. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235-242
51. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* **68**, 352-367
52. Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., Bateman, A., and Eddy, S. R. (2015) HMMER web server: 2015 update. *Nucleic Acids Res* **43**, W30-W38
53. Katoh, K., and Standley, D. M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**, 772-780
54. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165
55. Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313
56. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457-462
57. Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., and Ostell, J. (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* **44**, 6614-6624

FOOTNOTES

The crystal structure of *C. saccharobutylicum* LsrB has been deposited in Protein Data Bank under PDB 6DSP with the description "LsrB from *Clostridium saccharobutylicum* in complex with AI-2".

The abbreviations used are: autoinducer-2, AI-2; (S)-4,5-dihydroxypentane-2,3-dione, DPD;, (2S,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran-borate, S-THMF-borate; (2R,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran, R-THMF; Kyoto Encyclopedia of Genes and Genomes, KEGG; Root mean square deviation, RMSD; Protein Data Bank, PDB; Isothermal titration calorymetry, ITC; Phosphorylated AI-2, P-AI-2; 3-hydroxy-2,4-pentadione-5-phosphate, P-HPD; 3,4,4-trihydroxy-2-pentanone-5-phosphate, P-TPO; dihydroxyacetone phosphate, DHAP; dithiothreitol, DTT; tobacco etch virus, TEV; nickel-nitrilotriacetic, Ni^{+}-NTA.

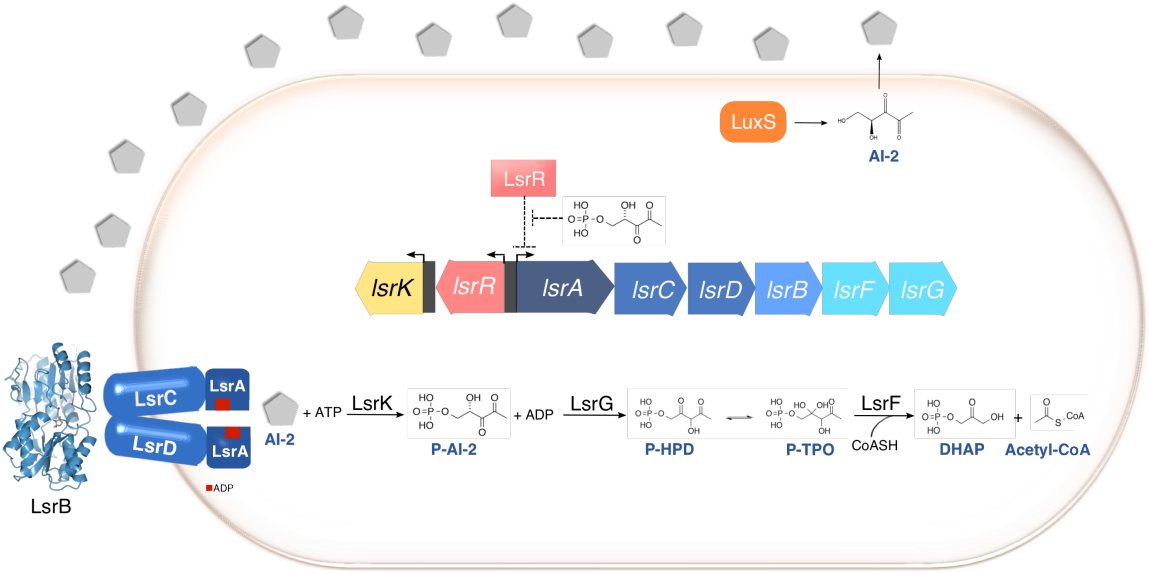

Figure 1. Schematic representation of the AI-2 internalization process. LuxS produces AI-2 as (4S)-4,5-dihydroxy-2,3-pentanedione. After synthesis, AI-2 is exported to the extracellular milieu and accumulates in proportion to bacterial density. Once AI-2 accumulates to a threshold concentration, it is internalized and phosphorylated by LsrK. Phosphorylated AI-2 (P-AI-2) can then bind the transcriptional regulator LsrR, blocking repression and leading to the expression of the genes in the *lsr* operon. Subsequent expression of the LsrB receptor and of the associated ABC transporter (*lsrACD*) promotes AI-2 internalization and consequent depletion of extracellular AI-2. In addition, phosphorylation by LsrK sequesters the signal inside the cell. In *E. coli*, P-AI-2 is further processed by LsrG which catalyzes the isomerization of P-AI-2 to 3-hydroxy-2,4-pentadione-5-phosphate (P-HPD), an isomer that exists in equilibrium with its hydrated form 3,4,4-trihydroxy-2-pentanone-5-phosphate (P-TPO). LsrF then catalyzes the transfer of an acetyl group from P-HPD to coenzyme A, creating dihydroxyacetone phosphate (DHAP) and acetyl-CoA (key metabolites used by the cell in metabolic pathways like glycolysis and the citric acid cycle).

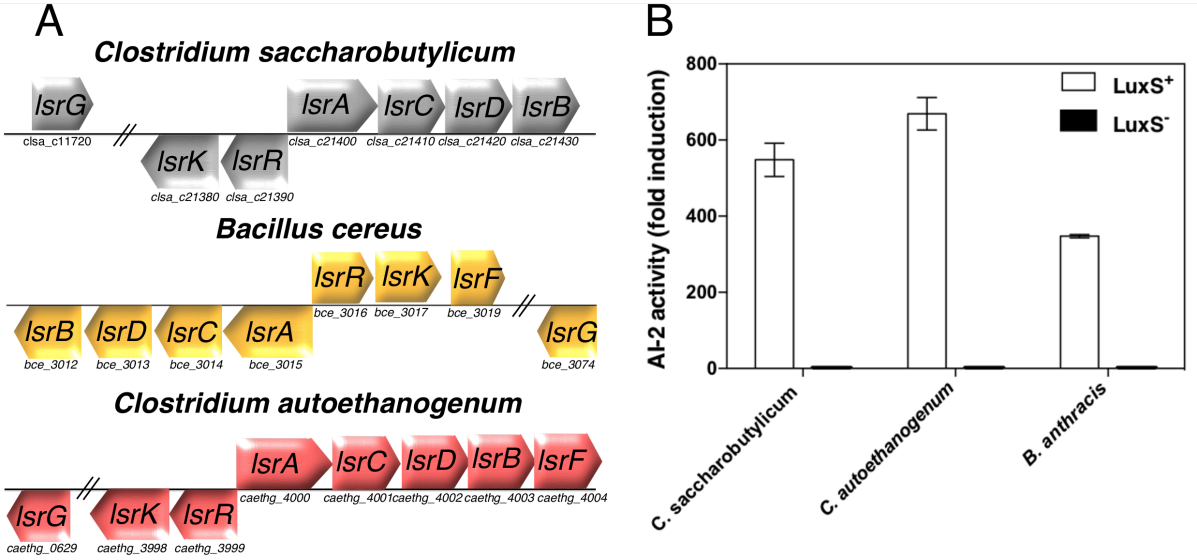

Figure 2. LsrB-like receptors from *C. saccharobutylicum* and *C. autoethanogenum* bind AI-2. (A) Comparison of the putative AI-2 transport and processing orthologs in *C. saccharobutylicum* (top) and *C. autoethanogenum* (bottom) with *B. cereus* (middle), which has a functional Lsr transporter system. (B) AI-2 binding was assessed by measurement of light production of *V. harveyi* MM32 after addition of ligand released from the pure putative LsrBs from *C. saccharobutylicum* and *C. autoethanogenum* expressed in a LuxS$^+$ *E. coli* BL21 strain (white bars). The LsrB receptor from *B. anthracis* was used as a positive control. As negative controls we tested the mentioned proteins expressed in a LuxS$^-$ mutant (black bars). Results are shown as fold induction relative to the light production induced by the growth medium.

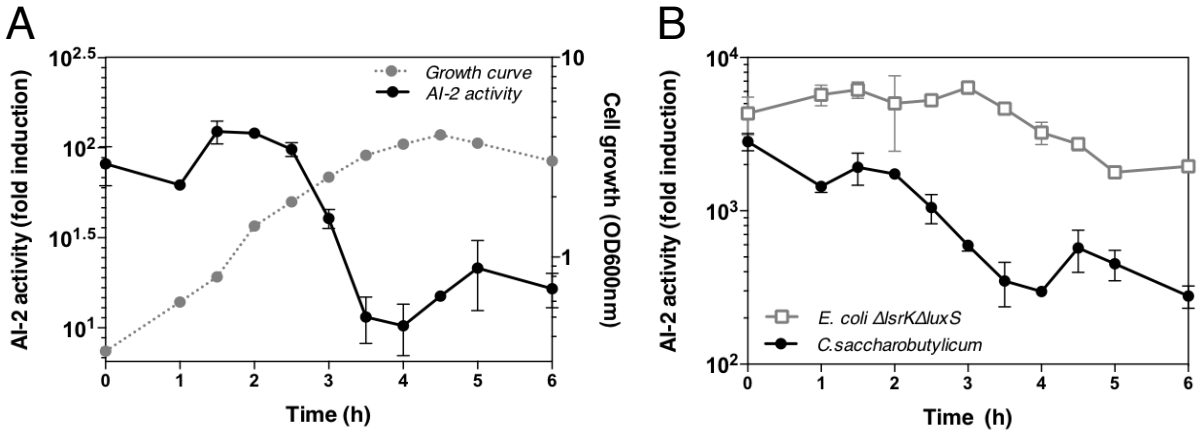

Figure 3. *C. saccharobutylicum* LsrB receptor internalizes AI-2. (A) AI-2 internalization profile of *C. saccharobutylicum* (black circles) and respective growth curve (grey circles). (B) AI-2 internalization profile of *C. saccharobutylicum* (black circles) and *E. coli* Δ*lsrK*Δ*luxS* (empty squares) after addition of 40 μM of AI-2 at time zero. Cell free supernatants were collected at the indicated time points and activity of extracellular AI-2 was measured by assessing light produced by *V. harveyi* MM32 in response to cell-free supernatants. AI-2 activity is reported as fold induction relative to light production induced by the growth medium. The internalization curves are representative of three independent experiments shown in Fig.S2. Error bars represent the standard deviation of three technical replicates. Given the different lag phases of *C. saccharobutylicum* growth it was not possible to join the biological replicates.

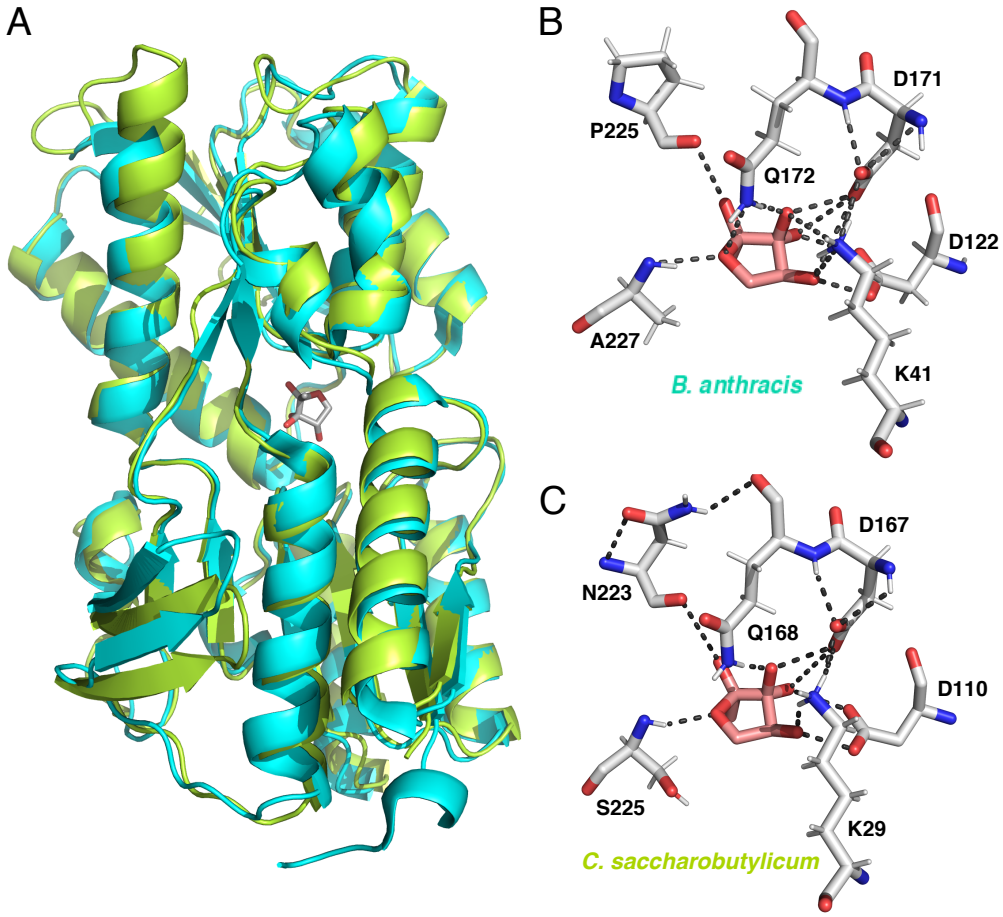

Figure 4. Crystal structure of the LsrB receptor from *C. saccharobutylicum* in complex with AI-2. (A) Ribbon diagram of *C. saccharobutylicum* LsrB (green) superimposed to the LsrB of *B. anthracis* (PDB ID: 4PZ0; blue). The structures are very similar with a root mean square deviation of 0.78 Å, calculated by PyMOL. Bound AI-2 is shown as sticks. (B) Hydrogen bonds between AI-2 and LsrB from *B. anthracis* and (C) LsrB from *C. saccharobutylicum*.

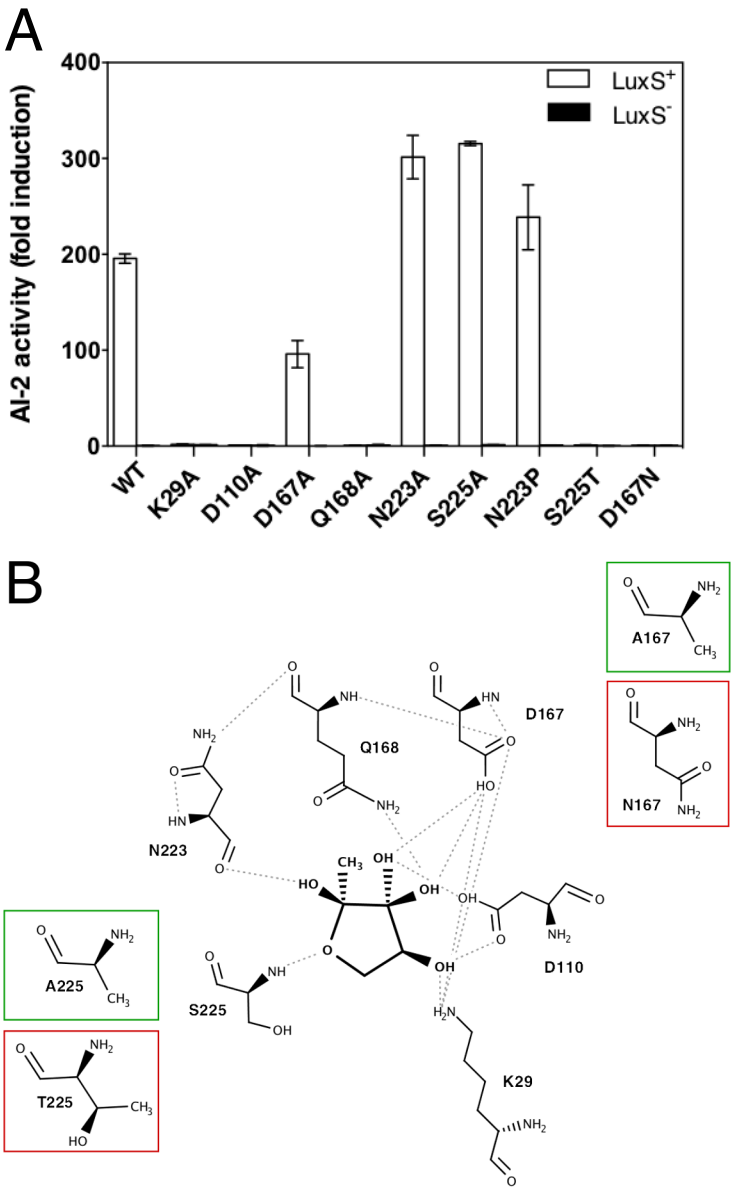

Figure 5. Determination of the amino acid residues essential for binding of AI-2 by *C. saccharobutylicum* LsrB. (A) Binding of AI-2 to wild-type and binding site mutants of CsLsrB. Binding assessed by measurement of light production of *V. harveyi* MM32 after addition of ligands released from purified proteins by thermal denaturation. Proteins were expressed either in $LuxS^+$ (3) or $LuxS^-$ (3) *E. coli* BL21 strains. Results are shown as fold induction relative to the light production induced by the growth medium. (B) Two-dimensional representation of the binding site of CsLsrB. Substitutions of aspartate 167 or serine 225 (which interfere with AI-2 binding) are boxed in red while substitutions that allow AI-2 binding are boxed in green.

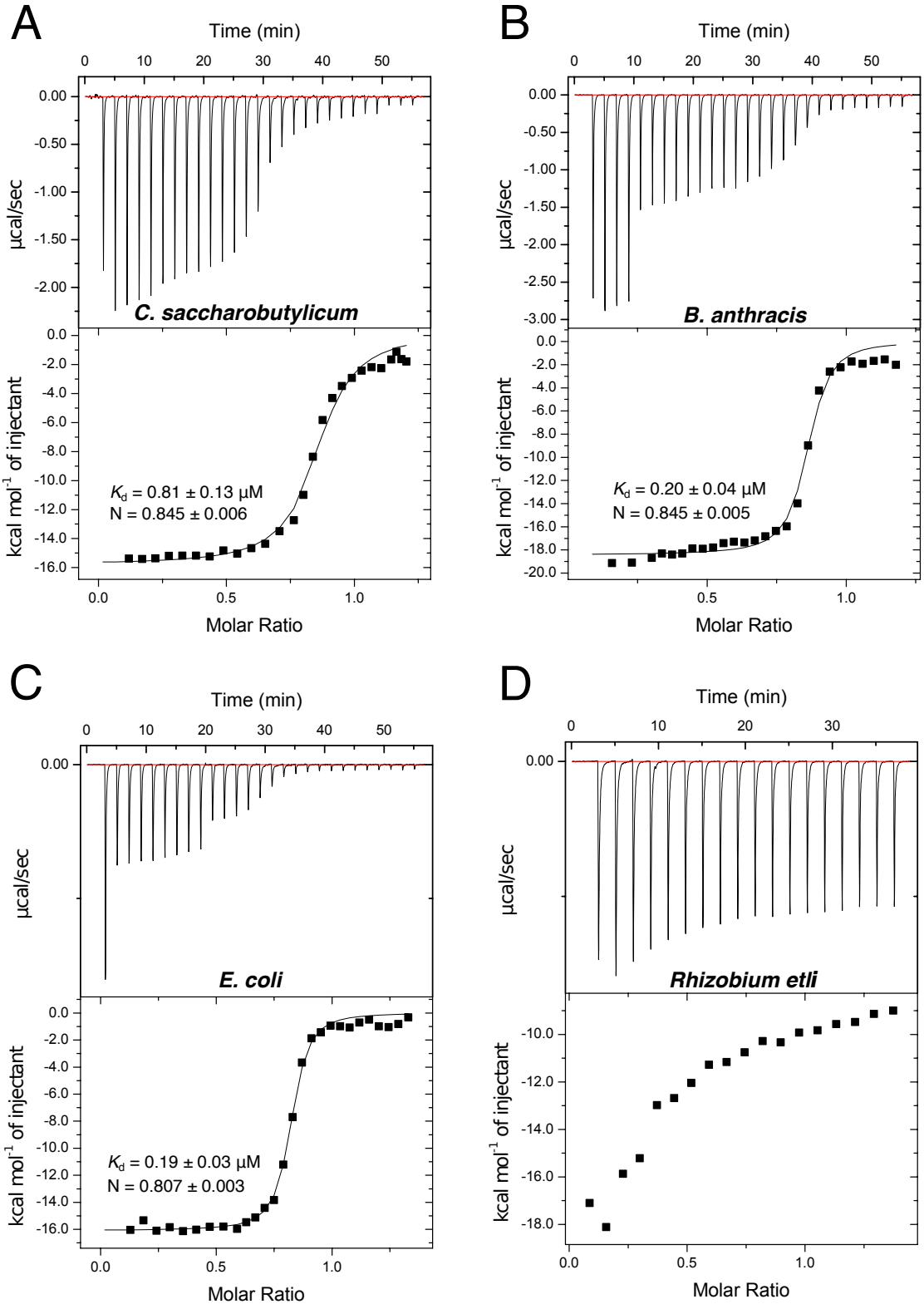

Figure 6. LsrB receptors bind AI-2 with submicromolar affinity. Binding curves of AI-2 to LsrB proteins from (A) *C. saccharobutylicum*, (B) *B. anthracis*, (C) *E. coli*, and (D) from an LsrB ortholog from *Rhizobium etli* as negative control. Upper graphs show the raw ITC data and lower graphs show the integrated heat of reaction. Dissociation constants were obtained by fitting a one-site binding model to the displayed curves, which are representative of three independent experiments.

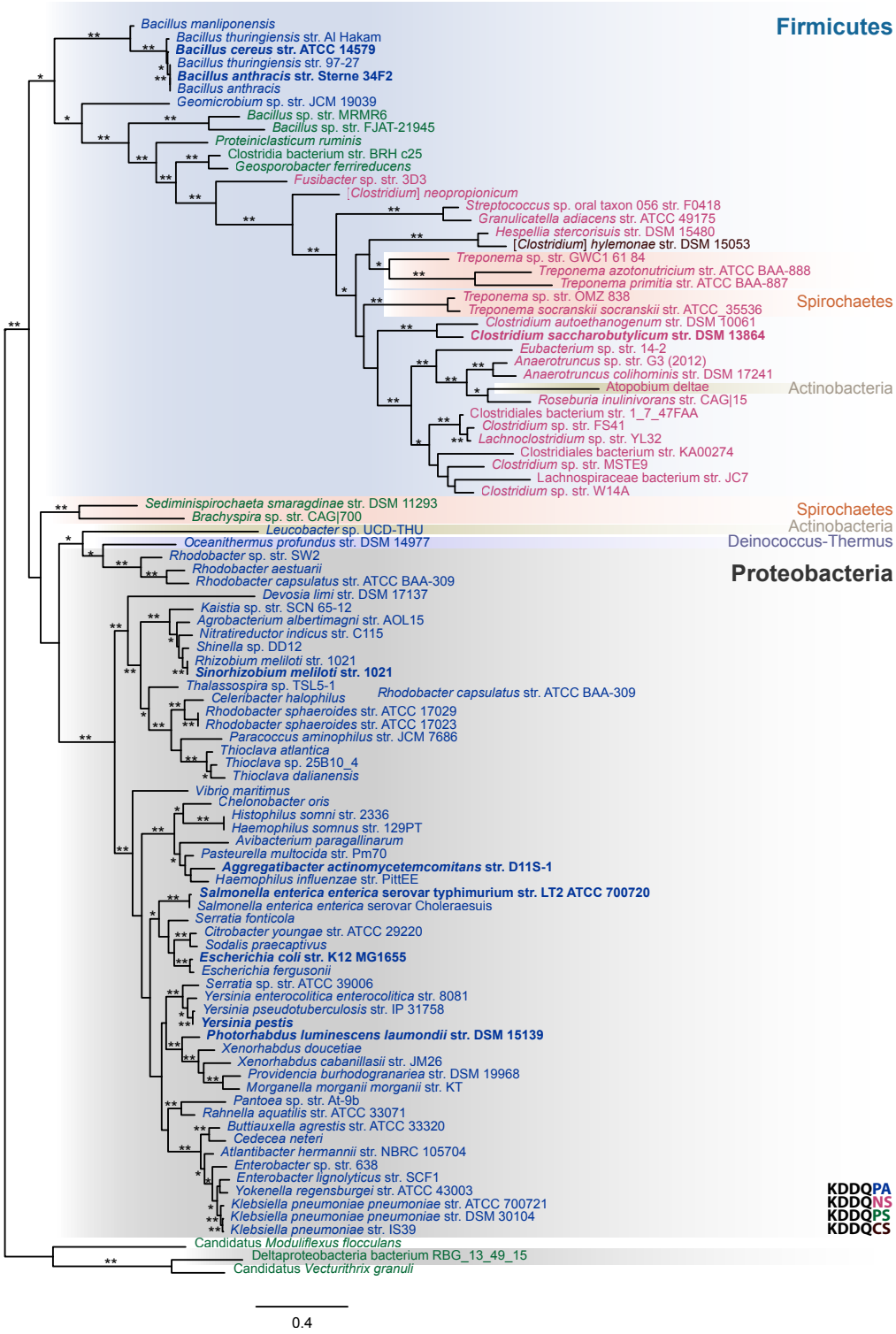

Figure 7. Phylogenetic history of LsrB protein sequences. LsrB proteins with canonical residues (KDDQPA) are depicted in blue, noncanonical residues with one substitution (KDDQPS) in green and noncanonical residues with two substitutions (KDDQNS) in red. In bold is our focal taxon *C. saccharobutylicum* and the organisms for which there are experimental studies supporting the functionality of the LsrB in at least one strain of these species [2, 6, 24-26, 49, 50]. Taxonomic classifications are shown on the right and shades of the same color were used to group bacterial species belonging to the phylum following NCBI. Analyses were done with maximum-likelihood (LG+I+G model of protein evolution) and the tree was rooted following results from supplementary Fig. S4. Numbers on the nodes are bootstrap support estimated in RaxML using the autoFC option.

Supplementary information for

Identification of novel autoinducer-2 receptors in Clostridia reveals plasticity in the binding site of the LsrB receptor family.

Inês M. Torcato, Meghann R. Kasal, Patrícia H. Brito, Stephen T. Miller, Karina B. Xavier

Corresponding authors: Stephen Miller and Karina Xavier
Emails: smiller1@swarthmore.edu and kxavier@igc.gulbenkian.pt

This PDF file includes:

Figs. S1-S4;
Tables S1-S3, S5 and S7-S8;
Experimental procedures for Fig.S1;

Other supplementary materials for this manuscript include the following:

Tables S4 and S6 (excel file)

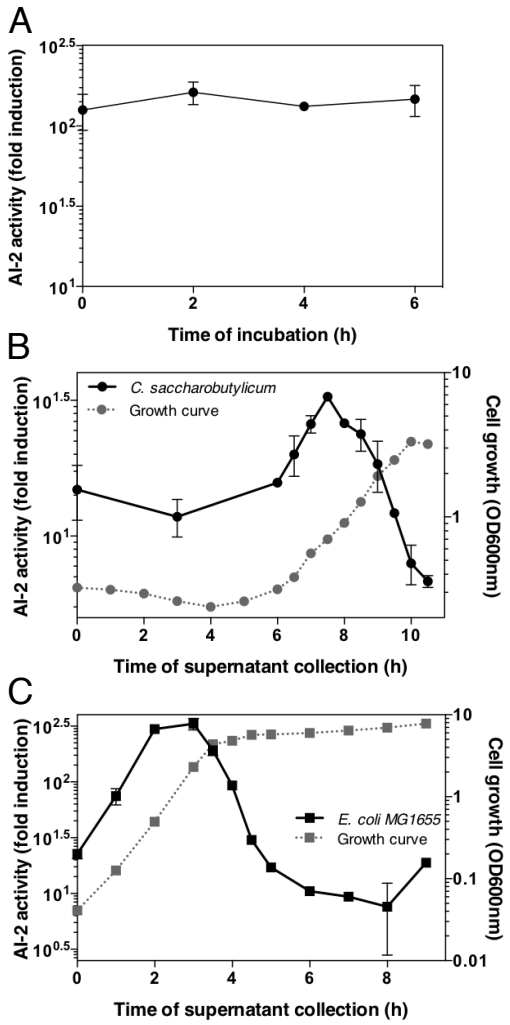

Fig. S1. AI-2 activity measured by *Vibrio harveyi* MM32. (A) activity of 40 μM AI-2 when incubated with supernatants of *C. saccharobutylicum* at 9h30 of growth. (B) AI-2 internalization profile in *C. saccharobutylicum* (black circles) and respective growth curve (grey circles). (C) AI-2 internalization profile in *E. coli* (black squares) and respective growth curve (grey squares). AI-2 activity reported as fold induction relative to light production induced by the growth medium. Error bars represent the standard deviation of three technical replicates.

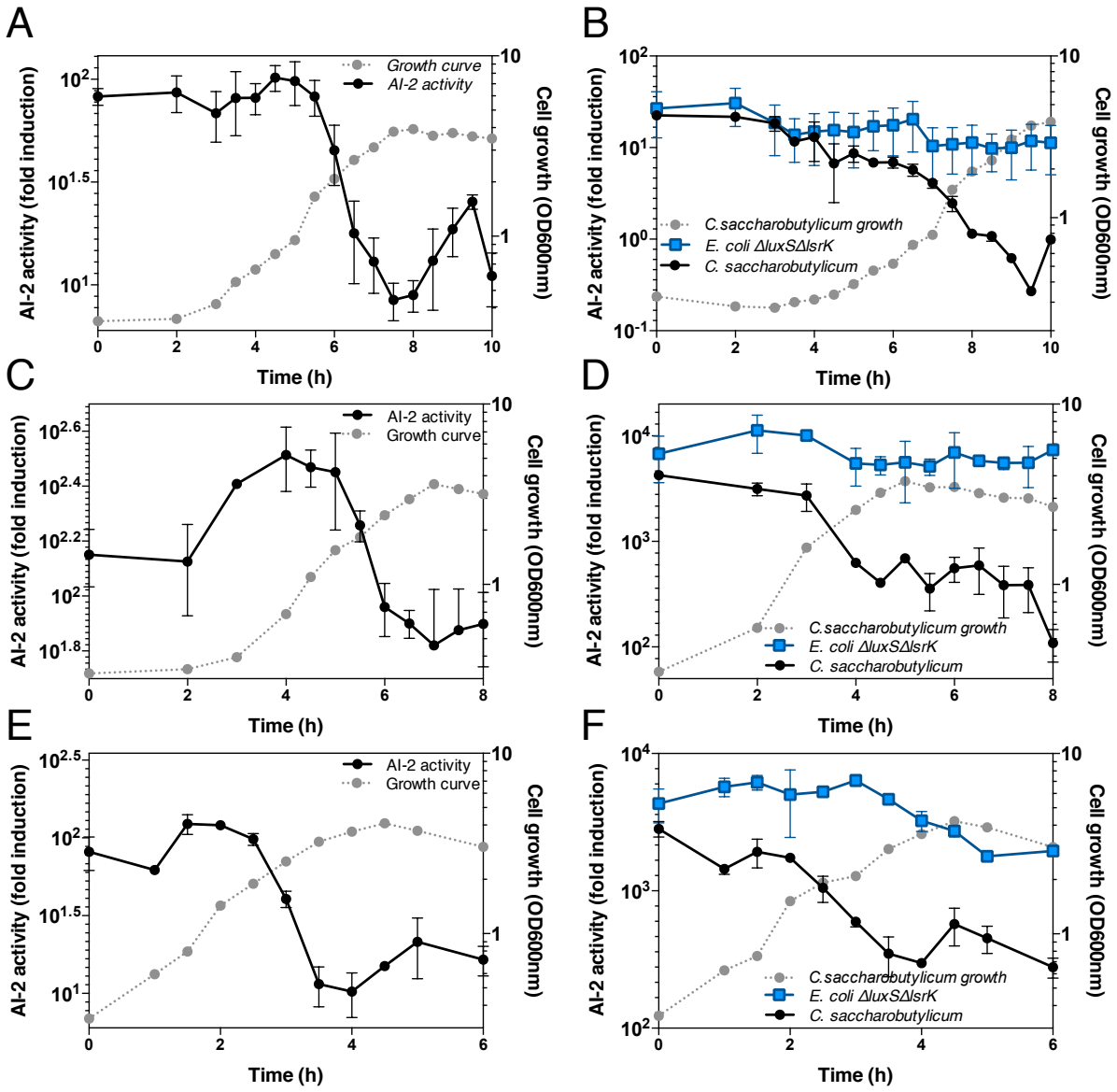

Fig. S2. AI-2 internalization by *C. saccharobutylicum*. (A), (C) and (E) internalization profile of *C. saccharobutylicum* (black circles) and respective growth curve (grey circles). (B), (D), (F) AI-2 internalization profile of *C. saccharobutylicum* (black circles) and *E. coli* Δ*lsrk*Δ*luxS* (blue squares) after addition of 40 μM AI-2 at time zero; *C. saccharobutylicum* respective growth curve is shown as grey circles. Cell free supernatants were collected at the indicated time points and activity of extracellular AI-2 was measured by assessing light produced by *V. harveyi* MM32 in response to cell-free supernatants. AI-2 activity is reported as fold induction relative to light production induced by the growth medium. The curves represent three independent experiments. Error bars represent the standard deviation of three technical replicates.

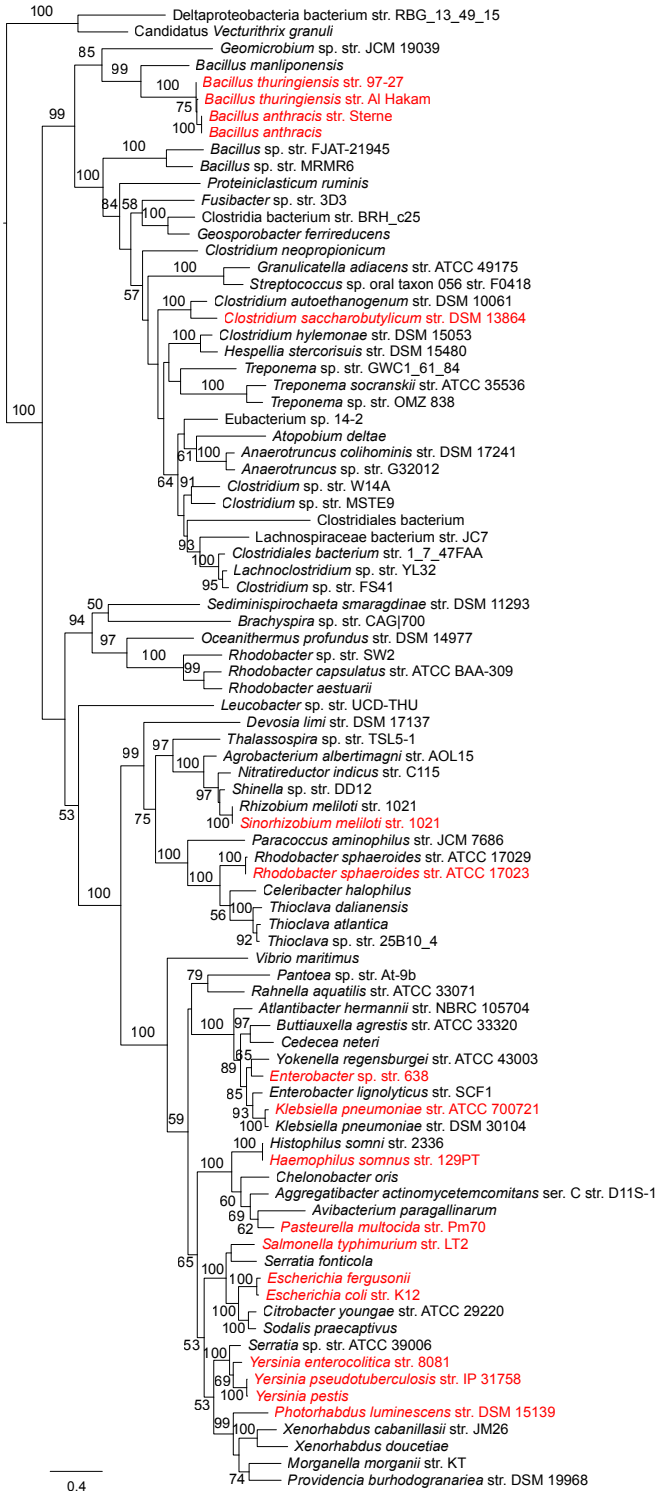

Fig. S3. Phylogenetic history of LsrA protein sequences. Analyses were done with maximum-likelihood (LG+I+G model of protein evolution) and the tree was rooted following previous results with a larger dataset that also includes the ATP-biding protein from the Rhamnose and the Ribose ABC-type transporters (data not shown). Numbers on the nodes are bootstrap support estimated in RaxML using the autoFC option. Only bootstrap support larger than 50% is shown. Scale bar indicates expected number of amino acids substitutions per site. In red we denote our focal taxon *C. saccharobutylicum*, taxa with LsrB sequences experimentally determined to be functional AI-2 receptors, and taxa from group I of [1].

Fig. S4. Phylogenetic history of LsrB protein sequences and its closely related homologous sequences. Analyses were done with maximum-likelihood (WAG+I+G model of protein evolution) in RaxML, and the tree was

rooted by midpoint rooting. Numbers on the nodes are bootstrap support estimated with the autoFC option in RaxML. Only bootstrap support larger than 50% is shown. Scale bar indicates expected number of amino acids substitutions per site. In red we denote our focal taxon *C. saccharobutylicum*, taxa with LsrB sequences experimentally determined to be functional AI-2 receptors, and taxa from group I of [1]. In green are taxa represented with Rhamnose-binding proteins.

Experimental procedures

Incubation of AI-2 in *C. saccharobutylicum* supernatants. *C. saccharobutylicum* culture at 9h30 of growth was collected and filtered using a 0.2 µm filter (Acrodisc, Pall Life Sciences) originating cell-free supernatant that was frozen at -20ºC, overnight. The supernatant was supplemented with synthetic AI-2 to a final concentration of 40 µM and incubated for 2h, 4h or 6h at 37ºC without agitation in a transparent, flat bottom 96 well plate (Costar 3596). After incubation, supernatants were frozen at -20ºC, overnight. AI-2 activity was measured using *V. harveyi* MM32 bioluminescence reporter assay as described for the internalization assays. Data presented corresponds to 5h of incubation with *V. harveyi* MM32 at 30ºC. Standard deviation was calculated from three technical replicates and propagation of uncertainty was used to calculate the standard deviation after normalization of the light produced by AI-2 by the light produced by the medium.

AI-2 internalization in *C. saccharobutylicum* and *E. coli.* Determination of AI-2 activity in *C. saccharobutylicum* was performed as described in the experimental procedures of the main text in the section of AI-2 internalization assays. Determination of AI-2 activity in *E. coli* K12 MG1655 cell cultures during growth was performed as previously described [1]. Briefly, *E. coli* was revived from the -80ºC frozen stock directly to 5 mL of broth (modified PY+X with arabinose and 100 mM MOPS buffer, pH 7.0; Table S1) and grown at 37ºC with 200 rpm of agitation until saturation. A dilution of approximately 100 times to 12.5 mL of fresh medium in a 125 mL erlenmeyer was performed and culture suspensions were collected at the indicated time points for optical density measurement at 600 nm in a UV-visible spectrophotometer (Helios delta, ThermoSpectronic, USA) and for detection of AI-2. For AI-2 activity assessment, the culture suspensions were filtrated originating cell-free fluids (Multiscreen filter plates, Millipore). The cell free supernatants were frozen at -20ºC overnight and AI-2 activity in these fluids was accessed in triplicate using *V. harveyi* MM32 bioluminescence reporter assay, as described above. Data presented corresponds to 6h of incubation with *V. harveyi* MM32, at 30ºC.

Table S1. Composition of the liquid growth medium employed. Modified PY+X with arabinose and 100 mM MOPS buffer, pH 7.0.

Component	Quantity
Trypticase peptone…….....	5 g/L
Peptone from meat (pepsin digested)…….....	5 g/L
Yeast extract……...	10 g/L
Arabinose……	5 g/L
L-Cysteine-HCl x H_2O……...	0.50 g/L
0.5 M MOPS buffer pH 7.0[a]……...	200 mL/L
Salt solution……	20 mL/L
$CaCl_2$ x 2 H_2O…….	0.25 g/L
$MgSO_4$ x 7 H_2O……...	0.50 g/L
K_2HPO_4……...	1 g/L
KH_2PO_4……...	1 g/L
$NaHCO_3$……..	10 g/L
NaCl…….....	2 g/L

[a]MOPS buffer was filter sterilized and added to the autoclaved medium.

Agar plates (Modified PY+X with galactose)
Composition of the solid medium was the same as in table S1 with the addition of 15 g/L of agar and of 5 g/L of galactose instead of arabinose. No MOPS buffer was added to this solid medium.

Table S2. Sequence identity of the Lsr orthologs and LuxS present in *C. saccharobutylicum* and *C. autoethanogenum* when compared with the ones from previously studied organisms.

Organism	KEGG identifier	Protein	% identity *C.saccharobutylicum*	KEGG identifier	% identity *C.autoethanogenum*	KEGG identifier
Bacillus cereus (bca)	BCE_3012	LsrB	39.2	CLSA_c21430	37.9	CAETHG_4003
	BCE_3015	LsrA	46.3	CLSA_c21400	47.1	CAETHG_4000
	BCE_3014	LsrC	43.9	CLSA_c21410	43.1	CAETHG_4001
	BCE_3013	LsrD	40.9	CLSA_c21420	46.7	CAETHG_4002
	BCE_3016	LsrR	39.7	CLSA_c21390	40.7	CAETHG_3999
	BCE_3017	LsrK	51.7	CLSA_c21380	52.9	CAETHG_3998
	BCE_3019	LsrF	-	-	40.2	CAETHG_4004
	BCE_3074	LsrG	33.3	CLSA_c11720	31.9	CAETHG_0629
	BCE_4946	LuxS	51.8	CLSA_c30740	30.2	CAETHG_0412
Bacillus anthracis Sterne (bat)	BAS2763	LsrB	39.2	CLSA_c21430	37.9	CAETHG_4003
	BAS2766	LsrA	46.0	CLSA_c21400	46.3	CAETHG_4000
	BAS2765	LsrC	42.4	CLSA_c21410	44.0	CAETHG_4001
	BAS2764	LsrD	43.4	CLSA_c21420	46.1	CAETHG_4002
	BAS2767	LsrR	41.1	CLSA_c21390	41.0	CAETHG_3999
	BAS2771	LsrF	-	-	40.5	CAETHG_4004
	BAS2825	LsrG	32.0	CLSA_c11720	29.0	CAETHG_0629
	BAS4687	LuxS	51.8	CLSA_c30740	30.2	CAETHG_0412
Eschericia coli K-12 MG1655 (eco)	b1516	LsrB	43.6	CLSA_c21430	40.2	CAETHG_4003
	b2687	LuxS	42.8	CLSA_c30740	32.5	CAETHG_0412
Salmonella serovar Typhimurium LT2 (stm)	STM4077	LsrB	40.1	CLSA_c21430	39.8	CAETHG_4003
	STM2817	LuxS	41.4	CLSA_c30740	33.1	CAETHG_0412

Table S3. Data collection and refinement statistics.

Wavelength	1181
Resolution range	50.59 - 1.368 (1.417 - 1.368)
Space group	P 43 2 2
Unit cell	90.78 90.78 182.81 90 90 90
Total reflections	1150236 (54053)
Unique reflections	160866 (7881)
Multiplicity	7.1 (6.9)
Completeness (%)	100.0 (100.0)
Mean I/sigma(I)	14.4 (1.6)
R-merge	0.056 (0.819)
R-meas	0.065 (0.896)
R-pim	0.033 (0.509)
CC1/2	0.999 (0.731)
Reflections used in refinement	160771 (15892)
Reflections used for R-free	2015 (202)
R-work	0.1599 (0.2199)
R-free	0.1765 (0.2271)
CC(work)	0.879 (0.764)
CC(free)	0.889 (0.750)
Number of non-hydrogen atoms	5327
macromolecules	4948
ligands	20
solvent	359
Protein residues	652
RMS(bonds)	10
RMS(angles)	1.07
Ramachandran favored (%)	96.76
Ramachandran allowed (%)	2.93
Ramachandran outliers (%)	0.31
Rotamer outliers (%)	0.19
Average B-factor	19.77
macromolecules	19.50
ligands	14.42
solvent	23.79

Statistics for the highest-resolution shell are shown in parentheses.

Table S4. See attached excel file.

Table S5. Primers used in this study.

Construct	5' (sense) sequence	3' (antisense) sequence
C. saccharobutylicum WT LsrB	CAGGGCTCCTTCACCTCAAGC	CGCCCACCCTTTTAGAAATC
C. saccharobutylicum LsrB K48A	GGAAATGTTACTGTAACATTTATTC CCGCACTTACAGGTAATGCATTTTT TGAATC	GATTCAAAAAATGCATTACCTGTAAGT GCGGGAATAAATGTTACAGTAACATT TCC
C. saccharobutylicum LsrB D129A	GTTACAGTTACTACATGGGCTTCTG ACGTAGATCCATCA	TGATGGATCTACGTCAGAAGCCCATGT AGTAACTGTAAC
C. saccharobutylicum LsrB D186A	TTATTCAAATGCAACAGTTACTGCT CAAAATTCATGGCAGGTAGAAG	CTTCTACCTGCCATGAATTTTGAGCAG TAACTGTTGCATTTGAATAA
C. saccharobutylicum LsrB Q187A	TTCAAATGCAACAGTTACTGATGC AAATTCATGGCAGGTAGAAGGG	CCCTTCTACCTGCCATGAATTTGCATC AGTAACTGTTGCATTTGAA
C. saccharobutylicum LsrB N242A	CACATTCTGATATTGACTTAATTAT ATGTGCTGATTCTACAGCACTTCCT GGA	TCCAGGAAGTGCTGTAGAATCAGCAC ATATAATTAAGTCAATATCAGAATGTG
C. saccharobutylicum LsrB S244A	TCTGATATTGACTTAATTATATGTA ATGATGCTACAGCACTTCCTGGAC	GTCCAGGAAGTGCTGTAGCATCATTAC ATATAATTAAGTCAATATCAGA
C. saccharobutylicum LsrB D186N	CATTATTCAAATGCAACAGTTACTA ATCAAAATTCATGGCAGGTAGAAG	CTTCTACCTGCCATGAATTTTGATTAG TAACTGTTGCATTTGAATAATG
C. saccharobutylicum LsrB S244T	GATATTGACTTAATTATATGTAATG ATACTACAGCACTTCCTGGACAAG	CTTGTCCAGGAAGTGCTGTAGTATCAT TACATATAATTAAGTCAATATC
C. saccharobutylicum LsrB N242P	TGTCTGCACATTCTGATATTGACTT AATTATATGTCCTGATTCTACAGCA CTTCC	GGAAGTGCTGTAGAATCAGGACATAT AATTAAGTCAATATCAGAATGTGCAG ACA

Table S6. See attached excel file.

Table S7. LsrB and Rhamnose-biding protein sequences added to the UniProt dataset

Species/strain	Kegg Acc	KEGG Orthology
Yersinia_pseudotuberculosis IP 31758	YpsIP31758_3524	K10555; AI-2 transport system substrate-binding protein
Yersinia enterocolitica subsp. enterocolitica 8081	YE0525	K10555; AI-2 transport system substrate-binding protein
Sinorhizobium meliloti 1021	SM_b21016	K10555; AI-2 transport system substrate-binding protein
Salmonella enterica subsp. enterica serovar Choleraesuis	SCH_3966	K10555; AI-2 transport system substrate-binding protein
Rhodobacter sphaeroides ATCC 17029	Rsph17029_3146	K10555; AI-2 transport system substrate-binding protein
Haemophilus somnus 129PT	HS_0054	K10555; AI-2 transport system substrate-binding protein
Haemophilus influenzae PittEE	CGSHiEE_08220	K10555; AI-2 transport system substrate-binding protein
Escherichia fergusonii	EFER_1561	K10555; AI-2 transport system substrate-binding protein
Bacillus thuringiensis Al Hakam	BALH_2660	K10555; AI-2 transport system substrate-binding protein
Bacillus thuringiensis 97-27	BT9727_2713	K10555; AI-2 transport system substrate-binding protein
Bacillus anthracis strain Sterne 34F2	BAS2763	K10555; AI-2 transport system substrate-binding protein
Rhizobium leguminosarum bv. viciae 3841	pRL110413	K10559; rhamnose transport system substrate-binding protein
Sinorhizobium medicae	Smed_0225	K10559; rhamnose transport system substrate-binding protein

Table S8. ATP-biding protein sequences added to the UniProt dataset.

Species/strain	Kegg Acc	KEGG Orthology
Yersinia_pseudotuberculosis IP 31758	YpsIP31758_3521	K10558; AI-2 transport system ATP-binding protein
Yersinia enterocolitica subsp. enterocolitica 8081	YE0528	K10558; AI-2 transport system ATP-binding protein
Sinorhizobium meliloti 1021	SM_b21019	K10558; AI-2 transport system ATP-binding protein
Salmonella enterica subsp. enterica serovar Choleraesuis	-	-
Rhodobacter sphaeroides ATCC 17029	Rsph17029_3149	K10558; AI-2 transport system ATP-binding protein
Haemophilus somnus 129PT	HS_0051	K10558; AI-2 transport system ATP-binding protein
Haemophilus influenzae PittEE	-	-
Escherichia fergusonii	EFER_1564	K10558; AI-2 transport system ATP-binding protein
Bacillus thuringiensis Al Hakam	BALH_2664	K10558; AI-2 transport system ATP-binding protein
Bacillus thuringiensis 97-27	BT9727_2716	K10558; AI-2 transport system ATP-binding protein
Bacillus anthracis strain Sterne 34F2	BAS2766	K10558; AI-2 transport system ATP-binding protein
Rhizobium leguminosarum bv. viciae 3841	pRL110412	K10562; Rhamnose transport system ATP-binding protein
Sinorhizobium medicae	Smed_0226	K10562; Rhamnose transport system ATP-binding protein

- means the gene is absent or not annotated

References:

1. Pereira CS, de Regt AK, Brito PH, Miller ST, & Xavier KB (2009) Identification of functional LsrB-like autoinducer-2 receptors. *Journal of bacteriology* 191(22):6975-6987.