Interaction of the TonB dependent transporter HasR with its cognate TonB-like protein HasB in a membrane environment

Kamolrat Somboon[1], Oliver Melling[1], Maylis Lejeune[2], Glaucia M.S. Pinheiro[2], Annick Paquelin[3], Benjamin Bardiaux[2], Michael Nilges[2], Philippe Delepelaire[3], Syma Khalid[1*], Nadia Izadi-Pruneyre[2*]

[1]School of Chemistry, University of Southampton, Southampton S017 1BJ, United Kingdom. [2]Structural Bioinformatics Unit, Department of Structural Biology and Chemistry, CNRS UMR3528, CNRS USR3756, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris, France. [3]Institut de Biologie Physico-Chimique, UMR 7099, CNRS Université Paris Diderot, 13 rue Pierre et Marie Curie, 75005, Paris, France.

*: corresponding author

Abstract

Energized nutrient import in bacteria needs the interaction between a TonB-dependent transporter (TBDT) and a TonB protein. The mechanism of energy and signal transfer between these two proteins is not well understood. They belong to two membranes separated by the periplasmic space and possess each a disordered and flexible region. Therefore, the membranes, their distance and geometrical constraints together with the protein dynamics are important factors for deciphering this trans-envelope system. Here we report the first example of the interaction of a TBDT with a TonB protein in the presence of both membranes. By combining molecular dynamics simulations in a membrane model, *in vitro* and *in vivo* phenotypic experiments we obtained the comprehensive network of interaction between HasR, a heme/hemophore receptor and its dedicated TonB protein, HasB.

Introduction

Gram-negative bacteria are delimited by an envelope composed of two membranes (Outer membrane and inner membrane, OM and IM) separated by the periplasmic space. Multiple macromolecular machines embedded in this envelope ensure and regulate bacterial exchanges with the environment. In contrast to small molecules that are internalised through porins by passive or facilitated diffusion, nutrients that are either larger than 600 Da or present at very low concentration, are transported *via* OM specific transporters by an energised process. As both the OM and the periplasm lack an energy source, these systems use the so-called TonB complex to couple substrate entry to the proton-motive force (pmf or the proton gradient across the IM). The TonB complex and specific TonB-dependent transporters (TBDTs) allow many bacteria to acquire essential nutrients, such as metal sources, vitamin B_{12}, some specific carbohydrates, etc[1,2]. The TBDT structural organization, conserved in a

wide variety of nutrient transporters, consists of a 22-stranded β-barrel anchored in the OM, with 11 extracellular loops that form the substrate binding site, and 11 periplasmic loops. The region upstream from this domain folds inside the barrel, forming the so-called "plug".[3]

The TonB complex comprises three IM proteins, TonB, ExbB and ExbD. TonB is the only protein of the complex interacting directly with the TBDT and thus transferring the energy, whereas ExbB and ExbD forming a proton-channel, are considered as the engine of the system [1]. The structure of the whole TonB complex and the full-length TonB protein are still unknown. TonB is composed of a N-terminus transmembrane helix inserted into the IM, an unstructured proline-riche domain in the periplasm followed by a C-terminal and globular domain that is the only identified domain of interaction with TBDT [1]. Only the structure of this globular periplasmic domain of TonB from *Escherichia coli*, either alone[4] or in complex with TBDT are known [5-6]. In the complex structures, TonB interacts with the TonB-box, a stretch of 7-10 conserved residues, exposed in the periplasm and located at the N-terminus of TBDT. It is assumed that upon this interaction and in the presence of energy, by either a rotation or a linear movement of TonB, some conformational changes of the TBDT occur, eventually opening a channel for the entry of the substrate.

The heme acquisition system (Has) is developed by commensal and pathogenic bacteria to acquire heme that is their major iron source in mammals [7-8]. The Has system was first identified in *Serratia marcescens*, an opportunistic pathogen. This system studied here as prototype of Has system and reconstituted in *E. coli*, is composed of the TBDT HasR functioning in synergy with a hemophore (HasA) or a heme carrier protein. Extracellular heme, either free or bound to host hemoproteins (like hemoglobin), is harvested by HasA and delivered to HasR to be internalized. The complex formed by HasA-HasR is then dissociated and the hemophore is recycled into the external medium. The two last steps of heme internalization and the recycling of HasA need energy. The energy is provided by HasB, a TonB paralog dedicated to the Has system [7-8].

HasR has the same barrel and plug structural organization as other TBDTs [9]. It belongs to a class of TBDT coupled to a signalling activity. In these TBDTs, the TonB-box is not at the extreme N-terminus, but instead it is connected to a periplasmic globular domain called the signalling domain or $HasR_{SD}$ in HasR [2,11]. $HasR_{SD}$ sends the signal of the presence of extracellular sources of heme (heme and HasA) to an anti-sigma/ECFσ factor pair which activates the expression of the *has* operon encoding the genes of the system. The expression of this operon is also regulated by Fur (Ferric uptake regulator), a transcriptional repressor of many iron acquisition systems [7].

$HasR_{SD}$ and the TonB-box were not resolved in the X-ray structure of HasR[9], most likely due to their dynamics. In a previous work, we combined the X-ray structure of HasR, the NMR structure of $HasR_{SD}$ with electron microscopy density maps and SAXS data, to obtain a

structural model of the full-length HasR in its free (apo-HasR) and HasA-heme loaded forms (holo-HasR). This model showed for the first time the structure of a full-length TBDT where the signalling domain was projected far from the β-barrel (at 80 and 70 Å in the apo- and holo-form, respectively)[11]. A similar location of the signalling domain with respect to the β-barrel was recently shown in the X-ray structure of FoxA, a siderophore TBDT from *Pseudomonas aeruginosa* [12].

HasR and HasB interact through their periplasmic exposed regions. In a previous work, we solved the structure of the C-terminal and periplasmic domain of HasB ($HasB_{CTD}$) [13]. This structure revealed differences with the known structure of *E.coli* TonB that could explain the specificity of HasB for HasR. We also studied the interaction of $HasB_{CTD}$ with a peptide corresponding to the TonB-box of HasR and showed that an ionic interaction between HasR (D100) and HasB (R169) is crucial for the stability of this interaction [13]. Furthermore, by using isothermal titration microcalorimetry (ITC), we showed that heme and HasA, although not required for the HasR-$HasB_{CTD}$ interaction *in vitro*, modulate this interaction [14]. This implies that structural changes associated with substrate binding on the extracellular side of HasR are transmitted to the periplasmic side of HasR containing its signalling domain and the region of interaction with HasB.

HasR and HasB are located in two different membranes, the OM and the IM that are separated by the periplasmic space. Therefore, to characterise the HasR-HasB interaction, it is critical to consider the constraints of the membranes and the distance between them. In this work, in order to develop a more thorough view of the HasR-HasB interaction, we have combined our structural and interaction data to model the protein complex in the context of *in silico* models of the inner and outer membranes. Molecular dynamics simulations of the HasR-HasB complex, revealed a previously unidentified network of interaction and the key role of some periplasmic loops in the complex formation and stability. We further validated the network of interactions by mutation, phenotypic and *in vitro* assays. This work represents the first example which shows the molecular details of the interaction of a TBDT with a TonB like protein with inclusion of both membranes.

Results

Molecular Dynamics of HasR-HasB complex

An initial model of HasR and HasB each in a membrane model was constructed based on their structural organization and guided by our previous experimental findings, as described in the methods. In this HasR model the linker was almost fully extended at a distance of 11.5 nm from the barrel. The $HasB_{CTD}$ is located ~ 6.8 nm from the HasR barrel (center of mass measurements). The HasB transmembrane domain is a single helix which is located in the model inner membrane. The two membranes are separated by ~ 12 nm, this is the shortest

distance (the distance from the center of the OM to the center of the IM is ~ 16nm), this was deemed to be a good compromise between the *in vivo* distance and keeping the simulation system size practical, especially given a larger distance between the two membranes would lead to extension of the HasB linker region but not likely impact the HasR-HasB interface.

Visual observations

On performing two initial simulations of the HasR-HasB model as described above, in one of the simulations we noted interaction of two periplasmic loops of HasR with $HasB_{CTD}$ (discussed in detail later), so we initiated 2 additional simulations of the HasR-HasB complex from a point at which the contact between the loops of HasR and HasB has already been established (Fig 1). Thus, overall we simulated the HasR-HasB complex as (1) HasR-HasB in which wild-types of both proteins are considered in the original model, (2) int-HasR-HasB in which the simulations are initiated after the contact with the loops has been made, and (3) $HasR_{L4}$-HasB in which L4 of HasR is mutated (this mutant is described below). 2 x 300 ns of the HasR-HasB and $HasR_{L4}$-HasB systems and 4 x 300 ns of the int-HasR-HasB systems were performed.

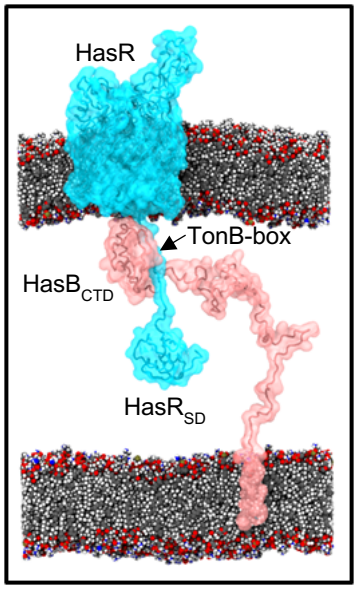


Fig 1. *The initial model of the HasR-HasB complex with each protein embedded in a model membrane at time = 0 ns. HasR is in cyan, HasB is pink, the lipids are grey, red, white and blue.* The outer and inner membrane are the upper and lower parts, respectively.

Secondary structure and structural integrity

Given the complexes we are simulating are models, albeit based on structural data and guided by experimental data, we first evaluated the structural integrity of the folded domains. The root mean square deviation (RMSD) from the initial model and the secondary structures were evaluated. In all cases we observe that the protein structural deviation is within

acceptable ranges based on simulations of similar proteins and there is no significant unfolding of the structured domains. The RMSD, root mean square fluctuation (RMSF) and secondary structure analysis of one simulation of each system are provided in the supplementary information.

Inter-protein motions

The long linker regions of HasR and HasB impart considerable flexibility to the complex. Therefore, we next explored the inter-domain and inter-protein movements. A useful way of filtering the large-scale motions is principal components analysis (PCA). We employed this method to characterise the motion along the first eigenvector for the HasR-HasB complexes in the HasR-HasB and int-HasR-HasB systems. Interpolation between the extreme conformations sampled in each case shows a distinct pattern in the inter-protein motions. The principle motion in the original HasR-HasB simulations is concomitant reduction in the distance between the $HasR_{SD}$ and the barrel, and the distance between $HasB_{CTD}$ and $HasR_{SD}$ as it is shown in Fig 2.

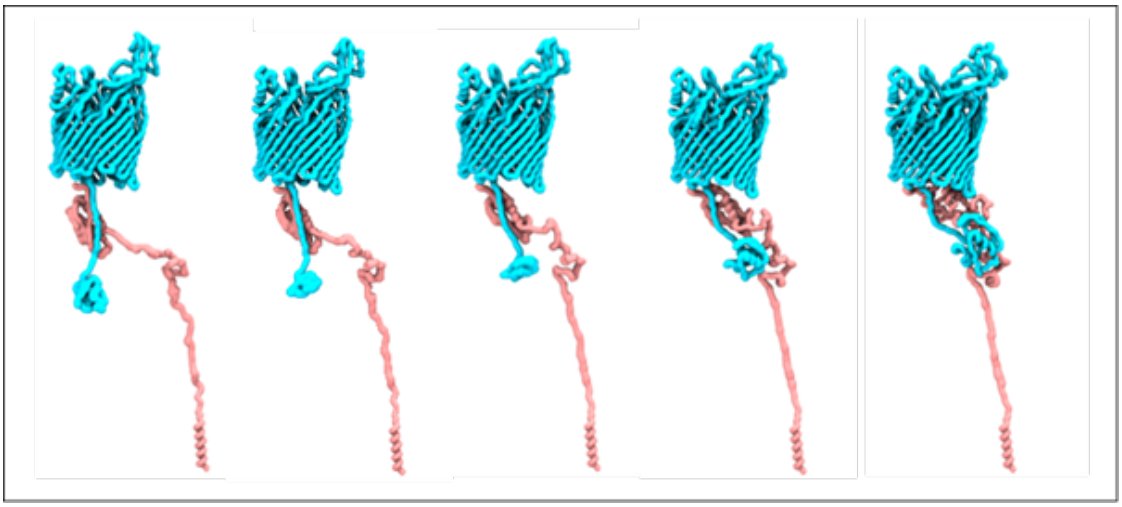


Fig 2. *Projection along the first principle component from the HasR-HasB simulation in which spontaneous interaction between HasB and HasR periplasmic loops are observed.* HasR is in cyan and HasB in pink. The membranes are not shown.

In simulations of the int-HasR-HasB system, the major motion is rearrangement of $HasB_{CTD}$ with respect to the HasR linker and $HasR_{SD}$ (since the large motion of $HasR_{SD}$ and $HasB_{CTD}$ had already occurred). These data strongly suggest that the HasR periplasmic region is considerably flexible and that the starting model required some inter-protein adjustments in particular with respect to interaction of the $HasR_{SD}$ with $HasB_{CTD}$ given the adjustment occurred spontaneously during the simulations.

In order to explore these movements in a more quantitative manner, we measured the HasR barrel-$HasR_{SD}$ and the HasR barrel-$HasB_{CTD}$ distances. In the HasR-HasB simulations there is a reduction in the HasR barrel-$HasR_{SD}$ distance from ~ 11 nm in the original model to ~7-7.5. nm by the end of both simulations. Similarly, the HasR barrel to $HasB_{CTD}$ distance is decreased from ~ 6.8 nm to ~ 4.5 nm in one simulation and to ~ 5.4 nm in the other simulation. In simulations of the int-HasR-HasB system, the distances remain at ~ 7.7 and 4.7 nm respectively across all four simulations. A striking feature of these measurements is the consistency in the int-HasR-HasB simulations. Measurements for all four independent simulations of each system stabilise to essentially the same value. This indicates that in terms of large-scale movement, the protein domains have reached a metastable state for these complexes.

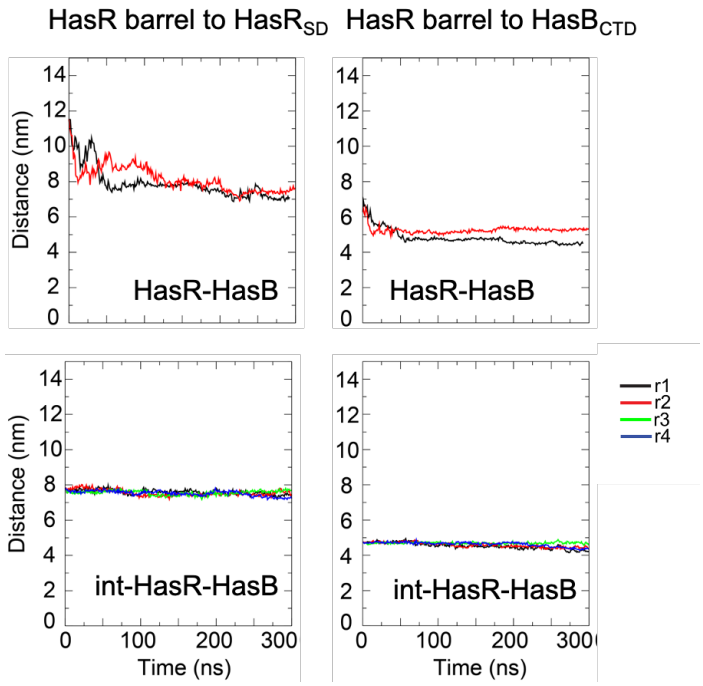


Fig 3. Intra- and inter-protein domain distances. *The panel on the left shows the minimum distance between the HasR barrel and the $HasR_{SD}$ as a function of time. The panel on the right shows the minimum distance between the HasR barrel and the $HasB_{CTD}$. In each case the center of mass of the protein domain is used for the measurements.*

Although the distance between $HasR_{SD}$ and $HasB_{CTD}$ is reduced in the HasR-HasB simulations and is maintained at near constant value in the int-HasR-HasB simulations, we did not observe any specific, stable interactions between these two domains. This can be explained by the fact that these domains are each linked to the rest of the protein with a long and dynamic disordered region. In order to reduce the distance between the $HasR_{SD}$ and $HasB_{CTD}$, the HasB linker region becomes less extended and more folded. We observe the formation and disruption of salt bridges and formation of hydrophobic contacts (Fig 4).

Although we do not observe the formation of secondary structural elements in the HasB linker, this may be due to the limited timescales possible with MD simulations.

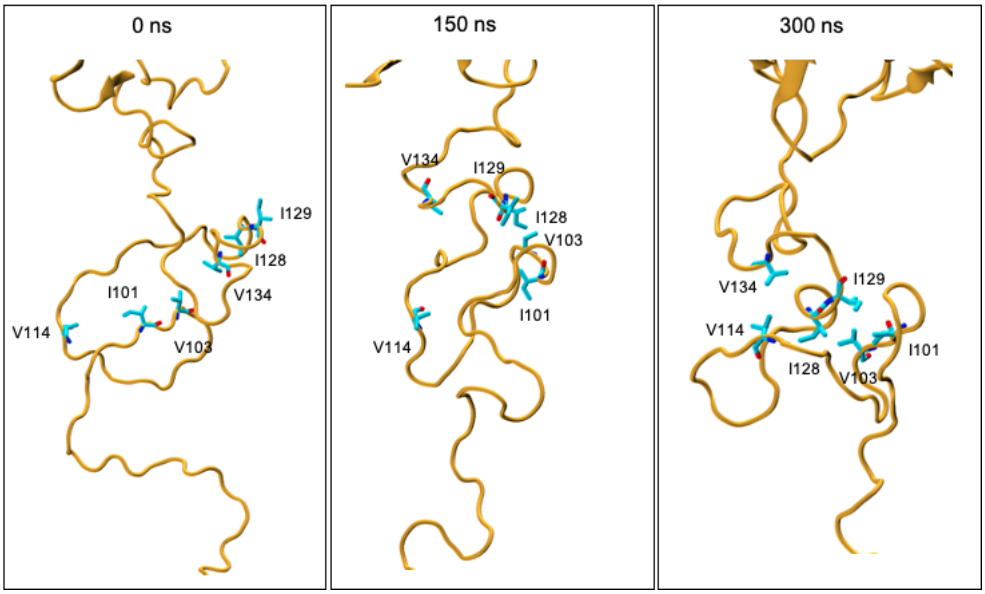


Fig 4: *Formation of hydrophobic contacts as the HasB linker becomes less extended. Key residues are labelled.*

Key salt bridge D100-R169

In three of the four int-HasR-HasB simulations we observed the salt bridge between D100 of HasR and R169 of HasB which was previously identified by our experimental work [13] (Fig 5). D100 is in the TonB-box of HasR, and R169 is in the $HasB_{CTD}$. Interestingly in some simulations the interaction is disrupted (defined as the minimum distance between the residues > 0.35 nm), but then reforms either for an extended period or for transitory periods (full details are in the supplementary information).

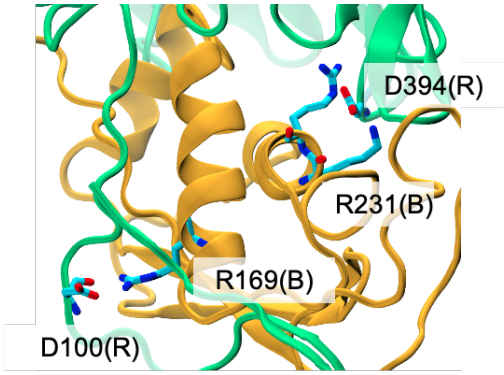


Fig 5: *Snapshot taken from t = 250 ns of one of the int-HasR-HasB simulations, showing the D100-R169 salt bridge intact. The D394-R231 salt bridge is also shown. HasR is in green and HasB in gold.*

Formation of an inter-protein β-sheet

In the initial model the residues around D100 in HasR (93-111) stand located within a suitable distance to the third β-strand of HasB (246-258) to form an inter-protein β-sheet, as suggested by our previous model of the interaction of $HasB_{CTD}$ with a peptide corresponding to the TonB-box[13]. We monitored the number of backbone hydrogen bonds between the two proteins in this region in each simulation. There are up to 5 hydrogen bonds present between residues 93-111 of HasR and residues 246-258 of $HasB_{CTD}$ in any one frame of HasR-HasB and int-HasR-HasB simulations. In the HasR-HasB simulations there is a reduction in the number of hydrogen bonds as the simulations proceed, such that by ~ 200 ns there are generally either none or just a single hydrogen bond present in this region. In the int-HasR-HasB simulations there are 2-8 hydrogen bonds throughout the simulations. The number of hydrogen bonds does decrease as some of the simulations proceed. Hydrogen bonding between backbone atoms is in some cases replaced by hydrogen bonds between side chains in this region. Examples include the A98-E249 backbone hydrogen bonding interaction which is replaced by A98 (backbone) to E249 (side-chain) in one of the int-HasR-HasB simulations. Full details are provided in the supplementary information. Thus, this region of interaction is labile and while the strands remain in close proximity to each other, the number of hydrogen bonds fluctuates. Overall, the inter-protein contacts are more stable in the int-HasR-HasB simulations, in other words once contact with HasR periplasmic loops has been established. We note here that an equivalent inter-protein β-sheet was reported between *E. coli* $TonB_{CTD}$ and the TonB-box of the TBDT FecA and BtuB [5,6].

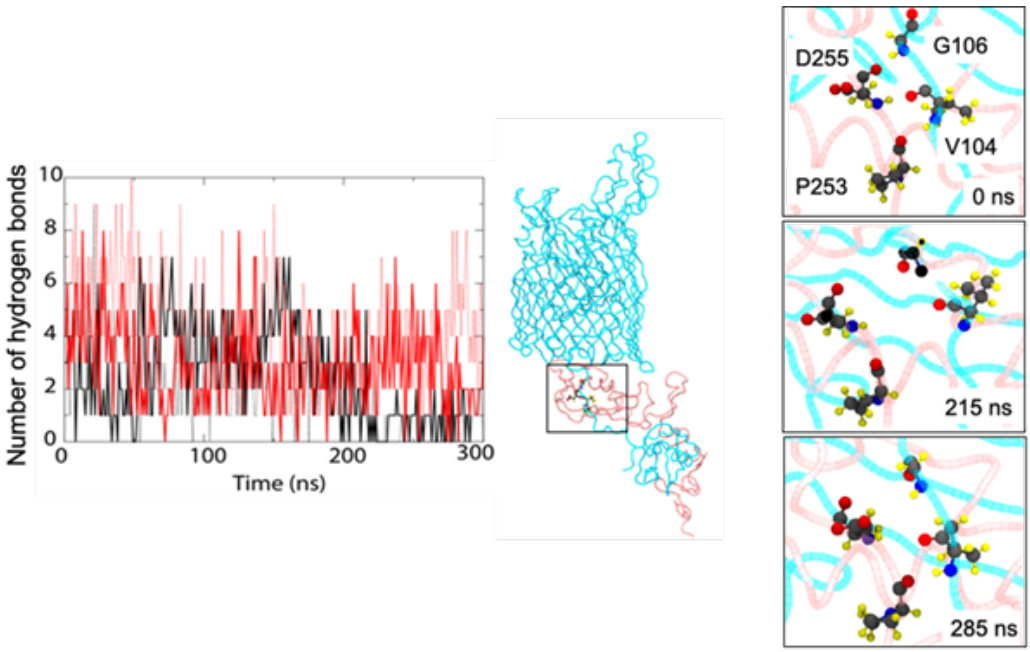


Fig 6. Hydrogen bonds within the inter-protein β-sheet. *The panel on the left shows the total number of hydrogen bonds between HasR 99-111 and HasB 246-258 as a function of*

time. The panel on the right (a zoom of the central panel) shows four residues (V104 and G106 of HasR and P253 and D255 of HasB) highlighted from one of the int-HasR-HasB simulations as an example of the lability in the interactions in this region. At time = 0 ns, backbone hydrogen bonding is seen between V104 and D255, G106 and D255, V104 and P253. At 215 ns the inter-protein hydrogen bonds are broken, by 285 ns the hydrogen bond between V104 and P253 is reformed and the four residues are in general closer together than at 215 ns.

Interaction of HasR periplasmic loops with $HasB_{CTD}$

As already discussed, at the start of the HasR-HasB simulations there is a large inter-protein motion in which the distance between the $HasR_{SD}$ and the $HasB_{CTD}$ is reduced. In one of the simulations this movement results in interaction of HasB with periplasmic loops L1 and L4 of HasR, the loop numbering being from the N- to the C-terminus. The interactions formed spontaneously within about 20 ns for L1 and 50 ns for L4 and then remained stable for the duration of the simulation. The int-HasR-HasB simulations were initiated from when these loops have already made contact with $HasB_{CTD}$, and these interactions are maintained for the duration of all four simulations (4 x 300 ns). The L1-HasB interactions are largely through hydrophobic contacts of A238 and P239 of L1 with L189 of HasB.

Once the interactions with L1 are formed, the HasR linker and $HasB_{CTD}$ move with respect to each other and then $HasB_{CTD}$ contacts L4 *via* several salt bridges. The most prevalent are between either E393 or D394 of HasR L4 with either K230 or R231 of HasB (Fig 5 and 7). Salt bridges between E393 and R223 or K230 are also observed. While the L4-$HasB_{CTD}$ interaction exists for the duration of the 5 x 300 ns of the simulations in which it is observed, no other individual salt bridge is observed for the duration of any of these simulations. Instead the salt bridges are observed to form, break and often reform throughout the simulations, suggesting either a protein-protein interface stabilised by non-specific interactions or that the simulations are not long enough to achieve a stabilised interface between the two proteins. Having said that the persistence in the overall number of contacts in all five simulations between the proteins in this region does suggest that L4 is important for the interaction of HasR with the $HasB_{CTD.}$ This is further reinforced in the simulation on $HasR_{L4}$-HasB system, $HasR_{L4}$ being a mutant of HasR in which L4 (P_{392}EDVDWLD) is replaced by the sequence ATSA. The size of four residues is the minimum required to form a periplasmic loop. In these simulations the distance between centers of mass of the HasR barrel and $HasB_{CTD}$ is ~ 5.6 nm (compared to ~ 4.5 nm in the simulations in which contact between L4 and $HasB_{CTD}$ is observed) and L4 does not come within hydrogen-bonding distance of $HasB_{CTD}$ during these simulations. Interestingly in the absence of L4-HasB interaction, the D100-R169 is not maintained.

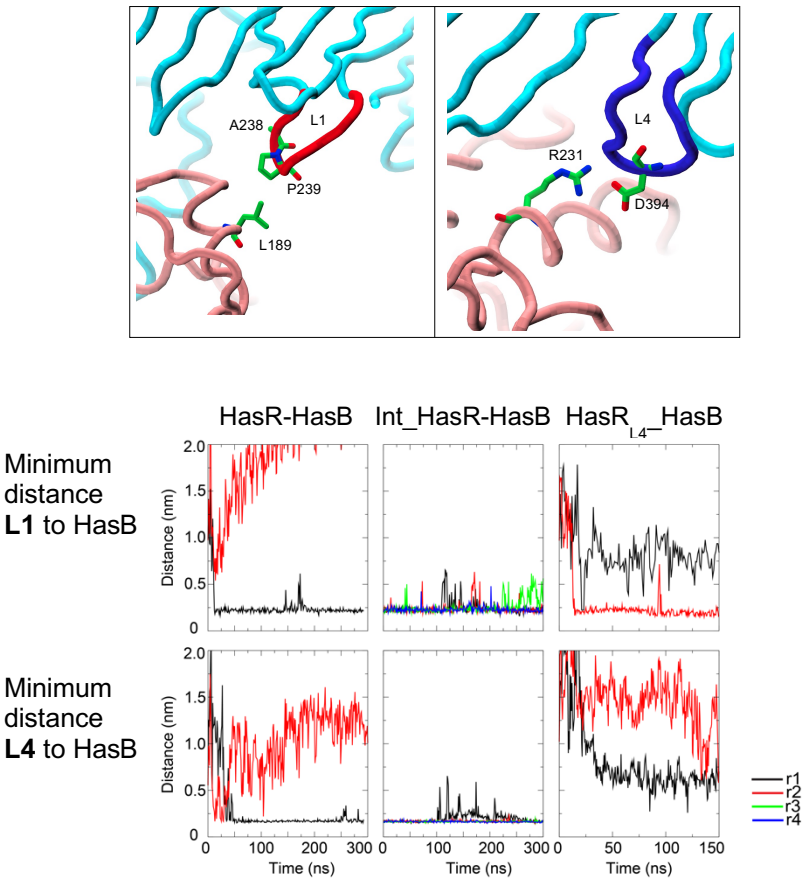


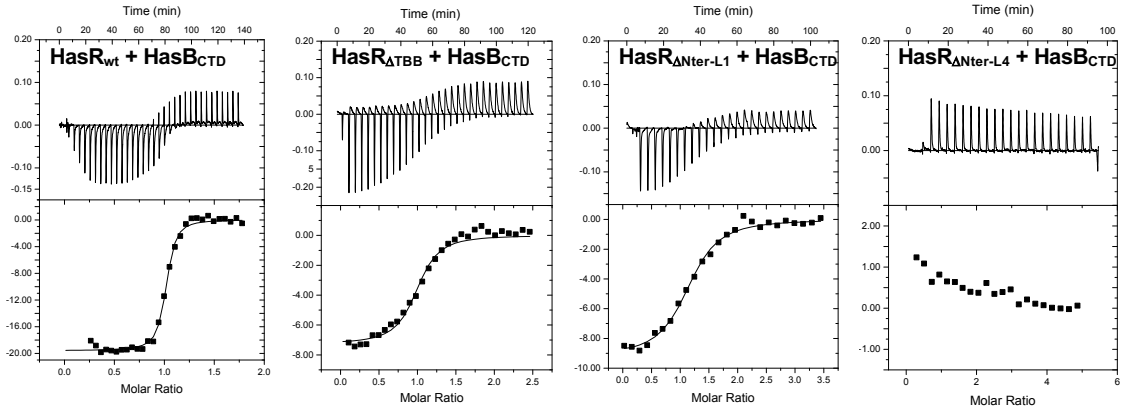
Fig 6: HasR L1 and L4 interaction with HasB*. Top: L1 and L4 are coloured in red and blue, respectively. Some key residues are labelled. Bottom: The evolution of the minimum distance between L1 or L4 to HasB during different simulations are shown. The distance from the closest atom between the loops and HasB was measured.*

Experimental validation of the role HasR L1 and L4 in the HasR-HasB interaction

In a previous work, we studied a HasR mutant in which the TonB-box was deleted ($HasR_{\Delta TBB}$). Although $HasR_{\Delta TBB}$ was not functional *in vivo*, it was still able to bind to $HasB_{CTD}$ *in vitro,* but with significantly less affinity [13] (Fig 7). We concluded that besides the TonB-box, other regions of HasR are recognized by HasB, although the TonB-box seems to be the main region of interaction.

Here in order to understand the contribution of the HasR periplasmic loops 1 (L1) and 4 (L4) in the HasR-HasB interaction, we used variants of HasR containing only the β-barrel and the plug. These variants lacking the N-terminal region and thus the TonB-box ($HasR_{\Delta Nter}$) allowed to measure the only contribution of HasR periplasmic loops in the interaction with HasB. Two different variants of $HasR_{\Delta Nter}$ in which the sequences of L1 (P_{232}GKE) or L4 (P_{392}EDVDWLD) were substituted by ATSA, were used. These mutants $HasR_{\Delta Nter\text{-}L1}$ and $HasR_{\Delta Nter\text{-}L4}$ were prepared and purified by using the same protocol as for the wild type, indicating that they were

well folded and inserted into the OM. The ITC experiment with $HasR_{\Delta Nter\text{-}L4}$ and $HasB_{CTD}$ displayed only a small endothermic signal or a heat of dilution. On the contrary, $HasR_{\Delta Nter\text{-}L1}$ interacts with $HasB_{CTD}$ with comparable thermodynamic parameters of interaction as $HasR_{\Delta TBB}$ (Fig 7). The substitution of HasR Loop1 has no additional effect than the deletion of the TonB-box on the interaction with HasB, whereas the HasR Loop 4 substitution has a drastic effect on the interaction.



Binding partners	N	Ka (M^{-1})	ΔH ($kcal.mol^{-1}$)
$HasR_{wt}$ + $HasB_{CTD}$	1.03	7.5 (± 0.2) x 10^7	-19.1 ± 0.24
$HasR_{\Delta TBB}$ + $HasB_{CTD}$	0.98	3.9 (± 0.8) x 10^6	-7.3 ± 0.18
$HasR_{\Delta Nter\text{-}L1}$ + $HasB_{CTD}$	1.1	3.03 (± 0.4) x 10^6	-9.2 ± 0.19

Fig7: ITC analysis of the interaction of $HasB_{CTD}$ with different variants of HasR. *Representative experiments are shown (top). In each case, the heat signal is shown together with the binding isotherm derived from the signal. Stoichiometries (N), affinity constants (Ka), Δ H of binding at 25 °C are shown (bottom).*

Then to assess the role of HasR L4 *in vivo* we monitored growth curves of heme auxotroph *E.coli K12* C600Δ*hemA* expressing the complete has locus (*hasISRADEB*) with *hasR* wild type or mutated in L4. A mutant in L3 was used as a control. This bacterial strain needs an external source of heme (here 0.4µM heme-BSA) and a functional Has system to grow. The complete *has* allows the expression of all Has proteins required for the secretion of HasA (*hasA, hasD, hasE*), heme internalisation (*hasR, hasB*) and the transcription regulation of the system (*hasI, hasS, hasB*). In this strain *E. coli tonBexbBexbD* were removed and *exbBexbD from S. marcescens was added* (HasB partners in the IM) in order to have a fully functional HasB.

The growth curves with the expression of the HasR wild type or mutated in L3 ($HasR_{L3}$) or L4 ($HasR_{L4}$) are shown in Fig 8. The HasR loop 4 substitution significantly affected the bacterial

growth curve (Fig 8A). The expression level of $HasR_{L4}$ mutant was slightly weaker than that of the wild type, although comparable to that of another mutant ($HasR_{L3}$) used as a control with a similar growth curve to that of the wild type (Fig 8B). With $HasR_{L4}$, the lag phase was significantly longer (9.5h versus 5.5h for the wild type and $HasR_{L3}$) and the OD600 did not reach the maximum obtained in the presence of wild type HasR.

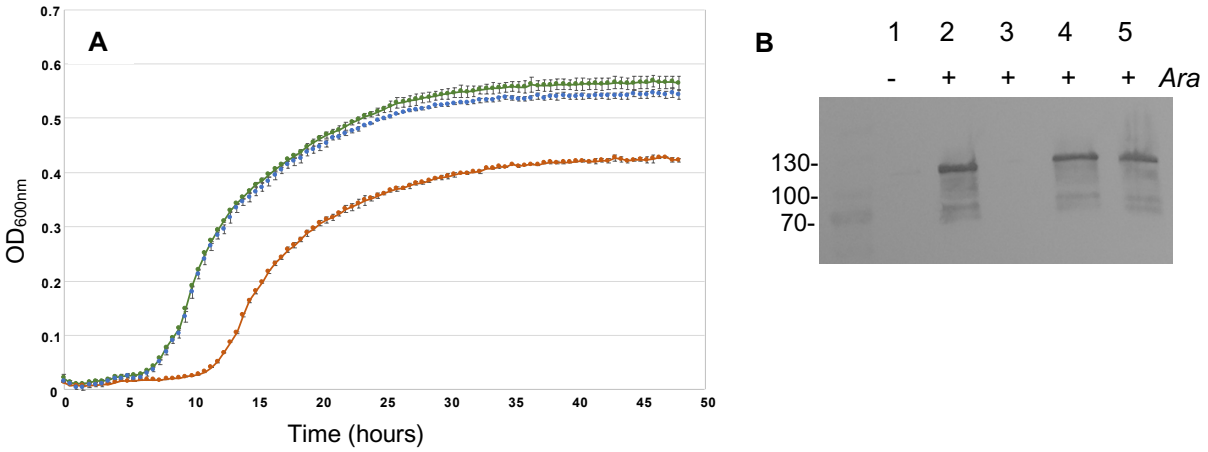


***Fig 8: Phenotypic analysis of HasR L4 substitution in vivo**. **A**: Growth curves of E. coli C600 ΔhemAΔtonBΔexbBD harbouring vectors encoding $exbBD_{sm}$ and the wild type has locus (blue) or derivative with HasR L3 or L4 substitution, in green and orange respectively. **B**: Immunodetection with polyclonal anti-HasR rabbit serum of HasR in C600ΔhemA transformed with either pBAD24 (negative control, lane 3), pFR2 ($HasR_{wt}$, lanes 1 and 2), pFR2: $HasR_{L3}$ ($HasR_{L3}$ mutant, lane 4) and pFR2: $HasR_{L4}$ ($HasR_{L4}$ mutant, lane 5). Cells were inoculated at $50mOD_{600nm}$ in LB medium containing δ-aminolevulinic acid (25µg/ml) at 37°C for 3hrs. An equivalent of 0.1 OD600nm was loaded on each lane. Arabinose (Ara) concentration was 40µg/ml. Molecular mass of standard weight markers is indicated on the left (in kDa).*

Discussion

The TonB-TBDT interaction involves mainly a salt bridge between the TonB-box of the transporter and TonB, stabilizing an inter-protein β-sheet [4,5,6,13]. Both salt bridge and the inter-protein β-sheet were shown important for the energy transfer from the TonB protein to the transporter and for the conformational changes of the plug domain allowing the nutrient internalization [15,16]. In our system, the deletion or charge inversion of the residues involved in this salt bridge (D100 from HasR and R169 from HasB) abolishes the HasR-HasB interaction *in vitro* and the heme acquisition by HasR *in vivo* [13,14]. In the simulations presented here, the salt bridge and the β-sheet remain stable during the simulations despite the high flexibility of the regions around. Moreover, the distance between HasR barrel and HasB decreases and

stays quite stable during the simulations. This suggests that HasR and HasB form a stable complex consistently with their nanomolar affinity [14].

In this study, by combining *in silico*, *in vitro* and *in vivo* results, we show the key role of TBDT perisplasmic loops in the interaction with a TonB protein and the sequence of the events. The interaction between HasR L1 and HasB occurs very early and remains stable. This interaction mostly involves hydrophobic contacts. Once they are formed, the HasR linker and HasB move with respect to each other (mostly through HasB rotation) such that $HasB_{CTD}$ contacts L4. The HasR-HasB interaction is then stabilized through several salt bridges between L4 and HasB K230R231, suggesting their role in the stability of the complex. In the absence of charged residues of L4 ($HasR_{L4}$), the salt bridge between the TonB box and HasB is not maintained during the simulations. The L4-HasB interaction seems to stabilize the HasR-HasB network of interaction.

The enthalpic signal observed in the ITC experiment reflects the polar interactions. The $HasR_{\Delta Nter\text{-}L4}$-$HasB_{CTD}$ did not show any enthalpic signal of interaction. The ΔH value (-7.3 $kcal.mol^{-1}$) measured for the $HasR_{\Delta TBB}$-$HasB_{CTD}$ interaction is compatible with the formation of at least two salt bridges that could correspond to the contribution of L4. In $HasR_{\Delta Nter\text{-}L4}$, the charged residues of L4 are replaced by neutral residues and the salt bridges cannot be formed. *In vivo*, the substitution of this loop significantly affected bacterial growth, suggesting its key role in the external heme uptake and/or the signalling through HasR, both activities are dependent on HasB.

If we compare the HasR-HasB interacting network observed in our simulations with the sequence and the structure of other TBDT-TonB complexes (2GSK, 2GRX, 6I97) that are available, the residues of HasR L4 and L1, notably the charged residues of L4 are well conserved in FoxA, BtuB and FhuA, although these loops are slightly shorter than in HasR. The HasB residues that contact L1 and L4 in our simulations are also well conserved and exposed at the surface of *E.coli* and *Pseudomonas aeruginosa* TonB (Fig 10). However, there are no simultaneous interactions between TBDT L1 and L4 with TonB in these complex structures. This may be due to the fact that in these studies, only the globular C-terminal part of TonB was used. In our full-length model of HasB inserted into the IM, the dynamics of its periplasmic proline-rich region would allow a better exploration of the exposed surface of HasR and the simultaneous interaction of HasB with L1 and L4. Therefore, we speculate that the HasR-HasB interacting mode revealed in this study can occur in other TBDT-TonB interactions.

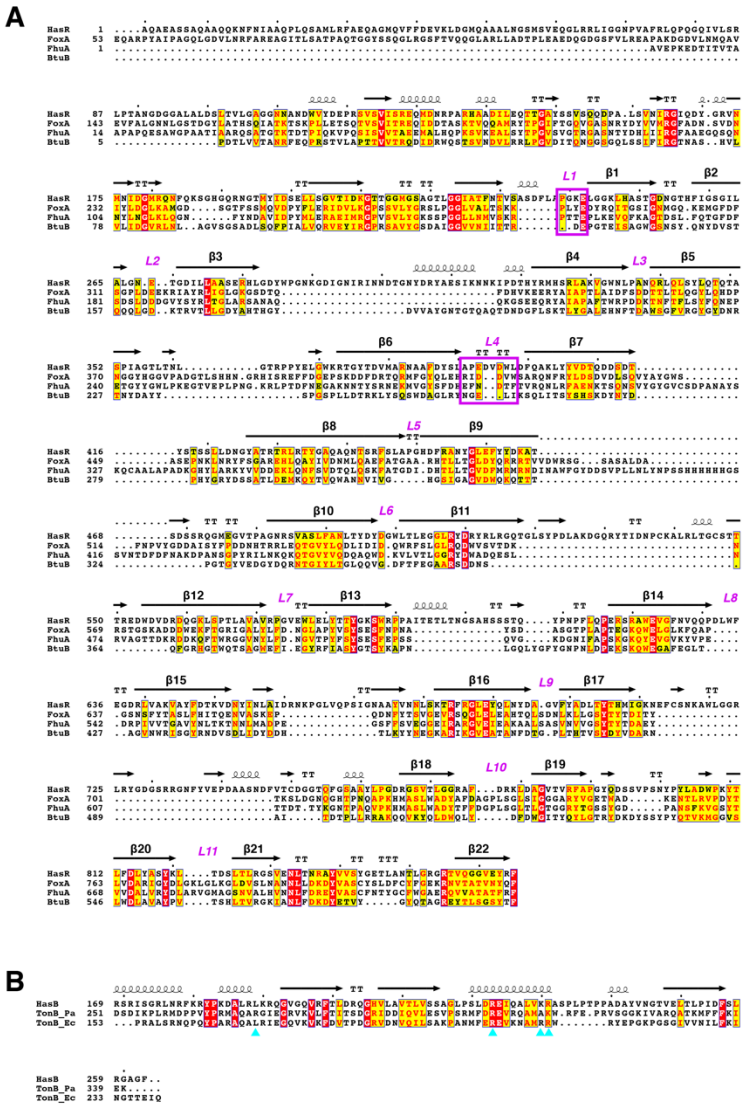


Fig 9: Structural alignment of TBDT structures. *(A) Alignment of TBDT HasR (PDB 3CSL), FoxA (PDB 6I97), FhuA (2GRX) and BtuB (2GSK) sequences. Positions of the 22 β-strands of the β-barrel are labelled along with the 11 periplasmic loops. Magenta boxes highlight the corresponding sequences for the periplasmic loops L1 and L4. (B) Alignment of HasB (PDB 2M2K) and TonB from Pseudomonas aeruginosa (PDB 6I97) and Escherichia coli (PDB 2GRX and 2GSK). Residues of HasB involved in HasR interaction from MD simulations are marked (cyan triangle).*

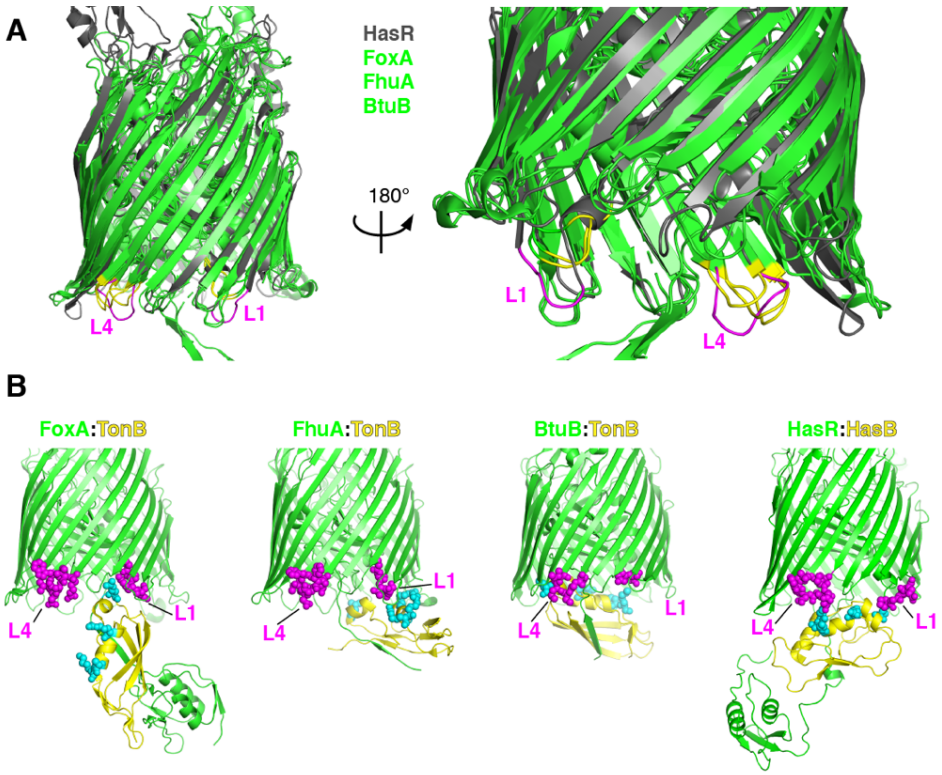


Fig 10: Periplasmic loops in TBDT in complex with TonB. (A) *Superposition of TBDT structures HasR (grey), FoxA, FhuA and BtuB (green). Loops L1 and L4 of HasR are coloured in magenta while the corresponding residues in FoxA, FhuA and BtuB are coloured in yellow.* ***(B)*** *Transporters (green) in complex with the C-terminal domain of TonB (yellow) are depicted in cartoon. The residues of loops L1 and L4 are shown in magenta spheres. Residues of TonB equivalent to the ones from HasB interacting with HasR in MD simulations (i.e. HasB L189, R223, K230 and R231) are shown as cyan spheres.*

Materials and Methods

Construction of the protein complex model

The HasR structure used as the starting point was a model that combined the X-ray structure of the HasR barrel and plug (PDB 3CSL) [9], the NMR structure of $HasR_{SD}$ (PDB 2M5J)[17] , SAXS data and the electron microscopy map of HasR (EMD-3978), as detailed by Wojtowicz *et al.*[11] In a previous work[13], a structure had been generated between the TonB-box of HasR and the C-terminal domain of HasB (PDB 2M2K) based on NMR experimental data (chemical shift perturbation and distance constraints) and by using the docking program Haddock[18]. This structure included a 21-residue peptide of the TonB-box, from G92 to A112, along with the C-terminal domain of HasB, N167 to F263 inclusive. This model was used to replace the equivalent region in the HasR starting structure by using the PyMol software package (*PyMol Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.*) The HasR mutant ($HasR_{L4}$)

was made by removing the L4 residues (P_{392}EDVDWLD) of HasR, before using Modeller to add the ATSA sequence. The H++ web-server[19-21] was then used to add hydrogen atoms to the new residues.

The HasB structure was modelled using the NMR structure of its C-terminal domain (residues 133 to 263 PDB 2M2K). The PEP-Fold web-server[22-27] was used to generate a model of the HasB transmembrane helix. The Modeller software (https://salilab.org/modeller)[28,29] was used to generate the disorder linker between the transmembrane helix and the C-terminal part and MODLOOP[30], a module of Modeller, was used to model the missing peptide bonds when connecting the various portions of the proteins.

Positioning the protein complex in the membranes

The CHARMM-GUI web-served[31-36] was used to assemble the lipid bilayers, both as squares with side-length 16.5 nm. The inner membrane was made up of two leaflets of 1-pamitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE). The outer membrane similarly had a lower leaflet of POPE but an upper leaflet of 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) as the tail lengths of DLPC are similar to those of lipopolysaccharide (which is found in the outer leaflet *in vivo*, but requires long simulations to achieve convergence and thus was not used here as our interest was in the protein complex rather than specific details of protein-outer membrane interactions)[20]. The two membranes were subjected to a 5000-step energy minimisation using the steepest descents algorithm.

To align the HasR barrel in the outer membrane, co-ordinates were taken from the MemProtMD database[26] and used for alignment to ensure the lipid headgroups were in the correct position relative to the protein. The location of the HasB helix in the inner membrane was estimated by using the location of the tryptophan residues, given there was no structural information available for this section[38].

Simulations involving HasR and HasB were set up as follows; once the proteins had been inserted into the membranes, any overlapping lipids were removed prior to energy minimisation. The CHARMM-GUI webserver was used to generate the necessary topology and parameter files for all models and structures reported in this study.

Molecular Dynamics Simulations

The systems were solvated with the TIP3P water model. K^+ and Cl^- ions were added to achieve an overall concentration of 150 mM charge neutral systems. The systems were then subjected to a short energy minimisation of 5000 steps, again using the steepest descent algorithm, to remove any steric clashes, followed by an equilibration protocol as shown in Table 1. This protocol, with slow relaxing of the restraints on the protein was necessary to allow the lipids to equilibrate around the proteins.

All simulations were carried out with the GROMACS 2019.6[39] version (www.gromacs.org) and CHARMM36[40] forcefield. The long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME)[41] method, whereas the nonbonded interactions and the short-range electrostatics were cut off at 1.2 nm, applied using the potential shift Verlet cut off scheme. The temperature was maintained at 303.15 K using the Nose-Hoover thermostat[42,43] with a 5 ps coupling constant. To maintain a constant pressure at 1 atm, the Parrinello-Rahman barostat[44] was applied in a semi-isotropic fashion, again with a coupling constant of 5ps. The LINCS[45] algorithm was applied to all atoms to allow the time step of 2 fs. Table 2 shows the summary of production runs, all of which were run in the NPT ensemble.

Table 1: Equilibration Protocol

Step	Time Step (fs)	Total Time (ns)	Protein Backbone Restraints ($kJ\ mol^{-1}\ nm^{-2}$)	Ensemble
1	1	0.25	2000	NPT
2	1	0.5	2000	NPT
3	1	0.75	2000	NVT
4	1	1.0	2000	NVT
5	1	1.5	1500	NPT
6	1	2.0	1500	NVT
7	1	2.5	1000	NPT
8	1	3.0	1000	NVT
9	2	4.0	500	NPT
10	2	5.0	500	NVT
11	2	6.5	250	NPT
12	2	8.0	250	NVT
13	2	10.0	100	NPT
14	2	12.0	100	NVT

Table 2: Summary of the Equilibrium MD Simulation Systems

System	Temperature (K)	Simulation Length (ns)
HasR + HasB	303.15	300 (x2)
Int_HasR + HasB	303.15	300 (x4)
$HasR_{L4}$+ HasB	303.15	150 (x2)

Strains and Plasmids

The *E. coli* strains and plasmids used in this study are listed in Supplementary Table 1. The plasmids for the expression of HasR mutants, pFR2 derivatives were prepared as follows: pFR2 plasmid (lab collection) was amplified by PCR with two couples of oligonucleotides (delP3-5 and delP3-3, on the one hand, and delP4-5 and delP4-5, on the other that contained a *Spe*I site at their 5' extremity). The PCR product was digested with *Spe*I, self-ligated and transformed into *E. coli* XL1-Blue. Recombinant plasmids were isolated and correct ones selected after sequencing, and the insert bearing the mutation reinserted in an otherwise wild-type pFR2 plasmid. The pFR2delP3 and pFR2delP4 plasmids were thus obtained, replacing for delP3 335-LPANQ-339 by LATSAQ and for delP4 391-APEDVDWLDF-400 by AATSAF. The mutants were also cloned into the pAMhasISRADEB context, by exchanging the *Fse*I-*Bst*Z17I fragments between pFR2 mutants and pAMhasISRADEB.

Protein expression and purification

HasR wild type and mutants were expressed in 16 L fermentor and purified as previously described [46,11]. All of the purification steps were performed at 4°C or 12°C in the presence of a protease inhibitor cocktail (*cOmplet, Roche)*. Protein concentrations were estimated using UV absorbance and the following calculated or determined molar absorption coefficients: 19150 $M^{-1}\cdot cm^{-1}$ at 277 nm for HasA [47] 140000 $M^{-1}\cdot cm^{-1}$ at 280 nm for HasR[46] and 10000 $M^{-1}\cdot cm^{-1}$ at 280 nm for $HasB_{CTD}$ [14].

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed at 25 °C using a MicroCal VP titration calorimeter (MicroCal-GE Healthcare) under the same conditions as described previously[14]. The buffer was 20 mM sodium phosphate, pH 7, 0.08% Zwittergent 3–14 for all titrations. All protein samples were dialyzed in this buffer before the experiments. Consecutive aliquots of 5-10 µL of $HasB_{CTD}$ at 80-120 µM were added into the ITC cell containing HasR wild type or mutant proteins at 4-15 µM. The heat of dilution of protein injections was determined either by injecting the ligand into the buffer alone or by injecting the ligand into the cell after the saturation of the protein binding

site. The value obtained was subtracted from the heat of reaction to give the effective heat of binding. The resulting titration data were analysed using Microcal-ORIGIN software package. The molar binding stoichiometry (N), association constant (Ka; Kd =1/Ka) and enthalpy changes (ΔH) of binding were determined by fitting the binding isotherm to a model with one set of sites. All experiments were done in duplicate.

Bacterial growth test

Growth tests in liquid medium was performed with *E. coli* C600 Δ*hemA*Δ*tonB*Δ*exbBD* strain harbouring vector pAMhasISRADEB encoding the whole has locus (*hasISRADEB)* and including either the wild type hasr or a mutant. The expression of this vector is under Fur regulation, induced in iron deficiency conditions. An additional *vector* pBAD24exbBDsm *encoding E*xbB and ExbD from *S. marcescens*, under the control of the P_{araBAD} promoter was also present. A few colonies of this strain were first inoculated in 4ml of LB medium at 37°C with 100µM dipyridyl (an iron chelator), 4µg/ml arabinose and the corresponding antibiotics. Once the culture reached an OD_{600nm} of ca. 1.2-1.5, it was diluted and inoculated in 48 well Greiner plates, in the same medium to which was added 0.4µM He-BSA, as a heme source. The initial OD_{600nm} of the cultures was 0.001. Each well contained 300µl of growth medium. Duplicates of each strain were made and the plate was incubated at 37°C with vigorous shaking (500rpm) in a Clariostar Plus Microplate reader. OD_{600nm} was recorded every 30 minutes for 70 hours. All experiments were performed in triplicate. In the expression test of HasR and his mutants, δ-aminolevulinic acid, a heme synthesis precursor was added to the culture at 25µg/ml.

References

1. H. Celia, N. Noinaj, S. Zakharov, *et al., Nature,* 2016, 538, 60–65.
2. V. Braun. *FEMS. Microbiol. Rev*.,1995, 16, 295-307.
3. *N. Noinaj, M. Guillier, T.J. Barnard, S.K. Buchanan. Annu. Rev. Microbiol., 2010, 64,43–60.*
4. A. Garcia-Herrero, R.S. Peacock, S.P. Howard, H.J Vogel. *Mol. Microbiol.* 2007, 66,872-89.
5. *P.D. Pawelek, N. Croteau, C. Ng-Thow-Hing, et al., Science. 2006, 312:1399–402.*
6. *D.D. Shultis, M.D. Purdy, C.N Banchs, M.C. Wiener. Science. 2006, 312,1396–9.*
7. C. Wandersman, and P. Delepelaire. *Mol. Microbiol.* 2011, 85, 618-31.
8. W. Huang, A. Wilks. *Annu. Rev. Biochem.* 2017, 86, 799-823.

9. S. Krieg, F. Huché, K. Diederichs, *et al., Proc. Natl. Acad. Sci. U S A*., 2009, 27, 1045-50.
10. M.S. Rossi, A. Paquelin, J.M. Ghigo, C. Wandersman. *Mol. Microbiol*., 2003, 48(6):1467-80.
11. H. Wojtowicz H, A. Prochnicka-Chalufour, G.C. de Amorim *et al, Biochem. J.,* 2016, doi:1042/BCJ20160131.
12. I. Josts, K. Veith, H. Tidow. *Elife,* 2019 8:e48528. doi: 10.7554/eLife.48528.
13. G.C. de Amorim, A. Prochnicka-Chalufour, P. Delepelaire, et al., *PloS One*, 2013, doi: 10.1371/journal.pone.0058964.
14. J. Lefevre, P. Delepelaire, M. Delepierre, N. Izadi-Pruneyre. *J. Mol. Biol.,* 2008, 9;378(4):840-51.
15. J. Gumbart, M.C .Wiener, E Tajkhorshid**.** *Biophys J.,* 2007, 15;93(2):496-504.
16. Hickman, S., Cooper, R., Bellucci, L. *et al. Nat. Commun.* 2017, *Nat. Commun,* 8, 14804.
17. Malki I, Simenel C, Wojtowicz H, et al,. PLoS One. 2014, 11;9(4):e89502
18. G. C. P. van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, et al., *J. Mol. Biol.,* 2016, 428, 720-725.
19. J. Myers, G. Grothaus, S. Narayanan and A. Onufriev, *Proteins Struct. Funct. Genet.*, 2006, 63, 928–938.
20. J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic Acids Res.*, 2005. 33, W368–W371.
21. R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, 40, W537–W541.
22. J. Beaufays, L. Lins, A. Thomas and R. Brasseur, *J. Pept. Sci.*, 2012, 18, 17–24.
23. Z. Wang, J. Eickholt and J. Cheng, *Bioinformatics*, 2011, 27, 1715–1716.
24. Harpreet Kaur, Aarti Garg and G.P.S. Raghava, *Protein Pept. Lett.*, 2007, 14, 626–631.
25. J. Maupetit, P. Tuffery and P. Derreumaux, *Proteins Struct. Funct. Genet.*, 2007, 69, 394–408.
26. J. Maupetit, P. Derreumaux and P. Tufféry, *J. Comput. Chem.*, 2010, 31, 726–738.
27. A. C. Camproux, R. Gautier and P. Tufféry, *J. Mol. Biol.*, 2004, 339, 591–605.
28. A. Šali and T. L. Blundell, *J. Mol. Biol.*, 1993, 234, 779–815.
29. R. Sánchez and A. Šali, in *Protein Structure Prediction*, Humana Press, New Jersey, 2000, pp. 97–129.
30. A. Fiser and A. Sali, *Bioinformatics*, 2003, 19, 2500–2501.
31. B. R. Brooks, C. L. Brooks, A. D. Mackerell, *et al.*, *J. Comput. Chem.*, 2009, 30, 1545–1614.
32. J. Lee, X. Cheng, J. M. Swails, *et al., J. Chem. Theory. Comput.*, 2016, 12, 405–413.
33. M. G. Martin, *Fluid Phase Equilib* , 2006, *Elsevier,* DOI:10.1016/j.fluid.2006.07.014.

34. S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, 29, 1859–1865.
35. J. Lee, D. S. Patel, J. Ståhle, S. J. *et al.*, *J. Chem. Theory Comput.*, 2019, 15, 775–786.
36. S. Jo, J. B. Lim, J. B. Klauda and W. Im, *Biophys. J.*, 2009, 97, 50–58.
37. J. Parkin, M. Chavent and S. Khalid. *Biophys J.*, 2015, 4;109(3):461-8
38. A. J. Situ, S. M. Kang, B. B. Frey, W. *et al., J. Phys. Chem., 2018*, 122, 1185-1194.
39. D. Van Der Spoel, E. Lindahl, B. Hess, G. *et al.*, *J. Comput. Chem.*, 2005, 26, 1701–1718.
40. J. Huang and A. D. MacKerell, *J. Comput. Chem.*, 2013, 34, 2135–2145.
41. *E. Fadrná, K. Hladecková, J. Koca, J. Biomol. Struct. Dyn., 2005, 23:151-62.*
42. W. G. Hoover, *Canonical dynamics: Equilibrium phase-space distributions*, 1985, vol. 31.
43. S. Nosé, *Mol. Phys.*, 1984**,** 52, 255–268.
44. M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, 52, 7182–7190.
45. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, 18, 1463–1472.
46. N. Izadi-Pruneyre N, F. Huche*, G.S. Lukat-Rodgers et al., J. Biol. Chem.,* 2006, 281, 25541-50.
47. N. Izadi, Y. Henry, J. Haladjian *et al.*, *Biochemistry*, 1997, 36,7050-7.

Acknowledgement: This work was funded by the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS), the French Agence Nationale de la Recherche ANR HemeStockExchange, the Fondation pour la Recherche Médicale (Equipe FRM 2017M.DEQ20170839114). We thank Christophe Thomas from the platform of production and purification of recombinant proteins at the Institut Pasteur for bacterial cell production in fermentors, the platform of molecular biophysics at the Institut Pasteur for providing access to the microcalorimeter and Loïc Hellio for HasB sample preparation. We thank Riccardo Pellarin for helpful discussions.