Studying dynamics without explicit dynamics: a structure-based study of the export mechanism by AcrB

Méliné Simsir, Isabelle Broutin, Isabelle Mus-Veteau, Frédéric Cazals

HAL Id: hal-02532366
https://inria.hal.science/hal-02532366v2

Studying dynamics without explicit dynamics: a structure-based study of the export mechanism by AcrB

Méliné Simsir [*][†] Isabelle Broutin [‡] Isabelle Mus-Veteau [§] Frédéric Cazals[¶]

September 14, 2020

Keywords: conformational changes, allostery, transporters, molecular distances, clustering, interfaces.

Acronyms: lRMSD: least Root Mean Square Deviation; $RMSD_{Comb.}$: combined Root Mean Square Deviation.

---

[*]Université Côte d'Azur, Inria, France
[†]Université Côte d'Azur, CNRS, IPMC, France
[‡]Université de Paris, CNRS, Laboratoire CiTCoM, 75006, Paris, France
[§]Université Côte d'Azur, CNRS, IPMC, France
[¶]Université Côte d'Azur, Inria, France. Correspondence: Frederic.Cazals@inria.fr

Abstract

RND family proteins are transmembrane proteins identified as large spectrum drug transporters involved in multi-drug resistance. A prototypical case in this superfamily, responsible for antibiotic resistance in selected gram negative bacteria, is AcrB. AcrB forms a trimer using the proton motive force to efflux drugs, implementing a functional rotation mechanism. Unfortunately, the size of the system (1049 amino-acid per monomer and membrane) has prevented a systematic dynamical exploration, so that the mild understanding of this coupled transport jeopardizes our ability to counter it.

The large number of crystal structures of AcrB prompts studies to further our understanding of the mechanism. To this end, we present a novel strategy based on two key ingredients which are to study dynamics by exploiting information embodied in the numerous crystal structures obtained to date, and to systematically consider subdomains, their dynamics, and their interactions. Along the way, we identify the subdomains responsible for dynamic events, refine the states (A,B,E) of the functional rotation mechanism, and analyze the evolution of intramonomer and intermonomer interfaces along the functional cycle.

Our analysis shows the relevance of AcrB's efflux mechanism as a template within the HAE1 family but not beyond. It also paves the way to targeted simulations exploiting the most relevant degrees of freedom at certain steps, and also to a targeting of specific interfaces to block the drug efflux.

Our work shows that complex dynamics can be unveiled from static snapshots, a strategy that may be used on a variety of molecular machines of large size.

1 Introduction

1.1 Drug efflux: the example of AcrB

Resistance-nodulation-cell division (RND) transporters.   The periplasm of gram negative bacteria may be seen as a buffer zone helping manage cytotoxic substances. Such compounds, which are metabolic wastes or substances entering via porins, should be expelled outside the cell. Of particular importance are resistance-nodulation-cell division (RND) transporters, a superfamily of membrane transporters involved in activities ranging from multidrug resistance to trafficking of lipids. As opposed to channels, transporters do not open simultaneously both *communication* sides. A prototypical case in this superfamily, representative of the HAE1 family, is the AcrA-AcrB-TolC transporter [1]. While AcrB is the inner membrane transporter, TolC is the outer membrane channel, with AcrA the adaptor protein linking AcrB and TolC. The size of the transporter and its amino acids composition make it possible to transport a wide range of compounds. Importantly, these transport mechanisms follow the concentration gradient, with the molecule expelled towards the high concentration side. The process requires external energy to trigger the necessitated conformational changes. In the case of AcrB, this energy comes from the proton motive force (PMF): in short, AcrB performs a coupled transport, trading one proton flowing along the negative gradient from the periplasm into the cytoplasm, against a substrate molecule transported from the cytoplasm/inner membrane to the outside of the cell across the periplasm (via AcrA and TolC).

Structure of AcrB.   AcrB is a transmembrane trimeric protein which monomer involves 1049 amino acids which are usually decomposed into three domains (Fig. 1(Top right)) and 19 subdomains (SI Table 2). AcrB was first crystallized in 2002 by Murakami's team in a symmetric conformation [2]. This first crystal structure did not involve any substrate, and yielded symmetric monomers. The three domains are called the transmembrane domain (TM), the porter domain (PD), and the funnel domain. Within the trimer, the AcrB monomers form three layers parallel to the plane of the membrane: the transmembrane domains shapes a ring in the inner membrane, while the porter and funnel domains form membrane proximal and distal (respectively) domains in the periplasm.

   The TM domain involves 12 α-helices and contains the proton relay site. It can be split into two membrane insertions stretches (TM1-6 and TM7-12). The porter domain (also named pore domain, located in the periplasm, is responsible for the drug efflux. It comprises four topologically homologous subdomains (each consisting of two $\beta - \alpha - \beta$ sandwiches) working in pairs: PC1 and PC2 sandwich the substrate entrance site which faces the periplasm, while PN1 and PN2 sandwich the exit site. The funnel domain, also located in the periplasm, contains the last lock of the RND drug pathway, namely the exit gate. This domain, interacting with the AcrA β-barrel domain decomposes into two subdomains denoted DN and DC. Finally, the last sub-domain is composed of one alpha helix in the cytoplasm, denoted α-helix.

Mechanism.   While the first crystal structure was substrate free [2], the notable event was the (almost concomitant) release of crystal structures obtained with substrates (doxorubicin or minocy-

clin) [3], [4]. Both publications proposed a three state alternating site functional rotation mechanism, whence the metaphor of a peristaltic pump. The two groups termed the three conformations of a monomer as *Access (A) - Binding (B) - Extrusion (E)* and *Loose - Tight - Open* respectively. Thereafter we will use the ABE terminology.

Only monomers in state B were crystallized with the drug, and the other states were inferred based on structural comparisons in terms of lRMSD with state B. Structural studies of trimers suggest that all combinations of state are not possible [5], and six states for trimers have been postulated (SI Fig. 8).

The analysis of crystal structures yielded a number of key insights. Briefly, the proton coupled efflux of one substrate molecule runs through steps involving three states [2, 4, 6, 1, 7, 8]. The three monomers cycle as $(A,B,E) \rightarrow (B,E,A) \rightarrow (E,A,B) \rightarrow (A,B,E)$. This cycling is made possible by proton binding and drug binding. As a prototypical scenario, consider that from [7]. A monomer binding a proton on the high concentration side triggers conformational changes. This makes substrate binding possible, and the trimer adopts the states (A, B, E (protonated)). When B binds a second proton, the trimer changes to states (A', B', E'). The conformational changes at play are such that the substrate is extruded from B', and the protein released from E' towards the low concentration side. Getting back to the overall mechanism, it has been noticed that the export across AcrB-AcrA-TolC is allosteric in nature [9], with structural changes in AcrB triggering a repacking of AcrA (to avoid leakage in the periplasm), which in turns triggers a synchronized opening of TolC.

However, several important steps are still under debate at the AcrB level. For example, since substrate transport and power generation are coupled, the preeminence of the former over the latter is unclear [8]. It is also unclear whether the protonation occurs in the B state or at a transition state between the B and E states. Finally, structuralists first considered the drug unbound in state A [3], but it was later shown as being drug-size dependent [10].

Modeling. Selected aspects of the efflux mechanism have been studied via molecular dynamics (MD), both coarse-grained [11] and all-atom [12] – up 3μs of simulation time for the latter [13]. See [14] for a survey. However, this strategy faces two major difficulties. The first one is the size of the system which prevents the discovery of large amplitude conformational changes. The second one is the correct handling of protonation states, which is not treated by classical force fields, and therefore requires manual intervention. Beyond structural aspects, models were also developed to study energetics. In [7], by quantifying both the entropic changes of the solvent and the enthalpic - entropic changes of the protomers, it is argued that the functional rotation mechanism is only possible due to the trimer configuration. A model based on the free energy landscape approach [15] has also been proposed [8]. Using selected electrochemical potential terms as well as assumptions on the cooperativity between the protomers, a sequence of elementary events is proposed, so that overall, the efflux complies with the first and second laws of thermodynamics.

4

1.2 Contributions

As shown by our review of previous work, the understanding of the efflux by AcrB presents shadow zones, with uncertainties on the preeminence of certain steps, the unclear role of subdomains in the coupling of substrate transport and power generation, as well as assumptions on the cooperativity between subdomains within and across monomers.

To partially bridge these gaps, this work develops a novel strategy based on two key ingredients which are (i) to study dynamics by exploiting information embodied in the numerous crystal structures of AcrB obtained to date, and (ii) to systematically consider subdomains, their dynamics, and their interactions. More specifically, we set the following goals both for AcrB and other RND proteins:

- **Classifying states of unlabeled monomers.** We ascertain whether monomers can be unambiguously ascribed to a state denoted A or B or E, using the least root mean square deviation (lRMSD) distance calculations based on C_α carbons. Likewise, we check whether any trimer can be ascribed to a particular configuration of states, e.g. ABE or AAA.

- **Identifying subdomains compatible with states ABE.** We study the dynamics of subdomains to identify a minimal set of subdomains consistent with the classification of monomers, based on two ingredients. The first one is the analysis of conformational changes of subdomains within states and across states, which are compared against the fluctuations observed in the crystal structures. The second one is the consideration of all possible clusterings based on all possible subsets of subdomains, under the so-called $\text{RMSD}_{\text{Comb.}}$ structural distance.

- **Refining monomer states.** To mine the existence of stable substates of states A, B, and E, we refine the clustering analysis for whole monomers using the subdomains identified as dynamic, under the so-called combined RMSD as distance measure.

- **Characterizing the evolution of interfaces between subdomains along state changes.** Interfaces between subdomains provide insights on the cooperativity of the efflux mechanism, and also hint at interfaces which might be targeted to block the efflux. We perform a thorough study of all interfaces and their evolution between subdomains across states.

From a methodological standpoint, we note that these analysis are in spirit analogous to the classification of hemoglobin quaternary structures [16], based on rigid superpositions of the α and β subunits, buried surface areas at the dimer/dimer interface, and the analysis of cavities. These approaches are illustrative of the *geometric* method, which, over the past half-century, proved instrumental to provide in particular structural and thermodynamic insights on individual molecules but also protein complexes [17, 18, 19, 20]. Letting alone the system and the specific tools used, the case of AcrB raises one novel difficulty though, namely the identification of relevant subdomains to explain the overall mechanism, a demanding task due in particular to the presence of loops/linkers.

2 Material and methods

2.1 Crystallographic data

Structures for AcrB. We focus on a set of wild-type AcrB structures with median resolution 3.32 Å(SI Fig. 9). We gathered in total 32 pdb files containing $n_{\text{trim.}}^{\text{Asym.}} = 23$ asymmetric structures: 17 files containing one trimer and 3 files containing two trimers (SI Table 3). Whence a total of $n_{\text{mono.}}^{\text{Asym.}} = 17 * 3 + 3 * 6 = 69$ monomers in asymmetric structures. Out of these 69 monomers, 43 of them come with a state label (label for short) A or B or E assigned by the authors (SI Table 4).

With a focus on state A, we also consider $n_{\text{mono.}}^{\text{Sym.}} = 12$ symmetric structures of AcrB (wild-type structures, resolution better than 6Å) (SI Table 3).

The asymmetric unit of each such crystal containing a single monomer, the final dataset comprises a total of $n_{\text{mono.}}^{\text{Tot.}} = 81$ monomers.

Domains and subdomains for AcrB. We used the segmentation of a monomer into subdomains and loops/linkers introduced in previous work [2, 3, 4] – shortening subdomains of up to three a.a. to ensure a minimal loop length of four a.a. Those loops were rendered consistent with secondary structure of AcrB using pdb 1iwg [2]. This resulted in 8 subdomains and 11 loops (SI Tab. 2).

Structures for other RNDs. To extend our results to other RND protein, we gather structures of RND having at least 5 different pdbs in order to have significant results. We gathered in total 6 pdb files containing $n_{\text{trim.}}^{\text{Asym.}} = 6$ for MexB (2 of them from cryo-EM), and 8 pdb files containing $n_{\text{mono.}}^{\text{Sym.}} = 8$ for CusA (SI Table 3).

Domains and subdomains for other RNDs. We tried to find in MexB (HAE1-RND) and CusA (HME-RND) subdomains equivalent to those of AcrB. To this end, we used structural definition of AcrB on 1iwg and searched for similar structural subdomains in 3kso [21] for CusA and in 3w9i [22] for MexB (SI Tab. 2).

2.2 Methods

In the following, we describe the methods used to investigate the goals introduced in section 1.2 for AcrB, using its states A, B and E. As detailed in SI (Sec. 7.1), all scripts were coded in python, based on packages of the Structural Bioinformatics Library (`http://sbl.inria.fr`, [23]), on Biopython PDB [24], and also several Scipy packages.

The corresponding package, called `Molecular_cradle`, is described at `https://sbl.inria.fr/doc/Molecular_cradle-user-manual.html`.

2.2.1 Classifying states of unlabeled monomers

Clustering structures. To classify states of unlabeled monomers within trimers, we resort to hierarchical clustering of monomers. To compare two monomers, we use the classical least root

mean square deviation (lRMSD), a global comparison based on C_α carbons, and the recently introduced combined RMSD or $RMSD_{Comb.}$ [25]. Given the decomposition of a structure into subdomains (SI Tab. 2), $RMSD_{Comb.}$ provides a weighted average of the lRMSD observed between these subdomains (the weight being the size i.e. the number of a.a.), stressing the role of local similarities. Practically, we use the implementation from the Structural Bioinformatics Library [23]–`https://sbl.inria.fr/doc/Molecular_distances_flexible-user-manual.html`.

Given a structural distance (lRMSD or $RMSD_{Comb.}$), consider the matrix which off diagonal entries are the $n_t \times (n_t - 1)/2$ distances obtained for all pairs of $n_{mono.}^{Asym.} = 69$ monomers.

We feed the resulting matrix to hierarchical clustering. A clustering run yields a dendogram, namely a binary tree which leaves are the individual monomers, with each internal node representing a merge between two clusters. The height of the internal node is called the *merge* or *fusion* value. Two methods provided equivalent performances: average linkage (AL), and Ward [26]. Practically, we present results with AL rather than Ward in all cases, since the fusion values are in distance units rather than squared distances units.

A cut in this dendogram defines a partition of the input data into clusters. Practically, we use two types of cuts: a topological cut (Defs. 1, and 3) and geometric cut (Section 2.2.2).

Propagating labels. The clusters obtained may be used to label unlabeled structures by propagating known labels to unlabeled structures within a cluster. To this end, we define:

Definition. 1 (Valid clustering of asymmetric structures with lRMSD .) *Consider a clustering of the $n_{mono.}^{Asym.}$ monomers, obtained by cutting a dendogram based on the lRMSD .*

Term a subtree anchored at an internal node of the dendogram homogeneous *if all state labels of its labeled leaves are identical i.e. are either A or B or E.*

The clustering is termed valid *with respect to a state provided that there exists an homogeneous subtree of size $n_{mono.}^{Asym.}$ containing all monomers for that state.*

A clustering valid with respect to the three states is simply termed valid.

Note that in our case, $n_{mono.}^{Asym.} = 69/3 = 23$. In practice, given a dendogram, we seek a cut so that definition 1 applies. The method to find such a cut plainly consists of a postorder traversal of the binary tree, as long as the subtree explored remains homogeneous.

2.2.2 Identifying subdomains compatible with states ABE

Conformational changes for subdomains. The crystal structures provide opportunities to understand which domains characterize the successive drug efflux steps. On the other hand, in performing structural analysis based on molecular distances (lRMSD and $RMSD_{Comb.}$), the precision on coordinates and the dynamics present in the crystal structures must be taken into account. For the former, from Cruickshank's formula [27, 28], the typical precision on atomic coordinates at a resolution of say 2.5Å lies in the range $[0.2, 0.4]$ Å. For the latter, atomic oscillation amplitudes are related to B factors by the formula $B = 8\pi^2 \bar{u}^2$. Yet, in a typical crystal, B factors may be affected by conformational disorder [29, 30]. For latter comparisons, we define the *mean displacement* for

a set of structures as the average value of $\bar{u}$ obtained using all C_α of the monomers in these structures. In practice, we use a single mean displacement computed for all (symmetric, asymmetric) structures.

Consider a particular subdomain (loop, TM region, etc), and the corresponding median lRMSD computed either over all pairs of structures within a state (A or B or E), or from two states i.e. (A and B) or (A and E) or (B and E). To single out significant distance values on a subdomain basis, we compare the median lRMSD obtained for a set of pairs (see below) against the average mean displacement per subdomain, computed for those pairs. To cope with the three states A, B and E, we perform six comparisons for each subdomain:

- one *intra state* comparison for each state A, B, E. The pairs processed are all pairs for that subdomain for the state considered.

- one *interstate* comparison for each pair of states i.e. (A, B), (A, E), and (B, E). The pairs processed are all pairs of subdomains for the two states considered.

A given comparison, be it intra state or inter state, is termed positive provided that the median lRMSD value is larger than the average mean displacement. In the following, we use the number of positive comparisons out of the six possible ones to qualify subdomains:

Definition. 2 (Static, dynamic, unstable subdomains.) *Consider the three intra state and the three inter states comparisons for a subdomain. Considering the number of positive comparisons, we define:*

- *(0/0) No positive comparison whatsoever:* **static subdomain***.*

- *(0/ $\geq$ 1) No positive intra state comparison, and at least one positive inter state comparison:* **dynamic subdomain***.*

- *($\geq$ 1/0) Positive intra state comparison(s) but no positive inter state comparison: irrelevant since not observed.*

- *($\geq$ 1/ $\geq$ 1) Positive intra and inter state comparison(s). A subdomain with at least one positive inter state comparison, for which at least one of the associated intra state comparison is negative:* **dynamic subdomain***. Otherwise:* **unstable subdomain***.*

Selecting subdomains. Assume that all monomers have been labeled (assigned a state) and clustered. We further our analysis by identifying those subdomains with significant dynamics, which account for valid clusterings. This requires two ingredients: first, considering all possible clusterings based on all possible subsets of subdomains; second, taking into account the relative size of subdomains by $\text{RMSD}_{\text{Comb.}}$ instead of the RMSD . We define:

Definition. 3 (Valid clustering of asymmetric structures with RMSD$_{Comb.}$.) *Consider a clustering of the $n_{mono.}^{Asym.}$ monomers found in asymmetric structures, obtained by cutting a dendogram based on RMSD$_{Comb.}$. A homogeneous subtree is called* maximal *if it is homogeneous and if the subtree anchored at its father node is not homogeneous.*

The clustering is termed valid *with respect to a state provided that the number of maximal homogeneous subtrees for that state is equal to one or two, with the additional condition in the latter case that the smallest subtree has at most two leaves. (A constraint meant to tolerate a number of outliers equal to 10%.)*

A clustering valid with respect to the three states is simply termed valid.

Note that in case a clustering is valid with respect one or two states (but not all three), the state(s) of interest are listed.

Similarly to Def. 1, given a dendogram, we seek a cut so that the definition applies, based on a postorder traversal of the binary tree.

2.2.3   Refining monomer states

As noticed above, we wish to single out those subdomains compatible with states. With a focus on a specific subdomain, we process the $n_{mono.}^{Tot.}$ monomers found in the asymmetric and symmetric structures.

We seek a valid clustering of these subdomains in two steps. First, we compute a dendogram using the lRMSD, and find homogeneous subtrees as explained previously, based on a postorder traversal. Second, we post-process these trees by cutting them into subtrees, in such a way that each subtree reported has the following property: the height of its root (the value at which the merge occurs in the dendogram) is less than the mean displacement.

2.2.4   Characterizing the evolution of interfaces between subdomains along state changes

To model contacts between two subdomains, we resort to the Voronoi based interface model, which defines an interface between two subdomains from the Voronoi diagram of the atomic solvent accessible model [31, 19]. More precisely, the Voronoi model identifies pairs of atoms, one on each subdomain, which are either directly in contact as their Voronoi cells are neighbor, or are contacting a common crystallographic water molecule. (Nb: Given the resolution of the crystal structures available, solvent water molecules were not incorporated for the calculation of the Voronoi models [31].) Consider two subdomains u and v. Practically, we distinguish two cases, namely an *intra monomer* interface $I_1(u,v)$ when u and v belong to the same monomer, and an *inter monomer* interface $I_2(u,v)$ when u and v belong to different monomers. For short, these interfaces are called intra and inter respectively. The number of atoms at such interfaces are denoted #*atoms* $I_2(u,v)$ and #*atoms* $I_1(u,v)$ respectively.

This number of atoms is known to be a robust linear estimator for the interaction area–each atom contributes 10 of buried surface area, which itself is a proxy for the *stability* of protein complexes [31, 19]. We also note that Voronoi models give a direct access to atomic packing properties [19], which are typically defined by the volume of atomic Voronoi regions, or by the

volume of the intersection between atomic balls and their Voronoi regions [32]. As packing defects concern almost exclusively solvent exposed atoms [19], our focus in the sequel is on interfaces.

Voronoi interfaces were computed using tools from the Structural Bioinformatics Library (SI Sec. 7.1). We exploit Voronoi interfaces as follows:

Definition. 4 (Intramonomer characteristic interface.) *Denote $\mathscr{S}$ the set of all monomers with a prescribed state. (Practically, $\mathscr{S} \in \{\mathscr{A}, \mathscr{B}, \mathscr{E}\}$, namely the sets of monomers associated with a state A, B or E.) Let $I_1(u \in m, v \in m)$ be the Voronoi interface between subdomains u and v in monomer $m \in \mathscr{S}$.*

The interface between subdomains u and v is termed characteristic *if it exists in a fraction $\geq f_1$ of all monomers in $\mathscr{S}$.*

We use $f_1 = 0.9$ in this work. This threshold is used to not discard interfaces involving small subdomains, in particular loops, may not be complete in the crystal structures. We proceed with:

Definition. 5 (Evolving interface for two subdomains within a state.) *Consider a characteristic interface $I_1(u, v)$ for a set of monomers $\mathscr{S}$. Consider the median of the number of atoms involved in all instances of this interface:*

$$med_S(u,v) = median\{\#atoms\ I_1(u \in m, v \in m), m \in \mathscr{S}\}. \tag{1}$$

The medians for states A, B and E are denoted $m_A = med_A(u,v), m_B = med_B(u,v), m_E = med_E(u,v)$.
The interface is termed evolving *across the three states provided that*

$$\frac{\max(m_A, m_B, m_E) - \min(m_A, m_B, m_E)}{\max(m_A, m_B, m_E)} \geq c(= 0.5) \tag{2}$$

We also study interface between subdomains belonging to different monomers – intermonomers interfaces. In order to compute interfaces for symmetric state crystals, we reconstructed the trimer from the asymmetric unit (a monomer).

Definition. 6 (Intermonomer characteristic interface.) *Consider a pair of states $(S,T) \in \mathscr{D} = \{(A,B),(A,E),(B,E),(A,A)\}$. Let $(\mathscr{S}, \mathscr{T})$ be the corresponding sets of monomers. An interface $I_2(u,v)$ between two subdomains u and v is termed* characteristic *provided that it exits in a fraction $\geq f_2$ of all pairs of such monomers.*

Note that practically, the sets $(\mathscr{S}, \mathscr{T})$ are neighboring monomers in the trimers studied. Practically, we also use $f_2 = 0.9$ in this work. We proceed with:

Definition. 7 (Evolving interface for two subdomains from two states.) *Consider an intermonomer characteristic interface for two subdomains u and v. The median of the number of atoms involved in all instances of this interface is defined by*

$$med_{(S,T)}(u,v) = median\{\#atoms\ I_2(u \in m, v \in n), (m,n) \in \mathscr{S} \times \mathscr{T}\}. \tag{3}$$

The medians for the four pairs of states are denoted $m_{AB} = med_{(A,B)}(u,v), m_{AE} = med_{(A,E)}(u,v), m_{BE} = med_{(B,E)}(u,v), m_{AA} = med_{(A,A)}(u,v)$. The interface is termed evolving *provided that*

$$\frac{\max(m_{AB}, m_{AE}, m_{BE}, m_{AA}) - \min(m_{AB}, m_{AE}, m_{BE}, m_{AA})}{\max(m_{AB}, m_{AE}, m_{BE}, m_{AA})} \geq c(= 0.2). \tag{4}$$

Note that the interfaces $(AB), (AE), (BE)$ correspond to asymmetric structures, with (AA) for symmetric ones.

Remark 1 *The previous methodology has been described for AcrB using its three states A, B and E. It adapts* mutatis mutandis *to other RNDs, taking for granted (i) a structural alignment method used to correspondences between subdomains. Practically, we use the* `Kpax` *aligner [33], and (ii) states equivalent to A, B and E.*

3   Results

We analyse the results for the four steps of our method, stressing in each case its systematic character using AcrB and its generic character using MexB. Moreover, we stress novel insights, if any for the mechanism.

3.1   Classifying states shows differences across RND

AcrB.   Using valid clusterings (Def. 1), we wish to propagate known labels of monomers to unlabeled monomers of the same cluster. Clustering all asymmetric AcrB monomers using the lRMSD indeed yields clearly separated clusters corresponding to the states A, B and E, from which the labeling of the 26 unlabeled monomers can be done unambiguously (Fig. 1, SI Table 4). All monomers being labeled, we note that the $n_{\text{trim.}}^{\text{Asym.}}$ trimers are in state ABE.

   To further this observation, we also clustered monomers using the lRMSD on the individual domains, obtaining a correct result for the TM and porter domains, but erroneous for the funnel domain (SI Table 5). This result is expected, as the funnel domain is known to be stable (in terms of conformations rather than interfaces) across states [4], which confuses the clustering algorithm. Clusterings obtained using $\text{RMSD}_{\text{Comb.}}$ for the three domains rather than the lRMSD also yields a valid clustering (Fig. 1, insert). Finally, in terms of lRMSD values, our studies confirm that state E is the most dissimilar, with states A and B relatively similar. Indeed, states A and B merge in the dendogram at 2.22Å, while state E merges and at 3.2Å.

MexB.   Within the RND superfamily, MexB is the most similar to AcrB. Our analysis focuses on the (solely) six PDB structures available (30 monomers). Since sequences of AcrB and MexB differ, lRMSD were computed using the `Kpax` aligner [33] from the `SBL`. Regarding subdomains, we used those found as dynamic for AcrB.

   We first build the dendogram of AcrB and MexB, based on the $\text{RMSD}_{\text{Comb.}}$ of subdomains (Loop2, Loop11, TM). This dendogram shows monomers of MexB incorporated to clusters of monomers of AcrB (Fig. 6 Top). MexB therefore exhibits three states which we denote M-A, M-B, and M-E. Yet, the correspondence with those of AcrB is not direct as evidenced by the proximity between groups of monomers in the dendogram. On the one hand, the states M-B and M-E are structurally close to those of AcrB. We confirm that this is not the case for M-A which clustering-wise differs from all clusters (A, B, and E) of AcrB.[34]

Remark 2 *We used the same analysis on CusA, resulting in the identification of two states, none of which corresponding to a state in AcrB (SI Fig. 14).*

3.2 Steps A, B and E are mainly due to relative motions between selected subdomains

AcrB. We noted above that distances (lRMSD , $RMSD_{Comb.}$) between two structures are affected by the crystal resolution, and are partly reflected by the B factor values–section 2.2.2. Using the relative values of distances and mean-displacements, we defined static, dynamic and unstable subdomains (Def. 2). The analysis of the corresponding scatter plot for all subdomains singles out five subdomains (Fig. 2):

- α-Helix (37 a.a.). With 3 intra and 3 inter state positive comparisons, and lRMSD values for inter and intra state comparisons which are comparable: α-Helix is unstable.

- TM (380 a.a.). With 0 intra and 3 inter state positive comparisons: TM is a dynamic subdomain, with changes between any two states. The highest median lRMSD being associated to inter state comparisons with E state.

- Loop2 (6 a.a.). With 0 intra and 2 inter state positive comparisons: Loop2 is dynamic. Moreover, since the lRMSD is large (1.75Å) for states (A, B), moderate (1.1Å) for states (B, E), and small (0.9Å) for states (A, E), state B differs from states A and E.

- Loop8 (11 a.a.). With 2 intra and 3 inter state positive comparisons, and a gap singled out by two inter state comparisons: Loop8 is also a dynamic subdomain. Note that median lRMSD associated to inter state comparisons containing state E are the highest of all (around 3.5Å).

- Loop11 (16 a.a.). One gets 1 intra and 2 inter state positive comparisons, with a gap associated with the intra statecomparison for state E. Median lRMSD for state E is very low (0.25Å), it increases for state B while staying stable (0.6Å), on contrary state A is unstable (1.2Å). Therefore, Loop11 is a dynamic subdomain as well.

AcrB structures have classically been analyzed based upon PN1, PN2, PC1, PC2, DN and DC subdomains ([3] and Fig. 2). We note that these subdomains are not dynamic by our criterion; that is, their interest in studying the mechanism lies in spatial rearrangements of *quasi-rigid* domains.

AcrB contains a large number of subdomains which vary in size and dynamics. We wish to identify those which yield a classification consistent with that of whole monomers (Sec. 3.1). Using subsets of the 5 subdomains identified, we replicate the clustering analysis (Sec. 2.2.2), based on the notion of valid clustering (Def. 3). Analysis is reported in Table 1.

We observe that α-Helix yield an incorrect clustering. This is expected, as we have seen that it does not exhibit any structural coherence event within states (all comparison inter - intra state above the line $y = x$, Fig. 2). On the other hand, the four subdomains (Loop2, Loop8,Loop11, TM) are compatible with the clustering of states (Fig. 3). The same holds by taking into account (Loop2, Loop8, Loop11) and TM separately.

Since RMSD values depend on size, and the size of subdomains varies drastically (TM is 380 a.a., which is almost 4 times more than PN1 and 20 times more than any loop; SI Table 2), we carry out an analysis based on regions of equivalent size. To do so, we split TM into 21 subdomains (12 TM helices and 9 loops, SI Sect. 7.5), three of which turn out to be dynamic (SI Table 7). However, none of these yields a clustering of higher quality than those previously obtained–in terms of separation.

MexB. Having identified the 3 alleged ABE states of MexB, we build the graph comparing median lRMSD values to mean displacement value only for subdomains with lRMSD above mean displacement value for at least one inter-state comparison (Fig. 6, Bottom left). The following differences can be noticed with respext to AcrB:

- The subdomains (Loop1,6,7,9, PN2, DN, DC) which are not dynamic for AcrB are so for MexB. The subdomains which stand out are identical (Loop2,8,11, TM).

- The individual subdomains show variable behaviors. Loop8, which is dynamic for AcrB (with the highest inter-state (A-E) lRMSD), is stable during the transition between state A and E for MexB. TM is dynamic for AcrB and MexB, yet, with the highest lRMSD in between B-E and the lowest between A-B in AcrB, and opposite behaviors for MexB.

Finally, we performed inter and intra state comparisons with respect to states A, B, E, M-A, M-B and M-E. In analyzing box plots, low 'variability' corresponds to a boxplot range smaller than 0.75Å(Fig. 6 Bottom). We note:

- **TM** between AcrB and MexB, B and E are very similar, A not (median RMSD above 1Å) (Fig. 6, third row, right plot)

- **Porter** low variability but high lRMSD values (Fig. 6, last row, right plot). The lowest lRMSD value being between E and M-E and B and M-B.

We conclude that MexB's B and E states are similar to their counterpart in AcrB. State A is different, but the differences still have to be understood.

3.3   The coherence between subdomains suggests the existence of substates

AcrB. We replicate the clustering analysis using all wild type monomers, including the $n_{\text{mono.}}^{\text{Sym.}} = 12$ symmetric structures. For the TM and porter domains, we perform two types of analysis: first, by clustering the structures using exactly one subdomain, using lRMSD ; second, by clustering the structures using several subdomains, using $RMSD_{\text{Comb.}}$. As previously, the clustering generates a dendogram (Fig. 4). Two clusters are defined as significantly different if their distance is above the mean displacement of the studied region ($lRMSD > \bar{u}$). The clusters are colored accordingly. For example, for the TM region, the clustering cutoff used is equal to $\bar{u} = 1$.

Consider for example the analysis of TM using $TM_{1\text{-}6}$ and $TM_{7\text{-}12}$ under $RMSD_{\text{Comb.}}$ (Fig. 4(Top row)). The dendogram features six clusters denoted $A'_{TM_{\text{comb}}}, A''_{TM_{\text{comb}}}, A'''_{TM_{\text{comb}}}, B'_{TM_{\text{comb}}}, B''_{TM_{\text{comb}}}$

and E_{TM_comb}, three of which refine state A and two of which refine state B. Note that these substates depend on which regions of TM used to perform the clustering. The same six clusters are obtained for the porter domain (Fig. 4(Bottom row)).

Finally, the same procedure applied to the funnel domain does not provide any coherent clustering – no further analysis was performed on this domain.

Remarkably, the clusters obtained using both types of analysis (lRMSD and RMSD$_\text{Comb.}$) on TM and porter are subsets of the clusters corresponding to the A, B, and E states. (NB: we omit outliers in the process: two structures in 5 cases, and up to 7 structures (Loop11, SI Fig10) in the worst case.) Moreover, the clusters obtained with TM and porter match one another, which allows to drop the dependence to the domain itself and call these substates of states A and B respectively.

The clustering involving substates provides insights on the dynamics of the TM (SI Fig. 10) and porter (SI Fig. 11) domains. Indeed, clusters of the porter domain are clearly separated (and are formed at higher merge values), while clusters for TM are mixed for state A. This owes to larger conformational changes in the porter domain.

Remark 3 *Note that the difference in quality criterion of structures, such as Resolution or R-free, are not significant between ' and '' clusters, while '' cluster is so (SI Fig. 12). However, the number of Ramachandran outliers varies significantly across the three states. A* coherent *location along the sequence of these outliers for the backbones of monomers corresponding to a substate could account for the clustering obtained. We are not aware though, of any method assessing the drifting of coordinates induced by such outliers.*

MexB. No substates were detected for MexB probably due to the small number of structures.

3.4 The systematic study of interfaces unveils a small number of dynamic interfaces

AcrB: intra-monomeric interactions. We characterize intra-monomeric interactions with Voronoi interfaces between subdomains (SI Table 8, Fig. 5(Top)).

The number of interfaces varies across states: 57 in A state, 68 in B state and 74 in E state. State A is the only one with no unique interface. One interface is shared between A and B ($I_1(PN1, Loop4)$), two between A and E ($I_1(PC1, Loop2)$, $I_1(Loop2, Loop3)$), twelve between B and E (Fig. 5(Top)), one is B specific ($I_1(Loop1, Loop5)$), and six are E specific ($I_1(Loop8, PN1)$, $I_1(PC2, TM)$, $I_1(Loop7, PC2)$,$I_1(Loop11, Loop7)$, $I_1(Loop8, Loop5)$, $I_1(PC1, PC2)$). With a focus on interfaces that change significantly, we note that 17 interfaces change significantly according to Def. 5 (Fig. 5, top diagram, interfaces tagged as *interface*), and 7 of them differs by at least 30 atoms (Fig. 5, top diagram, interfaces tagged as <u>interface</u>).

We note that state E has the largest interfaces since 6 of the above mentioned increase in state E, as seen with the following three cases: (i) PC1 and PC2 which only interact in E state (49 atom involved; Fig. 7); (ii) Loop8 and PC1 which interact in all states, yet more prominently in state E (state A and state B: circa 19 atoms, state E: circa 49 atoms); (iii) TM-Loop11 which interaction increases (state A: 30 atoms, state B: 42 atoms, state E: 63 atoms) (SI Table 8).

AcrB: inter-monomeric interactions. We proceed similarly for inter-monomeric interactions (SI Table 9, Fig. 5(Bottom)), with a focus on two classes of interfaces.

The first class relates to interfaces which are not found in every pair of states: $I_2(PC2, DC)$ exists in AB and AE (SI Fig. 13), $I_2(PN1, Loop3)$ exists in AE, AB and AA.

The second class contains interfaces found for any pair of states, yet, exhibit a large size change when moving from one pair of states to another pair of states. This class contains $I_2(PN2, PC2)$, $I_2(DC, PN1)$, $I_2(DN, PN1)$, $I_2(PN2, PN1)$, $I_2(TM, TM)$, $I_2(PN1, PN1)$. As seen from interface sizes and their evolution (SI Table 9), the loosest interface is between B and E and the tightest with A and B. For example, using median values, the interface between PN2 and PN1 involves 80 atoms for AB and 58 for BE; the interface between PN2 and PC2 involves 68 atoms for AB and 19 for BE.

MexB. Next, we computed the intramonomer interfaces of MexB's subdomains (SI Table 10). The interfaces tend to be larger in MexB than AcrB; also, for significant interfaces of AcrB, the size difference is less marked in MexB. Consider e.g. the interface between $I_1(Loop8, PC1)$: in AcrB, one moves from 19 atoms in state A to 49 atoms in state E; in MexB, one moves from 41 atoms in state A to 52 atoms in state E.

4 Discussion

4.1 Inclusive and hierarchical clustering provides a comprehensive understanding of states and substates

4.1.1 All asymmetric states are of the ABE type

Previous work on AcrB trimers had partly sorted out the states of trimers for two reasons. First, a number of monomers had not been assigned a state. Second, asymmetric trimers (TTO/BBE, LTT/ABB, LLT/AAB had been postulated ([5] and SI Fig. 8); among them two asymmetric states ABB and AAB were actually crystallized after stabilization thanks to cross-linking [35]. In unambiguously assigning a state to each monomer and confirming that all asymmetric states in current wild type known crystal structures are ABE, our clustering based analysis resolves both issues.

We also study the relationship between states and the dynamics of AcrB's three domains [2]. The conformational changes of these domains are not homogeneous. In fact, the funnel domain was described as stable, while the other two domains are mobile throughout the cycle [3, 4]. With higher lRMSD values between clusters, our analysis shows that the porter domain undergoes more significant conformational change than those of TM.

4.1.2 Putative substates might lead to diversity in simulations

Our clustering analysis based upon symmetric + asymmetric structures identifies 5 novel substates (3 for state A, 2 for state B). These findings rely structural differences present in the crystal structures, which are significant with respect to thermal fluctuations. We noted that substate A' is more

similar to B' than to A". Naturally, structural similarity does not entail kinetic proximity, as two meta-stable states on the potential energy landscape may be separated by high energy barriers or a succession of moderate barriers. Also, the connexion between structure and thermodynamics / kinetics is complex, as mildly different structures in terms of backbone lRMSD may have very different thermodynamic and/or kinetic properties. Indeed, meta-stable states assessed in structural terms (say based on lRMSD) do not provide any direct information on their statistical weights (thermodynamics), or on the transition rates between their basins (kinetics). Nevertheless, we analyze the putative connexions between substates, ligands, and thermodynamics.

Several structures of AcrB used in this study were obtained using various types of drugs / ligands, which may be classified by size (doxorubicin, minocyclin: small < 550g/mol; erythromycin, rifampicin: large), and /or by physico-chemical properties such as solubility with the octanol-water partition coefficient (puromycin, minocyclin: high $logP < 1.5$; erythromycin, rifampicin: low $logP > 2.5$). These molecules influence the crystal structure in several respects.
In terms of structure and thermodynamics, the ligands typically stabilize transient states. In our case, similar ligands can be found in substates A' and A" (doxorubicin, minocyclin, erythromycin). As for substate A"', even if two of the drugs crystallized were used in A' too (dequalinium and ethidium also used for 2gif and 2hrt), they weren't found in the later crystals. All the drugs in substate A"' were co-crystallized in a different location compared to A' and A" (central funnel). It is interesting to note that ethidium occupies a different recognition site in the central funnel, different from that of e.g. doxorubicin [36].
As for kinetics, the binding / efflux mechanism may differ [37]. As already noticed [10], drugs with higher molecular weight binds to A state before B state, while drugs with lower molecular weight bind to B state directly (or binding to A state is transient and has not been seen yet).

Also related to experimental conditions, we note that the crystallographers used different crystallization strategies: various buffer with different pH (4.6 to 8) which may alter the electrostatic interactions, different temperatures which may alter thermodynamics and kinetics, the use of crystallization helper proteins (DARpins) which may yield artifacts / non native conformations. This was already noted by Eicher et al. as a possible reason of structural differences between their structure and the structures obtained by another group [38]. Also, the use of DARpins does not correlate with clusters, suggesting they do not alter significantly the structures.

In any case, the substates identified also show that structures should be used with care when it comes to selecting states/conformations to launch simulations, as the starting point may bias the outcome. For example, computational studies tried to resolve protonation state of E monomer, but no agreement was found between [39] (used 2dhh, state A") and [40] (used 4dx5, state A').

4.2 Dynamics without explicit dynamics

4.2.1 Identification of four dynamic subdomains coherent with states

The dynamics of biomolecules generally encompasses two complementary aspects: first, large amplitude conformational changes typically involving rigid bodies whose relative positions change; second, (local) dynamics contributing to the entropic stabilization of meta-stable states. A fine

description of these two aspects is of course very hard, as dynamics span multiple time scales over which stabilization occurs [41].

Classical analysis focused on a number of domains (PN1, PN2, PC1, PC2, DN and DC [3]). In particular, the dynamics of several subdomains (PC1, PC2, and to a lesser extent loop F617) in terms of translocations were underlined by previous studies [42]. We note that these subdomains do not yield clusters coherent with the three states Our analysis of conformational changes versus fluctuations in crystal structures identifies subdomains whose relative positions change while their intrinsic dynamics are not significantly altered. These five subdomains are Loop2, Loop8, Loop11, TM, and α-helix (Fig. 2). α-helix was found to interact with AcrZ in selected AcrB structures [43]. To date, no equivalent protein was found in another RND. AcrZ is believed to stabilize AcrB and accompany the ABE cycle. Remarkably, all of them but α-helix yield a correct clustering of states under RMSD$_{\text{Comb.}}$ (Table 1). (We note that using RMSD$_{\text{Comb.}}$ is mandatory here to account for the size variation of these subdomains.) Thus, their conformational changes are characteristic of the ABE states.

4.2.2 Incidence of these subdomains on overall relative motions

We now delineate the role of the four subdomains (Loop2,Loop8, Loop11, TM) in the cycle A, B, E.

State A was previously termed a loose state [4], since monomers are not constrained by neighboring ones, and internal interactions are also slack. Our analysis strengthens this observation, as we note that state A minimizes the number of interfaces between subdomains within a monomer (44 compared to 52 and 54).

Loop2 is dynamic and undergoes its largest conformational change during this transition, and the dynamics of this loop is inherent to the switch from A to B state. In fact for Loop2, the clustering shows that state A is closer to state E than state B: lRMSD A to E: 0.9 Å, lRMSD A to B: of 1.75Å– data not shown. The movement of this Loop characterizes state B, and this movement is easily explained taking into account the other subdomains and the drug. While a movement of its neighboring subdomains might cause the dynamics, interaction with the drug is likely involved (2 of the 6 amino acids of that loop are part of the distal pocket (DP), SI Table 2). Being localized between PN1 and PN2, the dynamics of this loop is correlated to the significant decrease of the interface size between these two subdomains in state B. At that stage of the mechanism, the drug binding pocket must get larger in order to accommodate the drug, whatever the recognition site–out of three– exploited by the drug [10, 36, 44].

Loop8, which contains T676-loop described in the literature, is the linker between PC1 abd PC2. These two subdomains do not undergo conformational changes themselves (lRMSD less than the subdomains' mean displacement, Fig. 2), and their tighter interactions owes to the movements of the connecting loops. The T676-loop regulates the entry into the access pocket (AP)[45], which is pivotal for substrate transport. Molecular dynamics studies identified key amino acids interacting with the drug [14]. These amino acids are also said to be involved in the opening and closure of the PC domains. This is made possible by the oscillatory movement of Loop8 between the A and B states [45]. But the main changes undergone by this loop and captured by our clustering (correct clustering only for state E) is the conformational change responsible for the movement

of PC1 and PC2 happening during the transition B to E. A change also characterized in [45]. The creation of this new interface is responsible for the closure of the entrance to the pocket and the ejection of the drug up/out of the monomer.

For Loop11, we detect the coil to helix transition, which role has been assessed by [46].

TM is continuously hydrated and yields correct clusters for all states in our study (Table 1). The closest states are A and B–yet different enough to be separated. This is consistent with the open water route from the periplasm in both states and opening of the cytoplasm route in state B [47]. Structures in the E state stand further apart in the clustering, due to the closure of the periplasm route. The fact that states A, B and E coalesce upon splitting TM into its constitutive helices shows that changes in the relative position of the helices rather than individual conformational changes are at play (Table 1). Changes such as upward movement of TM2, or lateral movement of TM8 toward TM10 have already been characterized. [48, 49] It has been speculated [47] that proton uptake in TM occurs in between states A and B, or in between states B and E, and that proton release occurs between states E and A. Proton binding to protomer leads to a decrease in free energy, thus changing to a new more stable conformation [8].

The dynamics of Loop2, Loop11 and TM result in an increasing number of interfaces between states A and state B. The additional dynamic of Loop8 leads to new interfaces between state B and E (appearance of ($I_1(Loop7, PC2)$, $I_1(Loop11, Loop7)$, $I_1(PC1, PC2)$), and enlargement of existing ones ($I_1(Loop11, Loop8)$, $I_1(Loop11, TM)$). This is a consequence of the collapse of the drug binding pocket upon drug exit. Loop8 and even more Loop11 are close enough to the TM domain to define interfaces (Fig. 7 Lower left inset). We also note that Loop1, Loop5, Loop7 PN1 and PC2, which are not dynamic by our criteria, contact TM all along the cycle.

The signal propagation / force transfer to go from one state to another might involve these subdomains. We hypothesize that the dynamics of Loop2 Loop8 Loop11 and TM correspond to the motion A to B to E to A, while the dynamic of F617 correspond to the migration of the drug or the substrates. Recent studies also described new entry pathways for ligands [44, 36], next to Loop1 and Loop7. This suggests that regions involved in substrate passage and those involved in the ABE cycle behave differently. These hypothesis deserve further investigations [50].

Summarizing, our work underlines that quasi-rigid rearrangements of 7 out of 8 subdomains, quasi-rigid rearrangements or no rearrangements of 8 out of 11 Loops, and the dynamics of the 4 remaining are enough to lead to significant change of interfaces needed for the drug efflux.

4.3 Interfaces of potential therapeutic interest

The interface area between two proteins interacting non-covalently typically lies in the range 1500 - 4000 $Å^2$ [19]. (Note that interaction area refers here to the buried surface area, that is the surface area of the two molecules represented in the solvent accessible model, which get buried upon complex formation.) Remarkably, a robust statistic is that all interface atoms contribute on average ~ 10 $Å^2$ to this interface [31]. The situation is different for protein - drug interactions, where the smaller molecular weight of the latter yields interface areas on the protein side in the range 300 - 1000 $Å^2$ [51]. In light of these values, interfaces (within a monomer, in-between monomers) featuring a change of at least 30 atoms theoretically stands a chance to be of interest to modulate

the interaction / the mechanism. As a complementary statistic stressing relative rather than absolute measures, we also inspect those interfaces satisfying a significant relative change, as specified by Eqs. (2, within a monomer) and (4, across monomers).

Our results show that a wide variety of interfaces could be targeted.

On the one hand, using Eq. (2) 17 intra-monomer interfaces change significantly of size throughout the change of state. We noticed that one of these interfaces $I_1(Loop3, PC1)$, is involved in the binding of P9D, an inhibitor co-crystallized with AcrB [22], via a phenylalanine enriched groove between Loop3 and PC1. Out of 17 interfaces, 7 of them exhibit a change of more than 30 atoms when the state changes (SI Table 8).

The interfaces $I_1(Loop8, PC1)$ and $I_1(Loop9, PC1)$, which are located in the *middle* of the drug binding pocket, exist in all states, yet vary in size. Of particular interest is the interaction of $I_1(DC, PN1)$: it also satisfies both selection criteria (Eq. 5, at least 30 atoms), and exhibits a large size change (state A: 50 atoms; state E: 7 atoms); yet, DC and PN1 are stable subdomains (Fig. 2). More generally, several of our interfaces are located in the drug binding pocket, and could therefore be targeted by inhibitors.

On the other hand we have the inter-monomer interfaces for which two of them increase of more than 30 atoms. The first one, $I_2(PN1, PN1)$, does not seem easily accessible, as it is buried in the center of the trimer, The second one, $I_2(PN2, PC2)$, is of interest (SI Fig. 13). An inhibitor targeting this interface could bind while a BE interface would occur, since it is at its smallest. Because the interface would not be able to get back to its AB or AE size, both monomers could be blocked in those states. And if as hypothesized in the literature there could be only one E monomer in a trimer, the efflux would be stopped.

4.4   Understanding the conformational changes of other RND proteins

Studying the mechanism of RND proteins is a challenge because of their size, the need of energy from the proton gradient, and the membrane. As a matter of fact, most of the studies about the efflux mechanism are on one model, AcrB. Here, we tried to further the application to other RND proteins. Structural differences across this superfamily force us to either focus on structurally similar families, such as only tripartite ones, or on proteins of the same family (such as HAE1, to which AcrB's belongs).

4.4.1   Classifying states stresses different mechanisms

We used RMSD$_{\text{Comb.}}$ on tripartite RND proteins (AcrB, MexB, CusA), obtaining clusters similar to those of AcrB for MexB only. For CusA, only symmetric structures were resolved, yielding two states via our analysis (SI Fig. 14). Each cluster is characterized by different ligand(s) (monovalent ion, bivalent ion, no ion), which in itself is a cause of structural differences. Overall, AcrB's mechanism as a template suits MexB, but is of lesser interest for other tripartite RND pumps as well as monomeric ones for which subdomains are too different .

4.4.2 MexB: Different relative motions induces different mechanism

The structural differences of the non-transmembrane part of the proteins and lack of diverse structures in HME family protein confined us to the HAE1 family for the study of the porter domain.

Applying AcrB's workflow to MexB succeeded, a fact which we ascribe to the decent number of structures and the high overall sequence identity (70%). Previous work reported the similarity between states B and E of AcrB, and M-B and M-E of MexB, and differences between states A and M-A [34, 22, 52, 53]. These analysis also reported a noticeable difference in the dynamics of Loop11 which has already completed its coil-to-helix transformation in state B in MexB, and that M-A is more constrained than state B .

Our work refines these findings. First, we note that MexB has 6 additional dynamic subdomains, when compared to AcrB. Second, we note that differences between A and M-A mainly reside in the porter domain (Fig. 6 upper boxplot), and that M-A is almost as constrained as state M-E in the porter domain. Third, the stability of Loop8 during the transition E-A might explain higher constrains in M-A compared to AcrB's state A.

Note that no state M-A with a drug bound was resolved. It would be interesting to check whether drugs binding in state A could bind to state M-A. If we presume that changes in the TM domain are due to the protonation, the lRMSD value and fusion values of the clusters could suggest that either the kinetics of the protonation differ, or the number of protons differs (this number is still under debate even for AcrB). [14, 39, 40, 54]

5 Outlook

The coupled transport performed by AcrB and more generally RND transporters is a complex mechanism involving large amplitude conformational changes, as well a complex energetics. While previous modeling work has essentially explored two veins, namely the analysis of selected events occurring on short time scales using molecular simulation, and the design of thermodynamic models, our work introduces a third tier aiming at unveiling global dynamics based on static structures. More specifically, given a set of crystal structures containing the different states of the system scrutinized, we undertake a systematic study of dynamics of subdomains, and also of interactions between subdomains. The method relies on the ability to cluster conformations, based on global RMSD or the combined RMSD applied to suitable sets of subdomains. Because the sets of coherent subdomains are unknown a priori, our method systematically inspects them all. Our study is semi-supervised since we propagate known labels to unlabeled structures within clusters. However, the unlabeled case would work similarly, assuming the existence of well separated clusters that would define the labels. The method is general to encompass the case of homologous proteins–if one can identify common subdomains via alignments. It can also be applied to molecular machines involving different proteins, since the $RMSD_{Comb.}$ structural distance can be cascaded and computed in a hierarchical fashion–a requirement to run clusterings and classify states but also identify subdomains compatible with states.

The comparison of structures can be considered as standard for simple systems, in particular two biomolecules forming a binary complex–one compares the unbound structures to the bound

ones. Geometric criteria have also shown useful to classify quaternary structures of hemoglobin, for which all four subunits are known to be relevant. Our work goes beyond by considering a large ensemble of structures for which the role of subdomains but also loops/linkers is unclear in the first place, when a systematic analysis at the subdomain level. We also note that our grouping strategy is analogous in spirit to the design of Markov state models from molecular dynamics simulations. States in Markov state models indeed group together *coherent* conformations.

The systematic character of our analysis holds promises in several directions.

From the structural standpoint, the ability to identify dynamic domains, or domains which relative positions change at a specific step of a complex mechanism makes it possible to launch dynamic explorations (based on molecular dynamics or Monte Carlo methods) focusing on those degrees of freedom (typically dihedral angles) relevant to that step.

From the thermodynamic standpoint, we noticed in our review of previous work that the models developed were hinging on hypothesis regarding particular interactions between subdomains. Our systematic analysis of such interactions paves the way towards enhanced such models. In particular, the ability to compute local partition functions in regions of the conformational space where such interaction properties hold might lead to a direct and reliable estimate of thermodynamic and possibly kinetic properties via master equations. Along the way, it would be beneficial to also incorporate into the models other interaction properties (H-bonds, salt-bridges, hydrophobic interactions, specific interactions involving side chains). This endeavor, however, will require a new generation of algorithms to perform reliable predictions of thermodynamic and kinetic properties.

From the application standpoint, the exhaustive study of interfaces between subdomains offers a putative list of interfaces to be targeted by drugs or peptides to alter the mechanism. For the case of AcrB, this targeting would aim at blocking the cycle of alternating states $A \rightarrow B \rightarrow E \rightarrow A$. Last but not least, we also believe that our systematic approach is a first step towards the development of mechanistic models explaining complex mechanisms. In physics, Newton's cradle is a simple example of a system that can be modeled as a graph whose vertices represent the elementary parts– the beads, and whose edges represent certain interactions between these parts at certain moments in time. A complex molecular machine could be modeled in a similar way, with vertices representing subdomains and edges interactions between them–e.g. evidenced by interfaces. Our systematic analysis makes it possible to create such graphs, and we anticipate that in a near future, several molecular machines, for which there exist significant databases of structures, will be analyzed this way. Examples of high interest include transporters such as AcrB, but also fusion proteins, polymerases, or the nuclear pore complex, to name a few.
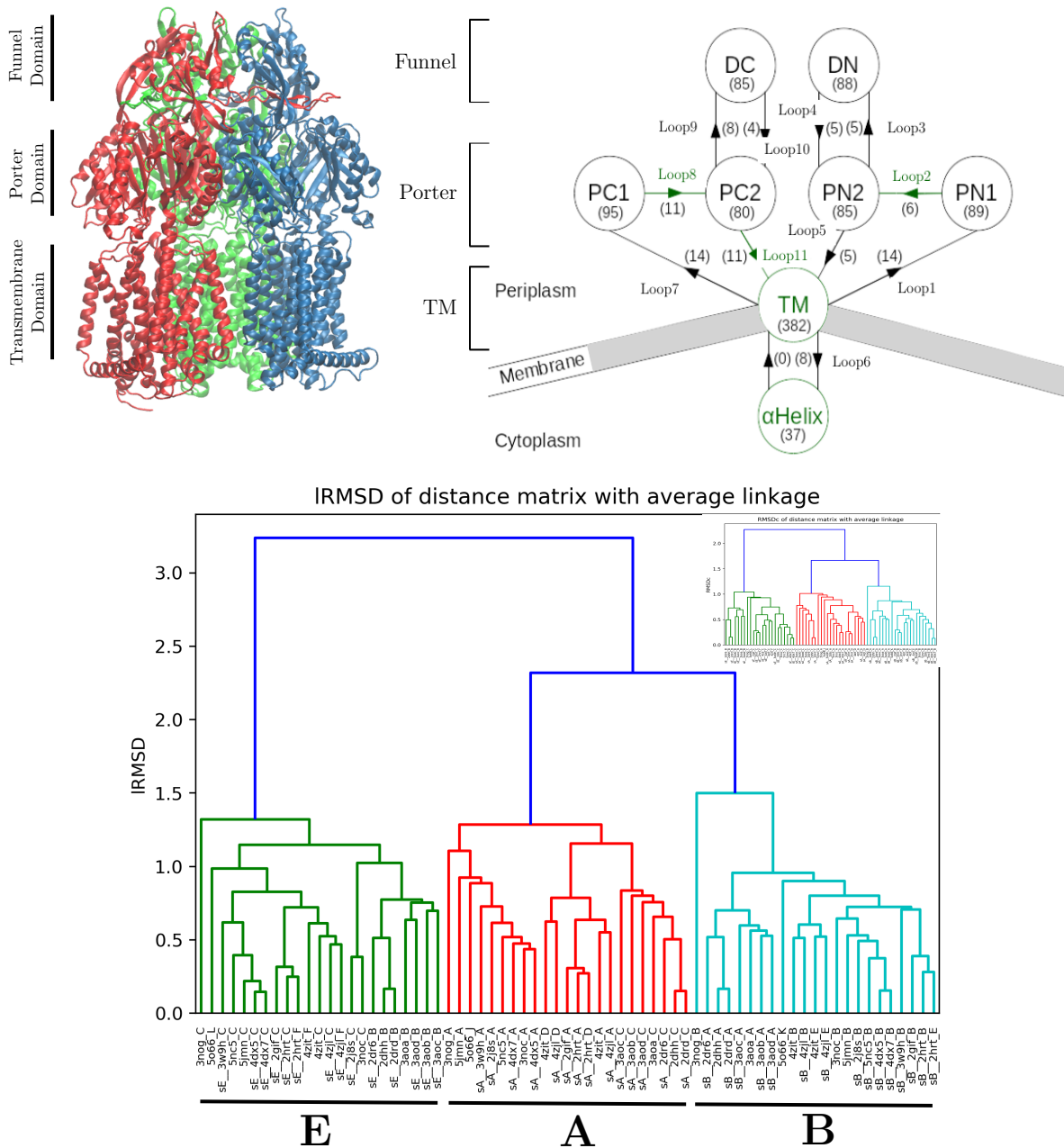
6   Artwork

Figure 1:

(Top left) Side and top views of the AcrB trimer (Top right) Topology of AcrB: subdomains defined from the sequence. Nodes represent subdomains–in parenthesis the number of a.a., and arc coil regions. Green regions are dynamics, as identified by our analysis–section 3.2 **(Bottom) Clustering the $n_{\text{mono.}}^{\text{Asym.}}$ asymmetric AcrB monomers using the lRMSD : clearly separated clusters allow labeling the 26 unlabeled monomers from labeled ones.** Hierarchical clustering with Average linkage was used. Main panel lRMSD ; inset RMSD$_{\text{Comb.}}$ with three domains (TM, porter, funnel). A leaf of the tree reads as follows: sE_3w9h_C: state E, PDB id 3w9h, chain id in PDB file C. Known labels are found in the same cluster: sE (green cluster), sA (red cluster), sB (cyan cluster). These labels allow labeling the remaining unlabeled structures. Note that RMSD$_{\text{Comb.}}$ yields fusion values significantly lower than those with lRMSD .
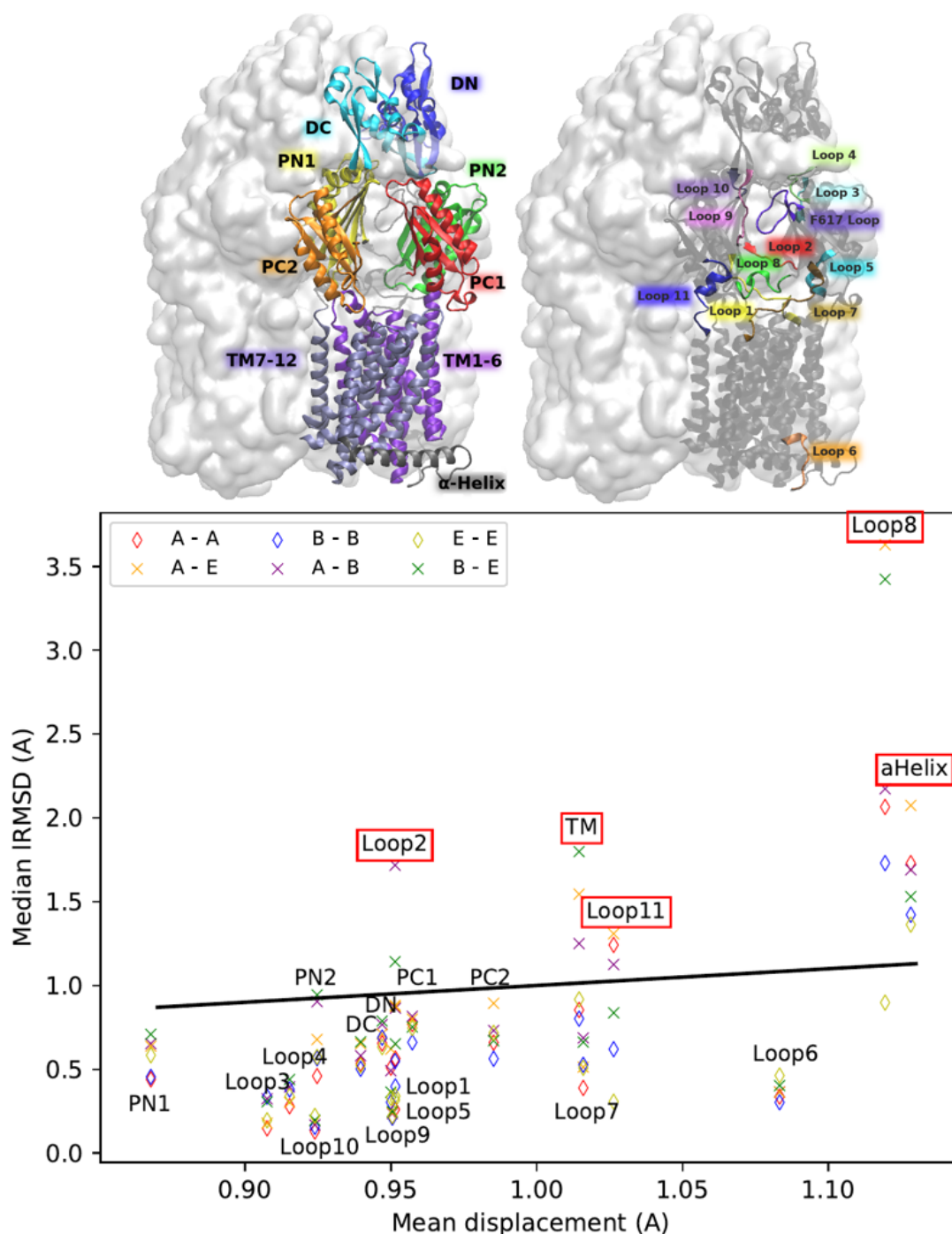
Figure 2: **(Top) The 19 subdomains in a monomer of AcrB (Bottom) Comparing the mean displacement in crystal asymmetric structures against the median lRMSD of subdomains for different states and transitions identifies 5 dynamic/unstable subdomains (Def. 2)–red boxes. Comparisons of monomers ascribed to the same state** A: red, B: blue, E: yellow; **Comparisons of monomers ascribed to different states:** A to E: orange, A to B: purple, B to E: green. The mean displacement is defined as $< u >$, with $B = 8\pi^2 \bar{u}^2$. The region of the scatter plot above the line $y = x$ identifies mobile subdomains.

24

Figure 3:

Clustering the AcrB asymmetric monomers restricted to subdomains (Loop2, Loop8, Loop11, TM, α-helix) (section 3.2) with RMSD$_{\text{Comb.}}$ yields three clusters corresponding to the A/B/E states. Green: cluster E, Cyan: cluster B, Red: cluster A.

	average
Loop11	**C(B) C(E)**
Loop2	**C**
Loop8	**C(E)**
TM	**C**
Loops from TM	X
Helices from TM	C(E)
TM5	C(E)
TMLoop3	**X**
TMLoop5	**X**
TMsplited	**X**
aHelix	X
Loop11 TM	C
Loop2 Loop11	C
Loop2 Loop11 TM	C
Loop2 TM	C
Loop8 Loop11	C(B) C(E)
Loop8 Loop11 TM	C
Loop8 Loop2	C(B) C(E)
Loop8 Loop2 Loop11	**C**
Loop8 Loop2 Loop11 TM5 TMLoop3 TMLoop5	C(B)
Loop8 Loop2 TM	C
Loop8 TM	C
Loop8 Loop2 Loop11 TM	**C**
Loop8 Loop2 Loop11 TM5	**C(E)**

Table 1: **Clustering selected subdomains of AcrB under RMSD$_{Comb.}$ versus the A, B, E states: Loop2, Loop8, Loop11, TM yield compatible clusterings.** A clustering yielding three clusters corresponding to the A/B/E states is plainly labeled C for Correct – X for incorrect; a clustering segregating state S only is denoted C(S). Note that TM5, TMLoop3 and TMLoop5 are Transmembrane subdomains, while the other subdomains are in the porter domain. Lines in bold are commented in the main text.
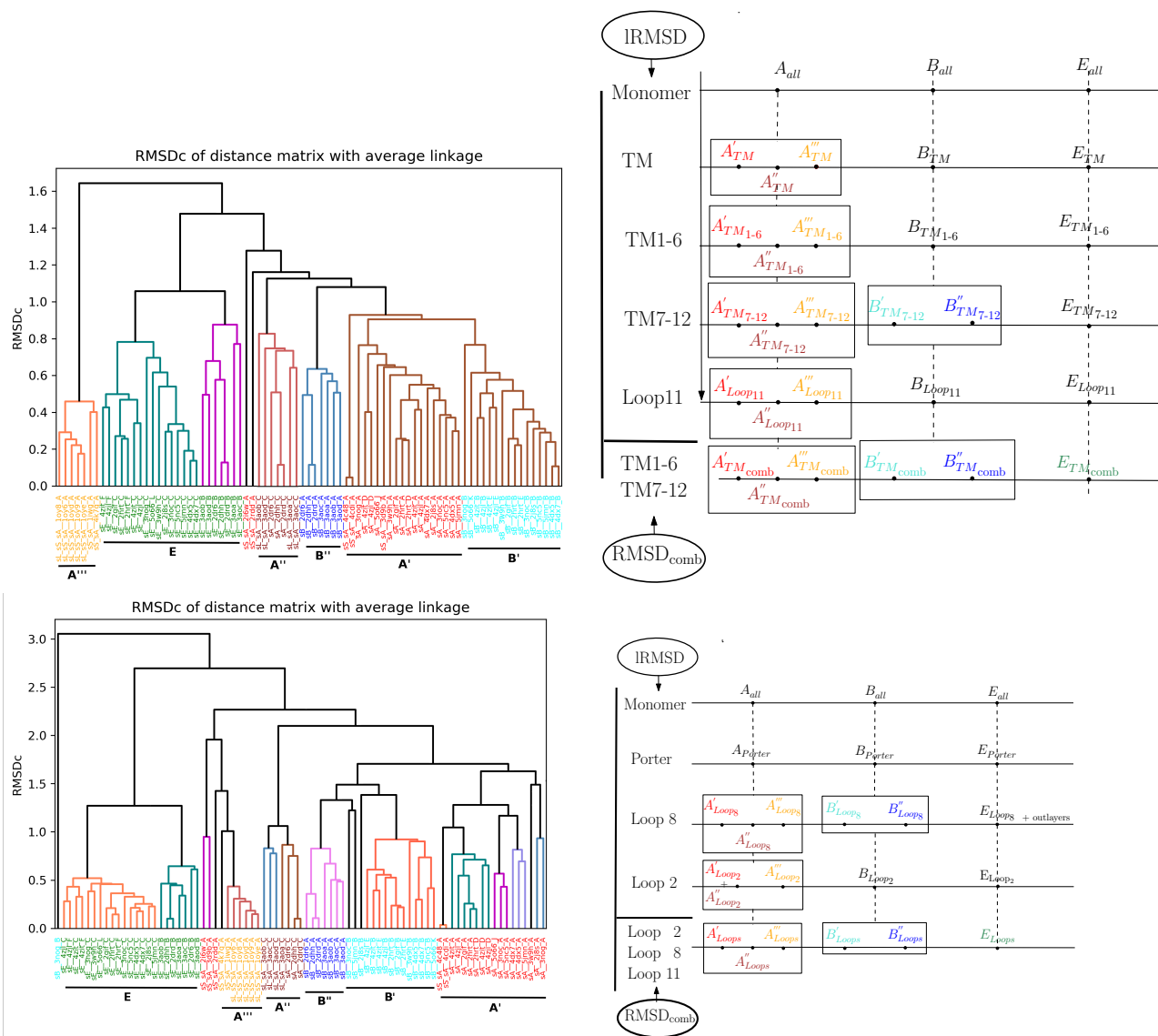
Figure 4: **AcrB: states A, B and E and the associated six substates for all (symmetric + non-symmetric) structures. (Left panels: dendograms) (Right panels: substates and subdomains)** Schematic representation of clusters from the dendogram, with columns for substates and lines for clustering conditions. **(Top row)** Using TM and selected subdomains. **(Bottom row)** Using porter and selected subdomains. Since the groups of structures obtained from both analysis are identical, we drop the dependency to the subdomains and simply refer to the five substates as $A^{'}, A^{''}, A^{'''}$ and $B^{'}, B^{''}$. See Section 3.3 and SI Figs. 10 and 11.
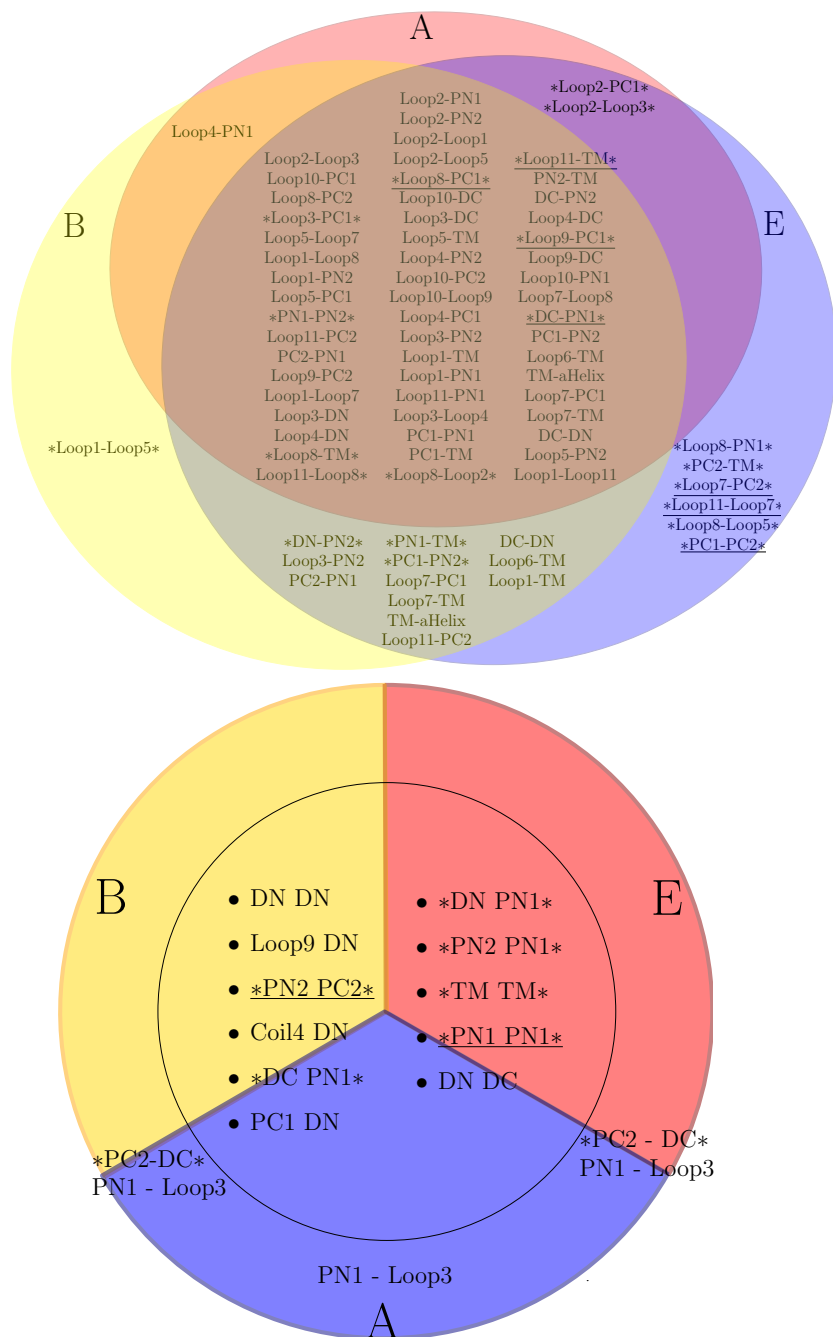
Figure 5: **AcrB: Evolution of interfaces between subdomains within a monomer and across monomers. (Top: interfaces within a monomer.)** Conventions used to tag an interface uses two criteria: *interface*: criterion from Eq. (2); <u>interface</u>: interface size differs by at least 30 atoms. **(Bottom**: Interfaces between two monomers.) Interfaces listed within the inner circle are found in the three pairs of interfaces i.e. AB, AE and BE. Interfaces listed along a radius are found in the two states apart from that radius. The interface PN1-Loop3 pertains to symmetric structures only. Tagging an interface name uses two criteria: *interface*: criterion from Eq. (4) <u>interface</u>: interface sizes differ by at least 30 atoms across two conditions.
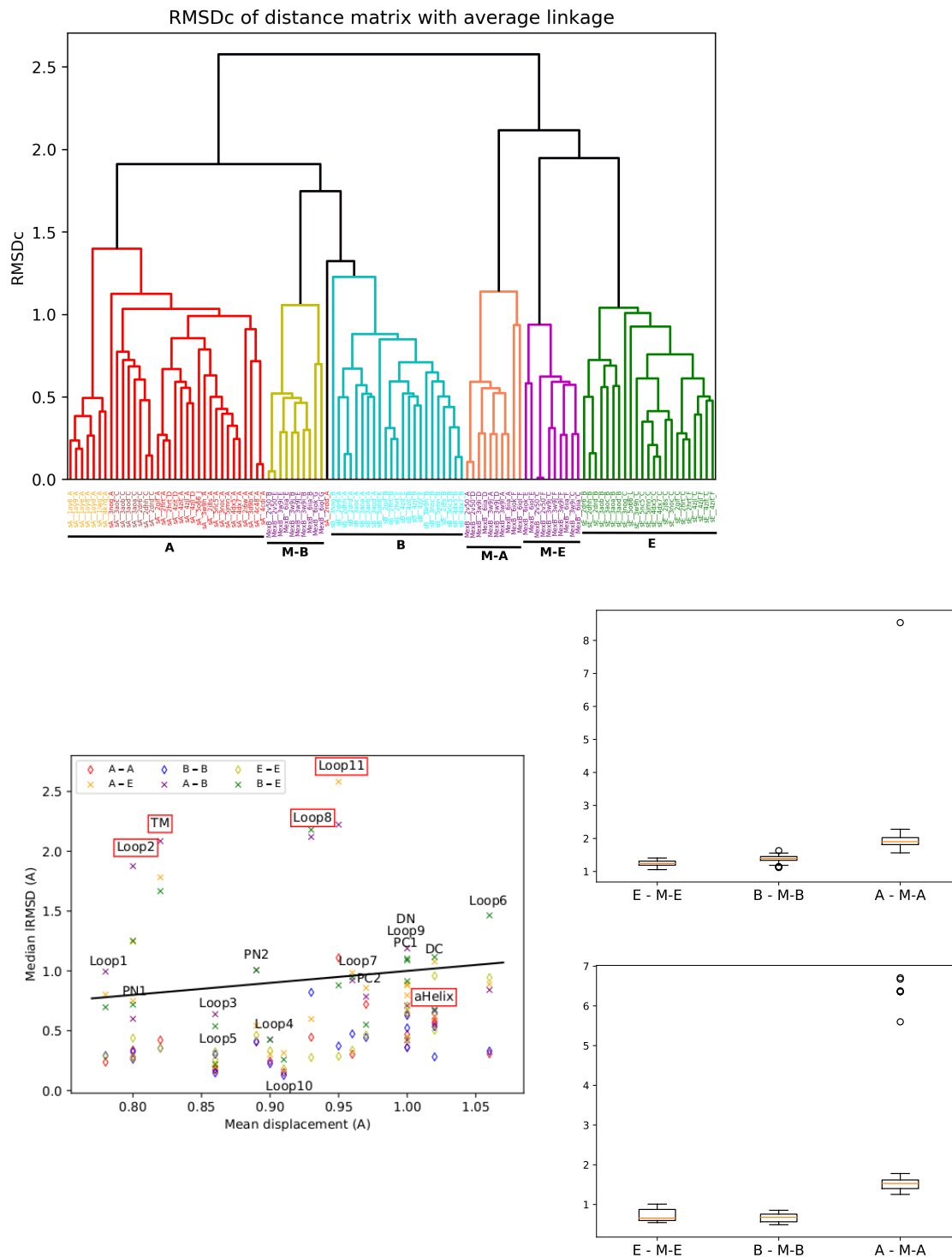
Figure 6: **Analysis of MexB's states. (First row) Clustering with RMSD$_{\text{Comb.}}$ for AcrB and MexB.** RMSD was computed on domains (TM, porter, funnel) **(Second row, left) Characterization of dynamic subdomains of MexB.** Mean square displacement in crystal asymmetric structures versus median lRMSD. **(Second row, right) States of AcrB versus states of MexB.** On a per state basis, boxplot of lRMSD values (Å) for pairs of domains from AcrB and MexB. Top: Porter Domain; bottom: TM Domain.

PC2

PC1

- recognition of substrates at affinity site
- loose state with least internal interfaces between subdomains
- 3 identified substates adopted by subdomains in a monomer

A'

A

A" A"

- TM : small global change (deprotonation)
- Loop 2
- Loop 11: helix to coil
  ↳ reopening of the entrance
- Loop 8

Decompression of PC1 and PC2 (-> reopening of drug binding pocket and its entrance)

- Loop 2 dynamic (++)
  ↳ no interface, thus mobile subdomain
- Loop 11 start to change to helix
- TM : small global changes

$-drug$
$-H^+$

$+drug$

- $B \rightarrow E$ due to protonation
- increase of interfaces between main porter domain (such as PC1 and PC2)
- closure of entrance of drug binding pocket (DBP) and zip-like closure of DBP

E

$+H^+$

B'

B" B

- $A \rightarrow B$ due to drug accommodation
- Reduction of interfaces between main porter subdomains
- Opening of drug binding pocket (DBP)
- 2 identified substates adopted by subdomains

- TM: small global change (protonation)
- Loop2
- Loop 11: helix -> closure of the entrance
- Loop 8
⇒ Compression of PC1 and PC2 triggers the closure of drug binding pocket
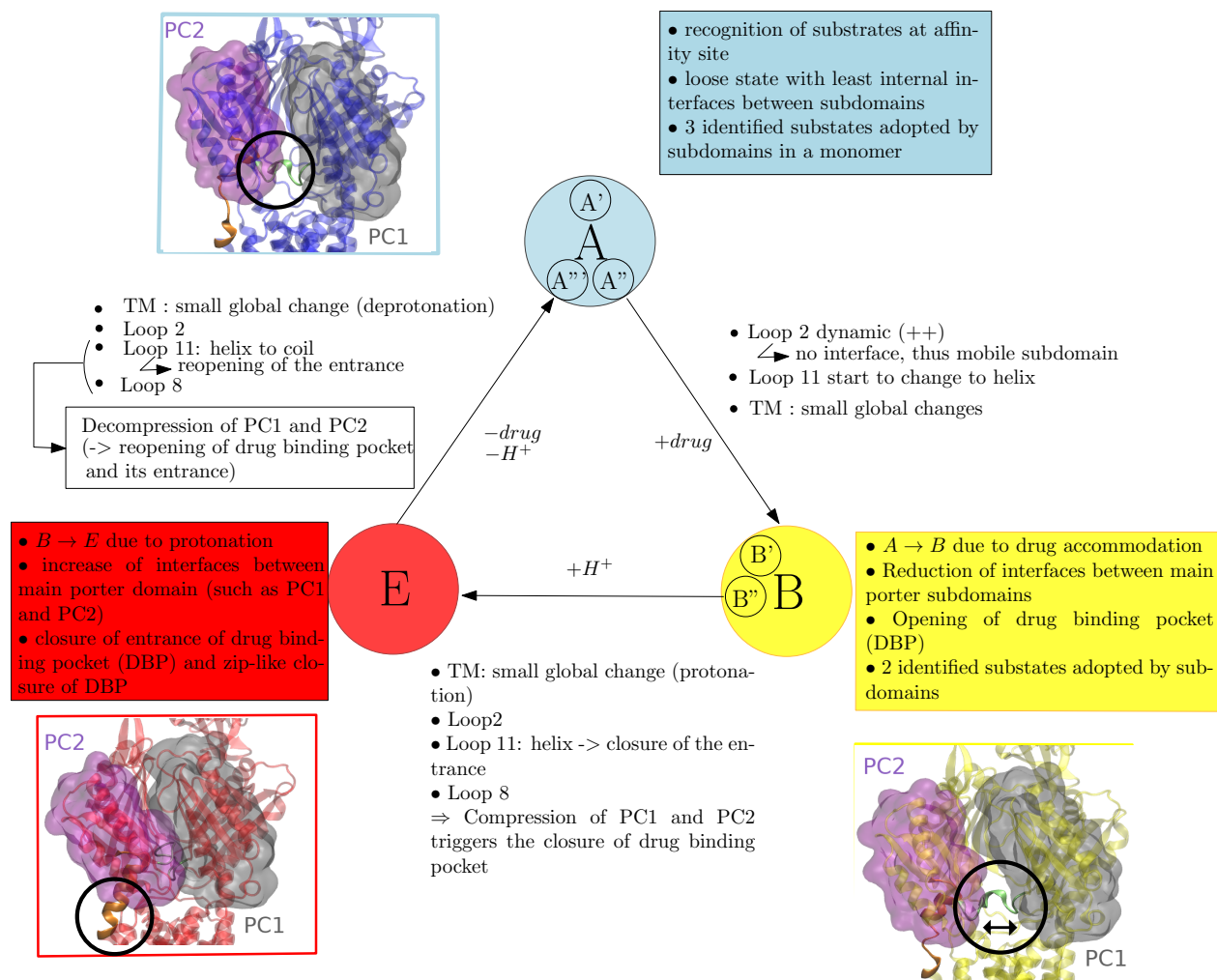
PC2

PC1

PC2

PC1

Figure 7: **Positioning of the main events highlighted in this work with respect to the (A,B,E) cycle**. Our analysis singles out two types of events (stressed by circles in the insets): the dynamic behavior of a subdomain (within a state, in-between two states), and the evolution of interfaces between two subdomains (within a monomer, across monomers). Dynamic subdomains between each state are in bullet list. Deduced effect on structure are written next to the subdomains. Loop11 undergoes a coil to helix conformational change, which contribute to the closing of the entrance to the drug binding pocket. Illustrations: relative movement of PC1 and PC2 due to Loop8. and definition of the interface $I_1(PC1, PC2)$. Starting from state A, both subdomains get further apart in state B before colliding in E state. The median values for the number of interface atoms are: state A: 0, state B: 0, state E: 49. Loop11 is represented in orange and Loop2 in green.

References

[1] A. Yamaguchi, R. Nakashima, and K. Sakurai. Structural basis of RND-type multidrug exporters. *Frontiers in microbiology*, 6:327, 2015.

[2] Satoshi Murakami, Ryosuke Nakashima, Eiki Yamashita, and Akihito Yamaguchi. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*, 419(6907):587–593, 2002.

[3] Satoshi Murakami, Ryosuke Nakashima, Eiki Yamashita, Takashi Matsumoto, and Akihito Yamaguchi. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature*, 443(7108):173, 2006.

[4] M. Seeger, A. Schiefner, T. Eicher, F. Verrey, K. Diederichs, and K. Pos. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science*, 313(5791):1295–1298, 2006.

[5] M. Seeger, K. Diederichs, T. Eicher, L. Brandstatter, A. Schiefner, F. Verrey, and K. Pos. The AcrB efflux pump: conformational cycling and peristalsis lead to multidrug resistance. *Current drug targets*, 9(9):729–749, 2008.

[6] K. Pos. Drug transport mechanism of the acrb efflux pump. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1794(5):782–793, 2009.

[7] H. Mishima, H. Oshima, S. Yasuda, and M. Kinoshita. Statistical thermodynamics for functionally rotating mechanism of the multidrug efflux transporter AcrB. *The Journal of Physical Chemistry B*, 119(8):3423–3433, 2015.

[8] X. Zhang, Min M. Liu, and L. Han. Energy coupling mechanisms of AcrB-like RND transporters. *Biophysics reports*, 3(4-6):73–84, 2017.

[9] Z. Wang, G. Fan, C. Hryc, J. Blaza, I. Serysheva, M.Schmid, W. Chiu, B. Luisi, and D. Du. An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *eLife*, 6, 2017.

[10] R. Nakashima, K. Sakurai, S. Yamasaki, K. Nishino, and A. Yamaguchi. Structures of the multidrug exporter AcrB reveal a proximal multisite drug-binding pocket. *Nature*, 480(7378):565, 2011.

[11] Xin-Qiu Yao, Hiroo Kenzaki, Satoshi Murakami, and Shoji Takada. Drug export and allosteric coupling in a multidrug transporter revealed by molecular simulations. *Nature communications*, 1:117, 2010.

[12] Attilio V Vargiu and Hiroshi Nikaido. Multidrug binding properties of the acrb efflux pump characterized by molecular dynamics simulations. *PNAS*, 109(50):20637–20642, 2012.

[13] Venkata Krishnan Ramaswamy, Attilio V Vargiu, Giuliano Malloci, Jürg Dreier, and Paolo Ruggerone. Molecular rationale behind the differential substrate specificity of bacterial rnd multi-drug transporters. *Scientific reports*, 7(1):8075, 2017.

[14] A. Vargiu, V. Ramaswamy, G. Malloci, I. Malvacio, A. Atzori, and P. Ruggerone. Computer simulations of the activity of rnd efflux pumps. *Research in microbiology*, 169(7-8):384–392, 2018.

[15] X. Zhang, Y. Zhao, J. Heng, and D. Jiang. Energy coupling mechanisms of MFS transporters. *Protein Science*, 24(10):1560–1579, 2015.

[16] S. Dey, P. Chakrabarti, and J. Janin. A survey of hemoglobin quaternary structures. *Proteins: Structure, Function, and Bioinformatics*, 79(10):2861–2870, 2011.

[17] F. Richards. The interpretation of protein structures: total volume, group volume distributions and packing density. *Journal of molecular biology*, 82(1):1–14, 1974.

[18] C. Chothia and J. Janin. Principles of protein-protein recognition. *Nature*, 256:705–708, 1975.

[19] J. Janin, R. P. Bahadur, and P. Chakrabarti. Protein-protein interaction and quaternary structure. *Quarterly reviews of biophysics*, 41(2):133–180, 2008.

[20] S. Marillet, P. Boudinot, and F. Cazals. High resolution crystal structures leverage protein binding affinity predictions. *Proteins: structure, function, and bioinformatics*, 1(84):9–20, 2015.

[21] Feng Long, Chih-Chia Su, Michael T Zimmermann, Scott E Boyken, Kanagalaghatta R Rajashankar, Robert L Jernigan, and W Yu Edward. Crystal structures of the cusa efflux pump suggest methionine-mediated metal transport. *Nature*, 467(7314):484–488, 2010.

[22] R. Nakashima, K. Sakurai, S. Yamasaki, K. Hayashi, C. Nagata, K. Hoshino, Y. Onodera, K. Nishino, and A. Yamaguchi. Structural basis for the inhibition of bacterial multidrug exporters. *Nature*, 500(7460):102, 2013.

[23] F. Cazals and T. Dreyfus. The Structural Bioinformatics Library: modeling in biomolecular science and beyond. *Bioinformatics*, 7(33):1–8, 2017.

[24] B.A. Chapman and J.T. Chang. Biopython: Python tools for computational biology. *ACM SIGBIO Newsletter*, pages 15–19, 2000.

[25] F. Cazals and R. Tetley. Characterizing molecular flexibility by combining lRMSD measures. *Proteins: structure, function, and bioinformatics*, 87(5):380–389, 2019.

[26] F. Murtagh. A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26(4):354–359, 1983.

[27] DWJ. Cruickshank. Remarks about protein structure precision. *Acta Crystallographica Section D: Biological Crystallography*, 55(3):583–601, 1999.

[28] D. Blow. *Outline of crystallography for biologists*. Oxford University Press on Demand, 2002.

[29] A. Kuzmanic, N. Pannu, and B. Zagrovic. X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nature communications*, 5:3220, 2014.

[30] O. Carugo. How large b-factors can be in protein crystal structures. *BMC bioinformatics*, 19(1):61, 2018.

[31] F. Cazals, F. Proust, R. Bahadur, and J. Janin. Revisiting the Voronoi description of protein-protein interfaces. *Protein Science*, 15(9):2082–2092, 2006.

[32] F. Cazals, H. Kanhere, and S. Loriot. Computing the volume of union of balls: a certified algorithm. *ACM Transactions on Mathematical Software*, 38(1):1–20, 2011.

[33] D. Ritchie, A. Ghoorah, L. Mavridis, and V. Venkatraman. Fast protein structure alignment using Gaussian overlap scoring of backbone peptide fragment similarity. *Bioinformatics*, 28(24):3274–3281, 2012.

[34] Gaby Sennhauser, Magdalena A Bukowska, Christophe Briand, and Markus G Grütter. Crystal structure of the multidrug exporter mexb from pseudomonas aeruginosa. *Journal of molecular biology*, 389(1):134–145, 2009.

[35] M. Seeger, C. Von Ballmoos, T. Eicher, L. Brandstätter, F. Verrey, K. Diederichs, and K. Pos. Engineered disulfide bonds support the functional rotation mechanism of multidrug efflux pump AcrB. *Nature structural & molecular biology*, 15(2):199–205, 2008.

[36] Martijn Zwama, Seiji Yamasaki, Ryosuke Nakashima, Keisuke Sakurai, Kunihiko Nishino, and Akihito Yamaguchi. Multiple entry pathways within the efflux transporter acrb contribute to multidrug recognition. *Nature communications*, 9(1):1–9, 2018.

[37] Shi Chen, Rafal P. Wiewiora, Fanwang Meng, Nicolas Babault, Anqi Ma, Wenyu Yu, Kun Qian, Hao Hu, Hua Zou, Junyi Wang, Shijie Fan, Gil Blum, Fabio Pittella-Silva, Kyle A. Beauchamp, Wolfram Tempel, Hualiang Jiang, Kaixian Chen, Robert Skene, Y. George Zheng, Peter J. Brown, Jian Jin, Cheng Luo, John D. Chodera, and Minkui Luo. The dynamic conformational landscapes of the protein methyltransferase setd8. *bioRxiv*, 2018.

[38] T. Eicher, H-J Cha, M.A Seeger, L. Brandstätter, J. El-Delik, J.A Bohnert, W.V Kern, F. Verrey, M.G Grütter, K. Diederichs, et al. Transport of drugs by the multidrug transporter AcrB involves an access and a deep binding pocket that are separated by a switch-loop. *Proceedings of the National Academy of Sciences*, 109(15):5687–5692, 2012.

[39] T. Yamane, S. Murakami, and M. Ikeguchi. Functional rotation induced by alternating protonation states in the multidrug transporter AcrB: all-atom molecular dynamics simulations. *Biochemistry*, 52(43):7648–7658, 2013.

[40] Thomas Eicher, Markus A Seeger, Claudio Anselmi, Wenchang Zhou, Lorenz Brandstätter, François Verrey, Kay Diederichs, José D Faraldo-Gómez, and Klaas M Pos. Coupling of remote alternating-access transport mechanisms for protons and substrates in the multidrug efflux pump acrb. *Elife*, 3:e03145, 2014.

[41] S.A. Adcock and A.J. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews*, 106(5):1589–1615, 2006.

[42] Zhiwei Feng, Tingjun Hou, and Youyong Li. Unidirectional peristaltic movement in multisite drug binding pockets of acrb from molecular dynamics simulations. *Molecular BioSystems*, 8(10):2699–2709, 2012.

[43] Dijun Du, Zhao Wang, Nathan R James, Jarrod E Voss, Ewa Klimont, Thelma Ohene-Agyei, Henrietta Venter, Wah Chiu, and Ben F Luisi. Structure of the acrab–tolc multidrug efflux pump. *Nature*, 509(7501):512–515, 2014.

[44] Heng-Keat Tam, Viveka N Malviya, Wuen-Ee Foong, Andrea Herrmann, Giuliano Malloci, Paolo Ruggerone, Attilio V Vargiu, and Klaas M Pos. Binding and transport of carboxylated drugs by the multidrug transporter acrb. *Journal of Molecular Biology*, 432(4):861–877, 2020.

[45] Nadine Fischer and Christian Kandt. Porter domain opening and closing motions in the multi-drug efflux transporter AcrB. *Biochimica Et Biophysica Acta (BBA)-Biomembranes*, 1828(2):632–641, 2013.

[46] Martijn Zwama, Katsuhiko Hayashi, Keisuke Sakurai, Ryosuke Nakashima, Kimie Kitagawa, Kunihiko Nishino, and Akihito Yamaguchi. Hoisting-loop in bacterial multidrug exporter AcrB is a highly flexible hinge that enables the large motion of the subdomains. *Frontiers in microbiology*, 8:2095, 2017.

[47] N. Fischer and C. Kandt. Three ways in, one way out: Water dynamics in the trans-membrane domains of the inner membrane translocase AcrB. *Proteins: Structure, Function, and Bioinformatics*, 79(10):2871–2885, 2011.

[48] Yead Jewel, Quyen Van Dinh, Jin Liu, and Prashanta Dutta. Substrate-dependent transport mechanism in acrb of multidrug resistant bacteria. *Proteins: Structure, Function, and Bioinformatics*, 88(7):853–864, 2020.

[49] Jessica Kobylka, Miriam S Kuth, Reinke T Müller, Eric R Geertsma, and Klaas M Pos. Acrb: a mean, keen, drug efflux machine. *Annals of the New York Academy of Sciences*, 1459(1):38–68, 2020.

[50] Christian S. Seifert and F. Gräter. Force distribution reveals signal transduction in e. coli hsp90. *Biophysical Journal*, 103(10):2195–2202, 2012.

[51] A. Cheng, R. Coleman, K. Smyth, Q. Cao, P. Soulard, D. Caffrey, A. Salzberg, and E. Huang. Structure-based maximal affinity model predicts small-molecule druggability. *Nature biotechnology*, 25(1):71, 2007.

[52] Keisuke Sakurai, Seiji Yamasaki, Kaori Nakao, Kunihiko Nishino, Akihito Yamaguchi, and Ryosuke Nakashima. Crystal structures of multidrug efflux pump mexb bound with high-molecular-mass compounds. *Scientific reports*, 9(1):1–9, 2019.

[53] Kenta Tsutsumi, Ryo Yonehara, Etsuko Ishizaka-Ikeda, Naoyuki Miyazaki, Shintaro Maeda, Kenji Iwasaki, Atsushi Nakagawa, and Eiki Yamashita. Structures of the wild-type mexab–oprm tripartite pump reveal its complex formation and drug efflux mechanism. *Nature communications*, 10(1):1–10, 2019.

[54] Yead Jewel, Jin Liu, and Prashanta Dutta. Coarse-grained simulations of conformational changes in the multidrug efflux transporter AcrB. *Molecular BioSystems*, 13(10):2006–2014, 2017.

[55] S. Loriot and F. Cazals. Modeling macro-molecular interfaces with Intervor. *Bioinformatics*, 26(7):964–965, 2010.

[56] G. Sennhauser, P. Amstutz, C. Briand, O. Storchenegger, and M.G Grütter. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS biology*, 5(1):e7, 2006.

[57] N. Monroe, G. Sennhauser, M.A Seeger, C. Briand, and M.G Grütter. Designed ankyrin repeat protein binders for the crystallization of AcrB: plasticity of the dominant interface. *Journal of structural biology*, 174(2):269–281, 2011.

[58] A. Ababou and V. Koronakis. Structures of gate loop variants of the AcrB drug efflux pump bound by erythromycin substrate. *PloS one*, 11(7):e0159154, 2016.

[59] C. Oswald, H.-K Tam, and K.M Pos. Transport of lipophilic carboxylates is mediated by transmembrane helix 2 in multidrug transporter AcrB. *Nature communications*, 7:13819, 2016.

[60] W Yu Edward, Gerry McDermott, Helen I Zgurskaya, Hiroshi Nikaido, and Daniel E Koshland. Structural basis of multiple drug-binding capacity of the acrb multidrug efflux pump. *Science*, 300(5621):976–980, 2003.

[61] Li-Wei Hung, Heung-Bok Kim, Satoshi Murakami, Goutam Gupta, Chang-Yub Kim, and Thomas C Terwilliger. Crystal structure of acrb complexed with linezolid at 3.5 å resolution. *Journal of structural and functional genomics*, 14(2):71–75, 2013.

[62] Susanna Törnroth-Horsefield, Pontus Gourdon, Rob Horsefield, Lars Brive, Natsuko Yamamoto, Hirotada Mori, Arjan Snijder, and Richard Neutze. Crystal structure of acrb in complex with a single transmembrane subunit reveals another twist. *Structure*, 15(12):1663–1673, 2007.

[63] Debanu Das, Qian Steven Xu, Jonas Y Lee, Irina Ankoudinova, Candice Huang, Yun Lou, Andy DeGiovanni, Rosalind Kim, and Sung-Hou Kim. Crystal structure of the multidrug efflux transporter acrb at 3.1 å resolution reveals the n-terminal region with conserved amino acids. *Journal of structural biology*, 158(3):494–502, 2007.

[64] David Veesler, Stéphanie Blangy, Christian Cambillau, and Giuliano Sciara. There is a baby in the bath water: Acrb contamination is a major problem in membrane-protein crystallization. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 64(10):880–885, 2008.

[65] Chih-Chia Su, Feng Long, Michael T Zimmermann, Kanagalaghatta R Rajashankar, Robert L Jernigan, and W Yu Edward. Crystal structure of the cusba heavy-metal efflux complex of escherichia coli. *Nature*, 470(7335):558–562, 2011.

[66] Chih-Chia Su, Feng Long, Hsiang-Ting Lei, Jani Reddy Bolla, Sylvia V Do, Kanagalaghatta R Rajashankar, and W Yu Edward. Charged amino acids (r83, e567, d617, e625, r669, and k678) of cusa are required for metal ion transport in the cus efflux system. *Journal of molecular biology*, 422(3):429–441, 2012.

7 Supporting information

7.1 Software tools

Tools from the SBL**.** The following software tools from the Structural Bioinformatics Library (http://sbl.inria.fr, [23]) were used:

- Molecular distances measures (lRMSD , RMSD$_{Comb.}$) [25]: https://sbl.inria.fr/doc/Molecular_distances_flexible-user-manual.html
  Executables: sbl-rmsd-flexible-conformations.exe and sbl-rmsd-flexible-proteins-kpax.exe

- Voronoi interfaces [55, 23]: https://sbl.inria.fr/doc/Space_filling_model_interface_finder-user-manual.html and https://sbl.inria.fr/doc/Space_filling_model_interface-user-manual.html.
  Executable: sbl-bif-domainsW-atomic.exe

Other tools. Handling PDB files was done using Biopython PDB, see [24] and https://biopython.org/wiki/The_Biopython_Structural_Bioinformatics_FAQ The Scipy packages hierarchical clustering (https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html) and spatial distance (https://docs.scipy.org/doc/scipy/reference/spatial.distance.html) were also used.

Complete workflow. The workflow used in this paper is available in the Molecular_cradle package of the Structural Bioinformatics Library (http://sbl.inria.fr), as detailed in the following user manual https://sbl.inria.fr/doc/Molecular_cradle-user-manual.html.

7.2 Structure: overview

See Table 2 and Figure 8

7.3 Structural data used

See Figure 9 and Table 3.

7.4 Results – Classifying states shows differences across RND

See Table 4 and Table 5.

7.5 Results – Steps A, B and E are mainly due to relative motions between selected subdomains

See Table 6 and Table 7.

RND	Domains	Subdomains	Amino acids	Size	Key residues reported in [14]
		TM	1-28 331-494 540-558 876-1044	380	I27, K334, I337, H338, V341, D407, D408, K940
		TM1-6	1-28 331-494	191	
	TM	TM7-12	540-558 876-1044	189	
		Loop6	495-502	8	None
		α-Helix	503-539	37	None
		PN1	43-131	89	S46, S79, T91, Q89, Q124, Q125, S128, E130)
		Loop1	29-42	14	None
		Loop2	132-137	6	S134, S135, F136
		PN2	138-177 279-325	87	Q176, L177, G290, K292
		Loop3	178-182	5	F178, S180
		Loop4	272-278	7	E273, N274, D276, I277
AcrB		Loop5	326-330	5	Y327
	Porter	Loop7	559-572	14	D566
		PC1	573-668	96	M573, M575, Q577, F610, V612, F615, F617, R620, T624, F628, M662, F664, F666, N667, L668
		Loop8	669-679	11	P669, V672, E673, L674, T676
		PC2	680-717 816-859	82	D681, R717, E826, L828, S836, E842
		Loop9	718-725	8	N719
		Loop10	812-815	4	None
		Loop11	860-875	16	L868, Q872
	Funnel	DN	183-271	89	None
		DC	726-811	86	Y758
		TM	1-28 330-494 539-556 876-1044	380	
	TM	Loop6	495-509	15	
		α-Helix	510-538	29	
		PN1	43-130	88	
		Loop1	29-42	14	
		Loop2	131-137	7	
		PN2	138-177 279-325	87	
		Loop3	178-182	5	
		Loop4	273-278	6	
MexB		Loop5	325-329	5	
	Porter	Loop7	557-572	16	
		PC1	573-668	96	
		Loop8	667-679	13	
		PC2	680-717 816-859	82	
		Loop9	718-725	8	
		Loop10	812-815	4	
		Loop11	860-875	16	
	Funnel	DN	183-272	89	
		DC	726-811	86	
		TM	1-33 328-503 537-555 871-1044	402	
CusA		Loop2	121-137	17	
		Loop11	860-875	15	

Table 2: **Description of used subdomains.** The subdomains in bold are those responsible for conformational changes.
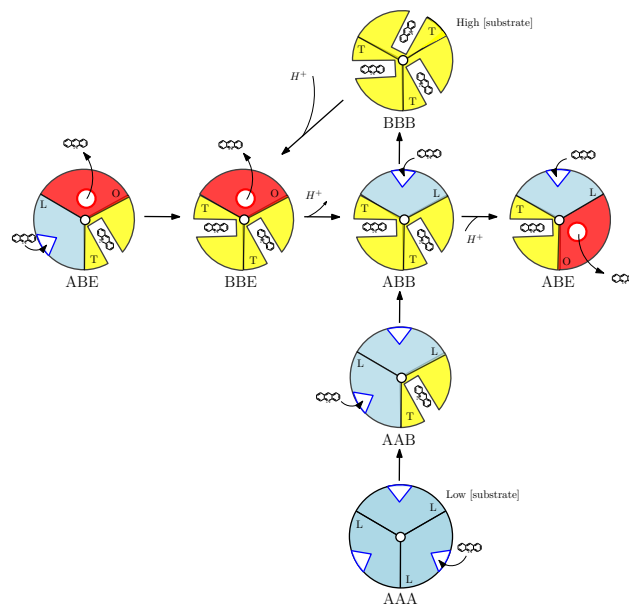
Figure 8: **Mechanism for the trimer** (Adapted from [5]) Schematic representation of the AcrB alternating site functional rotation transport mechanism extended by postulated intermediate steps. The conformational states loose (L), tight (T), and open (O) are colored blue, yellow and red, respectively. The lateral grooves in the L and T monomer indicate the substrate binding sites. The different geometric forms reflect low (triangle), high (rectangle), or no (circle) binding affinity for the transported substrates. Both states LTO at the far left and far right are identical.

Protein	Symmetry	article	pdb	Xtal or EM	Wild or mutate	State	Resolution (Å)
AcrB	Asymmetric	[3]	2dhh	xtal	wild	ABE	2.8
			2drd	xtal	wild	ABE	3.1
			2dr6	xtal	wild	ABE	3.3
		[4]	2gif	xtal	wild	ABE	2.9
			2hrt	xtal	wild	ABE (x2)	3
		[56]	2j8s	xtal	wild	ABE	2.5
		[57]	3noc	xtal	wild	ABE	2.7
			3nog	xtal	wild	ABE	3.34
		[10]	3aoa	xtal	wild	ABE	3.35
			3aob	xtal	wild	ABE	3.35
			3aoc	xtal	wild	ABE	3.35
			3aod	xtal	wild	ABE	3.3
		[38]	4dx5	xtal	wild	ABE	1.9
			4dx7	xtal	wild	ABE	2.25
		[22]	3w9h	xtal	wild	ABE	3.05
		[58]	4zit	xtal	wild	ABE (x2)	3.296
			4zjl	xtal	wild	ABE (x2)	3.47
		[59]	5jmn	xtal	wild	ABE	2.5
		[9]	5o66	EM	wild	ABE	5.9
			5nc5	EM	wild	ABE	3.2
	Symmetric	[60]	1oy6	xtal	wild	A	3.68
			1oy8	xtal	wild	A	3.63
			1oy9	xtal	wild	A	3.8
			1oyd	xtal	wild	A	3.8
			1oye	xtal	wild	A	3.48
		[61]	4k7q	xtal	wild	A	3.5
		[2]	1iwg	xtal	wild	A	3.5
		[62]	2rdd	xtal	wild	A	3.5
		[63]	2i6w	xtal	wild	A	3.1
		[64]	3d9b	xtal	wild	A	3.42
		[43]	4cdi	xtal	wild	A	3.7
			4c48	xtal	wild	A	3.3
MexB	Asymmetric	[34]	2v50	xtal	wild	M-ABE	3
		[22]	3w9i	xtal	wild	M-ABE(x2)	2.79
			3w9j	xtal	wild	M-ABE(x2)	3.15
		[52]	6iia	xtal	wild	M-ABE	2.91
		[53]	6iol	EM	wild	M-ABE	3.76
			6iok	EM	wild	M-ABE	3.64
CusA	Symmetric	[21]	3k0i	xtal	wild	C-1	4.116
			3k07	xtal	wild	C-2	3.521
			3kso	xtal	wild	C-1	4.367
			3kss	xtal	wild	C-1	3.88
		[65]	3ne5	xtal	wild	C-2	2.898
		[66]	3t53	xtal	wild	C-2	3.37
			3t51	xtal	wild	C-2	3.9
			3t56	xtal	wild	C-2	3.42

Table 3: **Crystal structures used in this study.**

A monomers	B monomers	E monomers
2dhh C	2dhh A	2dhh B
2drd C	**2drd A**	**2drd B**
2dr6 C	**2dr6 A**	**2dr6 B**
3aoa C	**3aob A**	**3aob B**
3aod C	**3aod A**	**3aoc B**
3aob C	3aoa A	3aoa B
3aoc C	**3aoc A**	**3aod B**
4dx5 A	**2hrt E**	**2hrt F**
4dx7 A	**2hrt B**	**2hrt C**
3noc A	**2gif B**	**2gif C**
5nc5 A	4zit B	**4zjl F**
2j8s A	**4zjl B**	**4zjl C**
3w9h A	4zit E	4zit F
2hrt D	**4zjl E**	4zit C
2hrt A	**4dx5 B**	**2j8s C**
2gif A	**4dx7 B**	3noc C
4zjl D	**5nc5 B**	**4dx5 C**
4zjl A	**2j8s B**	**4dx7 C**
4zit A	5jmn B	5jmn C
4zit D	3noc B	5nc5 C
5o66 J	**3w9h B**	**3w9h C**
3nog A	5o66 K	5o66 L
5jmn A	3nog B	3nog C

Table 4: **Conformations of trimers in asymmetric structures.** The structures are those described in Table 3. Each entry features the pdbid followed by the label of a chain. Bold labels for chains indicate that the corresponding state (A, B, E) has been annotated by the authors of the crystal structure.

	average
lrmsd monomer	C
rmsdc monomer	C
rmsdc monomer withLoops	C
lrmsd funnel	X
lrmsd porter	C
lrmsd TM	C
rmsdc subdomain	C
rmsdc subdomain nodock	C
rmsdc subdomain withLoops	C

Table 5: **Clusterings attempted: correct (C) and erroneous (X).** See Def. 1. For cluster's correctness written like C(X), the cluster is able to segregate one state against the other two.
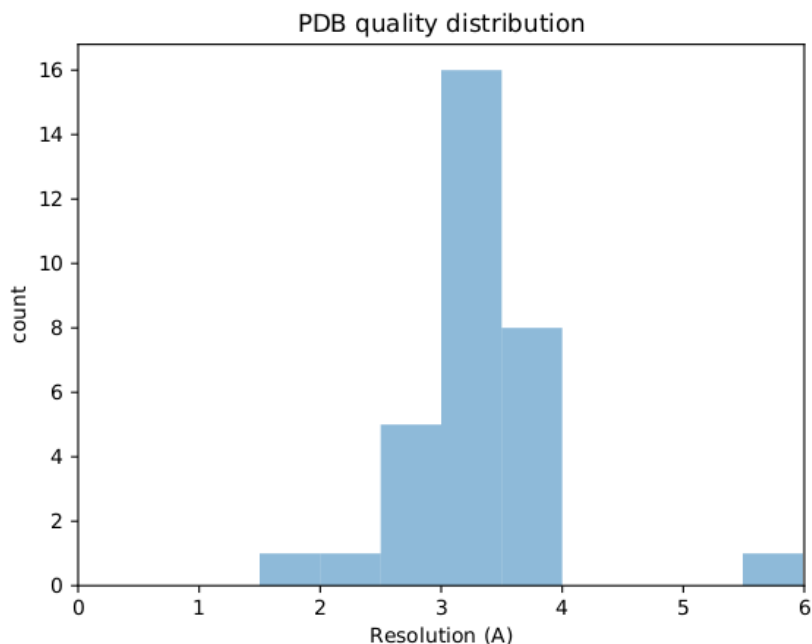
Figure 9: **Distribution of resolution of AcrB's pdb.** Median resolution is 3.32Å. See Table 3.

In the following, we replicate the analysis of section 3.2 for the subdomains of TM. More precisely, we split TM into 12 TM helices and 9 loops (SI Table 6), and study the dynamics of these subdomains (median lRMSD above their mean displacement just as before (Tab. 7)).

Using the same criterion as above (Def. 2), this analysis singles out three novel structural elements:

- TM5 (18 a.a.) : With 0 intra and 2 inter state positive comparisons, a dynamic subdomain.

- TMLoop3 (24 a.a.): With 0 intra and 1 inter state positive comparison, a dynamic subdomain.

Subdomains	Amino acids	Subdomains	Amino acids	Subdomains	Amino acids
TM1	10-28	TM8	872-888	TMLoop3	414-438
TM2	337-356	TM9	899-918	TMLoop4	458-465
TM3	366-385	TM10	925-943	TMLoop5	491-538
TM4	392-413	TM11	973-992	TMLoop6	889-898
TM5	439-457	TM12	999-1018	TMLoop7	919-924
TM6	466-490	TMLoop1	357-365	TMLoop8	944-972
TM7	539-555	TMLoop2	386-391	TMLoop9	993-998

Table 6: **Description of used subdomains for AcrB's split TM.**

Subdomain	Amino acids	Mean displacement (Å)	Median lRMSD (Å)	Compared state
TM5	439-457	1	1.204	State A - E
			1.188	State B - E
TMLoop3	414-438	1.08	1.217	A - E
TMLoop5	491-538	1.11	1.54	A - A
			1.633	A - B
			1.96	A - E
			1.337	B - B
			1.495	B - E
			1.447	E - E

Table 7: **Mean displacement in the crystal structures and median lRMSD of selected Transmembrane subdomains comparing different states.** Only the subdomains which median lRMSD is higher than the mean displacement are reported here.

- TMLoop5 (47 a.a. comprising the α-Helix): 3 intra and 3 inter state positive comparisons and no clear gap, this subdomain is unstable.

The eight subdomains just identified are used in the subsequent analysis.

We observe that TMLoop3 and TMLoop5 yield an incorrect clustering (Table 1). This is expected for TMLoop5, as we have seen that it does not exhibit any structural coherence event within states (all comparison inter - intra state above the line $y = x$, Fig. 2). For TMLoop3, we note that the dynamics are restricted to states A and E, and that the difference between the mean displacement and the median lRMSD is small (Table 7).

For TM5, we observe that clustering with Loop8 Loop2 and Loop11 yields a valid clustering for state E only (last line in Table 1) while clustering (Loop8 Loop2 and Loop11) plus whole TM yields a valid clustering (penultimate line, Table 1).

Remark 4 *Since F617-loop was identified as a key element in the efflux mechanism [42], and is localized in our PC1 subdomain, we tried to regenerate the cluster with this loop. Alas, the clusters were all erroneous. This can be explained because the lRMSD of this loop is below its mean displacement, which is probably because of its oscillatory movement [42].*
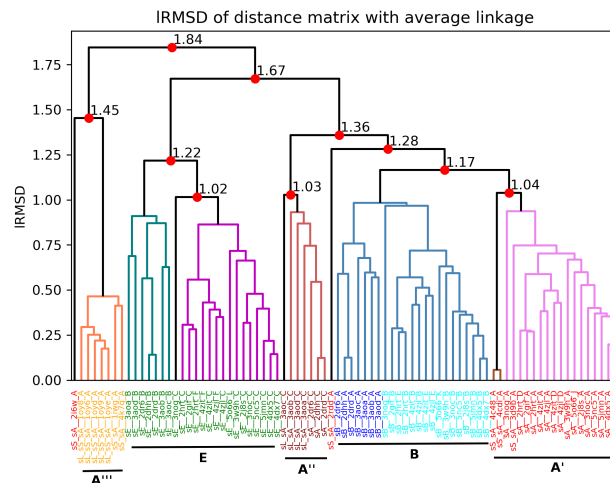
7.6 Results – Refining monomer states

See Figure 10, Figure 11 and Figure 12.

7.7 Results – The systematic study of interfaces unveils a small number of dynamic interfaces
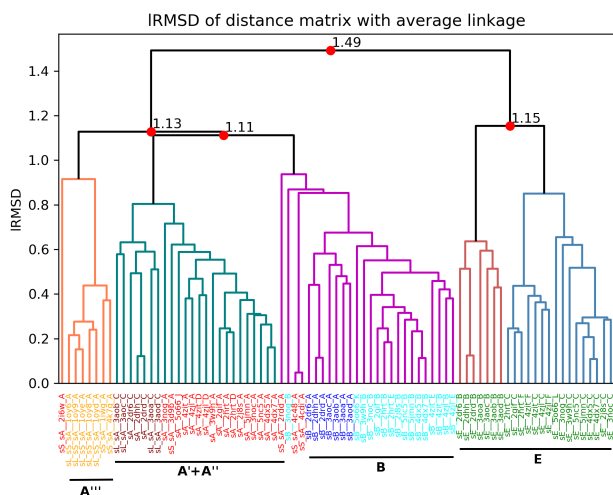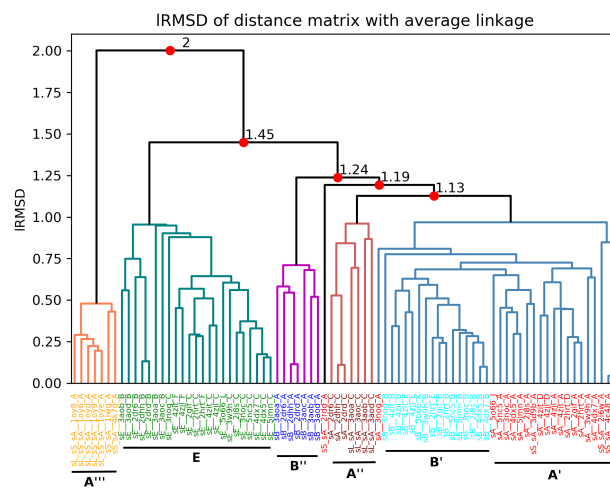
See Figure 13, Table 8 and Table 9.
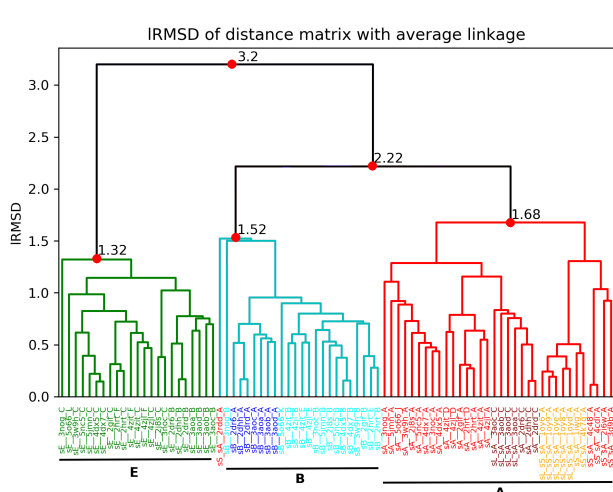
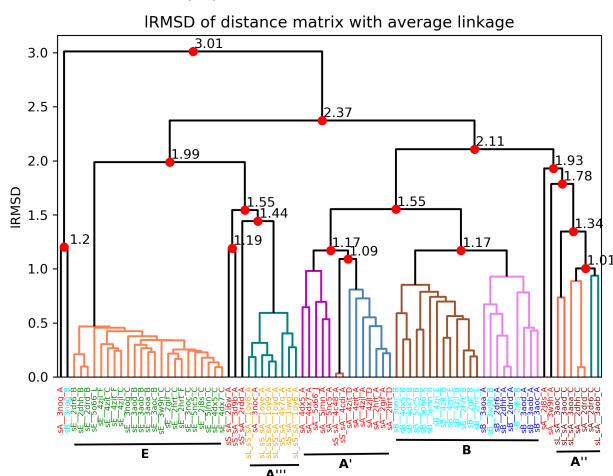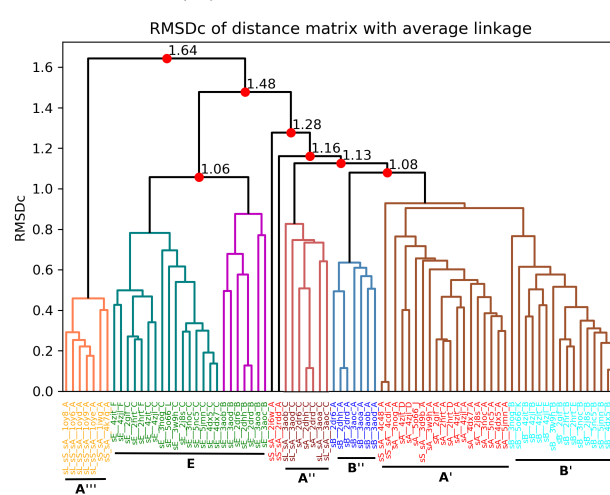Figure 10: **Clustering including symmetric structures: using whole monomers and subdomains of TM.**
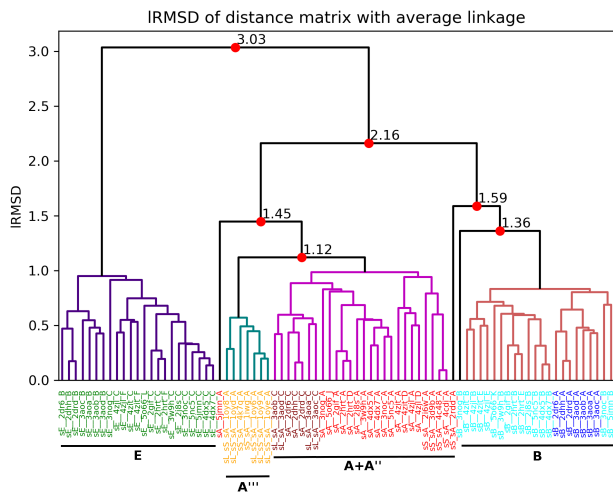
(A) Porter lRMSD
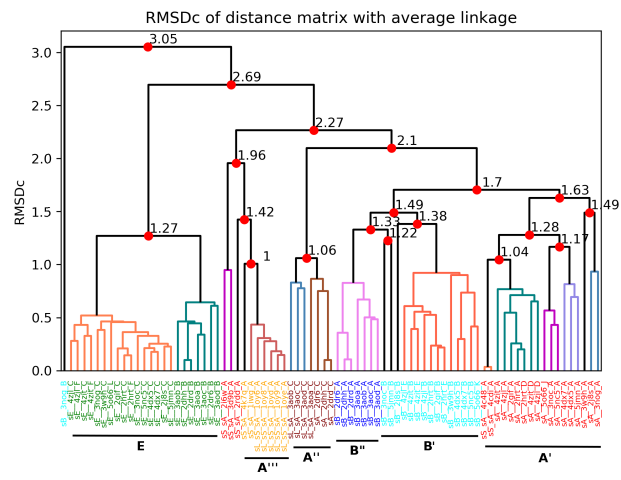
(B) Loop8 Loop2 Loop11 RMSD$_{\text{Comb.}}$

(C) Loop8 lRMSD
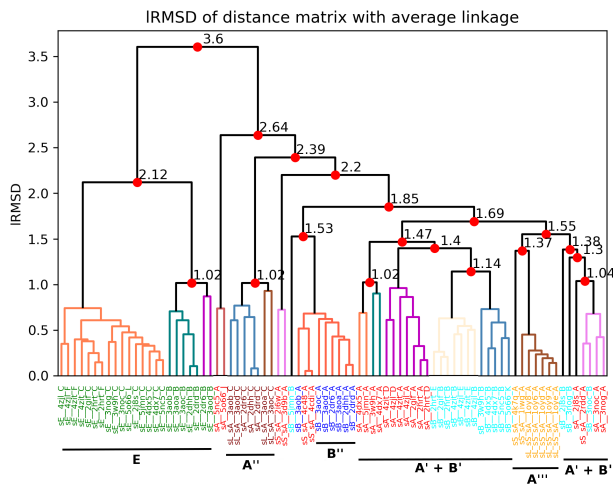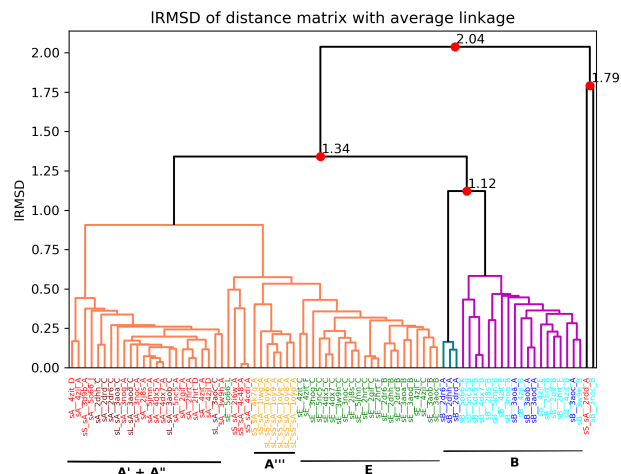
(D) Loop2 lRMSD

Figure 11: **Clustering including symmetric structures: using Porter and its subdomains.**
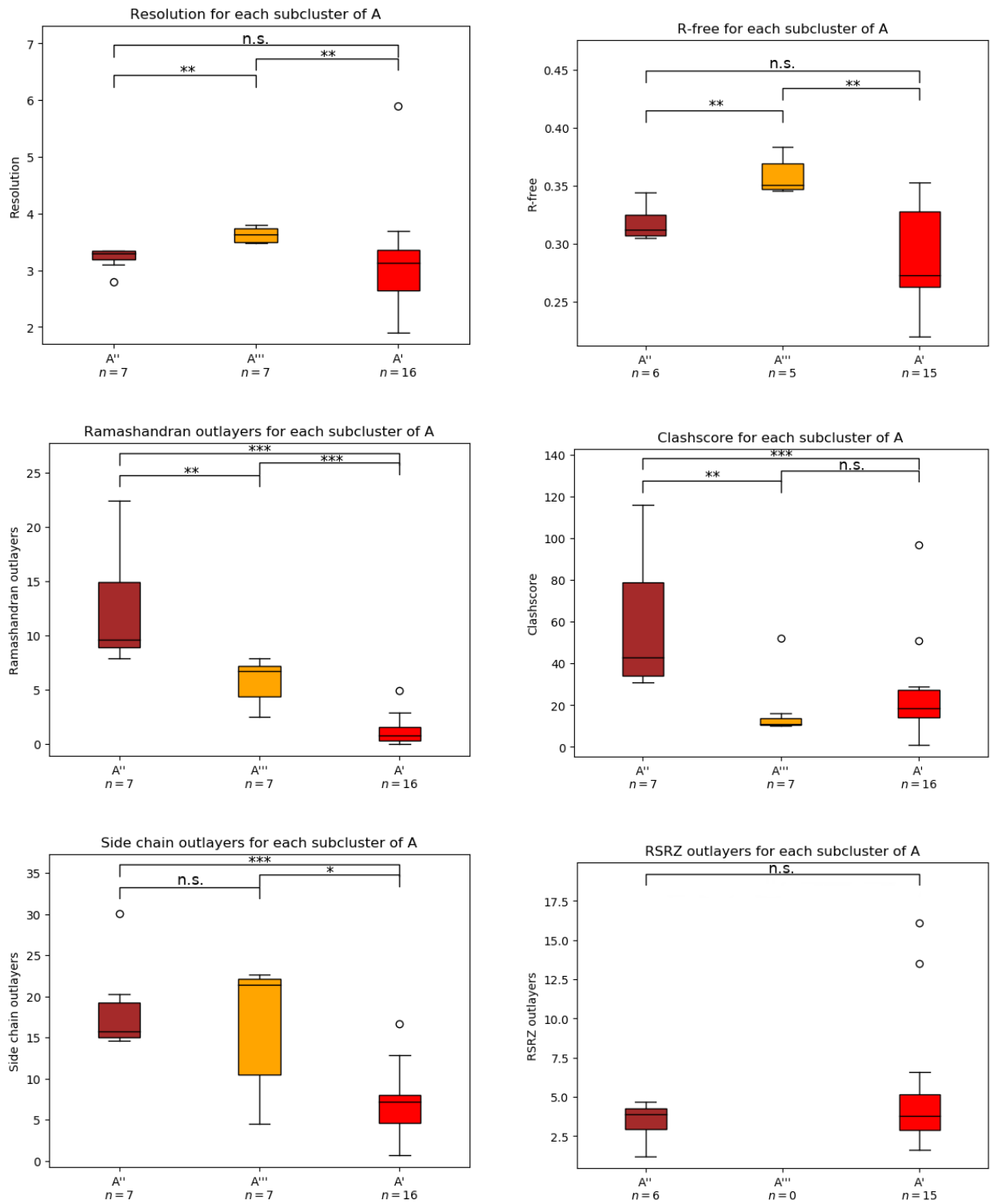
Figure 12: **Comparison of pdb quality scores.** Mann-Whitney test were performed to check differences. Size of the sample was specified for every cluster.

AB BE AE

Figure 13: **Evolution of two intermonomer interfaces:** $I_2(DC, PC2)$ **and** $I_2(PN2, PC2)$. (PDBid: 4dx5) Monomer colors read as follows: blue - state A, yellow - state B), red - state E. Interface $I_2(DC, PC2)$, median values for the number of interface atoms: AB: 19, BE: 0, AE: 21; Interface $I_2(PN2, PC2)$, median values for the number of interface atoms: AB: 68, BE: 19, AE: 67.

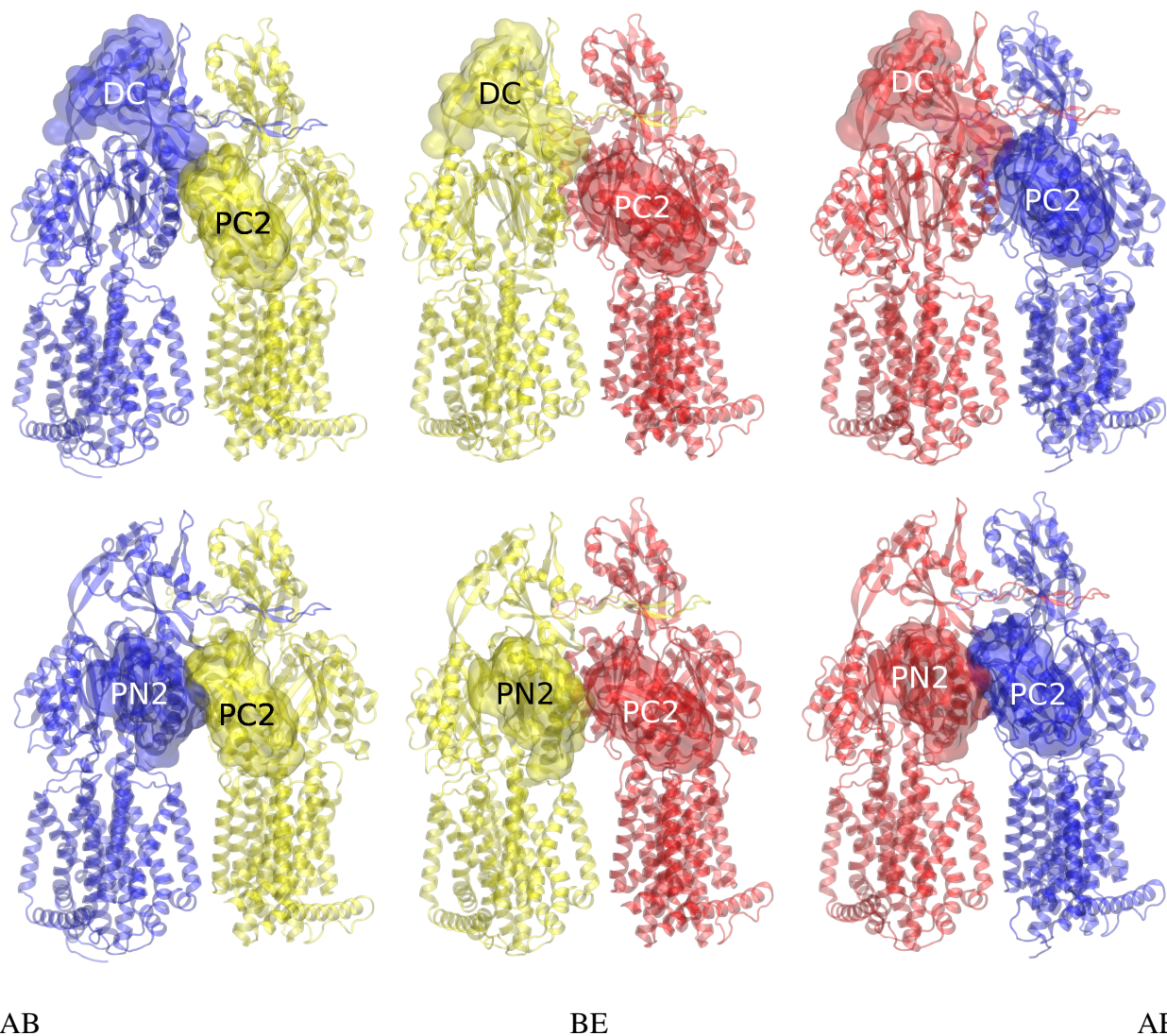|	aHelix	DC	DN	PC1	PC2	PN1	PN2	Loop1	Loop2
TM	210 / 212 / 206	o	o	37 / 35 / 30	**0 / 0 / 7**	**0 / 7 / 13**	16 / 22 / 15	144 / 152 / 149	o
aHelix		o	o	o	o	o	o	o	o
DC			332 / 324 / 328	o	o	**50 / 49 / 7**	27 / 24 / 25	o	o
DN				o	o	o	**0 / 4 / 4**	o	o
PC1					**0 / 0 / 49**	o	104 / 82 / 101	o	**24 / 0 / 18**
PC2						29 / 24 / 18	104 / 108 / 103	o	o
PN1							**32 / 7 / 18**	68 / 61 / 65	14 / 21 / 20
PN2								44 / 35 / 34	92 / 73 / 84
Loop1									7 / 11 / 12
Loop2									o

	Loop3	Loop4	Loop5	Loop6	Loop7	Loop8	Loop9	Loop10	Loop11
TM	o	o	34 / 34 / 35	138 / 144 / 139	157 / 151 / 161	**9 / 11 / 5**	o	o	**30 / 42 / 63**
aHelix	o	o	o	o	o	o	o	o	o
DC	30 / 30 / 30	24 / 23 / 26	o	o	o	o	38 / 39 / 31	18 / 17 / 17	o
DN	25 / 25 / 25	32 / 34 / 33	o	o	o	o	o	o	o
PC1	**23 / 5 / 21**	60 / 55 / 53	46 / 25 / 45	o	126 / 126 / 123	**19 / 18 / 49**	**24 / 35 / 59**	13 / 10 / 14	79 / 80 / 80
PC2	o	o	o	o	**0 / 0 / 36**	32 / 36 / 43	70 / 66 / 63	27 / 29 / 30	34 / 35 / 34
PN1	o	**7 / 8 / 0**	66 / 67 / 63	o	o	**0 / 0 / 6**	o	33 / 36 / 35	o
PN2	92 / 89 / 94	31 / 35 / 34	**0 / 10 / 0**	o	16 / 13 / 13	35 / 31 / 33	o	o	6 / 6 / 6
Loop1	o	o	21 / 22 / 22	o	o	**11 / 20 / 34**	o	o	o
Loop2	5 / 5 / 5	o	o	o	o	o	o	o	o
Loop3		48 / 28 / 46	o	o	o	o	o	o	o
Loop4			o	o	o	o	o	o	o
Loop5				o	26 / 15 / 26	**0 / 0 / 5**	o	o	o
Loop6					o	o	o	o	o
Loop7						42 / 55 / 32	o	o	**0 / 0 / 35**
Loop8							o	o	34 / 24 / 41
Loop9								49 / 47 / 49	o
Loop10									o

Table 8: **Interfaces between subdomains in a monomer of AcrB .** For any two subdomains, we computed the interface size, in number of atoms, for the A/B/E states – whence three numbers in each matrix entry. Symbol o hints at no interaction in any of the three states. Tagging uses two criteria: *interface*: criterion from Eq. 2 <u>interface</u>: interface size differs by at least 30 atoms.

Interactions	Nb of atoms (AB)	Nb of atoms (AE)	Nb of atoms (BE)	Nb of atoms (AA)
PN1 Loop3	4	4	**None**	4
DN DN	**6.0**	7	7	8
Loop9 DN	**7.0**	**8.0**	12	10
PN2 PC2	68	67	<u>**19.0**</u>	65
Loop4 DN	29	29	31	31
PC2 DC	19	21	**None**	19
DC PN1	47	49	55	49
PC1 DN	73	73	72	77
DN PN1	85	87	**63.0**	90
PN2 PN1	80	65	**58.0**	74
TM TM	100	103	87	**78.0**
PN1 PN1	121	<u>**90.0**</u>	109	115
DN DC	244	230	230	233

Table 9: **Interfaces between subdomains from different monomers of AcrB .** An interface is refered here only if it can be found in all crystal structures featuring the two states of interest. The number of atoms presented is the median value for all such interfaces. Tagging uses two criteria: *interface*: criterion from Eq. 4 <u>interface</u>: interface size differs by at least 30 atoms.

7.8 Discussion – Understanding the conformational changes of other RND proteins

See Figure 14 and Table 10.

Figure 14: **RMSD_{Comb.} between AcrB and CusA.** RMSD was computed on selected subdomains (Loop2, Loop11, TM). See Section 4.4.1

Table 10: **Interfaces between subdomains in a monomer for MexB.** For any two subdomains, we computed the interface size, in number of atoms, for the A/B/E states – whence three numbers in each matrix entry. Symbol o hints at no interaction in any of the three states. Tagging uses two criteria: *interface*: criterion from Eq. 2 *interface*: interface size differs by at least 30 atoms.

	aHelix	DC	DN	PC1	PC2	PN1	PN2	Loop1	Loop2
TM	233 / 229 / 232	o	o	31 / 26 / 27	**0 / 0 / 3**	**7 / 15 / 19**	28 / 27 / 26	150 / 137 / 154	5 / 5 / 6
aHelix		o	o	o	o	**56 / 58 / 4**	o	o	o
DC			309 / 315 / 311	o	**5 / 6 / 0**	**30 / 29 / 14**	25 / 22 / 30	o	o
DN				o	o	86 / 88 / 89	6 / 9 / 6	o	o
PC1					**41 / 0 / 73**	o	111 / 84 / 108	o	**14 / 0 / 14**
PC2						o	o	o	o
PN1							**35 / 5 / 12**	72 / 63 / 69	33 / 36 / 38
PN2								35 / 37 / 25	110 / 87 / 93
Loop1									**11 / 25 / 13**
Loop2									o

	Loop3	Loop4	Loop5	Loop6	Loop7	Loop8	Loop9	Loop10	Loop11
TM	o	o	26 / 28 / 27	102 / 162 / 158	162 / 164 / 163	**0 / 8 / 9**	o	o	104 / 96 / 110
aHelix	o	o	o	o	o	o	o	o	o
DC	27 / 24 / 28	20 / 21 / 22	o	o	o	o	58 / 55 / 58	8 / 8 / 7	o
DN	31 / 28 / 29	35 / 32 / 34	o	o	o	o	o	o	o
PC1	**18 / 7 / 18**	35 / 43 / 38	**48 / 18 / 47**	o	119 / 120 / 122	**41 / 20 / 52**	61 / 56 / 48	12 / 12 / 10	o
PC2	o	o	o	o	**24 / 9 / 27**	30 / 32 / 30	53 / 54 / 55	30 / 28 / 26	o
PN1	o	**10 / 16 / 0**	o	o	**25 / 11 / 16**	44 / 42 / 48	7 / 9 / 12	55 / 54 / 57	o
PN2	89 / 92 / 89	25 / 36 / 28	56 / 51 / 53	o	o	39 / 27 / 38	o	o	o
Loop1	**3 / 0 / 3**	o	**0 / 9 / 0**	o	o	o	o	o	**10 / 5 / 5**
Loop2	o	o	20 / 17 / 20	o	o	o	o	o	**0 / 0 / 3**
Loop3		**39 / 15 / 41**	o	o	o	o	o	o	o
Loop4			o	o	o	o	o	o	o
Loop5				**0 / 52 / 45**	o	**0 / 5 / 6**	o	o	o
Loop6					o	o	o	o	o
Loop7						40 / 44 / 45	o	o	**12 / 10 / 33**
Loop8							o	o	o
Loop9								43 / 46 / 45	23 / 23 / 31
Loop10									o
Loop11									o

51