Functional Motions Modulating VanA Ligand Binding Unraveled by Self-Organizing Maps

Guillaume Bouvier, Nathalie Duclert-Savatier, Nathan Desdouits, Djalal Meziane-Cherif, Arnaud Blondel, Patrice Courvalin, Michael Nilges, Thérèse Malliavin

Functional motions modulating VanA ligand binding unraveled by self-organizing maps

November 6, 2013

Guillaume Bouvier[1*], Nathalie Duclert-Savatier[1*], Nathan Desdouits[1], Djalal Meziane-Cherif[2], Arnaud Blondel[1], Patrice Courvalin[2], Michael Nilges[1], Thérèse E. Malliavin[1]

(1) Institut Pasteur, Unité de Bioinformatique Structurale; CNRS UMR 3825; Département de Biologie Structurale et Chimie; 25, rue du Dr Roux, 75015 Paris, France
(2) Institut Pasteur, Unité des Agents Antibactériens; 25, rue du Dr Roux, 75015 Paris, France

(*) contributed equally to the work.
Corresponding author : Thérèse E. Malliavin; Unité de Bioinformatique Structurale; CNRS UMR 3825; Institut Pasteur, 25, rue du Dr Roux, 75015 Paris, France; E-mail: terez@pasteur.fr; Phone: +33 1 40 61 34 75

Keywords : D-alanyl:D-lactate ligase, D-alanyl:D-alanine ligase, antibiotic resistance, vancomycin, molecular dynamic simulation, docking, classification, self organizing map.

1 Abstract

The VanA D-Ala:D-Lac ligase is a key enzyme in the emergence of high level resistance to vancomycin in *Enterococcus* species and Methicillin-Resistant *Staphylococcus aureus*. It catalyzes the formation of D-Ala-D-Lac, a surrogate peptidoglycan precursor with low affinity for vancomycin, that can replace D-Ala-D-Ala, which is subject to sequestration by vancomycin. Therefore, VanA appears as an attractive target for the design of new antibacterials to overcome resistance.

The catalytic site of VanA is delimited by three domains and closed by an ω-loop upon enzymatic reaction. The aim of the present work was: (i) to investigate the conformational transition of VanA associated to the opening of its ω-loop; and (ii), to relate this transition with the substrates or products binding propencities. Molecular dynamics trajectories of the VanA ligase of *Enterococcus faecium* with or without a disulfide bridge distant from the catalytic site, revealed differences in the ω-loop conformations with a slight opening. Conformations were clustered with an original machine learning method, based on self-organizing maps (SOM), which revealed four distinct conformational basins. Several ligands related to substrates, intermediates or products were docked to SOM representative conformations with the DOCK 6.5 program. Classification of ligand docking poses, also performed with SOMs, clearly distinguished ligand functional classes: substrates, reaction intermediates and product. This result illustrates the acuity of the SOM classification and supports the quality of the DOCK program poses. The protein-ligand interaction features for the different classes of poses will guide the search and design of novel inhibitors.

2 Introduction

Spreading of antibiotic-resistant bacterial pathogens is ever continuing, and the absence of new antibiotics in development pipelines is seriously threatening the future of public health. Vancomycin is a widely used glycopeptide antibiotic for the treatment of infections caused by multi-drug resistant Gram-positive pathogenic bacteria. However, resistance emerged in *Enterococcus* species and now spreads to other bacteria including *Staphylococcus aureus*, causing serious problems in the clinic.[1]

Vancomycin acts by inhibiting peptidoglycan synthesis. The antibiotic interacts with the D-Ala-D-Ala *terminus* of N-acetyl-muramyl-L-Ala-D-Glu-L-Lys-D-Ala-D-Ala late peptidoglycan presursors, hence sequestering the D-Ala-D-Ala dipeptide, and inhibiting the activity of the transpeptidases.[2] Resistance to vancomycin results mainly from the production of modified precursors ending with D-Ala-D-Lac, which exhibits 1000 fold lower binding affinities to vancomycin than D-Ala-D-Ala precursors (Figure 1). Synthesis of D-Ala-D-Lac requires the presence of a ligase with an altered specificity (VanA)[?] that acts at a critical step, thus, reprogramming peptidoglycan synthesis. As a result, it appears as a target of choice to develop new antibiotics. Inhibitors have been discovered[?] on the related enzyme, the D-Ala-D-Ala ligase.

The X-ray crystallographic structure of the D-Ala:D-Lac ligase, VanA from *Enterococcus faecium* (PDB entry: 1E4E) (Figure 2), and that of the D-Ala:D-Ala ligase TtDdl from *Thermus thermophilus* (PDB entry: 2YZG) display similar features. These enzymes are divided in three domains: N-terminal ([A2-G121] and [M1-G104], in blue), central ([C122-S211] and [A105-L192] in red and yellow) and C-terminal ([G212-A342] [S193-T319] in black and green), respectively, in 1E4E and in 2YZG structures. The ω-loop (in green in Figure 2) is part of the C-terminal domain. It encompasses the residues

[L236-A256] in 1E4E and the residues [Y218-A234] in 2YZG.[6–8] The region opposite to the ω-loop in the structure (yellow in Figure 2) is called "opposite domain" in the present work. It is composed of residues [A149-Q208] in 1E4E and [V131-K190] in 2YZG and folds in a two layer β sandwich. The substrates bind to a large pocket located at the interface between N-terminal, central, and C-terminal domains. In 1E4E, the ω loop closes the pocket and prevents ATP hydrolysis. Conversely in TtDdl, this flexible loop displayed various extensions in structures obtained with different reaction intermediate co-crystals (PDB entries: 2YZG, 2YZN, 2ZDG, 2ZDH, 2ZDQ, 2YZM). Cysteines 52 and 64 form a disulfide bridge in crystal structure 1E4E (Figure 2). The bridged form is called VanA_{SS} in the current work.[6,8]

The conformational transition of the ω loop, infered from the different ligase structures, is expected to play a key role in substrates binding, and should thus be studied in view of a VanA inhibitors development. Conformational transitions of biomolecules have been extensively studied by molecular modeling,[?] but the impact of these transitions on ligand docking have been investigated less systematically. Here, we propose to use an Artificial Neural Network, the self-organizing maps (SOMs),[?] to simultaneously characterize conformational transitions and ligand interactions. SOMs have been used in the past for the *in silico* screening of chemical compounds for drug discovery,[?,?] for the prediction of compounds selectivity,[?] for the detection of new bioactive molecules,[?,?] for the re-scoring of docking poses,[?,?] and for various clustering of conformational ensembles,[?,?,?,37] and protein fragments.[?] A detailed overview of the use of self-organizing maps in the framework of molecular modeling and structure-based drug design, has recently been published.[?]

The purpose of the present work is: (i) to describe the first steps of the ω-loop opening

through the analysis of the protein internal dynamics and (ii) to correlate the conformations sampled along this transition with the binding of ligands displaying various biological functions. The protein conformational transitions were analyzed with molecular dynamics simulations, while ligand binding was investigated by molecular docking calculations. The clustering methods, based on self-organizing maps (SOM) were developed to cluster protein conformations as well as to classify the ligand poses.

3 Materials and Methods

3.1 Preparation of simulated systems

All systems (see Table 1) were setup from the PDB X-ray crystallographic structures 2YZG, 2ZDH and 1E4E corresponding respectively to:

(i) the D-Ala:D-Ala ligase apo from *Thermus thermophilus* HB8 with open ω-loop,

(ii) the D-Ala:D-Ala ligase from *Thermus thermophilus* HB8 with closed ω-loop with ADP and D-Ala in its binding pocket[6] and,

(iii) VanA, the D-Ala:D-Lac ligase from *Enterococcus faecium* BM4147, containing ADP and phosphinate (1(S)-aminoethyl-(2-carboxypropyl)phosphoryl-phosphinic acid).[8]

The PDB structure 1E4E was used to produce the systems VanA_{SS}.lig bearing an ADP, a phosphinate inhibitor (PHY) and two Mg^{+2} ions in the catalytic site, and a C52-C64 disulfide bridge. The ligands were removed from 1E4E to build the corresponding apo system, VanA_{SS}. Then, the VanA_{SS} disulfide bridge was reduced to build the VanA system. Similarly, the TtDdl_{closed}.lig and the TtDdl_{closed} systems were built from the 2ZDH structure with or without the ADP, D-Ala and Mg^{+2} ions, respectively. Finally, the TtDdl_{open} system was built from the 2YZG structure.

Hydrogen atoms were added with the LEaP[9] module of AMBER 10.[10] The FF99SB force field[11] was used. The systems were neutralized with Na^+ counter-ions. The organic molecules were parametrized with Antechamber[12] and the **G**eneral **A**MBER **F**orce **F**ield (GAFF).[13] Explicit TIP3P[14] solvent water molecules were added to the systems in a cubic box under periodic boundary conditions with a buffer zone of 10Å. The system components are given in Table 1.

3.2 Molecular Dynamics Simulations

The **S**imulated **A**nnealing with **N**MR-**D**erived **E**nergy **R**estraints (Sander) module from AMBER 10[15] was used to perform five rounds of minimizations composed of steepest descent followed by conjugate gradient algorithms. Harmonic restraints were applied on the protein atom position with the reference set to the final position of the previous round and a force constant of 100, 50, 25, 10 and 5 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ in each round, respectively. Then, the systems were thermalized to 298 K for 20 ps with **M**olecular **D**ynamics (MD) at constant volume, by making use of the weak-coupling algorithm[16] and harmonic restraints of 25 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ on the solute atom positions. Thus, six to seven equilibration rounds were performed with a Langevin thermostat with a collision frequency $\gamma = 2$ ps^{-1}. One 5 ps MD round at constant volume was followed by four 2.5 ps and one 10 ps constant pressure MD rounds. Harmonic restraint force constants were 25, 25, 20, 15, 5 and 2.5 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$, respectively. Finally a last MD round of 60 ps was performed without any restraints.

Molecular **D**ynamics (MD) trajectories were recorded over 20 to 30 ns with the **P**article **M**esh **E**wald **M**olecular **D**ynamics (PMEMD)[?,17] module from AMBER 10. A cutoff of 10 Å was used for Lennard-Jones interaction calculations. Long-range elec-

trostatic interactions were calculated with the **P**article **M**esh **E**wald (PME) protocol.[17] The simulations were performed at a pressure of 1 atm and a temperature of 298 K under the control of a Berendsen thermostat with a coupling time of 2 ps.[16] The SHAKE algorithm[18] kept all covalent bonds involving hydrogens rigid so integration time step of 2 fs was used for all MD simulations. Atomic coordinates were saved every picosecond. D-Ala:D-Lac ligase MD trajectories were recorded seven to nine times with different initial random seeds. The D-Ala:D-Ala ligase trajectories were recorded only once.

3.3 Conformational analysis of the molecular dynamic simulations using self-organizing maps

Self-**O**rganizing **M**aps (SOM),[19,20] which are unsupervised neural networks, were used to cluster the 50 000 conformations sampled during the "VanA" and "VanA_{SS}" MD simulations. Conformations were encoded as follow: the $n \times n$ pairwise square Euclidean distance matrix D was calculated for n C_α atoms of the protein. Then, to compress the information, the covariance matrix, C of the lines versus columns of D was calculated:[21]

$$C_{i,j} = \frac{1}{n} \sum_{k=1}^{n} \sum_{l=1}^{n} (d_{i,k} - \bar{d_i})(d_{l,j} - \bar{d_j}) \tag{1}$$

where $\bar{d_i} = \frac{1}{n} \sum_{j=1}^{n} d_{i,j}$. As C describes a 3D object, its eigenvalues beyond the first four are null. Hence, the eigenvectors of C, $N_{i=1,..,4}$, corresponding to the four first eigenvalues, were kept applied to D; $D \cdot N_{i=1,..,4}$. This compression in $n \times 4$ matrices gives a conformational descriptor, which conserves information.

These descriptors were used to train a periodic Euclidean self-organizing map (SOM). Most commonly used SOMs are 2D SOMs, which are defined by three-dimensional ma-

trices. The first two dimensions, 2D, lengths are chosen by the user, here 50×50, and define the map size. As these dimensions are chosen to be periodic, the map is a toroid. The third dimension has the length of the input vectors, or descriptor, here: $4n$, and each vector along the third dimension is called a neuron.

The self-organizing maps were initialized with a random uniform distribution covering the range of values of the input vectors. At each step, an input vector is presented to the map, and the neuron closest to this input, the Best Matching Unit (BMU) is updated. The maps were trained in two phases. Guillaume: peux tu revoir cela: During the first phase, the 50 000 input vectors are presented to the SOM in random order to avoid mapping bias with a learning parameter of 0.5, and a radius parameter of 36, as explained in Reference 40.

During the second phase, the learning and radius constants were decreased exponentially from starting values 0.5 and 36, respectively, during 10 cycles of presentation of all the data in random order.

Hence, to delineate clusters on the SOMs, the conventional **U**nified **d**istance **m**atrix (U-matrix) is a useful tool. For each neuron ν on the map, a corresponding U-matrix element is calculated as the mean Euclidean distance between the neuron ν and its eight immediate neighbors:

$$\text{U-height}(\nu) = \frac{1}{8} \sum_{\nu \in N(\nu)} d(\nu, \mu) \tag{2}$$

where $N(\nu)$ is the set of neighbors, and $d(\nu, \mu)$ is the Euclidean distance between the vectors μ and ν. The resulting 50×50 U-matrix reveals the topological organization of the map, and can be used to draw the contours of clusters by applying a threshold distance value.

SOMs distribute data on the map so that points which are close or far in the descriptor

space are also close or far, respectively, on the map. However, they also distribute the data as evenly as possible on the map. This action enforces the similarity between neighboring neurons in the final map. If the system topology is poorly compatible with a projection on a tore, some distent conformational basins will be projected on close regions of the SOM, resulting in large conformational variations between close neurons, which, coupled with the enforcement of similarity just described, induce the formation of empty nodes.

The maps convergence was assessed quantitatively, by running 80 independent SOM calculations. Each calculation started from a different random map, and the comparison of the resulting maps was performed using the following flooding algorithm, inspired by the watershed algorithm[?] used in image processing. This algorithm works on the topology of the U-matrix. It starts from the global minimum and flood the map according to the landscape of the U-matrix. The maps are then reordered according to the order of the flooding process. The maps were compared by calculation of the average correlation between the reordered neurons. Since these averages were within the interval 0.98-1.0 (data not shown OR FIGURE corrdist.pdf), the map convergence was considered as effective, and the maps valid.

Representative conformations extracted from the SOMs clusters are available from the authors upon request.

3.4 Flow analysis of the self-organizing maps

The molecular dynamic trajectory evolution can be followed for each time step t by its position $(i,j) = \Phi(t)$ on the SOM. The ensemble time steps that project on neuron (i,j) is called $\{\tau_{i,j}\}$, and the total number of these steps is noted $f_{i,j}$. The local mean transfer vector field is then defined by the average SOM index difference for times t in $\tau_{i,j}$ to the

next steps $t+1$:

$$\mathbf{v}_{i,j} = \frac{1}{f_{i,j}} \sum_{t \in \tau_{i,j}} \frac{\Phi_{t+1} - \Phi_t}{\|\Phi_{t+1} - \Phi_t\|} \tag{3}$$

only defined for non-empty neurons where $f_{i,j}$ is non zero.

3.5 3D Self Organizing Maps

Similarly to the 2D SOM described in the previous sections, 3D self-organizing maps were built to describe the docking position of the ligands atoms.

The input of that SOM procedure was the set of 3D coordinates for individual atoms of the ligand along the molecular dynamics trajectory. The 1170000, 121199 and 275000 input vectors for ADP, D-Ala and D-Lac, respectively, were used to train three independent 3D SOMs. The self-organizing maps were initialized with a random uniform distribution of ligand coordinates and trained in two phases. During the first phase, input vectors are presented to the SOM in random order. Guillaume, peux tu donner le taux de decroissance... Initial radius and learning parameters were set to 7.5 and 1.0, respectively, and decreased exponentially to 0 during the training process. As described before, the SOM convergence was checked by multiple independent training runs.

Different sizes (ADP: 25x17x18; D-Ala: 18x17x11; D-Lac: 17x15x13) were chosen for the 3D SOMs, corresponding respectively to resolutions of 10.28, 60.98 and 38.51 neurons/Å^3, and thus to: 2.2, 3.9 and 3.4 neurons/Å. Such resolutions in neurons correspond to a precision of 0.25-0.5 Å in atomic coordinates, similar to the estimated positional error in X-ray crystallographic structures at about 2.5 Å resolution.

3.6 Docking procedure

The ATP, D-Ala, D-Lac, D-alanylphosphate (D-Ala(P)), the phosphinate (PHY, transition state inhibitor) and D-Ala-D-Lac, which are involved in, or interfere wit,h VanA enzymatic activity, were formated in mol2 with Chimera 1.4[22] and MarvinSketch 5.1[23] for docking.

UCSF DOCK 6.5[24–26] was used to perform ligand docking on representative $VanA_{SS}$ MD conformations selected by 2D SOM analysis. These structures were those having their structure descriptor closest in Euclidean distance from that of a populated neuron. Chimera[22] was used to add hydrogens, check atom assignment, and assign partial charges in line with the AMBER-ff99SB force field. It was also used to produce mol2 format files for the ligands and the selected conformations of the receptor. The DMS software program[27,28] generated the molecular surface of the receptor using a radius probe of 1.4 Å. Then, spheres were calculated around the receptor with the DOCK 6.5 command 'sphgen' with radius probe values varying between 1.4 Å and 4 Å. Spheres within a radius of 10 Å around the geometric center of the crystallographic ligands (ADP, PHY) found in 1E4E were selected. The grid encoding van der Waals and electrostatic interactions was pre-calculated with the "grid" tool[29] in a box containing the selected spheres. The DOCK program builds up to 500 flexible ligand docking poses, on the pre-calculated "grid" interaction map. The ligand poses were then re-scored with the implementation of the Hawkins **M**olecular **M**echanics **G**eneralized **B**orn **S**urface **A**rea (MM-GBSA) score,[30–34] implemented in UCSF DOCK 6.5.

The best scoring solution was kept for each protein - ligand pair. The binding pocket was defined by residues: E14, E15, V18, H98, G99, E103, S126, C129, M130, K132, T135, Y136, K170, P171, S174, G175, S176, S177, F178, V180, E213, I239, F240, R241, I242,

H243, Q244, R289, D291, L302, N303, E304, V305, N306, T307, P309, G310, S315, R316 and Y317.

4 Results

4.1 Concerted ω-loop / opposite domain motions correlate with the presence of the disulfide bridge

The global **R**oot **M**ean **S**quare **D**eviation (RMSD) from the initial structure for the Cα atoms stabilized at about 2.2 Å for the eight independent VanA trajectories (Figure 3a) and the ten $VanA_{SS}$.lig MDs (Figure 3c). By contrast, the seven $VanA_{SS}$ MDs (Figure 3b) displayed heterogeneous behavior. Curve with the smallest drift for this system, in black, is similar to that observed for VanA, whereas that with the largest drift, in red, increased up to 3.5Å after 17 ns (Figure 3b). Hence, the presence of the C52-C64 disulfide bridge correlated with a destabilization of VanA conformations.

The contributions of the different regions (C-terminal, central, N-terminal, opposite domains and ω-loop) to the RMSD were analyzed on the trajectories with the largest global RMSD drifts recorded for VanA, $VanA_{SS}$ and $VanA_{SS}$.lig (Figure 3d-f). A similar analysis was performed for the D-Ala:D-Ala ligase systems $TtDdl_{open}$, $TtDdl_{closed}$ and $TtDdl_{closed}$.lig (Figure 3g-i, Table 1). The ω-loop always displayed the largest drift, except for $TtDdl_{closed}$.lig (Figure 3i). The systems displaying the smallest ω-loop drifts were VanA and $TtDdl_{closed}$ (Figure 3d & h).

The large drifts of the ω-loop in MD simulations are in good agreement with the large conformation differences observed in X-ray structures.[6] Indeed, the ω-loop covers the binding site entrance in 2ZDH and 1E4E, whereas it extends away from the core of the D-

Ala:D-Ala ligase in 2YZG. Correspondingly, the largest observed drift was for TtDdl$_{open}$, (2YZG), which also has an empty catalytic site, and is probably in an inactive functional state. Interestingly, among the three systems built from 1E4E, VanA$_{SS}$ (Figure 3e) and VanA$_{SS}$.lig (Figure 3f) presented large ω-loop drifts despite an initial closed conformation. Noticeably, the large global protein RMSD drift observed in the presence of the C52-C64 disulfide bridge (Figure 3b,c), is mostly due to the ω-loop motions (Figure 3e,f). In the presence of the substrates ADP and PHY, the disulfide bridge still destabilized the ω-loop, but to a lesser extent (Figure 3c,f). However, the presence of the ligands strongly reduced all protein region drifts when the ω-loop is wrapped (Figure 3i).

The other protein regions rarely drifted beyond 3Å. Nonetheless, the opposite domain (yellow curves), the central domain (red curves) and the C-terminal domain (black curves) drifted more when the ω-loop made large motions (Figures 3e-g).

To describe the relative displacement of the protein regions with respect to each other, a **P**rincipal **C**omponent **A**nalysis (PCA) was performed on the Cα atoms trajectories (Figure 4). More eigenvectors were necessary to account for 90 % of the motions of the "opened" systems, than for the "closed" systems, with 16 to 25 and 46 to 61 eigenvectors required, respectively. A large and strongly dominant eigenvalue was observed (Figure 4f) for the simulations VanA$_{SS}$ and VanA$_{SS}$.lig, which displayed strongly correlated motions, and large ω-loop drifts. The relative importance of the first eigenvalue was lower in the presence of ligands. For example, the first eigenvector of VanA$_{SS}$ and VanA$_{SS}$.lig, contributed respectively 18.4 % and 6.0 % to the global motion variance. For VanA and TtDdl$_{closed}$ simulations, no dominant motion was observed as the first eigenvalue accounted for 1.5 % to 3.0% of the global motion.

The projection of the first PCA mode on the protein structures (Figures 4a-e) showed

homogeneously distributed motions with relatively small amplitude in TtDdl$_{closed}$ and VanA closed state MD simulations (Figures 4c,d). By contrast, motions were mainly located in the ω-loop and the opposite domain for VanA$_{\mathrm{SS}}$, VanA$_{\mathrm{SS}}$.lig and TtDdl$_{open}$, (Figures 4a,b,e).

Hence, PCA analysis revealed the specific internal fluctuations of the ω loop and the opposite domain. As expected, these fluctuations are larger for structures bearing an open ω loop and no ligand in the catalytic site. It was more surprising to find that the C52-C64 disulfide bridge would also increase so significantly the ω motions.

4.2 Self-organizing maps suggest contours of free-energy basins

Self-**O**rganizing **M**aps (SOM)[19,20] were used to project the conformational space explored by the ligase during MD simulations onto a smaller, bi-dimensional and topologically organized space. A 50×50 SOM was trained to cluster the protein conformations along one trajectory of VanA and VanA$_{\mathrm{SS}}$, respectively. The VanA and VanA$_{\mathrm{SS}}$ MD trajectories analyzed here occupied distinct zones of the SOM.

As described in Materials and Methods, the U-matrix is a convenient visualization tool to reveal SOM topological features.[35,36] Closely related structures are grouped in the same valleys or basins with small inter-neuron distances colored in blue separated by ridges of large inter-neuron distances, defining their boundaries in red (see Figure 5a). The U-matrix, which gives an evaluation of the state density, can thus be interpreted as a qualitative marker of the free energy landscape of the protein conformational space within the sampled area. The landscapes of VanA and VanA$_{\mathrm{SS}}$ showed large blue patches of homogeneous structures separated by thin red barriers that would be expensive to cross and lower green walls that can be crossed occasionally. For VanA, which performed

limited exploration, there were few big clusters separated by low barriers (called 4 in Figure 5a). For $VanA_{SS}$, there were larger barriers roughly dividing the U-matrix into two main regions, the first bearing two sub-regions (depicted by 1 and 2 in Figure 5a) and the second displaying a higher degree of diversity (noted 3 in Figure 5a). The map showed that the VanA MD spanned a smaller physical space (mostly blue neurons) than $VanA_{SS}$ which formed at least two independent coherent tracts (basins 1,2 *versus* 3) with higher diversity in the second one according to intrinsic distance (cyan to green neurons).

To give a quantitative support to the interpretation of the SOM clusters as free-energy basins, MM-PBSA and MM-GBSA energies were calculated along the VanA and $VanA_{SS}$ trajectories, with the AMBER 12 package tools,[?] and projected on the SOM (Figure 7a,b). Although MM-PBSA and MM-GBSA energies displayed significant fluctuations, the agreed reasonably well with the SOM clustering, since energies were more uniform within basins than between them and the contiguous basins borders displayed higher energies. This relative correspondance between SOM clustering and the energies supports the U-matrix as qualitative marker of the free energy landscape.

Structural properties were then projected and visualized onto the 2D trained map. The projection of the RMSD from the first frame of each trajectory further corroborates the quality and the convergence of the clustering process (Figure 5b) and the relation with the conformational landscape. For $VanA_{SS}$ the two regions delineated by the U-Matrix displayed distinct RMSD values. The lower U-matrix zone (basins 1 and 2) corresponded to comparable drifts to that of VanA, while the higher U-matrix zone (basin 3) revealed conformations that had largely evolved from the initial structure (Figure 5b). The homogeneous RMSD pattern of VanA and the bipolar one of $VanA_{SS}$ are directly related to the U-matrix patterns.

In order to evaluate structural changes, the evolution of the β-strands secondary structure content was projected on the SOM (Figure 5c). The α helices were only marginally affected on both systems. Only a slight uncoiling was observed for VanA between 10 and 15 ns of simulation (data not shown). Up to 12 amino acids loose their β structure in VanA_{SS} and VanA trajectories as can be seen on the SOM projection (Figure 5c). The $\beta-6$ strand located in the N-terminal domain of the protein (Figure 2) lost three to four residues in VanA_{SS} and no more than three in the last part of VanA trajectory. The most affected β structures apart from $\beta-6$ were $\beta-14$ and $\beta-15$, close to the ω-loop (Figure 2). Indeed, while $\beta-15$ gained 3 residues in VanA, both $\beta-14$ and $\beta-15$ lost two β residues in the most-drifting part, the last 10 ns, of the MD trajectory of VanA_{SS}. This secondary structure variability agrees with the role of hinges played by these β strands during the opening motion of the ω-loop in VanA_{SS}, as can be seen in (Figures 4a,c).

The SOM appeared to produce a meaningful clustering of conformations. Since the original trajectory can be followed on the map, SOM could also be used to investigate how the protein evolves in the different parts of the map with the vectors field $\mathbf{v}_{i,j}$ (see Materials & Methods Eq. 3). The vector field gives the propensity of the mapped conformations to evolve in the given direction. The vectors field appeared to follow the gradient of the U-matrix. The vectors with low or null norms are mostly present in the bottom of the basins. These results substantiate the interpretation of the U-matrix as a marker of the free energy landscape. Large arrows indicate high net flow for densely populated regions, but could also be due to poor statistics in low density regions. Interestingly, some small vectors are also present on the U-matrix barriers between the closed and opened conformations of VanA_{SS}. These structures have the same probability to go to either basins, which, in practice, would correspond to the definition of a transition

states ensembles.[38] The trajectory of $VanA_{SS}$ (in pink in Figure 6) is characterized by two major basins. The first basin, subdivided in two sub-basins, 1 and 2, groups initial and then more equilibrated conformations of the closed state of $VanA_{SS}$ respectively. The second major basin, labeled "3", contains open states. The barrier between basins 2 and 3 is composed of low density neurons, with high convergent flows pointing to the transition states ensemble surrounded by stationary points. As already seen, VanA and $VanA_{SS}$ covered distinct conformational spaces except for a limited border area highlighted by brown diamonds in Figure 6.

The transition states between basins are defined as points of zero flow. Interestingly such points correspond to saddle point in the surface of the U-matrix. As explained in the Materials and Methods, flow is not defined at empty neurons, and thus points of zero flow close to empty neurons should not be picked up as saddle points. Saddle points detected in the present work (Figure 6) are located far from empty neurons.

The conformational clustering of the molecular dynamics simulations VanA and $VanA_{SS}$ show several basins corresponding to closed and open conformations of VanA. The detection of such conformations gives a more precise picture of the different steps of the interaction between VanA and the reaction substrates and should help to search for VanA inhibitors.

Hence, density metrics given by the U-matrix suggests that the basins could be interpreted or defined as free-energy basins, within the limits of the conformational sampling. Projection of β secondary structure evolution indicated that the β strands located close to the ω-loop hinges were the most variable ones. The analysis of the conformational flow defined populated regions during the ω loop opening that can be considered as transition states.

As described by 6 for TtDdl the opposite domain of $VanA_{SS}$ moves away from the binding cavity (Figure 5d, 4a). In contrast to the observed motions in TtDdl, only a sub-part of $VanA_{SS}$ central domain, the opposite domain, is involved in the opening motion during the course of the dynamics.

4.3 SOM classification of ligand poses related to their function

In the Ter-Ter mechanism of the ligases (Figure 1),[?] the ATP binds first. It is followed by a first D-Ala and then either D-Lac, or a second D-Ala for VanA or Ddl, respectively. The ligands (ATP, D-Ala, D-Lac, PHY, D-Ala(P), D-Ala-D-Lac) were docked individually on conformations representing each neuron of the 2D-SOM to relate conformations sampled along the ω loop opening and ligand binding propensity. A neuron was represented by the structure, either from VanA or $VanA_{SS}$, which had closest descriptor to that of the neuron after training.

One 3D self-organizing map, 3D-SOM, was built from the docking results for each ligand. The descriptors were the coordinates of all atoms of each ligand. Mapping the identity of the ligand, (ADP, D-Ala, *etc*...) on the resulting map indicated their respective consensus binding sites. The 3D-SOM was projected onto the 3D Cartesian coordinates simply using the neuron descriptor field (Figure 8). The respective ligand binding sites agreed with those observed in crystal structures of TtDdl in complex with ADP and D-Ala (2ZDH).[6] Interestingly, the binding sites identified by docking here for D-Lac overlapped with those of phosphinate, a transition state analog co-crystallized with VanA (1E4E) or that of D-Ala-D-Ala in TtDdl (2ZDQ).[6,8]

To further analyze the ligand docking specificity, the GB/SA docking scores (see Materials and Methods) were then projected onto the 2D SOM used to cluster the MD

conformations (Figure 9). Noticeably, VanA displayed binding trends that agreed with the enzymatic role of the ligand. In addition to the ligand binding site specificity observed with the 3D-SOM, the conformational ligand binding specificity could hence be established. For instance, ATP binds exclusively in the opened ω-loop conformation basin defined by the U-matrix (Figures 5 & 9a) and scored better than the reaction products ADP (data not shown). The second partner of the reaction, D-Ala, binds non-selectively to almost all the VanA_{SS} structures (Figure 9b), and less than half of the VanA structures. Not surprisingly, the product of the enzymatic reaction, D-Ala-D-Lac (Figure 9f), does not display binding selectivity. The acylphosphate, D-Ala(P) (Figure 1) corresponding to the phosphorylated form of the former D-Ala, and D-Lac only binds with a good score, to the same restricted region of the SOM map, corresponding to the third basin where the reaction takes place (see Figure 9c,d). Phosphinate mimicking the tetrahedral intermediate binds also with a good score to the third basin (see Figure 9e). Strikingly, the best phosphinate docking scores were observed on the conformations, that were delineated as the transition state ensemble between closed and open ω-loop states in section "Self-organizing maps suggest contours of free-energy basins".

5 Discussion

In the present work, we used MD simulations to investigate VanA conformational sampling, in particular the first opening steps of the ω-loop. Two main conformational basins were visited in the presence of a disulfide bridge between C52 and C64. Known ligands (substrates, products and intermediate alike) were docked on representative conformations issued from the clustering. This analysis showed a correlation between docked ligand binding energies and the protein conformation, which is in good agreement with

the Ter-Ter ordered mechanism of the ligase.

The MD simulations performed here indicated that the ω-loop opening mechanism of VanA is similar to that of the endogenous enzyme, TtDdl.[6,45] Indeed, the semi-open ω-loop conformation of VanA is similar to that observed in the 2ZDG TtDdl structure.[6] Furthermore, the correlation of the ω-loop and opposite-domain motions (Figure 4) in the MD simulations agreed with available structures data on ligases.[6] The similarity between consensus binding sites of ADP, D-Ala and D-Lac in representative conformations of VanA (Figure 8) and those observed in crystal structures of TtDdl (2ZDG, 2ZDH, 2ZDQ)[6] strongly supports that those two proteins make similar interactions with their substrates. This ligand binding similarity and the mechanistic similarities implied by MD simulations interestingly supports the idea that new inhibitors against both D-Ala:D-Ala and D-Ala:D-Lac ligase could be found and developed.

A recent study of the D-Ala:D-Ala ligase described a possible ω-loop opening mechanism in Ddl by **S**teered **M**olecular **D**ynamics (SMD).[45] Conformations extracted from this opening path were used in an initial screening, which allowed to identify experimentally validated inhibitors. This study highlighted the importance of the ω-loop opening conformational analysis in the quest for new ligase inhibitors. In addition, the importance of the ω-loop dynamics for the D-Ala:D-Lac ligase, was shown. Furthermore, the opposite domain motion is also crucial for the activity of the VanA ligase.

The clustering of molecular dynamics simulations, performed here using SOMs, was used to extract representative conformations. The representative conformations have different propensities to bind ligands at different stages of the enzymatic reaction (substrates, intermediate-like, products), as it was shown by the clustering of the ligand docking poses. These conformations are thus good candidates to perform virtual screen-

ing runs in the context of the development of new antibiotics able to overcome pathogenic resistance.

The new insights into the relationship between VanA conformational transition and predicted ligand interactions were made possible by the use of the self-organizing maps (SOMs).

The main advantage of the distance matrix based SOM, compared the usage of Cartesian coordinates,[37] is that the clustering is independent of any structural alignment. This is of major importance to cluster structures involving large conformational changes as in protein folding studies.[40]

However, distance matrices are highly redundant, and PCA compression[21] was used to reduce data size. Finally, the SOM algorithm, applied to PCA compressed distance matrices,[19,20] gives rise to a conformational clustering method that is independent of any choice of reference conformation, or any coordinate RMSD calculation.

Another advantage of self-organizing maps is that they provide a simplified description of the conformational space of a protein, without having to choose specific variables describing the principal motions.

However, a limitation in the interpretation of the U-matrix in terms of free energy landscape and transition state ensembles arises in the present study from the length of the molecular dynamics trajectories. 25 ns is a short time interval compared to the timescales usually simulated when one performs a full analysis of the free energy landscape for the system.[?] A quantitative analysis of the convergence of the trajectories in each basin determined from the SOM clustering was attempted by using the cosine content.[?] Values of 0.105, 0.813 and 0.929 are respectively obtained on the trajectory VanA and on the two time intervals of 0-16.2 and 16.2-25 ns of the trajectory $\mathrm{VanA_{SS}}$ (Figure 3b), before and

after opening of the loop ω. The small value obtained on the trajectory VanA agrees with the short timescale of the oscillatory motion observed for the ω loop in this trajectory. In contrast, along $\mathrm{VanA_{SS}}$, more complex dynamical behavior is observed, which is not dominated by one single motion. Because of this complexity, the motion timescales cannot be efficiently sampled during the short 25 ns trajectories recorded in the present work. As most of the trajectories are far from being converged, the prediction of free energy profiles from the conformational clustering by SOMs should thus be considered as being only qualitatively.

Nevertheless, in the particular case studied here, due to the existence of very relevant and different X-ray crystallographic structures from the TtDdl ligase, it was possible to obtain interesting insights into the free energy landscape of VanA.

The projection of the RMSD onto the SOM (Figure 5b) revealed a description of the conformational space dividing the set of conformations into distinct basins, in agreement with the global RMSD observation along MD trajectories (Figure 3). Furthermore, the transition structures between the basins can be detected by searching saddle points in the U-matrix.[41] Interestingly, these transition structures are favorable for the docking of the phosphinate tetrahedral-intermediate analog (Figure 9e).

The conformational clustering by SOMs gives a statistical picture of the MD simulation evolution. Through the preservation of the Boltzmann distribution by the molecular dynamics, the map resulting from the SOM algorithm contains information on the free-energy surface of the conformational space. One important feature of the SOM is to preserve the topological organization of the input space: closely related structures of the input space are grouped together in the SOM output space. Another trend is to distribute evenly data on the map so that apart from highly favorable or unfavorable zones the neu-

ron occupancy is homogeneously distributed. Hence, within the limits of the sampling completeness, SOMs seemed to provide a relevant delineation of free energy areas. The length of the simulations (25 ns) proved sufficient to offer significantly different docking specificities that could reflect the function of the ligands.

The relation between SOM clustering and conformation propensities suggests that SOMs could give a general framework for the definition of relevant reaction coordinates or collective variables allowing readily to project the evolution of MDs on the free energy topological map. Hence, SOM clustering appears attractive to analyze the conformational sampling in the framework of enhanced sampling methods.[42,43] The limits between free-energy basins are characterized by low populated areas, reflecting a low probability to access this conformation during the MD simulation. Since SOMs can be used to define free energy basins they readily allow the identification of the transition state ensembles by analysis of the flow given by the transfer vectors field, looking for null flow neurons implying an equiprobability to reach either close-by basin as described by Bolhuis and Ding[38,41] and Vanden-Eijnden.[44]

SOM analysis can also simply relate protein conformation to the ligand binding propensity by projecting of ligand docking scores on the conformational 2D-SOM. Pose classifications agreed with the ligand function, which supports the coherence of docking and scoring. These results validate the docking protocol as a specific tool to identify potential inhibitors of the D-Ala:D-Lac ligase. Furthermore, this analysis allows to choose the most relevant conformations to search for specific inhibitors by virtual screening. In the frame of the docking study on D-Ala:D-Lac ligase, taking into account the ω loop flexibility was essential to cluster ligands according to their functions, in agreement with results recently obtained by molecular docking on D-Ala:D-Lac ligase from *Leuconostoc*

mesenteroides.[?]

6 Conclusion

Molecular dynamics simulations of the D-Ala:D-Lac ligase was used to investigate the substrates binding mechanism. First, it appeared that the presence of a disulfide bridge between cysteines C64 and C52 induced the opening of the ω-loop and of the opposite domain, which is essential for unhindered entrance in the ligase catalytic site. Second, the development of an original clustering approach delineated the early steps of the opening mechanism and helped to identify representative conformations of this transition. The docking of known ligands on these representative conformations unraveled the relation between conformation and docking propensity in agreement with the ligand function.

We propose self-organizing maps as a general method for relating conformational transition of biomolecules and ligand docking poses.

This paves the way for the selection of appropriate binding site and pocket conformations for the search of D-Ala-D-Lac inhibitors. Furthermore, the conformation clustering can be related to the definition of system free-energy basins along MD simulations, and could thus be of interest in the frame of enhanced sampling and conformational free energy landscape simulations.

References

[1] Courvalin, P. Vancomycin resistance in gram-positive cocci. *Clinical Infectious Diseases* **2006**. 42, S25–S34.

[2] Reynolds, P. E. Structure, biochemistry and mechanism of action of glycopeptide antibiotics. *European Journal of Clinical Microbiology and Infectious Diseases* **1989**. 8, 943–950.

[3] Arthur, M.; Molinas, C.; Bugg, T.; Wright, G.; Walsh, C.; Courvalin, P. Evidence for in vivo incorporation of d-lactate into peptidoglycan precursors of vancomycin-resistant enterococci. *Antimicrobial agents and chemotherapy* **1992**. 36, 867–869.

[4] Reynolds, P. E.; Snaith, H. A.; Maguire, A. J.; Dutka-Malen, S.; Courvalin, P. Analysis of peptidoglycan precursors in vancomycin-resistant enterococcus gallinarum bm4174. *Biochemical Journal* **1994**. 301, 5.

[5] Arthur, M.; Reynolds, P.; Courvalin, P. Glycopeptide resistance in enterococci. *Trends in microbiology* **1996**. 4, 401–407.

[6] Kitamura, Y.; Ebihara, A.; Agari, Y.; Shinkai, A.; Hirotsu, K.; Kuramitsu, S. Structure of D-alanine-D-alanine ligase from Thermus thermophilus HB8: cumulative conformational change and enzyme-ligand interactions. *Acta Cryst D* **2009**. 65, 1098–1106.

[7] Meziane-Cherif, D.; Saul, F.; Haouz, A.; Courvalin, P. Structural and Functional Characterization of VanG D-Ala: D-Ser Ligase Associated with Vancomycin Resistance in Enterococcus faecalis. *J Biol Chem* **2012**. 287, 37583–37592.

[8] Roper, D.; Huyton, T.; Vagin, A.; Dodson, G. The molecular basis of vancomycin resistance in clinically relevant enterococci: crystal structure of D-alanyl-D-lactate ligase (VanA). *Proc of the Natl Acad of Sci* **2000**. 97, 8921–8925.

[9] Case, D.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005**. 26, 1668–1688.

[10] Case, D.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J Computat Chem* **2005**. 26, 1668–1688.

[11] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins* **2006**. 65, 712–725.

[12] Wang, J.; Wang, W.; Kollman, P.; Case, D. Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc* **2001**. 222, U403.

[13] Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. Development and testing of a general AMBER force field. *J Comp Chem* **2004**. 25, 1157–1174.

[14] Jorgensen, W. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J Am Chem Soc* **1981**. 103, 335–340.

[15] Pearlman, D.; Case, D.; Caldwell, J.; Ross, W.; Cheatham, T.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free

energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**. 91, 1–41.

[16] Berendsen, H.; Postma, J.; Gunsteren, W. V.; DiNola, A.; Haak, J. Molecular dynamics with coupling to an external bath. *J Chem Phys* **1984**. 81, 3684–3690.

[17] Nam, K.; Gao, J.; Darrin, M. An efficient linear-scaling Ewald method for long-range electrostatic interactions in combined QM/MM calculations. *J Chem Theor Comput* **2005**. 1, 2–13.

[18] Ryckaert, J.; Ciccotti, G.; Berendsen, H. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comp Phys* **1977**. 23, 327–341.

[19] Kohonen, T. Self-Organized formation of topologically correct feature maps **1982**. 43, 59–69.

[20] Kohonen, T. *Self-Organizing Maps*. Springer Series in Information Sciences, Heidelberg, Germany., **2001**.

[21] Kloczkowski, A.; Jernigan, R.; Wu, Z.; Song, G.; Yang, L.; Kolinski, A.; Pokarowski, P. Distance matrix-based approach to protein structure prediction. *J Struct and Funct Genom* **2009**. 10, 67–81.

[22] Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **2004**. 25, 1605–1612.

[23] http://www.chemaxon.com/products/marvin/marvinsketch.

[24] Shoichet, B.; Bodian, D.; Kuntz, I. Molecular docking using shape descriptors. *J Comput Chem* **1992**. 13, 380–397.

[25] Meng, E.; Shoichet, B.; Kuntz, I. Automated docking with grid-based energy evaluation. *J Comput Chem* **1992**. 13, 505–524.

[26] Lang, P.; Brozell, S.; Mukherjee, S.; Pettersen, E.; Meng, E.; Thomas, V.; Rizzo, R.; Case, D.; James, T.; Kuntz, I. DOCK 6: Combining techniques to model RNA–small molecule complexes. *RNA* **2009**. 15, 1219–1230.

[27] Richards, F. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* **1977**. 6, 151–176.

[28] Connolly, M. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**. 221, 709–713.

[29] Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. A geometric approach to macromolecule-ligand interactions. *J Mol Biol.* **1982**. 161, 269–88.

[30] Srinivasan, J.; Cheatham, T.; Cieplak, P.; Kollman, P.; David, A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J Am Chem Soc* **1998**. 120, 9401–9409.

[31] Kollman, P.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, G.; Cieplak, P.; Srinivasan, J.; Case, D.; Cheatham, T. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* **2000**. 33, 889–897.

[32] Hawkins, G.; Cramer, C.; Truhlar, D. Pairwise solute descreening of solute charges from a dielectric medium. *Chem Phys Lett* **1995**. 246, 122–129.

[33] Hawkins, G.; Cramer, C.; Truhlar, D. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem* **1996**. 100, 19824–19839.

[34] Rizzo, R.; Aynechi, T.; David, A.; Kuntz, I. Estimation of absolute free energies of hydration using continuum methods: accuracy of partial charge models and optimization of nonpolar contributions. *J Chem Theo Comput* **2006**. 2, 128–139.

[35] Ultsch, A. Maps for the visualization of high-dimensional data spaces. In *Proc. Workshop on Self organizing Maps*. pages 225–230.

[36] Ultsch, A. *U*-matrix: a tool to visualize clusters in high dimensional data*. Fachbereich Mathematik und Informatik, **2003**.

[37] Fraccalvieri, D.; Pandini, A.; Stella, F.; Bonati, L. Conformational and functional analysis of molecular dynamics trajectories by Self-Organising Maps. *BMC bioinformatics* **2011**. 12, 158.

[38] Ding, F.; Dokholyan, N.; Buldyrev, S.; Stanley, H.; Shakhnovich, E. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys J* **2002**. 83, 3525–3532.

[39] Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J Mol Graph* **1996**. 14, 33–38.

[40] Spill, Y.; Bouvier, G.; Nilges, M. A convective replica-exchange method for sampling new energy basins. *J Comput Chem* **2013**. 34, 132–140.

[41] Bolhuis, P.; Chandler, D.; Dellago, C.; Geissler, P. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Ann Rev Phys Chem* **2002**. 53, 291–318.

[42] Lei, H.; Duan, Y. Improved sampling methods for molecular simulation. *Curr Opin Struct Biol* **2007**. 17, 187–191.

[43] Mitsutake, A.; Mori, Y.; Okamoto, Y. Enhanced sampling algorithms. *Methods Mol Biol* **2013**. 924, 153–195.

[44] Vanden-Eijnden, W.; Ren, W.; Vanden-Eijnden, E. Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem Phys Lett* **2005**. 413, 242–247.

[45] Hrast, M.; Vehar, B.; Turk, S.; Konc, J.; Gobec, S.; Janezic, D. Function of the D alanine: D-alanine ligase lid loop: a Molecular Modeling and Bioactivity Study. *J Med Chem* **2012**. 55, 6849–6856.

List of Figures

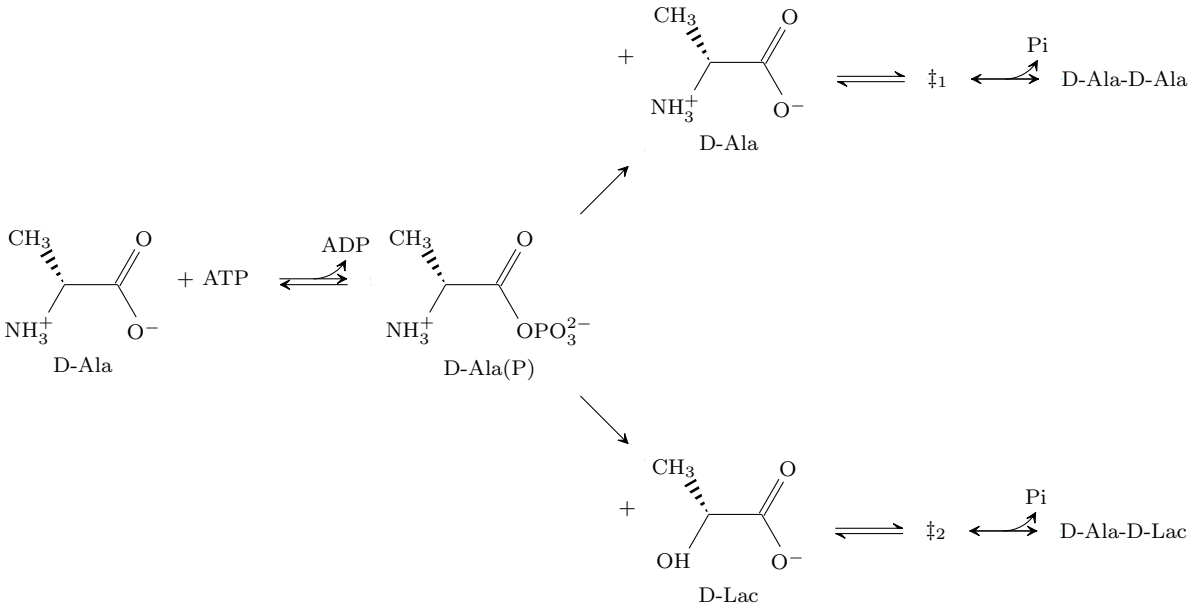


Figure 1: Enzymatic reaction of a D-Ala:D-Ala ligase (Ddl), upper branch, and D-Ala:D-Lac ligase (VanA) on the lower branch. The transition state analog, phosphinate (PHY), mimics the tetrahedral intermediate $\ddagger_2$.

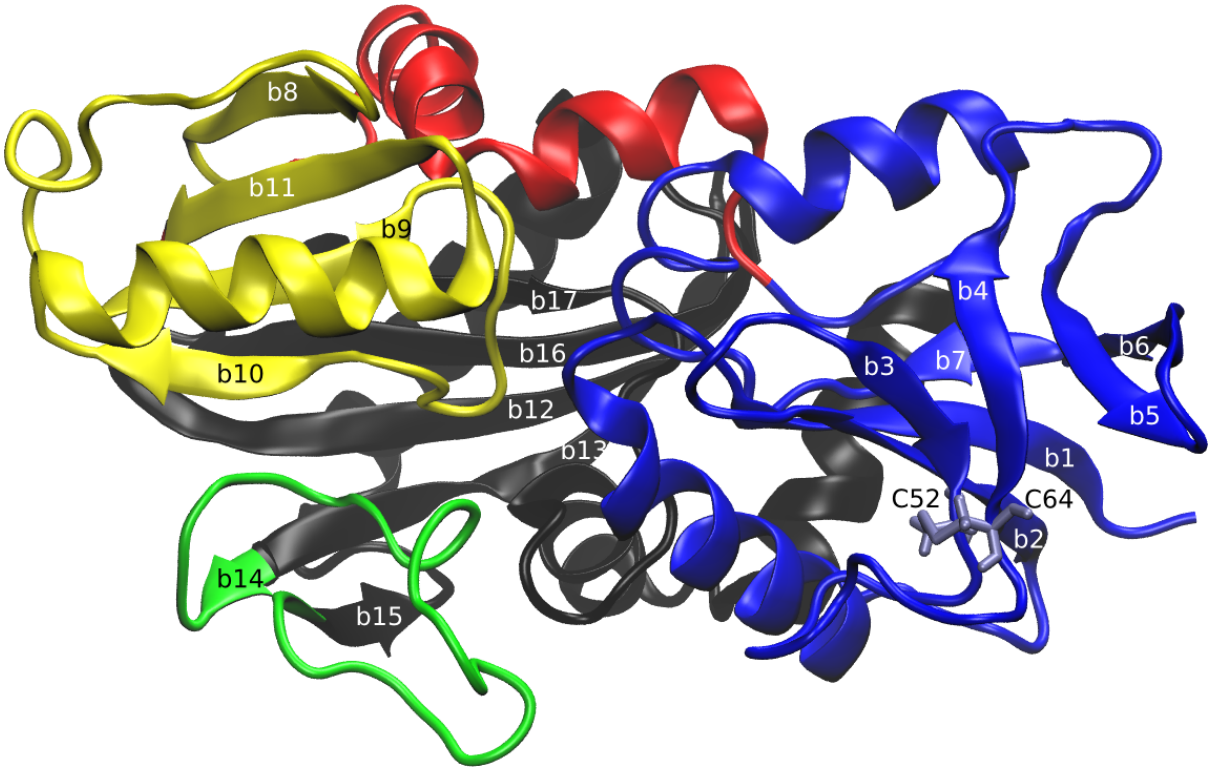


Figure 2: 3D X-ray crystallographic structure of VanA (PDB entry: 1E4E) colored according to its domains: the N-terminal [A2-G121] in blue, the C-terminal [G212-A342] in black, which includes the ω-loop [L236-A256] in green, and the central domain [C122-S211] in red, which includes the opposite domain [A149-Q208] in yellow. The disulfide bridge C52-C64, located in the N-terminal domain, is colored in pale blue (bottom right).

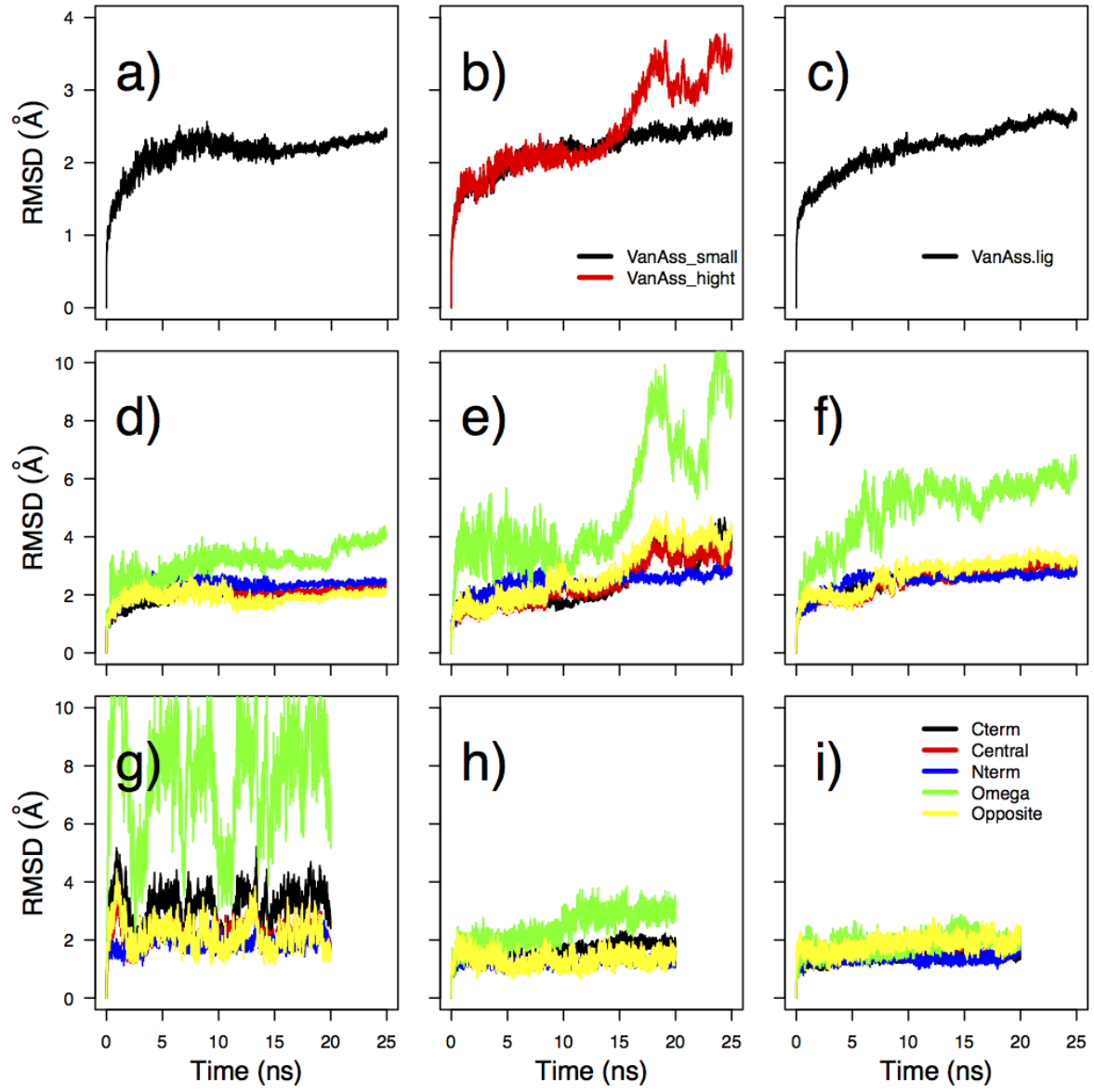
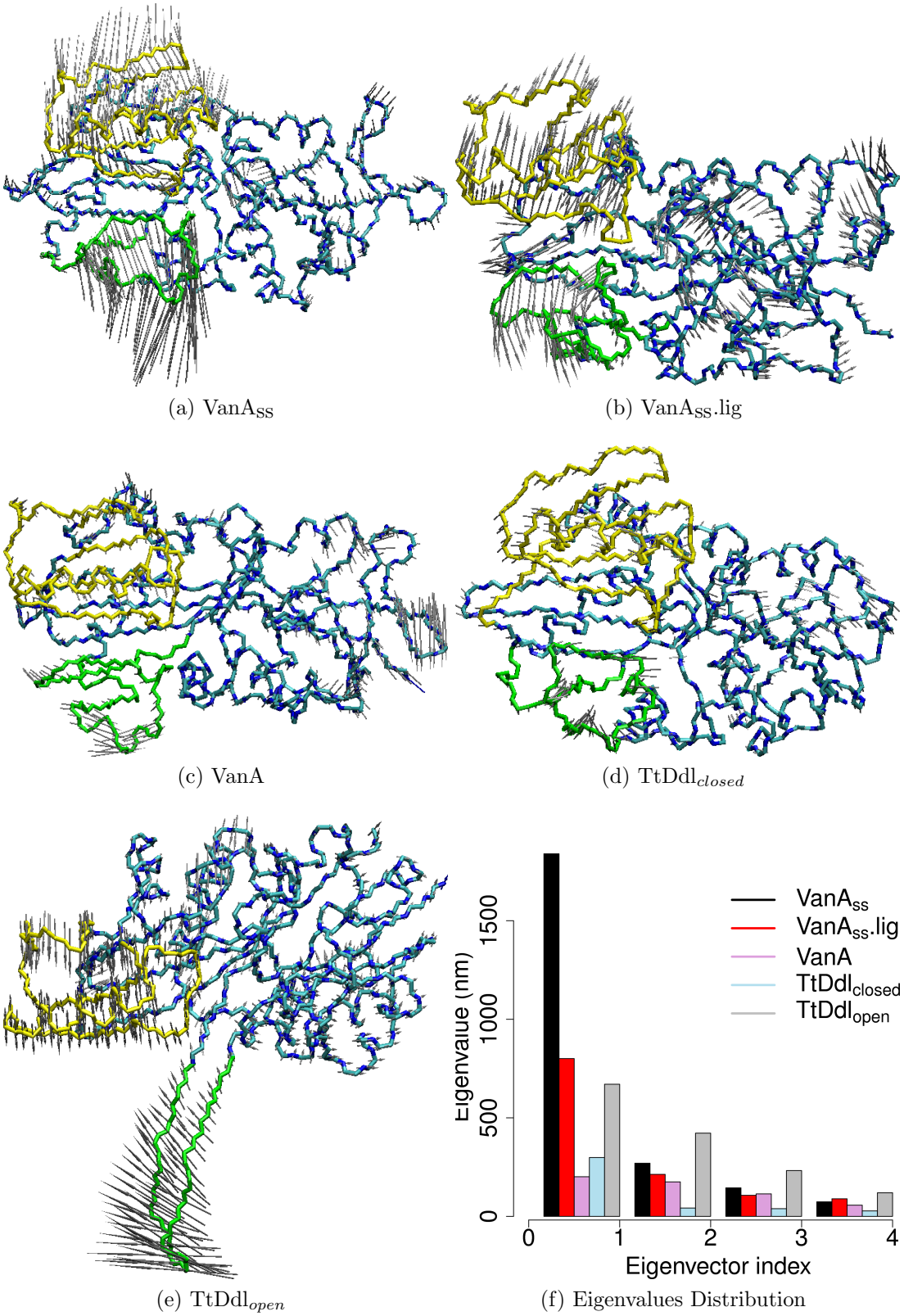


Figure 3: (a-c) Global conformational drifts, RMSD from the first conformations calculated on Cα coordinates for: a) VanA (averaged over 8 MD trajectories), b) VanA$_{SS}$ (averaged over 3 MD trajectories for the red curve and over 4 trajectories for the black one), c) VanA$_{SS}$.lig (averaged over 9 trajectories). (d-i) Drifts of ligases domains computed for one representative trajectory and colored according to the caption given on panel i. d) VanA, e) VanAss for the trajectory with the largest drift (VanAss_hight), f) VanA$_{SS}$.lig, g) TtDdl$_{open}$, h) TtDdl$_{closed}$, i) TtDdl$_{closed}$.lig.



(a) VanA$_{SS}$ (b) VanA$_{SS}$.lig

(c) VanA (d) TtDdl$_{closed}$

(e) TtDdl$_{open}$ (f) Eigenvalues Distribution

Figure 4: **P**rincipal **C**omponent **A**nalysis (PCA) of the Cα dynamics covariance matrix for MD simulations run on VanA and TtDdl. The ω-loop is colored in green, its opposite domain in yellow and the remaining parts of the protein in blue. (a-e): projection of the first mode on the 3D structures, f) eigenvalues distribution.

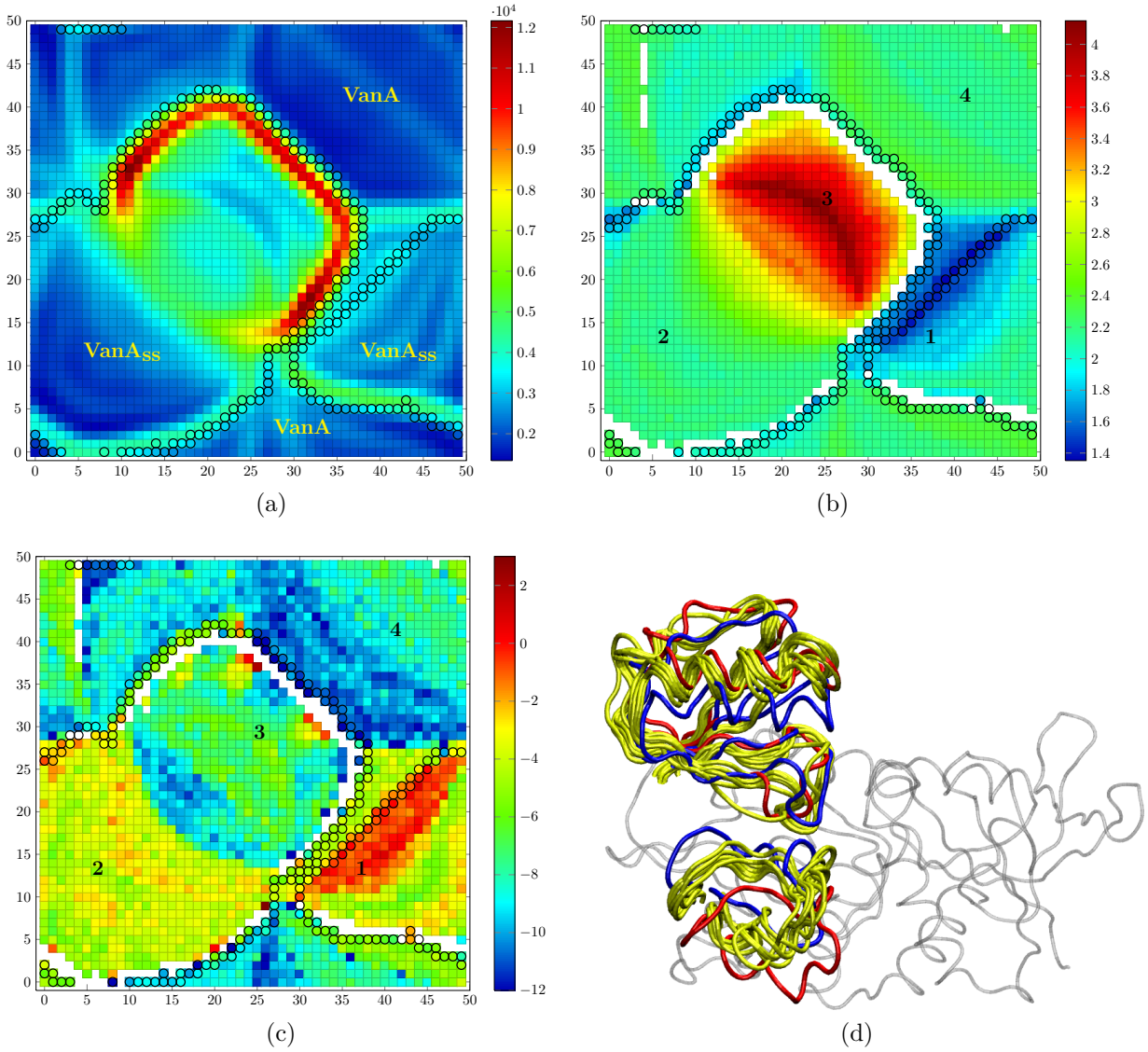


Figure 5: **a**) U-matrix for the SOM used to analyze VanA and VanA$_{SS}$ trajectories. The map is toric. Labels on the U-matrix show which system mapped the different SOM areas. Black circles mark the VanA trajectory border. **b**) Projection of the RMSD values (Å) relatively to the initial conformation. Numbers show the initial conformation region (1), the main low U-matrix VanA$_{SS}$ basin (2) and the high U-matrix one (3). **c**) Projection of β-strands content variation (current number minus initial one). **d**) Superposition of the first conformation (blue), of the last one (red) and of the transition states conformations (as defined by the flow analysis, yellow) of VanA$_{SS}$. The colors correspond to the color-map of the RMSD projection matrix.

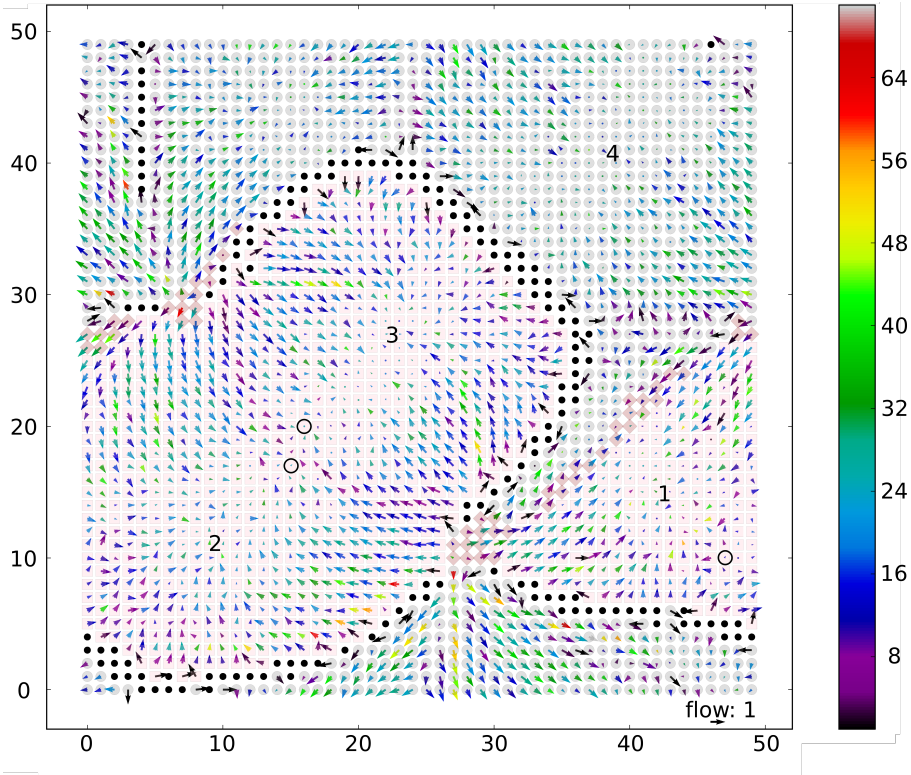
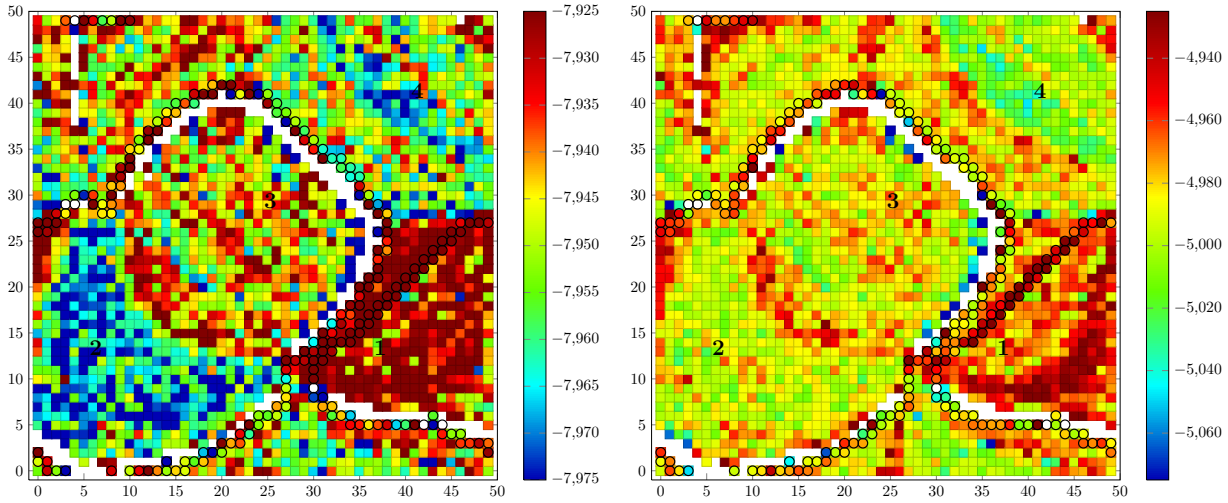


Figure 6: Flow analysis of the MD trajectories. VanA and VanA_{SS} trajectories are underlined by gray circles and pink square respectively. The intersection between the two trajectories is delimited by brown diamonds. The three basins of VanA_{SS} are numbered. The transition states ensembles of VanA_{SS} are pointed out with black circles. Black dots stand for unvisited neurons. The color code of the arrow gives the density ($f_{i,j}$) of each neuron, using the scale given at the right of the plot. The orientation of each vector indicates the resulting flow of the MD. The norms of the vectors are linked to the polarity of the corresponding flow. Zero-normed vectors are depicted by small black dots.



(a) MM-GB-SA (b) MM-PB-SA

Figure 7: Projection of the MM-PBSA (a) and MM-GBSA (b) energies (kcal.mol^{-1}) on the U-matrix obtained from the molecular dynamics trajectories VanA and VanA$_{SS}$. Black circles mark the VanA trajectory border.

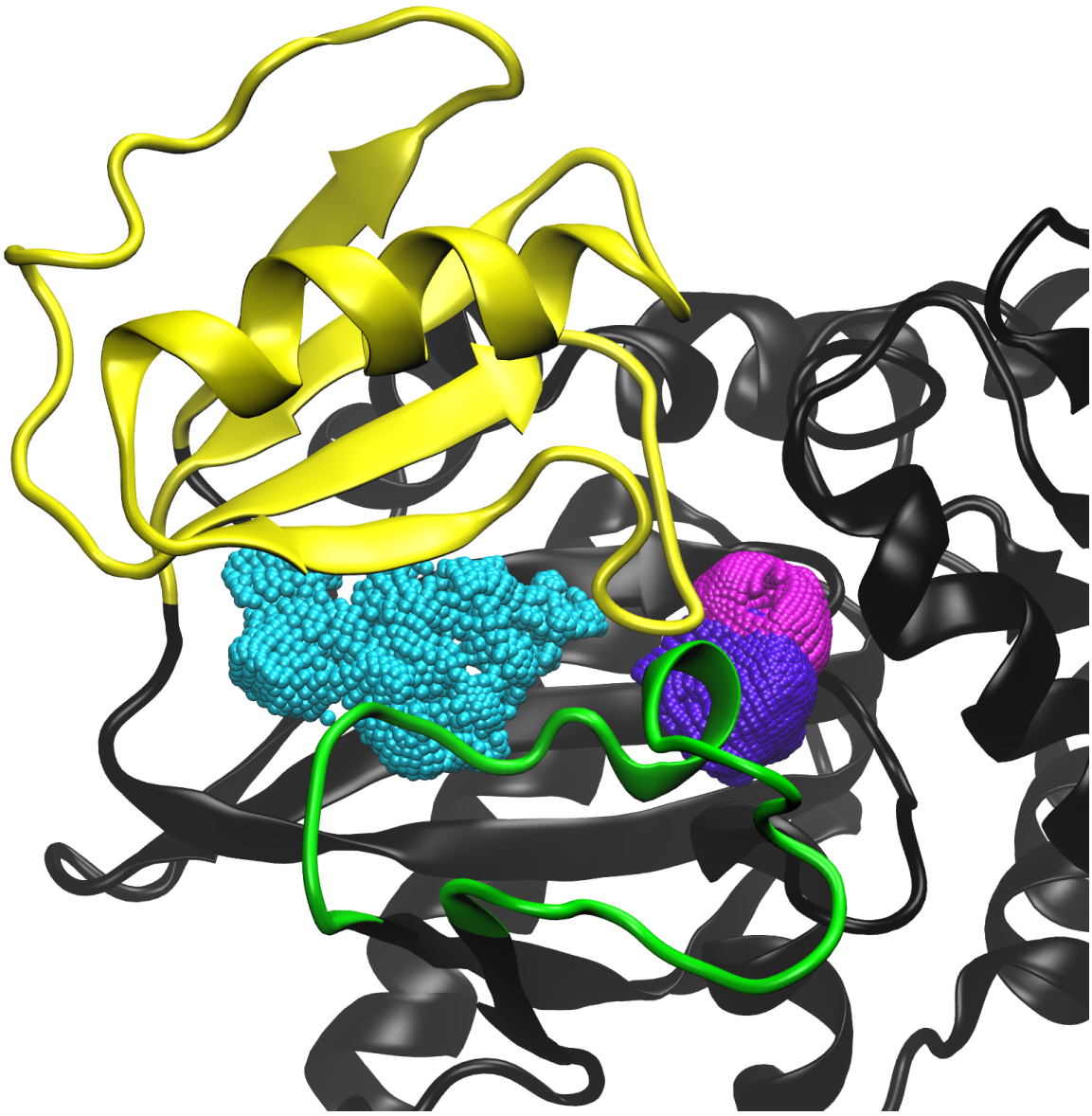
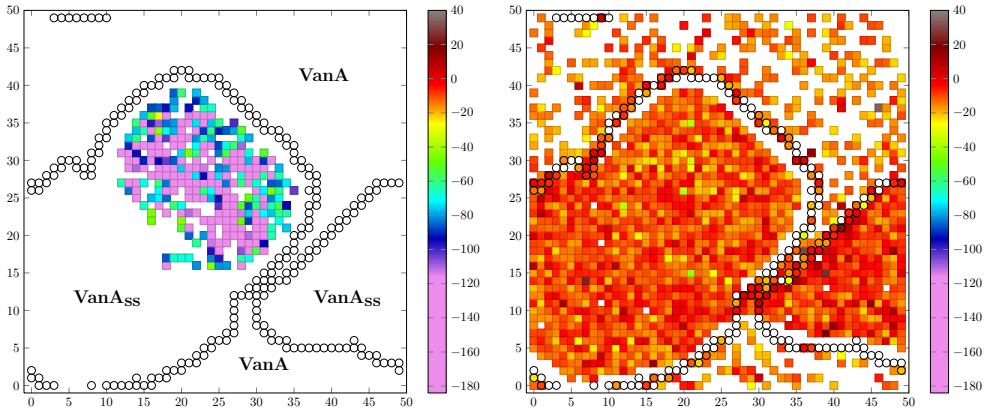
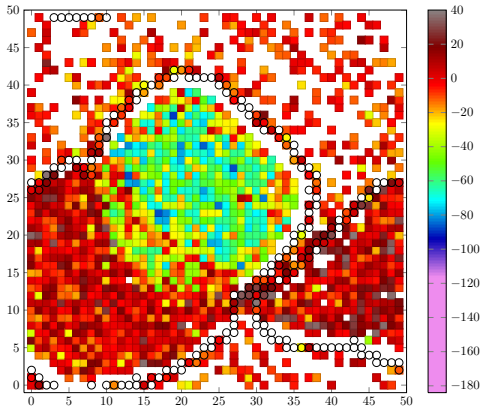


Figure 8: Binding sites calculated with the SOM 3D algorithm on the run docking poses. The ligand coordinates associated with each neurone is drawn as cpk, ADP binding site is colored in cyan, D-Ala in magenta and D-Lac in purple. The figure was prepared with VMD.[39]

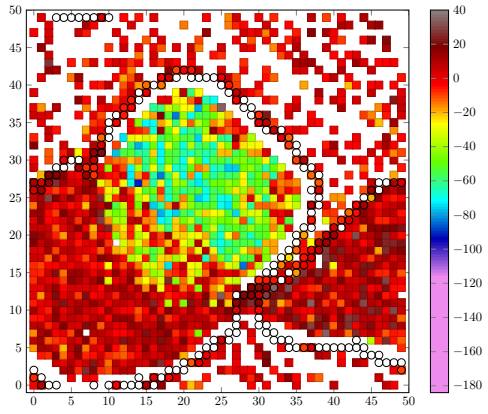


(a) ATP

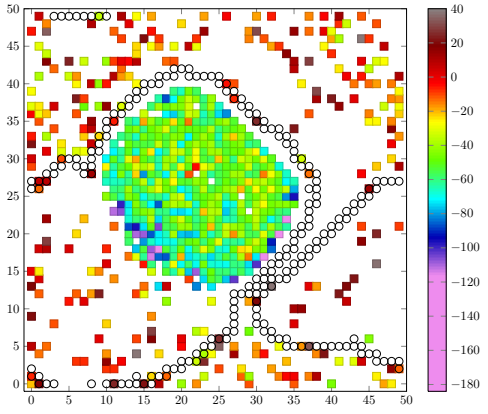
(b) D-Ala



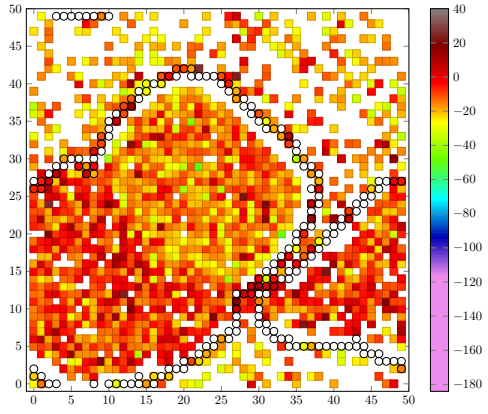
(c) D-Ala(P)



(d) D-Lac



(e) PHY



(f) D-Ala-D-Lac

Figure 9: Docking of key ligands involved in the ligase mechanism. VanA and VanA$_{SS}$ conformations were extracted from the SOM clustering. Black circles border VanA trajectory and the areas are labeled on ATP plot. The GBSA scores expressed in kcal.mol^{-1} were used to approximate the ligand binding free energy.

List of Tables

PDB	Name	Ligands	Counterions	Number of water molecules	Number of recorded trajectories
2YZG	$TtDdl_{open}$	-	12 Na^+	13366	1
2ZDH	$TtDdl_{closed}$	-	12 Na^+	10854	1
2ZDH	$TtDdl_{closed}$.lig	ADP, D-Ala, 2 Mg^{2+}	11 Na^+	10853	1
1E4E	VanA	-	5 Na^+	13585	8
1E4E	VanA.lig	ADP, PHY, 2 Mg^{2+}	4 Na^+	13582	9
1E4E	$VanA_{SS}$	-	5 Na^+	13585	7
1E4E	$VanA_{SS}$.lig	ADP, PHY, 2 Mg^{2+}	4 Na^+	13582	9

Table 1: Systems used for MD simulations

MD simulations	X-ray crystallographic structures	
Trajectory name	PDB	Ligands in the pocket
$TtDdl_{open}$	2YZG	-
$TtDdl_{closed}$	2ZDH	ADP, D-Ala, 2 Mg^{2+}
$TtDdl_{closed}$.lig	2ZDH	ADP, D-Ala, 2 Mg^{2+}
VanA	1E4E	ADP, PHY, 2 Mg^{2+}
VanA.lig	1E4E	ADP, PHY, 2 Mg^{2+}
$VanA_{SS}$	1E4E	ADP, PHY, 2 Mg^{2+}
$VanA_{SS}$.lig	1E4E	ADP, PHY, 2 Mg^{2+}

Table 2: X-ray structures used in MD simulations