Elucidating the structural and conformational factors responsible for the activity and substrate specificity of alkanesulfonate monooxygenase

V. Ferrarioa, P. Braiucaa,d, P. Tessaroa,b, L. Knapića, C. Gruberb, J. Pleissb, C. Eberta, E. Eichhorn and L. Gardossia*

aDipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Trieste, P.le Europa 1, 34127 Trieste, Italy; bInstitute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany; cGivaudan Schweiz AG, Ueberlandstrasse 138, 8600 Dübendorf, Switzerland; dSPRIN s.p.a., via Flavia 23/1, 34148 Trieste, Italy

Communicated by Ramaswamy H. Sarma

(Received 9 November 2011; final version received 15 February 2012)

The mechanism and substrate specificity of alkanesulfonate monooxygenase (SsuD) was investigated by combining molecular dynamics simulations, docking, and a comprehensive quantitative structure activity relationships (QSAR) analysis. The FMNH2 dependent monooxygenase undergoes a dynamic conformational change of the active site, passing from a closed to an open state. As a consequence, substrates have access to the active site and the cofactor is then regenerated by the associated oxidoreductase FMN reductase SsuE. Computational analysis of the interaction of SsuD with FMNH2 based on molecular docking and multiple 20 ns molecular dynamics simulations pointed out that the conformational change is mainly driven by salt bridge formation between Arg297 and Glu20 or Asp111. A set of substrates accepted by SsuD were described by means of ALMOND chemical descriptors and a partial least square (PLS) mathematical model was constructed. The PLS model correlates the structure of substrates and enzyme activity, namely kinetic properties ($k_{cat}/K_M$). Therefore, information coming from the PLS analysis goes beyond the simple ability of the enzyme to recognize the substrate, but includes the factors that affect the capacity of the enzyme to reduce the activation energy of the rate determining step of the reaction. The two principal components of the model are able to describe both steric and electronic factors and, more importantly, their interactions. Indeed, interactions of factors appear to affect significantly the ability of SsuD of transforming efficiently a substrate.

Keywords: alkanesulfonate monooxygenase; molecular dynamic simulation; 3D-QSAR; substrate specificity

Introduction

The work reported here aims at elucidating the structural and conformational factors responsible for the activity and substrate specificity of alkanesulfonate monooxygenase (SsuD). In bacteria, sulfur is mainly assimilated from inorganic sulfate via the cysteine biosynthetic pathway. In nature, where the levels of inorganic sulfate may be low, bacteria have to rely on organosulfur compounds such as sulfate esters, sulfamates, and sulfonates as sulfur sources. Under conditions of sulfate starvation, *Escherichia coli* synthesizes two systems, which cover the full range of uptake and desulfonation activities. These activities enable the growth of the microorganism with taurine (TauABCD proteins) and with a variety of aliphatic sulfonates (SsuEACDB proteins) as sole sources of sulfur (van der Ploeg, Eichhorn, & Leisinger, 2001). Sulfite liberated inside the cells from taurine and alkanesulfonates serves as sulfur for cell growth.

SsuD (alkanesulfonate monooxygenase EC 1.14.14.5) is a FMNH2-dependent monooxygenase, which catalyzes the oxygenolytic cleavage of the C–S bond of 1-alkanesulfonates by monooxygenation leading to the release of the corresponding aldehyde and sulfite. Catalysis strictly depends on oxygen and on FMNH2, the latter being provided by the associated oxidoreductase SsuE (FMN reductase, EC = 1.5.1.29) (Eichhorn, van der Ploeg, & Leisinger, 1999). SsuD or SsuD-like enzymes (namely those enzymes showing high sequence identity to *E. coli* SsuD) were found in *Bacillus subtilis* (van der Ploeg, Cummings, Leisinger, & Conerton, 1998), *Pseudomonas* strains (Kahnert et al., 2000; Kertesz, Schmidt-Larbig, & Wüst, 1999), and *Corynebacterium glutamicum* (Koch et al., 2005). These enzymes have the function of sulfur scavenging and they accept varying ranges of organosulfur compounds.

The structure of the substrate-free SsuD from *E. coli* has been solved previously (Eichhorn, Davey, Sargent, *Corresponding author. Email: gardossi@units.it*
Leisinger, & Richmond, 2002), but for a detailed mechanistic study the structure of an enzyme–substrate complex is crucial. SsuD enzymes are of great relevance in diverse bacteria for organosulfur assimilation. This is demonstrated by the numerous protein sequences having high identity with \textit{E. coli} SsuD and that can be derived also from open reading frames deposited in databases. Therefore, understanding substrate recognition and cofactor interaction is of high interest in order to clarify the distinctive role of this widespread enzyme in C–S bond cleavage.

In the attempt of elucidating the catalytic mechanism, previous studies of SsuD and SsuE from \textit{E. coli} addressed the transfer of the reduced flavin from the SsuE reductase to the oxygenase SsuD. Spectroscopic experiments pointed out interactions between SsuE and SsuD, suggesting a channeling mechanism for the transfer of the reduced flavin (Abdurachim & Ellis, 2006). Moreover, the alteration of the cofactor affinity of the SsuE in the presence of SsuD was also reported (Eichhorn et al., 1999). Zhan and co-workers proposed the formation of a C(4a)-hydroperoxylavin intermediate once FMNH$_2$ reacts with O$_2$ (Zhan, Carpenter, & Ellis, 2008). The complex would be directly involved in the desulfonation of alkanesulfonate, which leads to the release of sulfite and the aldehyde. The same intermediate was also observed in the mechanism of some flavin-dependent hydroxylases (Ballou, Entsch, & Cole, 2005).

Further contributions to the understanding of SsuD catalytic mechanism came from the comparison of a series of enzymes belonging to the same family of SsuD, namely the class C flavoprotein monoxygenases (van Berkel, Kamerbeek, & Fraaije, 2006). Flavoproteins of this class include enzymes involved in diverse biochemical reactions, like the biosynthesis of the antibiotics pristinamycin (Thibaut et al., 1995) and actinorhodin (Valton, Mathevon, Fontecave, Nivière, & Ballou, 2008), the degradation of natural aromatic compounds (Chaiyen, Suadee, & Wilairat, 2001), desulfurization of dibenzo-thiophene (Oldfield, Pogrebinsky, Simmonds, Olson, & Kulpa, 1997), nitrilotriacetate hydroxylation, (Uetz, Schneider, Snozzi, & Egli, 1992) and ethylenediaminetetraacetic acid (EDTA) degradation (Witschel, Nagel, & Egli, 1997). Among them, luciferase was studied extensively, both at biochemical and structural levels (Fisher, Thompson, Thoden, Baldwin, & Rayment, 1996) and such information was exploited for the interpretation of the structure of SsuD (Eichhorn et al., 2002). This enzyme presents a structural core with a TIM-barrel fold. The location of the active site was proposed at the end of the \textbeta-barrel (Eichhorn et al., 2002). The TIM-barrel fold is adopted by many enzymes for flavin binding and it has been found in the structure of all luciferases solved so far (Farber & Petsko, 1990). As a matter of fact, luciferase LuxAB, (Fisher et al., 1996) long-chain alkanemonooxygenase LadA (Li et al., 2008), F$_{420}$-dependent $N^5,N^{10}$-methylenetetrahydrodithynopterin reductase Mer (Shima et al., 2000), and F$_{420}$-dependent secondary alcohol dehydrogenase Adf from methanogenic\textit{Archaea} (Aufhammer et al., 2004) use reduced flavin as the flavin analog F$_{420}$ as cofactors. All these enzymes show high structural similarity (Li et al., 2008). In SsuD, luciferase, and LadA there are highly conserved amino acid residues. SsuD presents Cys54, His228, Arg297, and Tyr331, which correspond to Cys106, His44, Arg291, and Tyr110 in LadA and Cys14, His311, and Tyr63 in luciferase. The importance of these conserved residues for the mechanism of action was confirmed by mutagenesis studies: mutation of His44 (Fisher et al., 1996) in luciferase causes the loss of activity, whereas LadA loses its activity upon single mutations either of His311, Tyr63, or Cys14 (Li et al., 2008).

It must be underlined that SsuD enzymes are synthesized under sulfate-starvation conditions and they show a very low content in sulfur-containing amino acids. The observation that Cys54 of \textit{E. coli} SsuD is remarkably conserved suggests a relevant role of this amino acid in the catalytic mechanism. Recent mutagenesis studies indicate that Cys54 of \textit{E. coli} SsuD might contribute to the stabilization of the C(4a)-hydroperoxylavin intermediate formed during catalysis (Carpenter, Zhan, & Ellis, 2010). This is in agreement with the observation that the thiol of Cys106 of luciferase points towards position C(4a) of the isoalloxazine ring of the cofactor in the crystal structure of the FMN bound enzyme (Campbell, Weichsel, Montfort, & Baldwin, 2009). It was also demonstrated that Hg complexation of Cys54 in SsuD does not affect its activity upon single mutations either of Cys106, Tyr63, or Cys14 (Li et al., 2008).

A study concerning substrate specificity of SsuD reported that the pure enzyme catalyzed the conversion of pentanesulfonic acid to sulfite and pentaldehyde and was able to desulfonate a wide range of sulfonated substrates including C-2–C-10 unsubstituted linear alkanesulfonates, substituted ethanesulfonic acids, and sulfonated buffers (Eichhorn et al., 1999).

Since crystals of a SsuD-alkanesulfonate complex have not yet been obtained (Eichhorn et al., 2002), in the present study the mechanism of SsuD was investigated by computational methods. The analysis included the interactions of the enzyme with the substrates and with the reduced cofactor. The aim was to verify whether the preliminary binding of FMNH$_2$ is required for the subsequent recognition of the alkanesulfonate by a catalytically active conformation. Molecular docking and molecular dynamics were used for the calculation of enzyme–alkanesulfonate complexes, which were subsequently used for calculating structure activity relationships (QSAR) (Braiuca, Boscariol, Ebert, Linda, & Gardossi, 2006; Braiuca, Ebert, Basso, Linda, & Gardossi, 2009; Braiuca, Knacic, Ferrario, Ebert, & Gardossi,
The statistical analysis allowed pointing out the structural basis of the substrate specificity of the enzyme.

**Materials and methods**

The protein structures used for this study were retrieved from the Protein Data Bank (PDB) (Berman et al., 2000) (Id: 1m41 Eichhorn et al., 2002 for SsuD, 1luc (Gao & Ellis, 2005) for bacterial luciferase, 3b9o (Li et al., 2008) for LadA, and 1z69 (Aufhammer et al., 2005) for Mer). These initial structures were pretreated in Molecular Operating Environment (MOE) by removing the crystallographic water molecules present in the PDB file. Hydrogen atoms were added and their positions were optimized with an energy minimization procedure in the Amber99 force field in its MOE implementation. Subsequently, a minimization of the side chains was performed keeping the backbone atoms fixed.

Superimpositions of protein structures were performed using the MOE Superpose tool fixing the spatial position of the most preserved residues (His, Trp, Arg) in all the superimposed structures. The alignment was performed using the blosum30 matrix (Henikoff & Henikoff, 1992).

**GRID**

The GRID program (Goodford, 1985) is a computational procedure that calculates the interaction energies between the substrate and a small chemical probe (e.g., a functional group). Energies are calculated in all the nodes of a three-dimensional grid, which spans the structure of the substrates considered. The output is called Molecular Interaction Field (MIF), which can be visualized as an isopotential surface. The GRID analysis was performed on the SsuD protein, choosing a cage big enough to include the whole protein. The grid nodes were set every 0.5 Å. The probes used for the calculation of the MIFs were DRY (hydrophobic probe), WATER (H2O probe), and OS (sulfonic acid mimic probe).

**Docking**

The Docking procedure was performed using the docking module of MOE (2006).

In the case of FMNH2, an area of 1 nm radius was selected surrounding the active site, taking as references the coordinates of active His, Trp, and Arg residues. The force field used for the docking was MMFF94x, the partial charges of the atoms were calculated at the PM3 semi-empirical level, by means of the MOPAC7 program and the different poses were placed with the alpha PMI method and scored by means of the London dG scoring function (MOE, 2006). The final FMNH2 pose was finally located by evaluation of the scoring function (London dG) and the pose similarity with the FMN cofactor present in the PDB structure of LadA (PDB 3b9o).

The docking simulations of the substrates were performed in the same way considering a 1.2 nm radius selected area surrounding the FMN; the initial positions of the substrates were manually set in order to meet the three different criteria: (i) sulfonic group located in a OS probe interacting area; (ii) sulfonic group distance from FMNH2; and (iii) London dG scoring function. The substrate molecules were inserted by building each molecule with the MOE builder tool and subsequently minimized. For each substrate, the conformation presenting the highest score and fulfilling the structural requirements for the initiation of the enzymatic catalysis was chosen.

**Molecular dynamics**

The molecular dynamics simulations were performed using the software GROMACS 4 (Berendsen, van der Spoel, & van Drunen, 1995) with the GROMOS-96 53a6 force field (Oostenbrink, Villa, Mark, & van Gunsteren, 2004). The molecular dynamics simulations were performed using the software GROMACS 4 with the GROMOS-96 53a6 force field. The SsuD PDB crystal structure (PDB code 1m41) was implemented in the force field in gro file format by using the tool pdb2gmx of the GROMACS software package, which also adds the necessary hydrogens. The protein dimer was solvated with explicit SPC water in a virtual box of 1331 nm³. All the dynamics were performed in a NPT environment simulating the temperature of 300 K and keeping the pressure constant (Berendsen-thermostat and pressure (Berendsen et al., 1995), cut-off for electrostatic interaction was set to 1.4 nm and the limit for the van der Waals interactions set to 1.4 nm. For the minimization procedures, the Particle Mesh Ewald (PME) algorithm (Essmann et al., 1995) was used for the calculation of the electrostatic interactions setting the limit at 1.0 nm. The FMNH2 and the substrate molecules were parameterized in the GROMOS-96 53a6 force field by using the Dundee PRODRG2 server (Shuettelkopf & van Aalten, 2004) and manually refined in order to meet the correct force field definition. The FMNH2 and the substrate molecules were manually added in the gro file taking the coordinate from the docking results. The system was previously minimized with 10,000 step of steep descendent algorithm before every molecular dynamics calculation.

Molecular dynamics analysis were performed with GROMACS tools; distances measurements were calculated using g_mindist and computing the minimum distance between two residues; the results were visualized using Grace software (http://plasma-gate.weizmann.ac.il/Grace).

Each substrate conformation chosen for the construction of the data-set for the QSAR analysis (Cruciani, Clementi, & Baroni, 1993) was the one characterized by the lowest potential energy out of all the frames saved in the dynamics trajectories.
ALMOND

Four different probes were used for the calculation of the MIFs, namely the WATER, the DRY, the Ca++, and the TIP probes. The WATER probe describes and quantifies the dipolar interactions and the hydrogen bond formation; the DRY probe considers all the hydrophobic interactions; the Ca++ probe takes into account the interactions with charged groups (e.g. sulfonic group); and the TIP probe generates a MIF that is strictly dependent on the shape of the molecule (ALMOND v. 3.3.0).

The GRid INdependent Descriptors (GRIND) descriptors (ALMOND v. 3.3.0) are vectors connecting those points of the MIFs representing the most favorable interactions and they are generated under the condition of having the highest energy products.

The GRIND vectors can connect points belonging to the same MIF or to different MIFs and the ensemble of the vectors generates auto-correlograms and cross-correlograms, respectively.

The four probes used for the GRIND analysis generated four auto-correlograms and six cross-correlograms, corresponding to 600 independent variables (vectors). The GRIND vectors were further selected on the basis of their ability to discriminate between objects by using the tools of the ALMOND software, and consequently by discarding those vectors containing insufficient information in terms of ability to discriminate among substrates.

This operation resulted in the deletion of all the vectors of the Ca++/Ca++ auto-correlogram, because the Ca++ probe generates MIFs only in correspondence of the sulfonic group which is present in every object of the data-set. A further variable selection was accomplished by using the FFD variable selection algorithm (ALMOND v. 3.3.0), which allowed retaining only 217 vectors, corresponding to the most informative variables.

It must be underlined that before this procedure the predictivity of the model was <0.10.

Partial least square (PLS) models with 5 principal components were computed and validated by leave-one-out (LOO) method (ALMOND v. 3.3.0). QSAR substrate predictivity model is expressed in terms of experimental vs. predicted (on two principal components) $k_{cat}/K_m$.

Results and discussion

Analyzes of interactions between SsuD and cofactor

The crystal structure of SsuD was taken from the PDB (Berman et al., 2000). The structure is available since July 2002 under the entry code 1m41 and it has resolution of 2.3 Å. SsuD appears as a homodimer (subunits A and B), where each subunit consists in an eight-stranded α/β-barrel motif. This fold classifies SsuD as a member of the big (β/α)8 TIM-barrel family (Eichhorn et al., 2002). A GRID (Goodford, 1985) analysis of the enzyme surface indicated a prevalence of hydrophilic regions, which is in agreement with the cytosolic origin of the protein (see supplementary material Figure S1). Several hydrophobic areas are spread on the surface. The most relevant one is located at the entrance of the active site and it might facilitate the approach of the substrate to the catalytic site, while reducing the entropic penalty caused by the expulsion of water generated during catalysis.

SsuD is a FMNH2-dependent enzyme, but in the 1m41 (Eichhorn et al., 2002) PDB structure the reduced cofactor is not co-crystallized. Although the overall structure of the two subunits is similar, there are major conformational changes due to interactions of Arg297 with either Asp111 or Glu20. The switching of these interactions corresponds to a conformational transition from the closed to the open conformation and suggests a dynamic mechanism mediated by Arg297 (see supplementary material Figures S2 and S3). As a consequence, subunit A is in a closed conformation where the active site is not accessible by the solvent, while subunit B is in an open conformation.

Because the factors responsible for the conformational transition cannot be directly elucidated from the analysis of the cofactor-free apoprotein structures, a more detailed study of enzyme–cofactor interactions was performed. To model the structure of the SsuD-FMNH2 complex, SsuD was compared to the structures of two highly homologous FMNH2-dependent enzymes belonging to the (β/α)8 TIM-barrel family: bacterial luciferase co-crystallized with FMN (PDB 1luc) (Gao & Ellis, 2005) and long-chain alkane monoxygenase (LadA) co-crystallized with FMN (PDB 3b9o) (Li et al., 2008). Conversely, similarities of sequence and tertiary structure were analyzed. The three enzymes share a homologous organization of the active site and are expected to have similar catalytic mechanisms (Fisher et al., 1996; Li et al., 2008). In particular, the superimposition of structures 1luc, 3b9o, and 1m41 showed three highly conserved residues: His111, Trp196, and Arg127 for SsuD; His44, Trp194, and Arg107 for bacterial lucifer-
asc; and His311, Trp348, and Arg157 for LadA. The spatial position of the Arg residue is the most conserved (see supplementary material Figure S4) and previous studies on bacterial luciferase suggest a relevant role of Arg107 in coordinating and stabilizing the phosphate group of the cofactor (Campbell et al., 2009).

It is interesting to note that the analysis of the methylenetetrahydromethanopterin reductase (Mer) co-crystallized with its cofactor F420 (PDB 1z69) (Aufhammer et al., 2005) revealed the absence of the corresponding Arg residue, even though Mer belongs to the same (β/α)8 TIM-barrel family. As a matter of fact, the bulkier F420 cofactor can fit into the cofactor binding domain of Mer thanks to the bending of a loop that increases the space available (see supplementary material Figure S5).

The conserved Trp residue is located at the center of the active sites of all four enzymes considered and it creates a hydrophobic pocket to accommodate the cofactors (see supplementary material Figure S6).

In bacterial luciferase, it is evident that the indolic ring of Trp194 stabilizes the cofactor by interacting with its isoalloxazine ring (Fisher et al., 1996). Moreover, it is also known that the hydrophobicity of the cofactor accommodation cavity, mainly due to the Trp residue, is necessary for the creation of a water-free environment that prevents rapid breakdown of unstable flavin intermediate (Li et al., 2008).

The position of the conserved His differs in the three superimposed enzymes, which can be ascribed to their different substrate specificities. In bacterial luciferase and LadA, the conserved His is positioned near the cofactor and mutations of this residue causes the loss of activity (Fisher et al., 1996; Li et al., 2008).

The study of SsuD interactions with the cofactor was performed by docking the cofactor into the active sites of subunits A and B (Lengauer & Rarey, 1996), followed by MD simulations (Alder & Wainwright, 1959) of the enzyme-FMNH2 complexes. For the open (B) and the closed (A) subunits different docking poses were generated and 20 of them (for each subunit) were selected on the basis of London dG (MOE, 2006) scoring function. All poses obtained by docking FMNH2 into subunit B were discarded on the basis of London dG scoring function and because of their higher potential energy.
Finally, the best pose was selected by comparing three-dimensional coordinates of LadA co-crystallized with FMN with each of the 20 poses of the FMNH$_2$ docked into subunit A.

**MD simulations of the interactions between cofactor and active site**

The interactions between cofactor and active site were investigated by molecular dynamics simulations to analyze the role of the cofactor in stabilizing the two conformations of SsuD. Two initial structures for the MD simulations were created: an enzyme-cofactor complex where the cofactor was placed in the subunit A (closed conformation) and a free enzyme based on the structure of subunit B (open conformation).

The force field GROMOS-96 53a6 (Oostenbrink et al., 2004) was assigned, the protein was solvated with explicit water, minimized using the PME (Essmann et al., 1995) algorithm and a 20 ns molecular dynamics simulation was performed using GROMACS software (Berendsen et al., 1995). The trajectory of Arg297 and its interactions, either with Asp111 or Glu20, underlines the dynamic behavior of the enzyme active site. From the comparison of the two subunits, it appears that this movement is related to the presence of FMNH$_2$. In subunit A (closed conformation with docked cofactor), the distance between Arg297 and Asp111 reaches the equilibrium at the distance of 0.22 nm in approximately 10 ns. Afterwards, the subunit maintains that conformation for the rest of the simulation. The salt bridge between Arg297 and Asp111 in subunit A prevents the expulsion of the cofactor by moving a protein loop (from Thr108 to Glu114) and thus locking the cofactor inside the cavity (see supplementary material Figure S7). On the contrary, the distance between Arg297 and Glu20 fluctuates during the entire period of the simulation, changing between 0.2 and 1.0 nm (Figure 1), and indicating an absence of a salt bridge between these two amino acids.

In contrast, the dynamics of the cofactor-free subunit B (open conformation) shows a stable distance of 0.2 nm between Arg297 and Glu20 after the first 200 ps, which
Figure 3. Time dependence of the two relevant distances during four different 20 ns simulations. Subunit B: evaluation of the distances between Arg297 and Asp111 (in gray) and Arg297–Glu20 (in black) during the 20 ns molecular dynamics simulations. From top to bottom four typical cases that describe the conformational modifications that take place during the simulations in absence of cofactor: (1) closed, (2) open, (3) intermediate, and (4) switch between the closed and the open state.
is compatible with a salt bridge (Figure 2). Conversely, the distance between Arg297 and Asp111 fluctuates between 0.5 and 0.8 nm throughout the entire simulation, without reaching a constant value.

These observations suggest that the presence of the cofactor induces the stabilization of the subunit towards the closed conformation, whereas the cofactor-free subunit assumes a stable open conformation.

To confirm the two simulations of subunits A and B, 10 additional simulations of 20ns were performed for each subunit (full details are reported in the supplementary material Table SI). In all simulations of subunit A (with the docked cofactor and in closed conformation) the distance between Arg297 and Asp111 was compatible with salt bridge formation, while the distance between Arg297 and Glu20 increased sporadically. This observation suggests that the opening mechanism can still occur although the closed state is favored. On the contrary, the distance between Arg297 and Asp111 remained below 1 nm. The 10 simulations of the cofactor-free subunit B (open conformation) showed a different behavior. The two reference distances changed in such a way that both the closed and open conformations occur, but there was no favored conformation. Figure 3 reports four typical conformations obtained in the 10 MD simulations. In one simulation (indicated as MD 1 in Figure 3 and Table SI) the distance between Arg297 and Asp111 is about 0.3 nm, while between Arg297 and Glu20 it is above 1.0 nm, indicating that the active site is in the closed conformation and a salt bridge between Arg297 and Asp111 is formed. In a third simulation (indicated as MD 3 in Figure 3 and Table SI) both distances are about 0.3 nm showing that the subunit is in an intermediate state generated by a contraction of the crevice due to the absence of the bulky cofactor. A fourth simulation (indicated as MD 4 in Figure 3 and Table SI) displays both conformations: during the first 8 ns the active site is closed while, after a short intermediate state, it switches to the open conformation.

The global evaluation of all dynamic simulations suggests that the cofactor has a relevant effect on the stabilization of the closed active site state, although the opening mechanism can still happen. This can be related to the release of the oxidized cofactor, which needs to be regenerated after catalysis. When the active site is not binding the cofactor, there is no predominant stabilized conformation and the enzyme switches between open and closed states, ready to accept the cofactor. Such kind of a breathing motion is very common in complex proteins, (Chakravarty & Varadarajan, 1999; Merlino et al., 2005). Figure 4 shows three different conformations of subunit B observed during the 20 ns simulations. The salt bridge between Arg297 and Glu20 induces the open conformation (A), whereas the progressive approaching of Arg297 and Asp111 (C) translates into the closed conformation (B).

These simulations point out that Arg297, Asp111, and Glu20 are important residues for the enzymatic activity because they are involved in the stabilization of

![Figure 4](image-url)
the active site in the open and the closed conformations through formation of salt bridges. Furthermore, their movements have an essential role for cofactor binding and release. In conclusion, the presence of FMNH₂ cofactor stabilizes the closed state, while the empty active site showed to be in a dynamic equilibrium between the two states.

It must be underlined that during all the simulations the cofactor was never expelled from the active site, confirming the reliability of the structural model constructed by superimposing the SsuD and the LadA cofactor complexes. Therefore, the same model was used for the investigation of enzyme–substrate interactions (Figure 5).

**Understanding and modeling enzyme–substrate interactions**

SsuD catalyzes the oxidation of a broad range of structurally different aliphatic sulfonates having as general structure R–CH₂–SO₃H (Eichhorn et al., 1999). In order to study the structural basis of enzyme–substrate recognition, docking and MD simulations were performed on selected substrates reported in the literature (Eichhorn et al., 1999). The substrates and their experimentally measured $k_{cat}/K_M$ are reported in Table I. The data-set meets the necessary requirements in terms of structural diversity, $k_{cat}/K_M$ range and homogeneity of distribution, which are crucial for the generation of a consistent QSAR model (ALMOND v. 3.3.0). Data were then used for calculating a preliminary 3D-QSAR model (Braiuca et al., 2006; Braiuca, Knopic, et al., 2009b), which was exploited for extracting information on the structural features affecting enzyme specificity. The analysis was performed by means of the PLS method, which is based on principal components analysis.

Enzyme–substrate interactions were investigated by docking the substrates into the active site of the subunit A complexed with FMNH₂, which was obtained after 20 ns equilibration as described above (Figure 5). The generated poses were selected on the basis of the London dG scoring function (MOE, 2006).

Further on, the GRID sulfone/sulfoxide probe (OS probe) was applied to map the affinity of the active site for the sulfonate group (Figure 6) and to select only those poses, where the sulfonic group binds in the areas delineated by the OS probe MIFs (Figure 6). However, since the distance between sulfonic group and the FMNH₂ must be compatible with the catalysis, His11 appears to be the only basic residue able to interact with the sulfonic group while meeting all the other scoring criteria (see supplementary material Figure S8).

Following this scheme, one single pose for each substrate was chosen. The systems comprising the subunit A-FMNH₂ complex and the docked substrates were solvated with explicit water and energy minimized using the PME (Essmann et al., 1995) algorithm implemented in GROMACS (Berendsen et al., 1995) and the GROMOS-96 53a6 force field (Oostenbrink et al., 2004). Subsequently, short 300 ps molecular dynamics simulations were carried out for each enzyme-cofactor-substrate complex. Each MD simulation was analyzed and the substrate conformers with the lowest potential energy were selected and subsequently used for extracting information on the structural features that determine the substrate specificity of SsuD. For this purpose, the structures of the substrates were correlated with the respective activity by means of PLS. The mathematical model was calculated on the basis of molecular descriptors obtained by a GRID derived method named GRIND (GRID Independent Descriptors) (Pastor, Cruciani, Mc Lay, Pickett, & Clementi, 2000), implemented in the software ALMOND (ALMOND v. 3.3.0). This software was selected because it allows the use of a wide array of chemical probes, including the Ca⁺⁺ one. Unlike the classical 3D-QSAR approaches (e.g. CoMFA) (Todeschini & Consonni, 2009), the GRIND method presents the major advantage of being independent from any alignment of molecules/substrates, which is usually a primary source of error in this kind of studies (Puzyn, Leszczynski, & Cronin, 2010). The GRIND method transforms the information included in the MIFs into alignment independent descriptors (vectors) able to describe the relative spatial position of chemical groups throughout the molecules.
A PLS model was built up by correlating the GRIND descriptors with the experimental $k_{\text{cat}}/K_M$ values of the training set. The mathematical model explains 62% of the whole variance and has a predictive coefficient $q^2$ (calculated by means of the leave-one-out cross-validation procedure) of 0.86 on the second principal component.

Since the QSAR model here described is based on alignment free descriptors (GRIND), any direct connection between the structural/chemical features of molecules and their absolute position in the Cartesian space is lost. Nevertheless, the comparison of the different molecules of the data-set is feasible thanks to analysis of the PLS scores plot and the extraction of relevant information.

Table 1. Structures of substrates accepted by the SsuD and corresponding kinetic data taken from Eichhorn et al. (1999).

| Compound | General formula: $R$–CH$_2$–SO$_3$H, with $R$– | Experimental $k_{\text{cat}}/K_M$ (min$^{-1}$μM$^{-1}$) |
|----------|------------------------------------------------|--------------------------------------------------|
| 1        | ![Image 1](image1.png) | 6.7 |
| 2        | ![Image 2](image2.png) | 6.1 |
| 3        | ![Image 3](image3.png) | 6.0 |
| 4        | ![Image 4](image4.png) | 4.6 |
| 5        | ![Image 5](image5.png) | 3.2 |
| 6        | ![Image 6](image6.png) | 2.7 |
| 7        | ![Image 7](image7.png) | 1.8 |
| 8        | ![Image 8](image8.png) | 1.1 |
| 9        | ![Image 9](image9.png) | 0.6 |
| 10       | ![Image 10](image10.png) | 0.4 |

Figure 6. GRID analysis of the active site of SsuD. MIFs calculated with the OS probe are in green. His11 and FMNH$_2$ are highlighted in sticks mode, the other basic residues are in red sticks mode.
Figure 7. Variables loadings for the first (upper figure) and the second (lower figure) principal components. Variables are grouped in correlograms.
still possible by analyzing the loadings plots, which provide information on how the variables contribute to each principal component (Figures 7 and 8).

Variables with larger loading values better describe the activity of interest. It is important to underline that a positive loading indicates that the variable is directly correlated with $k_{\text{cat}}/K_M$. Conversely, variables negatively correlated with the activity are characterized by negative loadings.

The loadings plot for the first two principal components clearly shows that the first principal component discriminates the substrates mainly on the basis of their shape (TIP/TIP) (Figure 7). Indeed, the variance explained by the first PC is only the 24% of the whole variance and a PLS model based only on the first PC has a $q^2$ as low as 0.48.

The second principal component explains 38% of the whole variance and the loading plot (lower plot of Figure 7) indicates that all different types of descriptors contribute to this second PC and thus to substrate description and discrimination. The overall analysis of the PLS loadings plot clearly demonstrates how a satisfactory description of the substrates can be achieved only by accounting two principal components and conversely the largest part of all the relevant steric, electronic, and hydrophobic contributions and especially their interactions.

The analysis of the PLS score plot illustrates how each component is able to describe the different molecules (Figure 8).

The first principal component (Figure 8) is able to describe how enzyme specificity depends on the shape and length of substrate structure (TIP–TIP descriptors, see supplementary material Figure S9). When the only variability between molecules is the aliphatic chain length (as for molecules 1, 2, 7, and 9) the first principal component well indicates that longer aliphatic molecules are better substrates. On the other hand, substrates 8 and 10 are grouped because of their cyclic and polar structure (described by WATER/TIP variables, Figure 7 upper plot) and substrates 4, 5, and 6 are considered similar because of their globular shape (Figure 8). It is interesting to note that the second principal component includes the contributions of different variables describing not only the shape, but rather electronic properties and especially their interactions (confirmed by the high loading of all the correlograms in Figure 7). This latter point makes the interpretation more complex but it is evident that, differently from the first component, the aromatic
substances 3, 4, 5, and 6 are well described and discriminated by the second component on the basis of their ability of establishing H-bonds. As a consequence, the highly hydrophobic substrate 6 results to be the poorest one. On the other side, aliphatic substrates 1, 2, 9, and 7 are considered as a homogenous set of molecules since they do not differ in terms of electronic properties. Therefore, globally the two components are able to describe both steric and electronic factors and, more importantly, those interactions that affect the ability of SsuD of transforming efficiently a substrate. It must be underlined that molecular descriptors were correlated to kinetic properties ($k_{cat}/K_M$) and consequently such information goes beyond the simple ability of the enzyme of recognizing the substrate, but includes the factors that affect the ability of the enzyme to stabilize the relevant transition state of the reaction.

Conclusions
Molecular simulation methods have been applied to the study of SsuD action at molecular level, shedding light to enzyme–cofactor and enzyme–substrate interaction mechanisms. By combining docking and molecular dynamics, a comprehensive illustration of the structural features of SsuD was accomplished. The enzyme showed a two-state dynamic mechanism, switching between an open and a closed conformations of its active sites. Arg297, Asp111, and Glu20 govern the dynamical switch between the inactive and active conformations. The presence of FMNH$_2$ cofactor stabilizes the closed state, while the empty active site showed to be in a dynamic equilibrium between the two states. This mechanism has never been observed before and its function seems to regulate substrate access and cofactor regeneration. This represents a further example of a “breathing motion mechanism,” observable in many complex proteins.

The role of residues directly involved in the catalytic mechanism has been confirmed and clarified. His11 and Trp196 are crucial for stabilizing the sulfonic group of the substrates and the cofactor, respectively. The conformational analysis of a set of substrates performed by MD simulations was used for constructing a PLS model correlating substrate structures and enzyme activity. The model provided general guidelines for describing the physical–chemical features of substrates for being accepted by this enzyme. The output of the statistical analysis of the PLS model allows to withdraw information in a more rational and systematic procedure and to focus the attention also to the interactions of variables.

The analysis of the model suggests that the interaction between the substrate and the FMNH$_2$ molecule is crucial for the catalytic activity, whereas the nature of the substrate can influence the strength of this interaction and the correct substrate placement. So far, the lack of knowledge of SsuD mechanism and specificity has restricted the application of this enzyme. The substrates listed in Table 1 represent the only available data-set in the literature that meets the requirement necessary for the PLS construction. A more detailed elucidation of the mechanism of SsuD and the rational description of structural requirement of SsuD substrates provide new basis for a more extended exploitation of the catalytic potential of this class of enzymes, both for synthetic and environmental applications (Buhler, Witholt, Hauer, & Schmid, 2002). In this context, information here reported can be used as a reference also for substrate and enzyme engineering.

Acknowledgements
This work has received funding from the European Community’s Seventh Framework Programme under the FP7-KBBE-2008-2B grant agreement n° 227279. Thanks are due to Moldiscovery Inc. for providing the software.

Supplementary material
The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2012.674268.

References
Abdurachim, K., & Ellis, H.R. (2006). Detection of protein–protein interactions in the Alkanesulfonate Monooxygenase system from Escherichia coli. Journal of Bacteriology, 188, 8153–8159.
Alder, B.J., & Wainwright, T.E. (1959). Studies in molecular dynamics. I. General method. Journal of Chemical Physics, 37, 459.
ALMOND v. 3.3.0, www.moldiscovery.com, Molecular Discovery Ltd., Perugia, Italy.
Aufhammer, S.W., Warkentin, E., Berk, H., Shima, S., Thauer, R.K., & Ermler, U. (2004). Coenzyme binding in F$_{420}$-dependent alcohol dehydrogenase, a member of the bacterial luciferase family. Structure, 12, 361–370.
Aufhammer, S.W., Warkentin, E., Ermler, U., Hagemeier, C.H., Thauer, R.K., & Shima, S. (2005). Crystal structure of methylenetetrahydrodromethanopterin reductase (Mer) in complex with coenzyme F$_{420}$: Architecture of the F$_{420}$/FMN binding site of enzymes within the non-proeryl cis-peptide containing bacterial luciferase family. Protein Science, 14, 1840–1849.
Ballou, D.P., Entsch, B., & Cole, L.J. (2005). Dynamics involved in catalysis by single-component and two-component flavin-dependent aromatic hydroxylases. Biochemical and Biophysical Research Communications, 338, 590–598.
Berendsen, H.J.C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. Computer Physics Communications, 91, 43–56.
Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., … Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Research, 28, 235–242.
Braiuca, P., Boscariol, L., Ebert, C., Linda, P., & Gardossi, L. (2006). 3-D QSAR applied to the quantitative prediction of penicillin G amidae selectivity. Advanced Synthesis & Catalysis, 348, 773–780.
Braiuca, P., Ebert, C., Basso, A., Linda, P., & Gardossi, L. (2009). Computational methods to rationalize experimental strategies. Trends in Biotechnology, 27, 419–425.

Braiuca, P., Knapić, L., Ferrario, V., Ebert, C., & Gardossi, L. (2009). A three-dimensional quantitative structure-activity relationship (3D-QSAR) model for predicting the enantioselectivity of Candida antarctica lipase B. Advanced Synthesis & Catalysis, 351, 1293–1302.

Buhler, B., Witholt, B., Hauer, B., & Schmid, A. (2002). Characterization and application of xylene monooxygenase for multistep biocatalysis. Applied and Environment Microbiology, 68, 560–568.

Campbell, Z.T., Weichsel, A., Montfort, W.R., & Baldwin, T.O. (2009). Crystal structure of the bacterial luciferase/flavin complex provides insight into the function of the β subunit. Biochemistry, 48, 6085–6094.

Carpenter, R.A., Zhan, X., & Ellis, H.R. (2010). Catalytic role of a conserved cysteine residue in the desulfonation reaction by the alkanesulfonate monooxygenase enzyme. Biochimica et Biophysica Acta, 1804, 97–105.

Chaiyen, P., Suadee, C., & Wilairat, P. (2001). A novel two-component flavoprotein hydroxylase: p-hydroxyphenylacetate hydroxylase from Acinetobacter baumannii. European Journal of Biochemistry, 268, 5550–5561.

Chakravarty, S., & Varadarajan, R. (1999). Residue depth: A novel parameter for the analysis of protein structure and stability. Structure, 7, 723–732.

Cruciani, G., Clementi, S., & Baroni, M. (1993). H. Kubinyi (Ed.), 3D QSAR in drug design: Theory, methods and applications (pp. 567–582). Leiden: ESCOM.

Eichhorn, E., Davey, C.A., Sargent, D.F., Leisinger, T., & Richmond, T.J. (2002). Crystal structure of Escherichia coli alkanesulfonate monooxygenase SsuD. Journal of Molecular Biology, 324, 457–468.

Eichhorn, E., van der Ploeg, J.R., & Leisinger, T. (1999). Characterization of a two-component alkanesulfonate monooxygenase from Escherichia coli. Journal of Biological Chemistry, 274, 26639–26646.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., & Pedersen, L.G. (1995). A smooth particle mesh Ewald method. Journal of Chemical Physics, 103, 8577–8592.

Felts, G.K., & Petsko, G.A. (1990). The evolution of α/β barrel enzymes. Trends in Biochemical Sciences, 15, 228–234.

Fisher, A.J., Thompson, T.B., Thoden, J.B., Baldwin, T.O., & Rayment, I. (1996). The 1.5 Å resolution crystal structure of bacterial luciferase in low salt conditions. Journal of Biological Chemistry, 271, 21956–21968.

Gao, B., & Ellis, H.R. (2005). Altered mechanism of the alkanesulfonate FMN reductase with the monooxygenase enzyme. Biochemical and Biophysical Research Communications, 331, 1137–1145.

Goodford, P.J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. Journal of Medicinal Chemistry, 28, 849–857.

Henikoff, S., & Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences, 89, 10915–10919.

Kahnert, A., Vermeij, P.J., Wietek, C., James, P., Leisinger, T., & Kertesz, M.A. (2000). The ssu locus plays a key role in organosulfur metabolism in Pseudomonas putida S-313. Journal of Bacteriology, 182, 2869–2878.

Kertesz, M.A., Schmidt-Larbig, K., & Wüest, T. (1999). A novel reduced flavin mononucleotide-dependent methanesulfonate sulfonatase encoded by the sulfur-regulated ssu operon of Pseudomonas aeruginosa. Journal of Bacteriology, 181, 1464–1473.

Koch, D.J., Rückert, C., Rey, D.A., Mix, A., Pühler, A., & Kalinowski, J. (2005). Rational design of a Corynebacterium glutamicum pantothenate production strain and its characterization by metabolic flux analysis and genome-wide transcriptional profiling. Applied and Environment Microbiology, 71, 6104–6114.

Lengauer, T., & Rarey, M. (1996). Computational methods for biomolecular docking. Current Opinion in Structural Biology, 6, 402–406.

Li, L., Liu, X., Yang, W., Xu, F., Wang, W., Feng, L., … Rao, Z. (2008). Crystal structure of long-chain alkane monooxygenase (LaD) in complex with coenzyme FMN: Unveiling the long-chain alkane hydroxylase. Journal of Molecular Biology, 376, 453–465.

Merlino, A., Mazzarella, L., Carannante, A., Di Fiore, A., Di Donato, A., Notomista, E., & Sica, F. (2005). The importance of dynamic effects on the enzyme activity: X-ray structure and molecular dynamics of onconase mutants. Journal of Biological Chemistry, 280, 17953–17960.

Koch, D.J., Rückert, C., Rey, D.A., Mix, A., Pühler, A., & Kalinowski, J. (2005). Rational design of a Corynebacterium glutamicum pantothenate production strain and its characterization by metabolic flux analysis and genome-wide transcriptional profiling. Applied and Environment Microbiology, 71, 6104–6114.

Lengauer, T., & Rarey, M. (1996). Computational methods for biomolecular docking. Current Opinion in Structural Biology, 6, 402–406.

Li, L., Liu, X., Yang, W., Xu, F., Wang, W., Feng, L., … Rao, Z. (2008). Crystal structure of long-chain alkane monooxygenase (LaD) in complex with coenzyme FMN: Unveiling the long-chain alkane hydroxylase. Journal of Molecular Biology, 376, 453–465.

Merlino, A., Mazzarella, L., Carannante, A., Di Fiore, A., Di Donato, A., Notomista, E., & Sica, F. (2005). The importance of dynamic effects on the enzyme activity: X-ray structure and molecular dynamics of onconase mutants. Journal of Biological Chemistry, 280, 17953–17960.

MOE v. 2006.08, www.chemcomp.com, Chemcomp, Montreal, Canada.

Oldfield, C., Pogrebinsky, O., Simmonds, J., Olson, E.S., & Kulpa, C.F. (1997). Elucidation of the metabolic pathway for dibenzothiophene desulfurization by Rhodococcus sp. strain IGTS8 (ATCC 53968). Microbiology, 143, 2961–2973.

Oostenbrink, C., Villa, A., Mark, A.E., & van Gunsteren, W.F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. Journal of Computational Chemistry, 25, 1656–1676.

Pastor, M., Cruciani, G., Mc Lay, I., Pickett, S., & Clementi, S. (2000). GRid-Independent Descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. Journal of Medicinal Chemistry, 43, 3233–3243.

Puzyn, T., Leszczyński, J., & Cronin, M.T. (2010). Recent advances in QSAR studies, challenges and advances in computational chemistry and physics. New York: Springer.

Shima, S., Warkentin, E., Grabarse, W., Sordel, M., Wicke, M., Thauer, R.K., & Ermloff, U. (2000). Structure of coenzyme F420 dependent methylenetetrahydromethanopterin reductase from two methanogenicarchae. Journal of Medicinal Chemistry, 43, 3233–3243.

Shuettelkopf, A.W., & van Aalten, D.M.F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D, 60, 1355–1363.

Thibaut, D., Ratet, N., Bisch, D., Faucher, D., Debussche, L., & Blanche, F. (1995). Purification of the two-enzyme system catalyzing the oxidation of the D-proline residue of pristinamycin IIB during the last step of pristinamycin IIA biosynthesis. Journal of Bacteriology, 177, 5199–5205.

Todeschini, R., & Consonni, V. (2009). Molecular descriptors for chemoinformatics. Weinheim: Wiley-VCH.

Uetz, T., Schneider, R., Snozzi, M., & Egel, T. (1992). Purification and characterization of a two-component monooxygenase that hydroxylates nitroliotriacetate from “Chelobacter” strain ATCC 29600. Journal of Bacteriology, 174, 1179–1188.
Valton, J., Mathevon, C., Fontecave, M., Niviére, V., & Ballou, D.P. (2008). Mechanism and regulation of the two-component FMN-dependent monooxygenase ActVA-ActVB from *Streptomyces coelicolor*. *Journal of Biological Chemistry*, 283, 10287–10296.

van Berkel, W.J., Kamerbeek, N.M., & Fraaije, M.W. (2006). Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *Journal of Biotechnology*, 124, 670–689.

van der Ploeg, J.R., Cummings, N.J., Leisinger, T., & Connerton, I.F. (1998). *Bacillus subtilis* genes for the utilization of sulfur from aliphatic sulfonates Microbiology-SGM. *Microbiology*, 144, 2555–2561.

van der Ploeg, J.R., Eichhorn, E., & Leisinger, T. (2001). Sulfonate-sulfur metabolism and its regulation in *Escherichia coli*. *Archives of Microbiology*, 176, 1–8.

Witschel, M., Nagel, S., & Egli, T. (1997). Identification and characterization of the two-enzyme system catalyzing oxidation of EDTA in the EDTA-degrading bacterial strain DSM 9103. *Journal of Bacteriology*, 179, 6937–6943.

Zhan, X., Carpenter, R.A., & Ellis, H.R. (2008). Catalytic importance of the substrate binding order for the FMNH$_2$-dependent alkanesulfonate monooxygenase enzyme. *Biochemistry*, 47, 2221–2230.