

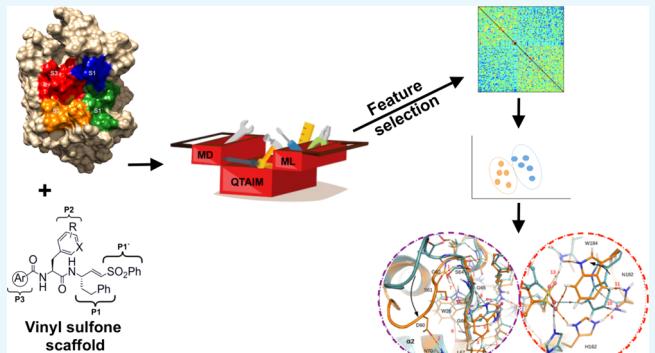
# Combining Charge Density Analysis with Machine Learning Tools To Investigate the Cruzain Inhibition Mechanism

Adriano M. Luchi,<sup>ID</sup> Roxana N. Villafaña, J. Leonardo Gómez Chávez, M. Lucrecia Bogado, Emilio L. Angelina,\*<sup>ID</sup> and Nelida M. Peruchena\*

Lab. Estructura Molecular y Propiedades, IQUIBA-NEA, Universidad Nacional del Nordeste, CONICET, FACENA, Av. Libertad 5470, Corrientes 3400, Argentina

## Supporting Information

**ABSTRACT:** *Trypanosoma cruzi*, a flagellate protozoan parasite, is responsible for Chagas disease. The parasite major cysteine protease, cruzain (Cz), plays a vital role at every stage of its life cycle and the active-site region of the enzyme, similar to those of other members of the papain superfamily, is well characterized. Taking advantage of structural information available in public databases about Cz bound to known covalent inhibitors, along with their corresponding activity annotations, in this work, we performed a deep analysis of the molecular interactions at the Cz binding cleft, in order to investigate the enzyme inhibition mechanism. Our toolbox for performing this study consisted of the charge density topological analysis of the complexes to extract the molecular interactions and machine learning classification models to relate the interactions with biological activity. More precisely, such a combination was useful for the classification of molecular interactions as “active-like” or “inactive-like” according to whether they are prevalent in the most active or less active complexes, respectively. Further analysis of interactions with the help of unsupervised learning tools also allowed the understanding of how these interactions come into play together to trigger the enzyme into a particular conformational state. Most active inhibitors induce some conformational changes within the enzyme that lead to an overall better fit of the inhibitor into the binding cleft. Curiously, some of these conformational changes can be considered as a hallmark of the substrate recognition event, which means that most active inhibitors are likely recognized by the enzyme as if they were its own substrate so that the catalytic machinery is arranged as if it is about to break the substrate scissile bond. Overall, these results contribute to a better understanding of the enzyme inhibition mechanism. Moreover, the information about main interactions extracted through this work is already being used in our lab to guide docking solutions in ongoing prospective virtual screening campaigns to search for novel noncovalent cruzain inhibitors.



## 1. INTRODUCTION

Chagas disease (CD), a major health issue in Latin America, is a neglected tropical disease caused by the flagellate protozoan parasite *Trypanosoma cruzi*. According to estimates by the World Health Organization, seven million people are chronically infected with the parasite and 7000 deaths per year are caused by CD. Because of massive migration, the disease has spread around the globe reaching nonendemic areas, where health service awareness of the condition is limited.<sup>1,2</sup>

Available chemotherapy for CD includes ineffective drugs for the chronic stage of the disease, leaving patients with only two palliative drugs, benznidazole and nifurtimox, introduced over 40 years ago. Furthermore, such drugs involve severe side effects, and drug resistance has been observed in some trypanosome strains. Thus, the discovery of new, safer, and more effective drugs to treat CD is required.<sup>3</sup>

Cruzain (Cz), the major cysteine protease of *T. cruzi* is a viable target for developing new drugs against CD because it is essential for parasite survival in the human stage of infection.<sup>4</sup>

Currently, 27 inputs are associated to this molecular target in the Protein Data Bank ([rcsb.org](https://www.rcsb.org)) where Cz has been cocrystallized with reversible and irreversible inhibitors.<sup>5</sup> Thereby, Cz presents itself as an attractive target for the development of potential therapeutics for the treatment of the disease by employing a structure-based approach.<sup>6,7</sup>

Among Cz inhibitors, those containing a vinyl sulfone warhead can exhibit good selectivity and a favorable prospective development despite the irreversible nature of inhibition. Jaishankar<sup>8</sup> synthesized and determined the inhibition constants against Cz of a series of vinyl sulfone

Received: June 27, 2019

Accepted: October 18, 2019

analogues closely related to K-777, a Cz inhibitor. They investigated how substitutions at P2 and P3 fragments of K-777 modify the activities against Cz.

In this work, we exploited the structure–activity relationship among the vinyl sulfone analogues described by Jaishankar<sup>8</sup> but from a structure-based perspective, that is, through the study of the molecular interactions at the enzyme binding site, in order to get some clues about the enzyme inhibition mechanism.

As a descriptor for molecular interactions in complexes of vinyl sulfones with Cz, the charge density value at the interaction critical point was employed. In the context of the quantum theory of atoms in molecules (QTAIM),<sup>9</sup> the mapping of the gradient vector field onto the complex electron charge density distribution gave rise to the topological elements of charge density. Among the topological elements, an interaction bond critical point (BCP) and the bond paths (BPs), which connect it to the interacting atoms, are unequivocal indicators of the existence of bonding interaction.

We have previously applied this theory to understand the action mechanism of human dihydrofolate reductase inhibitors,<sup>10,11</sup> BACE1 inhibitors,<sup>12,13</sup> D2 dopamine receptor ligands,<sup>14–18</sup> sphingosine kinase 1 (Sphk1) inhibitors,<sup>19</sup> and HIV-1 protease flap fragments,<sup>20</sup> among others.

QTAIM methodology allows detecting nondirectional interactions, for example, those involving  $\pi$  electrons in aromatic rings, among other weak and unusual contacts that otherwise would be missed in a merely geometrical analysis of the interactions.<sup>16</sup>

On the other hand, QTAIM analysis in biomolecular complexes (unlike small complexes in the gas phase) often gives rise to very dense and complex networks of interactions. The task of analyzing such intricate network of interactions becomes even more difficult when more than one of these networks must be analyzed simultaneously, for example, to extract structure–activity relationships from a set of Cz complexes with several inhibitors.

Therefore, the processing of such massive amount of data should not be done “by hand”, that is, by visual inspection of the molecular graphs by a human operator. If so, a lot of information “hidden” under the charge density data would be overlooked.

Accordingly, in this work we employed machine learning tools to automate the process of extracting information from charge density molecular graphs and to exhaustively exploit the charge density data.

We trained a support vector machine model with recursive feature elimination (SVM-RFE) that was able to discriminate between interactions present in complexes of the most active inhibitors (active-like interactions) and those that occur in the less active ones (inactive-like interactions).

Subsequently, the charge density-based correlation matrix describing how interactions are related to each other among the complexes was computed. This matrix, together with analysis of the molecular dynamic (MD) trajectories, revealed how interactions come into play together to trigger the enzyme into a particular conformational state. Most active inhibitors induce some conformational changes within the enzyme that lead to an overall better fit of the inhibitor into the binding cleft.

Analysis of intermolecular interactions revealed that backbone–backbone hydrogen bonds between the peptide-like inhibitor and enzyme and interactions with the Leu67 residue

play a key role in proper anchoring of the inhibitor to the Cz binding cleft. However, a quantitative structure–activity relationship could not be derived by considering only the intermolecular interactions between Cz residues and inhibitor atoms.

On the other hand, if intramolecular contacts involving protein residues are also analyzed with the help of the SVM-RFE model, it becomes clear that a more indirect mechanism of enzyme inhibition involving extensive conformational changes within the protein structure operates under the hood. Interactions at the S2 subpocket seem to be behind conformational changes occurring on the right wall of the binding cleft, while interactions at the S3 subsite mostly drive conformational changes on the left wall. Both conformational changes ultimately lead to rearrangements of residues at the S1' subsite that allows the proper positioning of the vinyl sulfone warhead, which in turn allows the formation of key backbone–backbone interactions between the inhibitor and binding cleft wall residues.

Moreover, residue rearrangements at the S1' subsite in complexes of most active inhibitors involve the formation of hydrogen bonds among residues of the catalytic triad that are considered as a hallmark of the substrate recognition event. This means that these high-affinity inhibitors are likely recognized by the enzyme as if they were its own substrate so that the catalytic machinery is arranged as if it is about to break the substrate scissile bond.

## 2. RESULTS AND DISCUSSION

**2.1. Compilation of a Structural Library of Cz–Inh Complexes with Activity Annotations.** Table 1 shows the inhibition constants against Cz of P2/P3-modified vinyl sulfones reported by Jaishankar,<sup>8</sup> while Scheme 1 shows substitution sites in vinyl sulfone analogues.

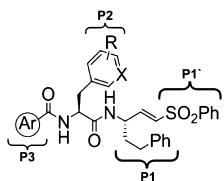
The crystal structure of Cz bound to the vinyl sulfone inhibitor K-777 (PDB ID 2OZ2) provides the structural basis for understanding inhibition of cruzain by vinyl sulfone

**Table 1. Cz Inhibition by Vinyl Sulfone Analogues Reported by Jaishankar<sup>a,b</sup>**

compound <sup>b</sup>	R	X	Ar	$K_i$ (nM)
9d	4-Me	CH	DHBD	19
7d	4-Me	CH	4-CF <sub>3</sub> Ph	45
6b	3-Me	CH	3,5-DiFPh	50
9b	3-Me	CH	DHBD	71
9a	H	CH	DHBD	80
7b	3-Me	CH	4-CF <sub>3</sub> Ph	92
7a	H	CH	4-CF <sub>3</sub> Ph	97
8c	3-CF <sub>3</sub>	CH	2-pyridyl	150
4c	3-CF <sub>3</sub>	CH	N-MePip	170
4a (K-777)	H	CH	N-MePip	220
8a	H	CH	2-pyridyl	250
8b	3-Me	CH	2-pyridyl	280
6d	4-Me	CH	3,5-DiFPh	350
6a	H	CH	3,5-DiFPh	980
8d	4-Me	CH	2-pyridyl	1700
4b	3-Me	CH	N-MePip	3300
4e	H	N	N-MePip	3600

<sup>a</sup>N-MePip: N-methyl piperazine; DHBD: 2,3-dihydro-1,4-benzodioxin-6-yl; and 3,5-DiFPh: 3,5-diFluorophenyl. <sup>b</sup>Compound naming was extracted from Jaishankar.<sup>8</sup>

**Scheme 1.** Vinyl Sulfone with Its Substitution Sites Named Ar, R, and X<sup>a</sup>



<sup>a</sup>Inhibitor parts that bind to S1, S1', S2, and S3 enzyme subpockets are named P1, P1', P2, and P3, respectively.

inhibitors.<sup>21</sup> Taking that structure as a template, complexes of Cz with the 17 vinyl sulfone analogues listed in Table 1 were built manually and then refined by performing MD simulations, as described in computational methods.

Naming letters (a, b, c, d, and e) and numbers (4, 6, 7, 8, and 9) represent compound series obtained by varying P2 and P3, respectively.

**2.2. Local Electron Charge Density As the Descriptor of Molecular Interactions in Cz–Inh Complexes.** To describe the molecular interactions in the modeled Cz–inhibitor complexes, the charge density topological analysis in the framework of the QTAIM was performed over the refined complexes.

Briefly, this analysis basically consists of the mapping of the gradient vector field onto the precomputed charge density of the complex,  $\nabla\rho$ . From this mapping, the charge density topological element arises. Among the topological elements, an interaction BCP and the BPs, which connect it to the interacting atoms, are unequivocal indicators of the existence of a bonding interaction.

As an example, Figure 1 depicts the BPs and BCPs associated to the noncovalent interactions (Cz–Inh as well as Cz–Cz and Inh–Inh interactions) in one of the complexes studied here.

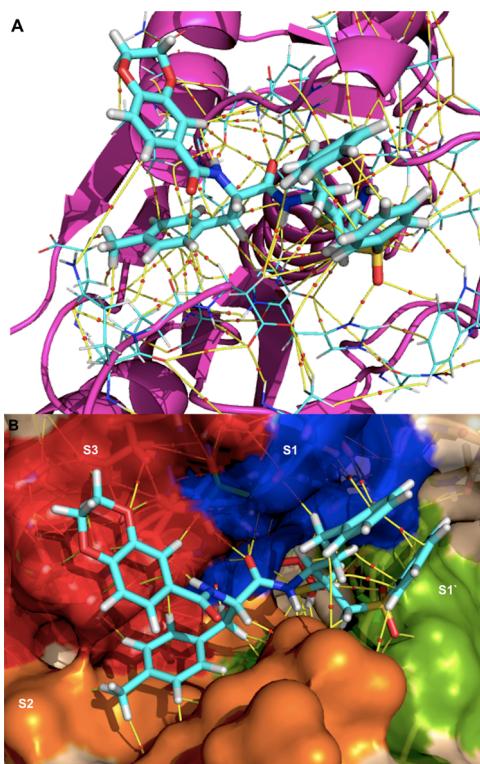
**2.3. Training an Interaction Classifier Based on the Charge Density Data.** At this point, we have at our disposal, the topological elements of the charge density describing the interactions in the Cz–Inh complexes and the activity data associated to the corresponding inhibitors.

Our goal is to take advantage of these data to find out favorable interactions (to stabilize the complex), which might explain the greater binding affinity of the more active inhibitors and the unfavorable (or less favorable) interactions that dominate the binding of the less active ones.

QTAIM analysis on biomolecular complexes often gives rise to very dense and complex networks of interactions. By inspecting Figure 1, it becomes evident that a comparative analysis of such intricate network of interactions for a set of Cz–Inh complexes cannot be performed by visual inspection of the molecular graphs by a human operator. If so, a lot of the information “hidden” under the charge density data would be overlooked.

Instead, in this work, we have applied machine learning tools to automate the process of extracting information from charge density molecular graphs and to exhaustively exploit the charge density data.

As explained by Fujita,<sup>22</sup> depending on the particular type of scientific question that needs to be answered, the predictive model can be more or less complex. Relatively simple linear models are more easily interpretable in terms of molecular interactions although their predictive power is limited. More



**Figure 1.** View of the intricate networks of interactions on the structure of the Cz–9d complex. Charge density topological elements describing the noncovalent interactions are depicted with small red circles (BCPs) and yellow lines connecting each BCP to both interacting atoms (BPs). Both intermolecular (Cz–Inh) and intramolecular (Cz–Cz and Inh–Inh) interactions are considered. The protein structure is depicted in the cartoon representation (A) and surface representation (B), where each surface color represents a different subpocket within the Cz binding cleft.

complex nonlinear models have greater predictive power but are more obscure or less interpretable.

In our case, our main interest was to shed light on the interactions implicated in the enzyme action mechanism by using a linear supervised model based on the complex interactions and their corresponding inhibitory activities.

Moreover, data sets where there are fewer observed entities than variables are becoming increasingly frequent, thanks to the growing ease of observing variables, together with the high cost of repeating observations in some contexts (e.g., DNA microarrays).<sup>23</sup> For example, Guyon<sup>24</sup> has built an SVM classifier to select a subset of genes biologically relevant to cancer from broad patterns of gene expression data, recorded on DNA microarrays. They used a relatively small number of training examples from cancer and normal patients.

It is well known that when the number of features is large and the number of training examples is comparatively small (as in the case of ref 24 and in our case), the risk of overfitting arises.

The overfitting problem can be reduced by measuring the feature importance and selecting the most discriminative feature subset. Elimination of redundant or irrelevant features can improve the model accuracy, the generalization capacity, and even the computational cost in some cases.<sup>25</sup>

SVM is a classification technique that uses support vectors to maximize the distance between the two classes. The coefficients of the model represent the vector coordinates

which are orthogonal to the hyperplane, and their direction indicates the predicted class. The absolute size of the coefficients (weights) can then be used to determine the feature importance for the data separation task.<sup>24</sup>

On the other hand, SVM-RFE is a backward feature selection algorithm based on SVM. SVM-RFE has been widely applied in many fields including genomics, proteomics, metabolomics, and other situations, where the data present a large number of features, and the samples are scarce.<sup>26</sup>

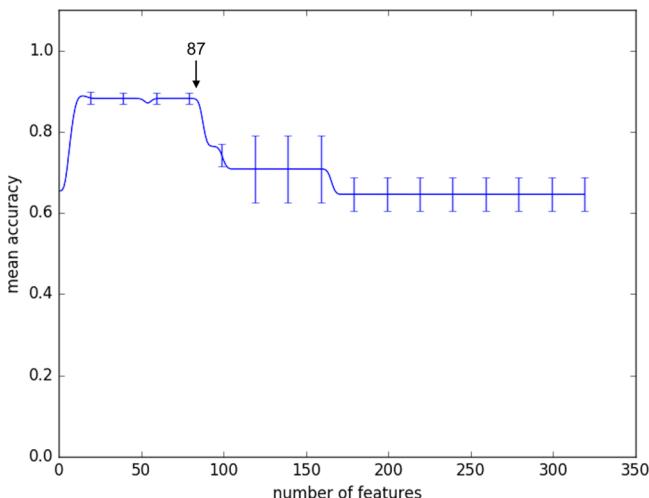
RFE begins with the entire set of features, creates the SVM model, and evaluates the accuracy. The least important predictors are erased, and the model is computed again.<sup>27</sup>

In this work, an SVM-RFE model was trained with the QTAIM-derived charge density information about molecular interactions from the 17 Cz–Inh complexes to select relevant features for the classification task, which might help to understand the enzyme action mechanism. Inhibitors were labeled as actives or inactives according to a decision threshold value of 170 nM of inhibitory activity, which ensures balanced classes.

SVM-RFE was built with a data set containing 319 interactions at the beginning, and then, the less relevant features were iteratively eliminated by a backward selection procedure.

The analysis of features that contribute to predictions only makes sense if the model reaches a reasonably high-performance level. Therefore, to monitor the accuracy of the model during the backward elimination of features, stratified two fold cross-validation was performed. In stratified cross-validation, the class distribution of each fold is preserved for the entire data set.<sup>28</sup>

**Figure 2** shows the cross-validation mean accuracy of the model as a function of the number of features selected by the



**Figure 2.** Iterative process of backward feature elimination and SVM model training with the remaining features. The mean accuracy of the SVM model is depicted as a function of the number of features. Error bars represent the variances of accuracy values among the folds.

SVM-RFE procedure. Also, the variance of the accuracy among the folds is depicted.

As can be seen in the figure, the mean accuracy of the model rises as the number of features drops up to 87, when the maximum performance is reached (87.75%). Also, note that the variance of the accuracy among the two folds decreases to a

minimum value on the plateau region between ~20 and 87 features.

Below ~20 features, the mean accuracy starts decreasing again, thus indicating that the classification model becomes too simple as to discriminate between compounds from active and inactive classes.

Therefore, for subsequent analysis of relevant features, we selected the SVM model trained with a subset of the best 87 features because further reduction of the number of features does not imply an increase in model performance.

The bar plot in **Figure 3** shows the top interactions (features) that were used by the final model to make the classifications. Only feature coefficients with absolute values greater than 2.0 are depicted in the figure.

The total height of stacked bars in **Figure 3** represents the interaction importance for the classification task while each category within the bar represents the charge density contribution of the two classes (active and inactive in orange and light blue, respectively) to the overall feature importance.

As can be seen in the figure, interactions with positive coefficients have overall greater contributions from compounds labeled as actives while the opposite is true for interactions with negative model coefficients, namely, their most important contributions come from compounds labeled as inactives.

Therefore, by using a simple and interpretable linear SVM classification model coupled with an RFE procedure, it is possible to extract useful information about what are the most important interactions to discriminate between active and inactive (or less active) compounds against Cz.

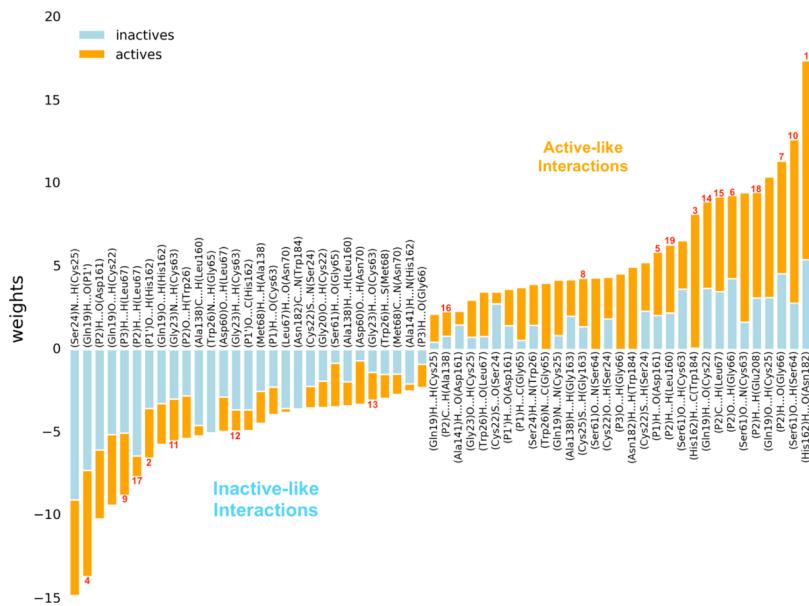
**2.4. Interaction-Based Correlation Matrix from Charge Density Data.** Although the trained classification model helps recovering the relevant interactions from charge density molecular graphs of Cz–Inh complexes, it does not necessarily provide information about how these interactions come into play together, namely, how they correlate to each other to bring the enzyme into a particular conformational state.

Different inhibitors might form different interactions which in turn might stabilize different conformational states of the enzyme. We wanted to know whether there could be a relationship between compound activity against Cz and the enzyme conformation stabilized. This information could be very useful to choose the appropriate target structure in future structure-based virtual screening campaigns.

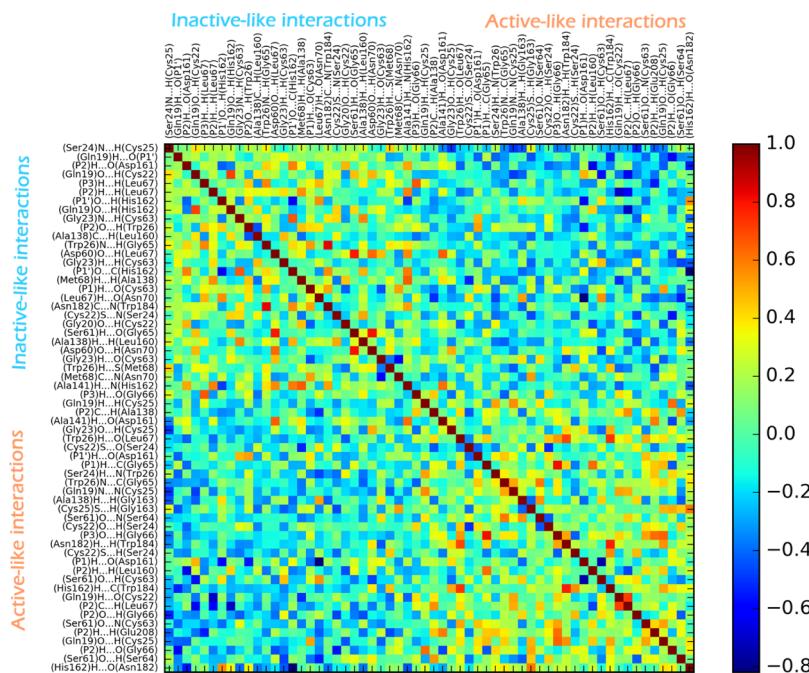
Accordingly, the correlation matrix describing how interactions are related to each other among the Cz–Inh complexes was computed from charge density data (**Figure 4**). Only interactions with importance greater than 2.0 in the SVM classifier were considered for the correlation analysis.

**Figure 4** shows that there is a clear anticorrelation (i.e., negative value) between active-like and inactive-like interactions, namely, between interactions that prevail in complexes of compounds labeled as actives and inactives, respectively. This means that as the first interactions become stronger, the last ones become weaker. This finding suggests that active and inactive (less active) compounds stabilize different conformations of Cz.

**2.5. Charge Density Molecular Graphs.** **Figure 5** shows the structural superposition of complexes Cz–6b and Cz–8d corresponding to compounds from active and inactive classes, respectively. Interactions that are either formed/broken (or just strengthened/weakened) in the comparison between both complexes are depicted through their corresponding charge



**Figure 3.** Top interactions (features) selected by the SVM model to make the class classifications. The numbers in red indicate the interactions discussed in the text.



**Figure 4.** Correlation matrix based on charge density data from interactions in Cz–Inh complexes.

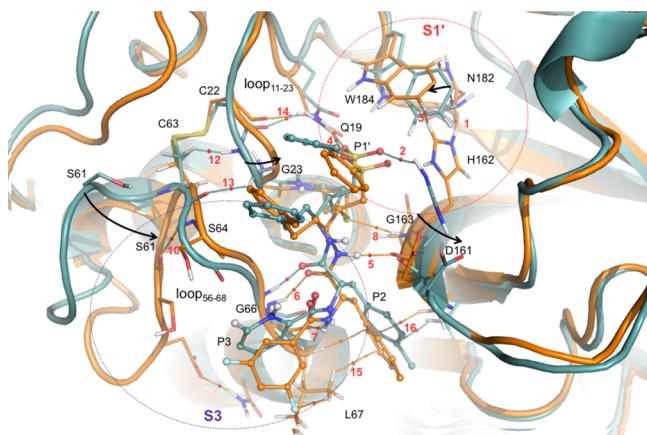
density topological elements (i.e., the BCPs and BPs). Charge density values for the discussed interactions are shown in Table S1 in the *Supporting Information*.

Among interactions that are prevalent in the most active group of Cz inhibitors, H-bond N–H...O=C between side chains of protonated His162 and Asn182 at the S1' enzyme subsite is the most relevant one for the classification task, according to the bar plot in Figure 3. This interaction can be identified as interaction 1 in the molecular graph of Figure 5 and in the Figure 3 bar plot.

As it is well known, interaction 1 facilitates the formation of the thiolate–imidazolium ion pair (Cys25) $S^- \cdots^+ H-N-($ His162) necessary for catalysis.<sup>29,30</sup> Therefore, it is remarkable that this interaction is formed by compounds from the

active class, such as compound **6b**, but not by compounds in the inactive class, such as compound **8d**. This means that compounds labeled as actives better mimic the enzyme substrate because they are able to accommodate the catalytic machinery as if it were about to cleave the substrate scissile bond.

In complexes of compounds from the inactive class, the His162 side chain is displaced away from Asn182 and twisted toward the inhibitor vinyl sulfone P1' moiety forming a strong (P1')S=O...H–N(His162) interaction which is one of the main features of the machine learning (ML) model among compounds labeled as inactives (interaction 2 in Figures 3 and 5). In these complexes, a nearby indole ring from Trp184



**Figure 5.** Structural superposition of Cz-6b (orange) and Cz-8d (light blue) complexes. Charge density topological elements for atomic interactions are also depicted: BPs connecting the nuclei are depicted in orange and light blue for Cz-6b and Cz-8d, respectively. BCPs are shown in small red spheres. Numbers in red indicate the most significant interactions (the same as Figure 3). Arrows indicate protein backbone displacement between Cz-8d and Cz-6b complexes.

occupies the space where residues His162 and Asn182 are going to interact in the active complexes.

Conversely, in complexes of compounds from the active class, opposite changes are observed: interaction 2 is weakened and His162 moves somewhat toward Asn182 to form interaction 1.

However, before interaction 1 can be established, the Trp184 ring must first vacate the region between residues His162 and Asn182. In doing so, Trp moves away from Asn182 and ends up right on top of the His162 ring where the Trp electron cloud forms a C–H···π stacking interaction with an His nonpolar hydrogen atom. This interaction, labeled as 3 in Figures 3 and 5, is also regarded by the SVM-RFE model as one of the main interactions among the active class of inhibitors.

We believe that these findings recovered with the help of an ML model are meaningful because Trp184 is a highly conserved residue among lysosomal cysteine proteases belonging to papain superfamily, and it was previously regarded as the “orchestrator” of the catalytic triad Cys25–His162–Asn182 because it is believed that it plays a critical role in the cleavage of the substrate by orienting the enzyme catalytic machinery.<sup>30</sup> Accordingly, based on our results and previous findings, we propose that Trp184 might act as a “switch” for interaction 1 formation.

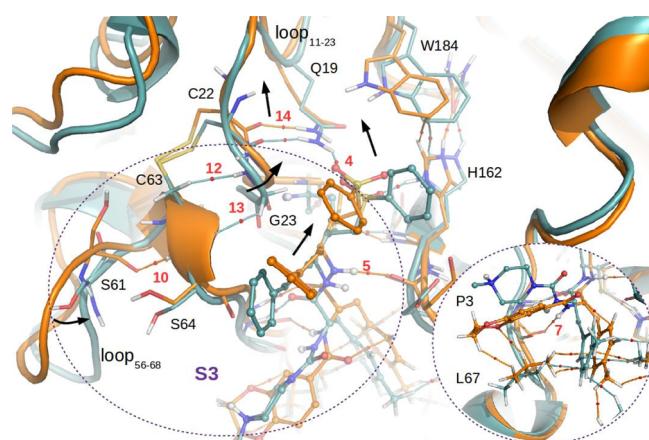
Continuing with analysis of relevant interactions at the S1' subsite, it can be seen in Figure 5 that the inhibitor sulfonic group is held in place at the entrance of the S1' subsite by two strong O···H interactions between both sulfonic O atoms and H atoms from the His162 imidazolium ring (interaction 2) and Gln19 side chain amide group (interaction 4). Both interactions are relevant features among the inactive class of inhibitors as evidenced in Figure 3. This means that these interactions are stronger in complexes of compounds from the inactive class and either are broken or become weaker in complexes of the active class.

It seems like when vinyl sulfones are strongly attached through interactions 2 and 4 as in the case of compounds labeled as inactives, the remaining inhibitor parts do not fit

well within the binding cleft, and so, they cannot establish other important interactions that help to properly attach the peptide-like backbone of the inhibitor. More concretely the backbone of residues P1 and P2 from the inhibitor do not fit properly into the narrow part of the binding cleft formed by backbone atoms from the enzyme S1 subsite (see below).

For the inhibitor to fit well into the enzyme binding cleft, it must be able to disturb the arrangement of residues within the S1' subsite. In other words, it must be able to either break or weaken interactions 2 and 4 that hold firmly the vinyl sulfone P1' moiety at the entrance of the S1' subocket.

These rearrangements involve shifting of His162 toward Asn182 and subsequent interaction 1 formation as explained above (Figure 5). It also involves retraction of the Gln19 side chain as discussed below. Rearrangements of the Gln19 side chain can be more clearly seen in Figure 6 which shows the structural superposition of complexes of compounds 9d and 4b from the active and inactive class, respectively.



**Figure 6.** Structural superposition of Cz-9d (orange) and Cz-4b (light blue) complexes. Charge density topological elements for atomic interactions are also depicted: BPs connecting the nuclei are depicted in orange and light blue for Cz-9d and Cz-4b, respectively. BCPs are shown with small red spheres. Arrows indicate protein backbone displacement between Cz-4b and Cz-9d complexes. Interactions of the P3 residue with Leu67 are highlighted in the bottom right snapshot.

Depending on the complexes analyzed, some inhibitors from the active class seem to push forward residues Gln19 and His162 so that the vinyl sulfone P1' moiety can penetrate a little more deeply into the S1' subsite (Figure 6). Thus, Gln19 and His162 might act as gatekeepers by selectively allowing the entrance to the P1' subsite to only the most active inhibitors.

After this rearrangement at the S1' subsite, the backbone of inhibitor residues P1 and P2 now fits well into the narrow region of the binding cleft. This is evidenced by the backbone–backbone interactions (P1)N–H···O=C(Asp161), (P2)C=O···H–N(Gly66), and (P2)N–H···O=C(Gly66) which are formed or enhanced in complexes of compounds labeled as actives and are some of the most relevant features among the active class, according to the SVM-RFE model (interactions 5, 6 and 7, respectively in Figures 3, 5, and 6). These interactions are also considered as a hallmark of the substrate recognition event in cysteine proteases.<sup>31</sup> Also, interaction (Cys25)S···H–N(Gly163) formation (interaction 8) helps to pull the inhibitor backbone (which is covalently bound to Cys25) toward the bottom of the narrow region of

the enzyme binding cleft, thus contributing to the overall better fit of the inhibitor which is observed in complexes of the most active compounds.

Because vinyl sulfone analogues reported by Jaishankar<sup>8</sup> differ only in P2 and P3 residues, the explanation about why some inhibitors are able to induce the required residue rearrangements within the P1' subsite and some other do not must be related in some way to interactions they establish at the S2/S3 subsites.

**2.5.1. Interactions at the S3 SubSite.** The P3 ring from compounds labeled as actives and inactives interacts in different ways with the key residue Leu67 at the S3 subsite (see Figures 5 and 6). Compounds from the active class have electron-rich groups at P3, and so, they tend to act as H-bond acceptors against the side chain of Leu67. This is evidenced, for example, by interactions such as Leu(67)C—H···π(P3) and Leu(67)C—H···F(P3) in which the electron cloud or fluorine lone pairs from the 3,5-difluorophenyl ring (compound 6b) act as acceptors (Figure 5). In the other case, oxygen lone pairs from the 2,3-dihydro-1,4-benzodioxin ring (compound 9d) act as the H-bond acceptor in interaction Leu(67)C—H···O(P3) (Figure 6). Unfortunately, these interactions are not recovered by the SVM-RFE model (if so, the model would be overfitting the charge density data) because there is no unique H-bond pattern to Leu67 (i.e., there are different H-bond acceptors).

On the other hand, compounds from the inactive class have electron-deficient P3 rings (2-pyridinium and N-methyl piperazine in series 8 and 4, respectively, see Table 1), and so, they only can form dihydrogen contacts with the Leu67 side chain, which are recovered by the ML model as one of the most important features among complexes of compounds labeled as inactives (interaction 9 in Figure 3).

From the mechanistic point of view, strong anchoring of the P3 ring to the Leu67 side chain in complexes of active compounds might pull the inhibitor toward the bottom of the binding cleft, thus allowing formation of backbone–backbone interaction 7 between the inhibitor P2 residue and Gly66, (P2)N—H···O=C(Gly66) (Figures 3, 5, and 6). As argued above, interaction 7 formation, together with interactions 5, 6, and 8, is an indicative of a good fit of the inhibitor backbone within the enzyme binding cleft.

Besides this direct effect of P3 interaction with Leu67 on the anchoring of the inhibitor backbone, there seems to be also an indirect mechanism by which P3 interactions at the S3 subsite influence the inhibitor binding mode.

In complexes of compounds labeled as actives, residues Ser61 and Ser64 from the same loop as Leu67 (i.e., loop<sub>S6–S8</sub>) form a C=O···H—N H-bond which stabilizes a closed turn between both residues. This (Ser61)C=O···H—N(Ser64) interaction (labeled as interaction 10 in Figures 3, 5, and 6) is recovered by the SVM-RFE model as the second most important feature among active-like interactions for the classification of compounds labeled as actives/inactives based on  $K_i$  values. It is likely that stability of interaction 10 is related at least in part with the type of interactions that the inhibitor P3 ring forms with the Leu67 side chain. An unstable dihydrogen bond pattern between P3 and Leu67 (i.e., through interaction 9) as in complexes of compounds labeled as inactives might perturb conformation of the loop<sub>S6–S8</sub>, thus leading to the observed breakage of interaction 10. Conversely, stable H-bonds between P3 and Leu67 like in complexes of compounds from the active class might help to hold more

firmly the loop, thus contributing to preserve the Ser61 → Ser64 turn in its closed form.

Moreover, conformation of the Ser61 → Ser64 turn seems to define how the loop<sub>S6–S8</sub> is going to interact with the surrounding protein structural elements like the nearby loop<sub>11–23</sub>. In complexes of compounds from the inactive class, there are several interactions recovered by the SVM-RFE model as inactive-like interactions that might help to maintain loops S6–S8 and 11–23 close together (i.e., (Cys63)H···N(Gly23), (Cys63)H···H(Gly23), and (Cys63)O···H(Gly23), labeled as interactions 11, 12, and 13, respectively). On the other hand, upon interaction 10 formation in complexes of compounds from the active class, there is a conformational rearrangement in the loop<sub>S6–S8</sub> that somehow causes the breaking of interactions 11, 12, and 13 that were holding both loops together. As the loop<sub>11–23</sub> moves away from the loop<sub>S6–S8</sub>, the first loop drags side chain of Gln19 through an interaction with the backbone of that loop (interaction 14, Figures 3, 5, and 6). While Gln19 is dragged backward, its side chain acquires a twisted conformation in which Gln19 gets further apart from the inhibitor. As a consequence, interaction 4 between the Gln19 side chain and inhibitor sulfonyl oxygen atom gets weakened. As discussed previously, rearrangement of the Gln19 side chain seems to be critical for proper positioning of the substrate within the Cz binding cleft and formation of backbone interactions 5, 6, 7, and 8.

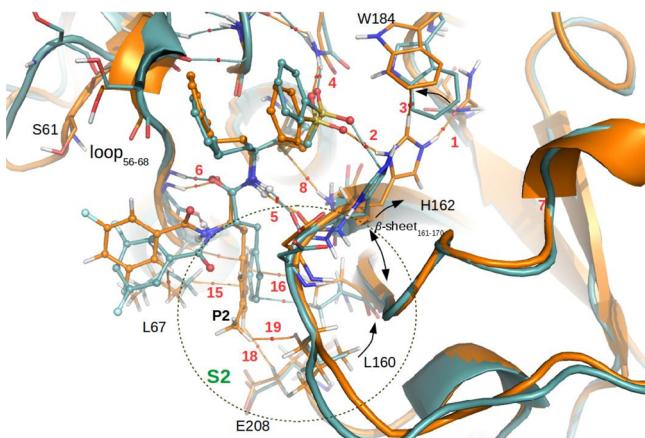
**2.5.2. Interactions at the S2 SubPocket.** Among all the subsites that encompass the Cz binding cleft, the only one which is deep enough to deserve the name of the subocket is S2. At the S2 subocket, the anchoring of compounds from the active class is mostly driven by π···H interactions between the P2 ring electron cloud and nonpolar hydrogens donated by Leu67 and Ala138 residues at both sides of the subocket. These interactions, named 15 and 16, respectively, have been selected by the SVM-RFE model as important features among compounds labeled as actives (see Figures 3, 5, and 6). On the other hand, compounds from the inactive class either do not form interactions 15 and 16 or they are much weaker. Instead, the P2 ring from these compounds forms dihydrogen contacts with Leu67 (interaction 17) which highlights the misplacement of the P2 ring within the S2 subocket.

Bringing together interactions analyzed for P2 and P3 residues, it is evident that Leu67 plays a key role in proper anchoring of both residues to the Cz binding cleft.

Figure 7 shows the structural superposition of complexes of compounds 6b and 6a from active and inactive classes, respectively. These compounds only differ in the substituent at the P2 residue, and so, they are suitable for studying differences in interaction patterns that can be directly attributed to the P2 structure.

Besides driving interactions with Leu67 and Ala138, most active compounds also form other interactions that it is worth noting. Thus, for example, interaction (P2)H···H(Glu208) between two nonpolar H atoms from P2 and Glu208 side chains, respectively, was selected by the SVM-RFE model as a relevant feature among inhibitors from the active class (interaction 18 in Figures 3 and 7). Glu208 lies at the bottom of the S2 subocket; hence, interaction of inhibitors from the active class with that residue indicates that they are able to reach such a distal region of the S2 subsite while inhibitors from the inactive class, in general, are not.

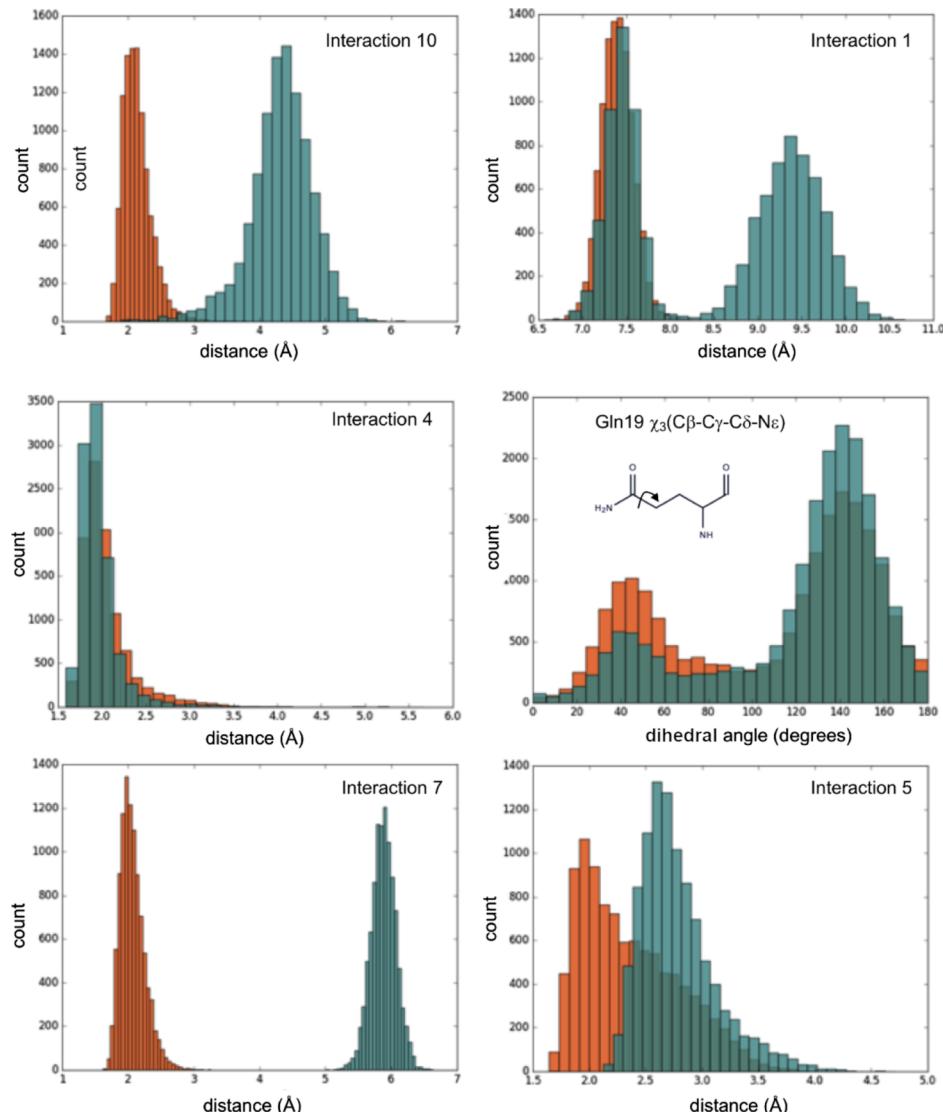
Close to Glu208, there is another residue, Leu160, which is also targeted by most active inhibitors through dihydrogen



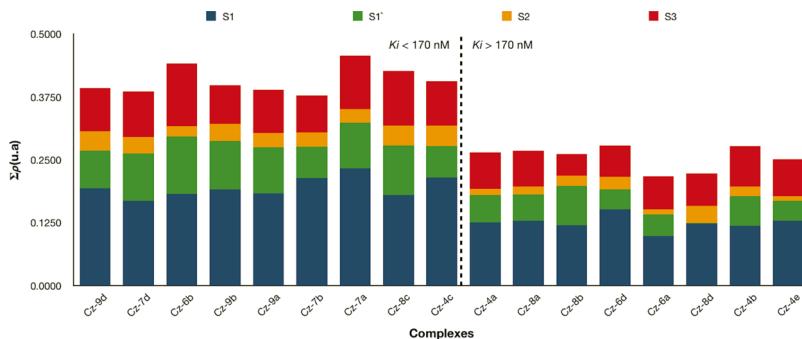
**Figure 7.** Structural superposition of Cz-6b (orange) and Cz-6a (light blue) complexes. Charge density topological elements for atomic interactions are also depicted: BPs connecting the nuclei are depicted in orange and light blue for Cz-6b and Cz-6a, respectively. BCPs are shown with small red spheres. Arrows indicate protein backbone displacement between Cz-6b and Cz-6a complexes.

interactions (P2)H···H(Leu160) which are also recovered by the ML model as a relevant feature among complexes of compounds labeled as actives (interaction 19 in Figures 3 and 7).

It is unlikely that attractive forces would be behind formation of these dihydrogen interactions as they are more suggestive of steric crashes between hydrophobic atoms from the ligand and enzyme. These subtle dihydrogen crashes usually are the footprints left by stronger repulsive forces that have been alleviated by displacements of the involved residues. Therefore, by inspecting these dihydrogen interactions, one can track back residue translocations or conformation changes that might have happened as a consequence of a former stronger steric crash. In particular, dihydrogen interactions 18 and 19 are the footprints of Glu208 and Leu160 side chain displacements, respectively, induced by substituents at the 3 or 4 position of the inhibitor P2 ring (see Table 1). In contrast, compounds that do not bear a substituent on the P2 ring, most of them belonging to the inactive class, do not reach the distal wall/bottom of the S2 subsite, and so, they do not form interactions 18 and 19. These interactions and in particular



**Figure 8.** Distance histograms for selected interactions from complexes Cz-6b (orange) and Cz-8d (cyan). Also, the histogram for the Gln19 side chain torsional angle is depicted. Distances corresponding to interaction 1 were measured between the center of mass of the involved residues.



**Figure 9.** Sum of the charge density values at the BCPs due to intra–intermolecular interactions in Cz–inhibitor complexes. Values are partitioned into four contributions corresponding to subpockets S1 (blue), S1' (green), S2 (orange), and S3 (red). From left to right, complexes are ordered in increasing values of  $K_i$ . Complexes are divided, with a dotted line, into two groups according to the decision threshold value ( $K_i$  170 nM) used in the SVM section. The compound nomenclature was extracted from Jaishankar.<sup>8</sup>

interaction 19 seem to be related with residue rearrangements at the S1 and S1' subsites. As the substituent at the P2 ring pushes away the side chain of Leu160, it also perturbs backbone interactions between the nearby  $\beta$ -sheet<sub>161–170</sub> and  $\beta$ -sheet<sub>135–139</sub> that are interacting in a hairpin-like motif. As a consequence, the backbone of the  $\beta$ -sheet<sub>161–170</sub> is partially “released” and residues at the end of that sheet, that is, Asp161, His162, and Gly163 experience a concerted backward movement that place them in a proper position as to form interactions 5 and 8 at the S1 subsite and triggers rearrangements at the S1' subsite involving His162 that ultimately leads to formation of interaction 1, as discussed previously.

Taking together the inhibitor interactions at the S2 subpocket and S3 subsite, the first ones seem to govern the conformational changes occurring on the right wall of the binding cleft (i.e., those involving the  $\beta$ -sheet<sub>161–170</sub>), while P3 residue interactions at the S3 subsite mostly drive the conformational changes on the left wall (i.e., those related to the loop<sub>56–68</sub>). Both conformational changes ultimately lead to rearrangements of residues His162 and Gln19 at the S1' site that allows the proper positioning of the vinyl sulfone warhead which in turn promotes formation of backbone–backbone interactions between the inhibitor and the binding cleft wall that are critical for inhibition.

Nevertheless, it should be kept in mind that the dissection of the inhibition mechanism problem by protein subsites might be an oversimplification because interactions at different subsites might be related to each other, namely, the conformational changes observed might depend not only on the substituents at P2 and P3 but also on the combination of both.

**2.6. Two End-State Conformational Model for Cz Supported by MD Simulations.** In Section 2.3, we separated interactions that are more prevalent in complexes of most active inhibitors (i.e., active-like interactions) from those that are more common in complexes of compounds from the inactive class (i.e., inactive-like interactions). Then, in Section 2.4, through the correlation analysis, we took a step further to conclude that active-like and inactive-like interactions stabilize two opposite conformations of the enzyme.

To further support this hypothesis, we looked at some active-like and inactive-like interactions along the MD trajectories of Cz–Inh complexes.

**Figure 8** depicts distance histograms from MD simulations of complexes Cz–6b and Cz–8d corresponding to several interactions regarded by the SVM-RFE model as important features for stabilization of either active or inactive end-state enzyme conformations. Also, a histogram for the Gln19 side chain torsional angle is depicted.

As evidenced in Figure 8, several interactions show a bimodal distribution of frequencies in which they are either formed or broken, which is in agreement with the two end-state conformational model proposed based on charge density analysis of selected structures from different MD simulations.

Distance distribution of interaction 1 in complex Cz–8d makes evident the two conformational states of residue His162. In that complex, His162 is roughly half of the time far away from Asn182 as in Cz conformation stabilized by less active inhibitors. On the other hand, in complex Cz–6b, His162 is close to Asn182 during the entire simulation time, thus favoring interaction 1 formation as in the conformation stabilized by most active Cz inhibitors.

Moreover, interaction 10 which is involved in loop<sub>56–68</sub> conformation is formed most of the time of the simulation in complex Cz–6b, thus stabilizing the closed form of the loop, whereas, the same interactions are mainly broken during the simulation of the Cz–8d complex.

As discussed previously, as a consequence of the loop<sub>56–68</sub> re-organization on going from complex of less active to most active Cz inhibitors, the loop<sub>11–23</sub> is also displaced upward and drags with it the Gln19 side chain through interaction 14 (not shown). Concretely, the dragging motion involves the twisting of the Gln19 side chain which is evidenced by the bimodal population of the  $\chi^3$  torsion angle where the twisted conformation is represented by the distribution around 40°. It can be seen in Figure 8 that the Gln19 side chain remains more time twisted in complex Cz–6b than in Cz–8d. Regarding interaction 4 between the Gln19 side chain amide and inhibitor sulfonyl oxygen atom, it remains formed all the time of simulations. However, the distance distribution is slightly displaced toward largest interaction distances in complex Cz–6b, which is likely a consequence of the lasting Gln19 side chain twisting that place it further apart from sulfonyl oxygens. As discussed previously, the weakening of interaction 4 might contribute to the overall better fit of the inhibitor within the Cz binding cleft.

Finally, distance distributions corresponding to backbone–backbone interactions 7 and 5 show that 6b is more firmly attached than 8d to the backbone of Gly66 and Asp161,

respectively, which is also in line with the previous charge density analysis on selected structures from different MD simulations.

**2.7. SubPocket Decomposition of the Binding Affinity.** Because charge density, as measured at the interaction critical point, is a local topological property, we can compute the contribution of a subset of such charge density values to the inhibitor total anchoring strength. In that way, we could know on which of the enzyme subpockets, the interactions with the inhibitor need to be improved.

Figure 9 shows the decomposition of the charge density values at the BCPs in Cz–Inh complexes by subpockets.

As can be seen in Figure 9, on going from the less active inhibitors to the most active ones, the inhibitor anchoring strength gets improved not in a particular subpocket but on all the enzyme subpockets. This finding is in line with our previous results. We have seen that substitution at P2 and P3 positions of the inhibitor not only induce changes in the S3 and S2 enzyme subpockets but also the entire enzyme binding cleft is aware of such substitutions. This strong communication between the different enzyme subpockets anticipates that the optimization of interactions separately on each of the enzyme subpockets might be difficult to achieve. Similarly, a fragment-based approach for drug design would also be challenging for the same reason. In a fragment-based pipeline, for discovery of novel Cz inhibitors, one would presumably start with a small fragment able to bind to the S2 subpocket (i.e., the easiest subpocket to target) and from there it would have to be enlarged toward the neighbor subpockets either by the fragment-growing or fragment-linking approach. Because of the strong inter-relationship between subpockets that we have shown throughout this work, there is no guarantee that on growing the S2 fragment toward S3, for example, the former interactions at S2 would be maintained.

### 3. CONCLUSIONS

In this work, we have calculated, analyzed, and summarized molecular interactions that arise from quantum calculations on complexes of Cz with 17 known inhibitors at the Cz binding site where the analysis of activity differences in terms of molecular interactions at the Cz binding cleft has not been described yet.

QTAIM provided topological elements of the charge density that describe the interactions in the Cz–Inh complexes. At this point, with more than three hundred interactions per complex, we trained a supervised learning classification model with RFE that discriminates between interactions present in complexes of the most-active inhibitors (active-like interactions) and those that occur in the less-active ones (inactive-like interactions). Moreover, the model also provided information about the interaction importances, namely, which are the most important interactions to discriminate between complexes of active and inactive (or less active) compounds against Cz.

Our model allowed us to point out 19 inter-/intramolecular main interactions that could explain the principal changes in the complexes under analysis.

Among the intermolecular interactions, backbone–backbone interactions 5, 6, 7, and 8 as well as interactions of inhibitor residues P2 and P3 with the Leu67 side chain play a key role in proper anchoring of most active inhibitors into the enzyme binding cleft. Unfortunately, no quantitative relationship was found between the structure and activity data when considering only intermolecular interactions.

By taking into account also intramolecular interactions and with the help of the SVM-RFE model to separate active-like from inactive-like interactions, a more indirect mechanism of enzyme inhibition involving extensive conformational changes within protein structure arises.

These protein conformational changes occur on both “walls” of the binding cleft promoted by intermolecular interactions at the S2 and S3 sites. Inhibitor interactions at the S2 subpocket trigger conformational changes on the  $\beta$ -sheet<sub>161–170</sub> (right wall), while interactions at the S3 subsite mostly drive conformational changes on the loop<sub>56–68</sub> (left wall). Both conformational changes ultimately lead to re-arrangements of residues His162 and Gln19 at the S1' site that allows proper positioning of the vinyl sulfone warhead and formation of key backbone–backbone interactions between the peptide-like inhibitor and binding cleft wall residues.

On the other hand, our study also allowed us to understand how important the role of the highly-conserved Trp184 is, enabling interaction 1 formation that leads to activation of the catalytic histidine. The “switching activity” of Trp184 is crucial for accommodation of the catalytic triad. Different interactions “orchestrated” by this residue determine activation/inactivation of the protein machinery.

In this regard, we have found that most active Cz inhibitors induce a conformation in which interactions considered as a hallmark of the substrate recognition event are present. Having isolated this “activated” Cz structure, we can use it in rigid docking experiments in the context of prospective virtual screening campaigns to “fish” highly active Cz inhibitors from compound databases.

Moreover, among relevant interactions that stabilize the “activated” Cz conformation, intermolecular interactions such as 5, 6, and 7 could be plugged into the docking algorithms to customize the scoring function and guide the docking predictions.

Finally, throughout this study, we also got a sense of the strong communication that exists between the enzyme binding cleft subpockets, the property that might help us to choose the best approach to follow in prospective screening campaigns. Probably a fragment-based approach is not the best choice in this case because of this property of high inter-relationship between subpockets.

All the collected information would be taken into account in the following prospective studies aimed to search novel Cz inhibitors.

### 4. COMPUTATIONAL DETAILS

**4.1. Simulation Protocol.** Jaishankar<sup>8</sup> synthesized and determined the inhibition constants against Cz of a series of vinyl sulfone analogues closely related to K-777.

Although the experimental structure of these vinyl sulfone analogues in the complex with Cz has not been determined yet (except for K-777, pdb id = 2OZ2), for peptide-like Cz inhibitors, a reasonably accurate initial guess of the inhibitor binding mode can be constructed “by hand” by placing each residue in the inhibitor sequence P1', P1, P2, and P3 into its own enzyme subpockets S1', S1, S2, and S3.

Initial coordinates of the complex were taken from the structure of Cz bound to K-777 (pdb id = 2OZ2).<sup>21</sup> By performing substitutions at P2 and P3 residues of K-777 to get the analogues reported by Jaishankar,<sup>8</sup> 17 closely related complexes were constructed and then refined by MD simulations.

All the Cz–inhibitor complex simulations were carried out with Amber14 software package<sup>32,33</sup> at 300 K temperature and extended up to 50 ns overall simulation time in a truncated octahedral periodic box of TIP3P water molecules. Amber ff14SB force field was used for proteins residues.<sup>34</sup> The antechamber software in the Amber-Tools package was used to generate ligand inhibitor parameters with GAFF force field and RESP charges.<sup>35</sup>

**4.2. Quantum Theory of Atoms in Molecules.** The structure of the potential energy minimum was selected from the MD trajectories of Cz–Inh complexes as a single representative structure upon which the charge density analysis was done.

Because accurate quantum mechanical calculations are still forbidden for full biomolecular complexes, reduced models were constructed from the potential energy minimum structures. A total of 28 residues (~570 atoms) were included in the reduced models: the vinyl sulfone inhibitor and the surrounding residues in a spherical volume of about 5 Å centered on the inhibitor atoms (Figure S1 in the Supporting Information shows the residues included in the reduced models).

The charge density was computed by density functional theory methodology with the M06-2x dispersion corrected hybrid functional and 6-31G(d) as the basis set, as implemented in Gaussian 09 package.<sup>36</sup> The topological analysis of charge density was then performed with Multiwfn software.<sup>37</sup>

**4.3. Support Vector Machines–Recursive Feature Elimination.** Charge density values associated to 319 noncovalent interactions per complex were used as features to train a linear SVM classifier. SVMs are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.<sup>38</sup>

If the data are not separable by a hyperplane, they can be mapped into feature spaces of higher dimensionality where linear separation of positive and negative examples might be possible (i.e., the so-called kernel trick). However, unlike linear models, SVM models trained on high-dimensional kernel spaces have black box character, and it is generally difficult to rationalize model performance.<sup>39</sup> Therefore, in this article, we restricted ourselves to linear SVM because our main interest was in uncovering relationships between the features (i.e., molecular interactions) and the biological activities to understand, ultimately, the enzyme inhibition mechanism. Nevertheless, it is important to bear in mind that the analysis of features that contribute to predictions only makes sense if the model reaches a reasonably high-performance level.

It is well known that when the number of features is large and the number of training examples is comparatively small, the risk of overfitting arises. Therefore, to overcome the problem of high dimensionality and scarce samples of our data set, SVM was coupled with the RFE algorithm during model training.

SVM-RFE is a feature selection algorithm based on backward elimination of features with lowest weights. In each iteration, the SVM model is trained with the current subset of features, the weight ( $|w|$ ) of each feature is calculated according to the SVM classifier, the features are ranked according to  $|w|$ , and then, the bottom-ranked features are eliminated.<sup>26</sup>

SVM-RFE and stratified twofold cross-validation were implemented with the help of the scikit-learn module of Python.<sup>40</sup>

**4.4. Dynamic Cross-Correlation Analysis.** The correlation matrix describing how interactions are related to each other among the Cz–Inh complexes was computed from charge density data obtained from QTAIM calculations. Only interactions with importances greater than 2.0 in the SVM-RFE classifier were considered for the correlation analysis.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acsomega.9b01934](https://doi.org/10.1021/acsomega.9b01934).

Residues that conform subpockets S1, S1', S2, and S3 from the cruzain binding cleft and local charge density values corresponding to interactions discussed in the main text (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: [emilioluisangelina@hotmail.com](mailto:emilioluisangelina@hotmail.com) (E.L.A.).  
\*E-mail: [arabeshai@yahoo.com.ar](mailto:arabeshai@yahoo.com.ar) (N.M.P.).

### ORCID®

Adriano M. Luchi: [0000-0003-2221-0458](https://orcid.org/0000-0003-2221-0458)  
Emilio L. Angelina: [0000-0002-6218-2765](https://orcid.org/0000-0002-6218-2765)

### Author Contributions

The manuscript was written through contributions of all the authors. All the authors have given approval to the final version of the manuscript. All the authors contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge SECyT-UNNE and CONICET for financial support. Geforce GTX Titan X GPU partially used for this research was donated by NVIDIA Corporation. The authors also acknowledge the computer resources at MareNostrum and the technical support provided by Barcelona Supercomputing Center (RES-BCV-2019-3-0015). The authors acknowledge the reviewers whose critical reading and suggestions helped to improve and clarify this manuscript.

## ■ REFERENCES

- (1) World Health Organization. Chagas disease (American trypanosomiasis). [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis)) (accessed Aug 6, 2019).
- (2) de Souza, A. S.; de Oliveira, M. T.; Andricopulo, A. D. Development of a Pharmacophore for Cruzain Using Oxadiazoles as Virtual Molecular Probes: Quantitative Structure-Activity Relationship Studies. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 801–816.
- (3) Wiggers, H. J.; Rocha, J. R.; Fernandes, W. B.; Sesti-Costa, R.; Carneiro, Z. A.; Cheleski, J.; da Silva, A. B. F.; Juliano, L.; Cezari, M. H. S.; Silva, J. S.; McKerrow, J. H.; Montanari, C. A. Non-Peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay. *PLoS Neglected Trop. Dis.* **2013**, *7*, No. e2370.
- (4) McGrath, M. E.; Eakin, A. E.; Engel, J. C.; McKerrow, J. H.; Craik, C. S.; Fletterick, R. J. The Crystal Structure of Cruzain: A Therapeutic Target for Chagas' Disease. *J. Mol. Biol.* **1995**, *247*, 251–259.

- (5) Bank, R. P. D. RCSB PDB: Homepage. <https://www.rcsb.org/> (accessed Aug 6, 2019).
- (6) Barbosa da Silva, E.; Dall, E.; Briza, P.; Brandstetter, H.; Ferreira, R. S. Cruzain Structures: Apocruzain and Cruzain Bound to S-Methyl Thiomethanesulfonate and Implications for Drug Design. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2019**, *75*, 419–427.
- (7) Herrera-Mayorga, V.; Lara-Ramírez, E.; Chacón-Vargas, K.; Aguirre-Alvarado, C.; Rodríguez-Páez, L.; Alcántara-Farfán, V.; Cordero-Martínez, J.; Nogueda-Torres, B.; Reyes-Espinosa, F.; Bocanegra-García, V.; Rivera, G. Structure-Based Virtual Screening and In Vitro Evaluation of New Trypanosoma Cruzi Cruzain Inhibitors. *Int. J. Mol. Sci.* **2019**, *20*, 1742.
- (8) Jaishankar, P.; Hansell, E.; Zhao, D.-M.; Doyle, P. S.; McKerrow, J. H.; Renslo, A. R. Potency and Selectivity of P2/P3-Modified Inhibitors of Cysteine Proteases from Trypanosomes. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 624–628.
- (9) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press, 1990.
- (10) Tosso, R. D.; Andujar, S. A.; Gutierrez, L.; Angelina, E.; Rodríguez, R.; Nogueras, M.; Baldoni, H.; Suvire, F. D.; Cobo, J.; Enriz, R. D. Molecular Modeling Study of Dihydrofolate Reductase Inhibitors. Molecular Dynamics Simulations, Quantum Mechanical Calculations, and Experimental Corroboration. *J. Chem. Inf. Model.* **2013**, *53*, 2018–2032.
- (11) Tosso, R. D.; Vettorazzi, M.; Andujar, S. A.; Gutierrez, L. J.; Garro, J. C.; Angelina, E.; Rodríguez, R.; Suvire, F. D.; Nogueras, M.; Cobo, J.; Enriz, R. D. The Electronic Density Obtained from a QTAIM Analysis Used as Molecular Descriptor. A Study Performed in a New Series of DHFR Inhibitors. *J. Mol. Struct.* **2017**, *1134*, 464–474.
- (12) Gutierrez, L. J.; Angelina, E.; Gyebrovszki, A.; Fülpö, L.; Peruchena, N.; Baldoni, H. A.; Penke, B.; Enriz, R. D. New Small-Size Peptides Modulators of the Exosite of BACE1 Obtained from a Structure-Based Design. *J. Biomol. Struct. Dyn.* **2017**, *35*, 413–426.
- (13) Barrera, Guisasola, E. E.; Gutierrez, L. J.; Andujar, S. A.; Angelina, E.; Rodríguez, A. M.; Enriz, R. D. Pentameric Models as Alternative Molecular Targets for the Design of New Antiaggregant Agents. *Curr. Protein Pept. Sci.* **2016**, *17*, 156–168.
- (14) Párraga, J.; Cabedo, N.; Andujar, S.; Piqueras, L.; Moreno, L.; Galán, A.; Angelina, E.; Enriz, R. D.; Ivorra, M. D.; Sanz, M. J.; Cortes, D. 2,3,9- and 2,3,11-Trisubstituted Tetrahydroprotoberberines as D2 Dopaminergic Ligands. *Eur. J. Med. Chem.* **2013**, *68*, 150–166.
- (15) Angelina, E.; Andujar, S.; Moreno, L.; Garibotto, F.; Párraga, J.; Peruchena, N.; Cabedo, N.; Villecco, M.; Cortes, D.; Enriz, R. D. 3-Chlorotyramine Acting as Ligand of the D2 Dopamine Receptor. Molecular Modeling, Synthesis and D2 Receptor Affinity. *Mol. Inf.* **2015**, *34*, 28–43.
- (16) Angelina, E. L.; Andujar, S. A.; Tosso, R. D.; Enriz, R. D.; Peruchena, N. M. Non-Covalent Interactions in Receptor–Ligand Complexes. A Study Based on the Electron Charge Density. *J. Phys. Org. Chem.* **2014**, *27*, 128–134.
- (17) Luchi, A. M.; Angelina, E. L.; Andujar, S. A.; Enriz, R. D.; Peruchena, N. M. Halogen Bonding in Biological Context: A Computational Study of D2 Dopamine Receptor. *J. Phys. Org. Chem.* **2016**, *29*, 645–655.
- (18) Parravicini, O.; Bogado, M. L.; Rojas, S.; Angelina, E. L.; Andujar, S. A.; Gutierrez, L. J.; Cabedo, N.; Sanz, M. J.; López-Gresa, M. P.; Cortes, D.; Enriz, R. D. Tetrahydroisoquinolines Functionalized with Carbamates as Selective Ligands of D2 Dopamine Receptor. *J. Mol. Model.* **2017**, *23*, 273.
- (19) Vettorazzi, M.; Angelina, E.; Lima, S.; Gonçalves, T.; Otevrel, J.; Marvanova, P.; Padrtova, T.; Mokry, P.; Bobal, P.; Acosta, L. M.; Palma, A.; Cobo, J.; Bobalova, J.; Csöllei, J.; Malik, I.; Alvarez, S.; Spiegel, S.; Jampilek, J.; Enriz, R. D. An Integrative Study to Identify Novel Scaffolds for Sphingosine Kinase 1 Inhibitors. *Eur. J. Med. Chem.* **2017**, *139*, 461–481.
- (20) Luchi, A.; Angelina, E.; Bogado, L.; Forli, S.; Olson, A.; Peruchena, N. Flap-Site Fragment Restores Back Wild-Type Behaviour in Resistant Form of HIV Protease. *Mol. Inf.* **2018**, *37*, 1800053.
- (21) Kerr, I. D.; Lee, J. H.; Farady, C. J.; Marion, R.; Rickert, M.; Sajid, M.; Pandey, K. C.; Caffrey, C. R.; Legac, J.; Hansell, E.; McKerrow, J. H.; Craik, C. S.; Rosenthal, P. J.; Brinen, L. S. Vinyl Sulfones as Antiparasitic Agents and a Structural Basis for Drug Design. *J. Biol. Chem.* **2009**, *284*, 25697–25703.
- (22) Fujita, T.; Winkler, D. A. Understanding the roles of the “Two QSARs”. *J. Chem. Inf.* **2016**, *56*, 269–274.
- (23) Jolliffe, I. T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc., A* **2016**, *374*, 20150202.
- (24) Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification Using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422.
- (25) Bolón-Canedo, V.; Alonso-Betanzos, A. Ensembles for feature selection: A review and future trends. *Inf. Fusion* **2019**, *52*, 1–12.
- (26) Lin, X.; Li, C.; Zhang, Y.; Su, B.; Fan, M.; Wei, H. Selecting Feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics. *Molecules* **2017**, *23*, 52.
- (27) Kuhn, M.; Johnson, K. *Feature Engineering and Selection: A Practical Approach for Predictive Models*, 1st ed.; CRC Press: Boca Raton, Florida, 2019; pp 250–270.
- (28) Zhou, X.; Tuck, D. P. MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data. *Bioinformatics* **2007**, *23*, 1106–1114.
- (29) Brinen, L. S.; Hansell, E.; Cheng, J.; Roush, W. R.; McKerrow, J. H.; Fletterick, R. J. A Target within the Target: Probing Cruzain’s P1’ Site to Define Structural Determinants for the Chagas’ Disease Protease. *Structure* **2000**, *8*, 831–840.
- (30) Arafat, K.; Świderek, K.; Moliner, V. Computational Study of the Michaelis Complex Formation and the Effect on the Reaction Mechanism of Cruzain Cysteine Protease. *ACS Omega* **2018**, *3*, 18613–18622.
- (31) Turk, D.; Günçar, G.; Podobnik, M.; Turk, B. Revised Definition of Substrate Binding Sites of Papain-like Cysteine Proteases. *Biol. Chem.* **1998**, *379*, 137–147.
- (32) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (33) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerrutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossaváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *Amber* 14, 2014.
- (34) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (35) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (36) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.;

Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09*, Revision A.02; Gaussian, Inc., 2016.

(37) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592.

(38) Li, H.; Liang, Y.; Xu, Q. Support Vector Machines and Its Applications in Chemistry. *Chemom. Intell. Lab. Syst.* **2009**, *95*, 188–198.

(39) Rodríguez-Pérez, R.; Vogt, M.; Bajorath, J. Support Vector Machine Classification and Regression Prioritize Different Structural Features for Binary Compound Activity and Potency Value Prediction. *ACS Omega* **2017**, *2*, 6371–6379.

(40) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.