

RESEARCH ARTICLE

Dissecting a novel allosteric mechanism of cruzain: A computer-aided approach

Lilian Hernández Alvarez^{1*}, Diego Enry Barreto Gomes^{2,3}, Jorge Enrique Hernández González¹, Pedro Geraldo Pascutti⁴

1 Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista Julio de Mesquita Filho, São José do Rio Preto, São Paulo, Brazil, **2** Diretoria de Metrologia Aplicada às Ciências da Vida (DIMAV), Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO), Rio de Janeiro, Brazil, **3** Institute de Chimie, Université de Strasbourg, Strasbourg, France, **4** Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

* lilianhern@gmail.com



Abstract

Trypanosoma cruzi is the causative agent of Chagas disease, a neglected infection affecting millions of people in tropical regions. There are several chemotherapeutic agents for the treatment of this disease, but most of them are highly toxic and generate resistance. Currently, the development of allosteric inhibitors constitutes a promising research field, since it can improve the accessibility to more selective and less toxic medicines. To date, the allosteric drugs prediction is a state-of-the-art topic in rational structure-based computational design. In this work, a simulation strategy was developed for computational discovery of allosteric inhibitors, and it was applied to cruzain, a promising target and the major cysteine protease of *T. cruzi*. Molecular dynamics simulations, binding free energy calculations and network-based modelling of residue interactions were combined to characterize and compare molecular distinctive features of the apo form and the cruzain-allosteric inhibitor complexes. By using geometry-based criteria on trajectory snapshots, we predicted two main allosteric sites suitable for drug targeting. The results suggest dissimilar mechanisms exerted by the same allosteric site when binding different potential allosteric inhibitors. Finally, we identified the residues involved in suboptimal paths linking the identified site and the orthosteric site. The present study constitutes the first approximation to the design of cruzain allosteric inhibitors and may serve for future pharmacological intervention. Here, no major effects on active site structure were observed due to compound binding (modification of distance and angles between catalytic residues), which indicates that allosteric regulation in cruzain might be mediated via alterations of its dynamical properties similarly to allosteric inhibition of human cathepsin K (HCatK). The current findings are particularly relevant for the design of allosteric modulators of papain-like cysteine proteases.

OPEN ACCESS

Citation: Hernández Alvarez L, Barreto Gomes DE, Hernández González JE, Pascutti PG (2019) Dissecting a novel allosteric mechanism of cruzain: A computer-aided approach. PLoS ONE 14(1): e0211227. <https://doi.org/10.1371/journal.pone.0211227>

Editor: Freddie Salisbury, Jr, Wake Forest University, UNITED STATES

Received: August 29, 2018

Accepted: January 9, 2019

Published: January 25, 2019

Copyright: © 2019 Hernández Alvarez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (grant number No. 2018/03911-8 to LHA and 2016/24587-9 to JEHG), and by Ministério da Educação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (grant number 031/2013 PRÓ-DEFESA 3 to LHA and PGP). The funders had no role in study design, data collection

Introduction

Cruzain is the major papain-like cysteine protease of *Trypanosoma cruzi*, the protozoan responsible for Chagas disease. This enzyme is indispensable for the survival and propagation of the

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

parasite and, therefore, is considered a potential drug target for the disease control [1–3]. Toxicity and inefficiency of the available chemotherapy [4–6], fueled the pursuit for alternative drugs, which in turn, has led to the discovery of many cruzain inhibitors. A tangible evidence of the latter, is the occurrence of twenty-five crystal structures of cruzain in complex with different competitive inhibitors in the Protein Data Bank (PDB) [1, 7–9]. On the other hand, numerous experimental studies and computational predictions have been performed to characterize the cruzain binding site and specificity, facilitating the design of active-site directed drugs [8–12]. However, the lack of interest of pharmaceutical industry in low profit products (which is the case of tropical diseases) [13], and the failure to find a “real drug” that reaches the production and distribution scales, have encouraged the search for new scaffolds of cruzain inhibitors, as well as different strategies of enzyme inhibition and, ultimately, novel therapeutic targets.

Allostery has usually been defined as the modulation of the protein function triggered by the interaction with a molecule (allosteric modulator) through a protein site (allosteric site) different from the functional one (orthosteric site) [14, 15]. Remarkably, it has been shown that allostery is not only a property of proteins, but also an inherent characteristic of most macromolecules [16, 17]. Since the proposal of the concerted Monod, Wyman, and Changeux (MWC) [18] and the sequential Koshland, Nemethy, and Filmer (KNF) models [19] in the 1960s, conformational changes were considered as a signature of allosteric modulation. However, a more modern concept of allostery takes into account the “conformation ensembles and population shift” rather than the classical two state model [20–24]. According to this concept, macromolecules exist as a population of equilibrium conformations, which has a certain dynamic distribution that may be altered by some environmental changes, such as covalent or noncovalent binding of ligands, pH, temperature, or ionic force [25]. The redistribution of the fraction of active/inactive protein conformations within the population promoted by the binding of an allosteric ligand leads to the modulation of the protein function [16, 25, 26].

In recent decades, the design of allosteric drugs has emerged as a promising research field in disease control [14, 15, 27]. Allosteric modulators offer many advantages that make them appropriate as drug candidates. A major benefit of these strategies, compared to those that perturb the active site directly, is that they offer a noninvasive or more specific protein control. Allosteric modulators neither interfere with the active site ligands nor do they obstruct its vicinity [17, 25, 28].

Several findings support the existence of allosteric modulation in the papain-like proteases [29–35]. One key example is the work of Novinec *et al.*, where several allosteric sites of human cathepsin K (HCatK) and the sector residues potentially mediating allosteric communication in the aforementioned protein family were identified employing the statistical coupling analysis (SCA) method (Fig 1) [36]. Both, in the previous work and in a more recent study, the co-crystallization of HCatK with allosteric inhibitors, positioned in a previously-predicted site 6, was reported (PDBID: 5J94 and 5JA7) [36–38]. On the other hand, there are other cases of allosteric modulation in several parasitic cysteine proteases such as falcipain 2 from *Plasmodium falciparum*, which is allosterically inhibited by heme, suramin and their derivatives, as well as a chalcone-like compound [33–35].

Interestingly, the role of glycosaminoglycans (GAGs) in allosteric modulation of papain-like cysteine proteases has been widely studied in the past decades. Proteases such as human cathepsins S (HCatS) and B (HCatB), brucipain, cathepsin L of *Leishmania mexicana* and even cruzain are allosterically modulated by GAGs [30, 32, 39–43]. However, no other biomolecules or chemical compounds exerting allosteric modulation on cruzain have been reported. So far, there is only an *in silico* work by Durrant *et al.* reporting the characterization of an adjacent subsite to the cruzain orthosteric site (Fig 1) [11]. However, it is expected that the predicted site will be useful for the design of more specific competitive inhibitors, not allosteric ones.

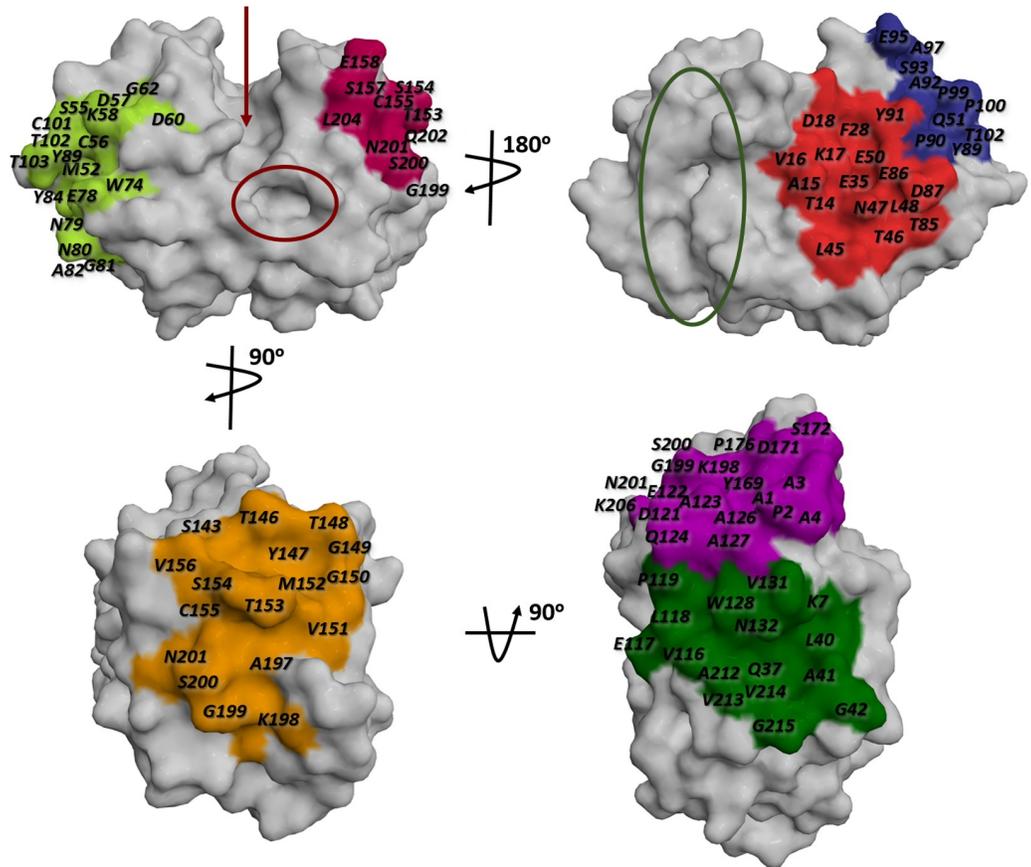


Fig 1. Mapping of cruzain potential allosteric pockets based on previous predictions for papain-like cysteine proteases superfamily. The allosteric sites were numbered from zero to seven, i. e., site 1 (lemon), site 2 (pink), site 3 (red), site 4 (blue), site 5 (orange), site 6 (violet) and site 7 (green). The position of the cruzain active site is indicated with an arrow and an adjacent subsite identified by Durrant *et al.* [11] is enclosed by red circle. The site 0 (glycosaminoglycan binding site in HCatK) is highlighted with a green circle.

<https://doi.org/10.1371/journal.pone.0211227.g001>

In recent years, the computer-aided drug design has been widely used in the field of rational discovery of allosteric inhibitors [37, 44–46]. Techniques such as SCA, molecular docking, molecular dynamics (MD) simulations and allosteric networks have emerged as a valuable complement to experimental methods for the study of allostery [16, 24, 47–49]. These approaches have also allowed the prediction of allosteric sites and the quantification of protein and/or ligand motions in full atomic detail, describing the molecules behavior at high resolution [28, 50–52].

Among the current *in silico* methodologies employed to study allostery, those based on the dynamical network analysis have proven particularly useful [24]. The latter approaches take advantage on the possibility that subtle adjustments in protein dynamics induce modifications of correlated-residue motions, thereby fine-tuning the signal propagation within the protein structure [24]. An insightful way of extracting information from the dynamical network representations is by clustering its nodes into communities, i.e., groups of highly intracorrelated residues that are loosely correlated with the remaining residues [53]. This analysis can further our comprehension on the dynamic perturbations triggered by the allosteric modulator in order to propagate the allosteric signal across the protein structure. In addition, residues that are crucial to the allosteric mechanism can be identified, thus providing a powerful way for intervention through mutagenesis [24, 53].

Here we present a combination of computational approaches, such as virtual screening (VS), MD simulations, binding free energy calculations and structure-based network analysis to propose novel allosteric sites and modulators of cruzain. To our knowledge, this is the first work where all crystal structures reported for this protease are compared in order to explore the diversity of its conformational space. An ensemble VS was conducted to cope with the flexibility of allosteric sites in the docking score. Moreover, a graph-based representation of protein structures was employed to investigate the correlated motions, allosteric propensities of protein residues and organization of residue-residue interaction networks. In general, we provide molecular-level insights into protein dynamics, transient pocket formation and enzymatic function, caused by ligand binding to the cruzain allosteric sites. The approach presented here provides a framework for the *in silico* design of allosteric drugs targeting this protease and other enzymes.

Materials and methods

Virtual screening

298 709 compounds from the files 21_p0.0.sdf and 21_p0.1.sdf contained in Lead Now library of ZINC database (<http://www.zinc.org>, accessed on March, 30th, 2016) were filtered in order to eliminate the potentially reactive and non-lead compounds. After this step, the protonation states of the remaining 209 422 compounds were defined according to the following criteria: *i*) all basic groups were protonated and *ii*) all acid groups were deprotonated. For simplicity, no tautomers were included in our compound library. Subsequently, each compound was parametrized using the MMFF94 force field and subjected to energy minimization (EM) until convergence, defined by an energy gradient tolerance of 0.05 kcal Å⁻¹ mol⁻¹. All previous steps were conducted with MOE platform [54]. Finally, the compound database was converted into pdbqt format with AutoDockTools [55].

Vs against the three representative structures of each selected allosteric site (Fig 2, step 3) were carried out in parallel with AutoDock Vina v1.12 software [56]. The number of generated poses was set to ten and docking simulation were initiated using a random seed. The exhaustiveness of the search and energy range between the best and the worst binding modes were set to the default values (8 and 3 kcal/mol, respectively) The grid box was defined according to the size of each cavity (S1 Fig), using AutoDockTools [55]. Twenty complexes with the highest Vina-score pose (S_{vina}) were selected per representative structure, thus generating 60 preliminary hits for each cavity (Fig 2, step 3). Subsequently, all these complexes were subjected to short MD simulations in order to re-score binding poses using Molecular Mechanics Generalized Born Surface Area (MM-GBSA) free energy calculation (Fig 2, step 4).

Ligand parametrization

Ligand parametrization was required in order to perform MD simulations of cruzain complexes. A less accurate parametrization using the AM1-BCC method [57] implemented in Antechamber was used for re-ranking the compounds selected from VS (Fig 2, step 4). After this intermediate rescoring step, a most rigorous charge derivation was conducted for the most promising compounds (e.g., top 5 compounds). The compounds selected for long simulations (Fig 2, step 6) underwent steepest descents EM until gradient convergence with the Universal Force Field (UFF) parameters using Avogadro [58]. The minimized structures were subsequently subjected to geometry optimization at B3LYP/6-31G(d,p) level using Gaussian 09 package [59]. The electrostatic potential (ESP) for each optimized structure was then generated by single-point (SP) calculations using the HF/6-31G(d) method and Merz-Kollman (MK) scheme [60]. The partial atomic charges were fitted to the ESPs through the Restricted Electrostatic Potential (RESP) fit implemented in the program antechamber of Amber14

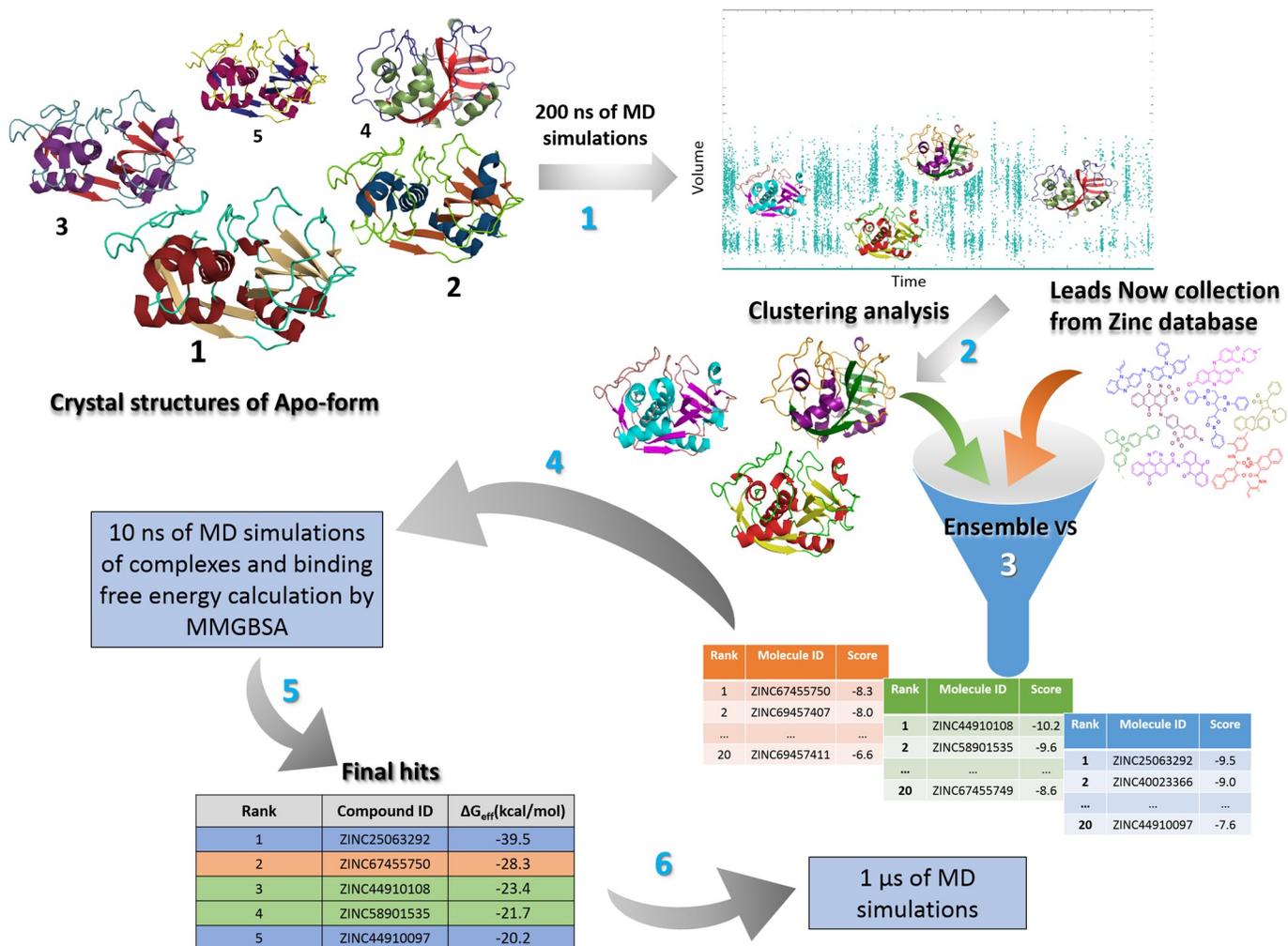


Fig 2. Workflow of structure-based identification of cruzain allosteric inhibitors. In the first step, five cruzain apo structures were selected and subjected to 200 ns of MD simulations. The second step represents the volume-based cluster analysis performed for each allosteric site, which produced three representative structures per pocket. The third step illustrates the ensemble VS accomplished for each cavity, where twenty hits were chosen per representative structure (60 per allosteric site). Subsequently, 10 ns MD simulations were carried out with these 60 complexes and the respective trajectories were employed for MM-GBSA binding free energy calculations. In this step, a final list of five hit compounds, ranked according to the MM-GBSA results, was obtained per allosteric site. Finally, 200 ns MD simulations were conducted to analyze the potentially mechanism of cruzain allosteric modulation exerted by these compounds.

<https://doi.org/10.1371/journal.pone.0211227.g002>

package [61]. Finally, the compounds were fully parametrized using the previously-determined partial charges and assigning the remaining parameters according to the Generalized Amber Force Field (GAFF) [62] implemented in Amber14 package [61].

Protein structures retrieval and systems setup and minimization

Twenty-five tridimensional (3D) structures of cruzain in complex with competitive inhibitors were obtained from Protein Data Bank (PDB) [63]. Subsequently, all cruzain structures (including each protein chain reported in each PDB file) were compared with each other through the calculation of pairwise Root Mean Square Deviation (RMSD) of the backbone atoms. This analysis was performed in order to find structures with highest structural differences after removing heteroatoms (Fig 2, step 1). Other parameters taken into account for the selection of cruzain structures were the experimental resolution (<2.5 Å) and their usage in

previous computational experiments, for comparison purposes (see [S1 Text](#)). The numbering scheme for cruzain residues was taken from their sequential order of appearance in the protein sequence (not the papain-numbering scheme present in PDB structures).

Protonation states of cruzain ionizable residues were determined at the main functional pH of the enzyme (pH = 5.5) using the PDB2PQR server [64]. Further preparation and simulations were performed with Amber14 [61]. Systems were solvated with explicit TIP3P water molecules [65] in a cubic box extending at least 10 Å from the solute surface, and treated with periodic boundary conditions. Negative net charges were neutralized by replacing water molecules with Na⁺ counterions. AMBER14SB force field [66] was selected for protein parametrization. EM was performed with pmemd.MPI program of Amber14 package [61]. This procedure was conducted in two steps: (i) solute atoms restrained with a harmonic constant of 10.0 kcal·mol⁻¹·Å⁻², and (ii) unrestrained system. In both cases, 1500 steps of steepest descent followed by 500 steps of conjugated gradient EM were performed.

To study the potential impact of the allosteric ligands on the functionality of the orthosteric site in terms of affinity, MD simulations with a mimetic substrate of cruzain were performed. Cruzain-peptide complexes were built following the strategy reported by Durrat *et al.* [11], i.e., by superimposing the average structure from MD simulations of the cruzain apo form with the crystal structure of human procathepsin K (PDB: 1BY8). According to the available information on specificity of cruzain binding site, the propeptide of procathepsin K was mutated using the mutagenesis plugin of Pymol v1.8 [67] in order to model the structure of an optimized substrate of cruzain. Previous studies have described that cruzain “prefers” basic and hydrophobic amino acids at P1 and P2 positions, being LEU and PHE the most frequent ones at P2 [68, 69]. Another work reported that cruzipain (naturally occurring form of cruzain), accommodates preferentially basic residues at P3-P5 and P3'-P5' positions, while there is no discrimination at P1' for residues PHE, LEU, ARG and ALA [68]. In addition, cruzain hydrolyzes substrates with SER, PHE, ALA or PRO at P2' with similar efficiency [69]. Based on this information, the final sequence generated for the heptapeptide was ACE-AKLF↓LAK-NME, where ACE and NME stand for acetyl and N-methyl amide capping groups, added to the N- and C-termini, respectively, in order to neutralize their net charges.

Molecular dynamics simulations

The equilibration procedure was conducted with pmemd.MPI and involved two sequential steps in the NVT and NPT ensembles, respectively, both keeping the solute heavy atoms restrained with a 10.0 kcal·mol⁻¹·Å⁻² restraint constant. During the 500 ps of NVT equilibration, the temperature was linearly increased from 10 to 300 K using the Berendsen thermostat [70]. The subsequent 500 ps of NPT equilibration was performed at a temperature of 300 K and a pressure of 1 bar, employing the Langevin thermostat and Berendsen barostat for temperature and pressure control, respectively [70, 71]. The particle mesh Ewald (PME) method was used to handle long-range electrostatic interactions [72] and the cut-off value for non-bonded interactions was set to 10 Å. Covalent bonds involving hydrogen atoms were constrained with the SHAKE algorithm [73]. A time step of 2 fs was used for numerical integration of the Cartesian equation of motion and coordinate files were recorded every 20 ps.

For cruzain apo-form, 200 ns of MD simulations were performed for each of five different structures selected in pairwise-RMSD analysis. Besides, the complexes involving the top-scoring ligands from the Vina-score ranking list were selected for shorter 10 ns MD simulations in complex with cruzain ([Fig 2](#), step 4). Finally, five independent 200 ns production runs were carried out for each complex selected from MM-GBSA results ([Fig 2](#), step 6). The pmemd.

cuda program [74] was employed to run the production simulations in identical conditions to those used during NPT equilibration step but without restraints.

Characterization and selection of cruzain cavities

Initially, each allosteric site previously-predicted by the SCA method [36] was mapped onto the cruzain structure for further analysis. Cruzain cavities were analyzed from the MD simulations of the apo form. The cavity volume must correspond to or exceed that of a ligand and the shape must allow its fit. In addition, the pocket physicochemical properties must complement those of the ligand [75]. For these reasons, the cavity volumes were calculated with POVME [76] throughout the simulation time, which provided an estimation of ligand accessibility for each cruzain identified pocket. Only cavities with a minimal volume greater than 200 \AA^3 were considered for the VS. Moreover, the most flexible cruzain regions were monitored through the calculation of per-residue Root Mean Square Fluctuation (RMSF) values. Finally, in order to characterize the electrostatic properties of each cavity, the APBS method available at PDB2PQR web-server was used to calculate the electrostatic potential on the protein surface (http://nbcrc-222.ucsd.edu/pdb2pqr_2.1.1/) and the results were visualized with Pymol v1.8 [67].

To provide a non-redundant set of cruzain pocket conformations for the ensemble VS (Fig 2, step 2), ten representative conformations per pocket were generated by volume-based clustering analysis, using the hierarchical-agglomerative algorithm implemented in cpptraj [77, 78]. Furthermore, the RMSD of pocket heavy atoms was calculated for each pair of representative conformations. Therefore, we considered the structures with the highest differences in terms of both, volume and pairwise RMSD values, while ignoring the cluster size. Our final selection was limited to three representative conformations (Fig 2, step 2). The data of pocket volume and pairwise RMSD values were normalized to keep both magnitudes comparable.

Trajectories analysis

The cpptraj module of Ambertools14 package [61, 78] was used to determine trajectory parameters such as RMSD, RMSF, average structure, interatomic distances and hydrogen bonds. The calculation of RMSD values for side chain and backbone atoms was performed employing the snapshots from production runs with respect to corresponding starting frame. Hydrogen bonds established between each ligand and cruzain were calculated with the default geometric definition of cpptraj. This geometric criteria establish a distance cutoff of 3.0 \AA between acceptor and donor heavy atom, and the acceptor-hydrogen-donor angle must be $\geq 135^\circ$ [61, 78]. Figures of protein cavities density and complexes were generated with Pymol v1.8 program [67].

Binding free energy calculations

MM-GBSA and MM-PBSA are computationally efficient methods for estimating the binding free energy (ΔG_{bind}) of protein-ligand complexes [79–82]. Here, the effective free energy (ΔG_{eff} , neglecting the configurational entropic contribution) values of all protein–ligand complexes were calculated using the MMPBSA.py program [83]. The single-trajectory approximation was used for this calculation, employing the MD simulations corresponding to the 60 complexes selected per cavity (Fig 2, step 4). The GB-neck2 model ($igb = 8$) with mbondi3 radii was used for estimating the polar solvation energy (ΔG_{GB}) [61]. The ΔG_{eff} values were averaged over all snapshots, thus generating a new score that allowed us to obtain a re-ranked list of the most promising candidates (Fig 2, step 5). For the final set of top-scoring compounds, average ΔG_{eff} values from simulation replicas were determined as a criterion of ligand stability into allosteric site. The standard error of the mean (*SE*) was calculated through the

following equation: $SE = \frac{SD}{\sqrt{N}}$, where SD is the standard deviation of the average ΔG_{eff} values obtained from the independent simulations (N).

Finally, per-residue effective free energy decomposition was carried out in order to determine the most important residues involved in cruzain-ligand interactions [84, 85]. Energetically-relevant residues, i.e., warm- and hot-spots, at the interfaces of the studied complexes were predicted using the energy decomposition protocol implemented in MMPBSA.py and the single trajectory approach. We defined as warm- and hot-spot residues those with an energy contribution (ΔG_{res}) to the total ΔG_{eff} value ranging from -1.0 to -0.4 kcal/mol and ≤ -1.0 kcal/mol, respectively, as defined elsewhere [86].

Coordination propensity calculation

The calculation of coordination propensity (CP) coefficient was performed with home-made scripts using the formula reported in previous works (Eq 1) [87, 88]. This methodology is based on the construction of a matrix whose elements, termed coordination propensity parameters, are calculated as follows:

$$CP_{i,j} = \langle (d_{i,j} - \langle d_{i,j} \rangle)^2 \rangle \quad (1)$$

where $d_{i,j}$ is the distance between the $C\alpha$ atoms of residues i and j in each snapshot and the brackets indicate the average distance over the trajectory.

The CP matrices obtained from independent replicas were averaged for each system. The values obtained for the apo form were subtracted from those obtained for the holo form ($\langle CP_{apo} \rangle - \langle CP_{holo} \rangle$) and reported for each complex.

Graph construction and calculations

Correlated motions between residues were calculated using the general correlation (GC) coefficients derived from mutual information theory [89]. In this approach, protein residue networks are defined as a set of residues; i.e., nodes, connected by edges (residue pair connections weighted by GC values). The GC matrix was calculated with g_correlation software of GRO-MACS v3.2.1 [89]. The GC of each pair of residues was obtained for every system, considering the motion of $C\alpha$ atoms and averaging over independent replicas for both, apo and holo cruzain systems. In order to highlight the principal changes in GC values of the analyzed systems, the matrices obtained for apo enzyme were subtracted from those obtained for holo complexes ($\langle GC_{holo} \rangle - \langle GC_{apo} \rangle$).

In this work, the network construction and the processing of GC and contact map matrices were made with Bio3d package [90], employing the methodology proposed by Sethi *et al.* [53]. The distance used for contact maps calculation was 5.0 Å, as previously used in other studies [91], in which it produced the smallest number of communities for the studied systems. The cut-off of GC value taken for network building was 0.5. For every system, the pairwise product between the average GC and contact map matrices corresponding to each simulation replica ($\langle GC_{ij} \rangle \cdot \langle Cmap_{ij} \rangle$) was used to generate the final average matrix employed in network community calculations. The graph community structure was built using the Girvan-Newman clustering method, which is essentially based on the edge betweenness partition criterion [92].

The path length was weighted using the GC coefficient according to the Eq 2 [53].

$$d_{i,j} = -\log|GC_{i,j}| \quad (2)$$

Here, $d_{i,j}$ is the distance between contacting nodes i and j . According to this analysis, the resulting graph will produce short distances for strongly correlated residues and longer

distances for residues with weak correlations. The optimal paths were identified using Dijkstra's algorithm available in NetworkX [93]. Finally, for a given pair of residues termed source and sink, 500 suboptimal pathways were calculated employing the Weighted Implementation of Suboptimal Pathways (WISP) method [94]. Residues CYS25 at the catalytic site and a hot spot of the allosteric site were selected as sink and source, respectively, for pathway analysis.

Construction of energy interaction networks based on cruzain ensembles from MD simulations

Protein energy networks (PENs) are a specific case within the protein structural networks (PSNs) approach [48, 95–100]. PENs construction is based on the calculation of pairwise interaction energies between protein residues in equilibrium ensembles, usually generated from MD simulations. These energies are considered the 'strength' between the nodes (residues) and are mapped as edges onto the protein structure. Subsequently, the strength values are generally normalized in order to weight or deduce the information transfer rate from one node to another. Finally, these data are used in network analysis such as shortest path identification and node centrality calculations.

In this work, the interaction energies were calculated employing the pair-wise energy decomposition function of MMPBSA.py program [83]. For network construction, we followed the protocol proposed by Vijayabaskar *et. al* [95], where only van der Waals and electrostatic interactions of non-continuous residues were taken into account. Furthermore, all energies were normalized between zero and one, the more negative values being closer to one, and the more positive ones, closer to zero. Subsequently the betweenness centrality was calculated with the igraph package [101]. The normalized data were transformed into path length through the following formula: $w_{ij} = -\log|Inorm_{ij}|$, where, $Inorm_{ij}$ is the normalized interaction energy between residue i and j and w_{ij} is the path length. Hence, the information transfer rate will be greater for the more favorable interactions, as the latter will have shorter path lengths.

Results

Transient pockets and functional gate in a conserved groove of papain-like cysteine proteases revealed during MD simulations

After mapping the putative allosteric sites of HCatK onto the surface of cruzain through structural alignment (Fig 1), the characterization of cruzain cavities was performed through an analysis of the enzyme conformational dynamics. The selection of cruzain structures for MD simulation of the apo form is described in detail in S1 Text. The results showed that site 1 and site 3 display a suitable volume size ($\geq 200 \text{ \AA}^3$) along simulation time (Fig 3); therefore, both were chosen for further VS experiments. Subsequently, three central structures per-site, satisfying the conditions described in Materials and Methods, were selected (Fig 4). The results of pairwise RMSD and volume comparison between all generated representative structures of site 1 and site 3 are shown in S3 Fig.

The time evolution of cavity volume indicated that site 3 is more stable than site 1 (Fig 3). However, the latter reached higher volumes, probably because it is flanked by a flexible fragment of cruzain loop(loop₈₄₋₁₀₉), homologue of the occluding loop of human cathepsin B (S4A Fig), that eventually can unveil the larger pocket. Moreover, the distribution of conformational space of site 1 showed three well-defined populations, being the structures with zero volume the most representative ones (Figs 3A and 4A). The above results suggest that site 1 constitutes a cruzain transient pocket. On the other hand, cruzain site 3 showed a conformational space characterized by two overlapping normal distributions of volume values (Fig 3B). However, there was an

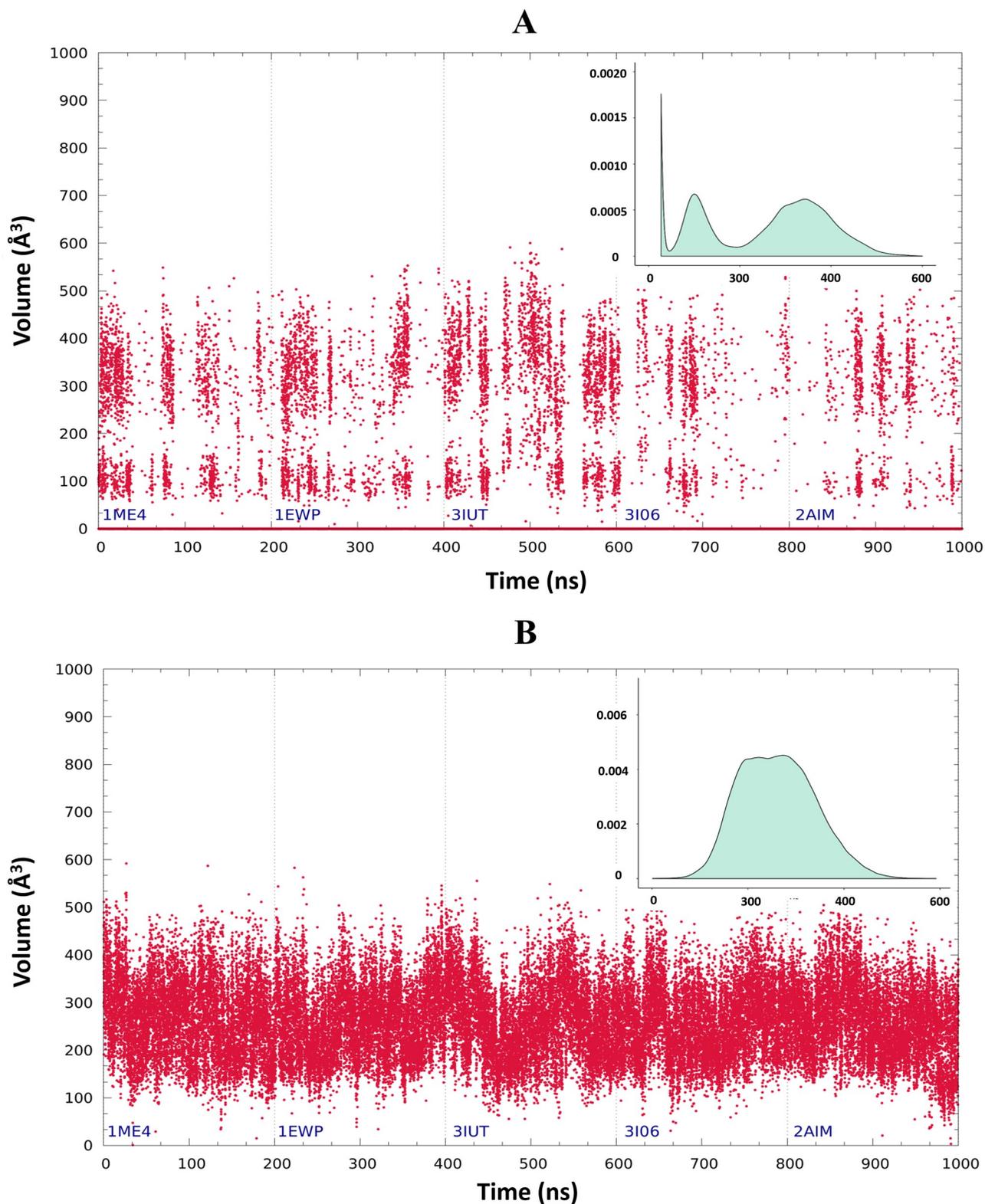


Fig 3. Dynamical analysis of cruzain sites 1 and 3. Time evolution of instantaneous volumes for site 1 (A) and site 3 (B). Graphs were subdivided according to the cruzain structure employed in each simulation (PDBID). Insets within each graph correspond to the volume probability histograms along the MD trajectory.

<https://doi.org/10.1371/journal.pone.0211227.g003>

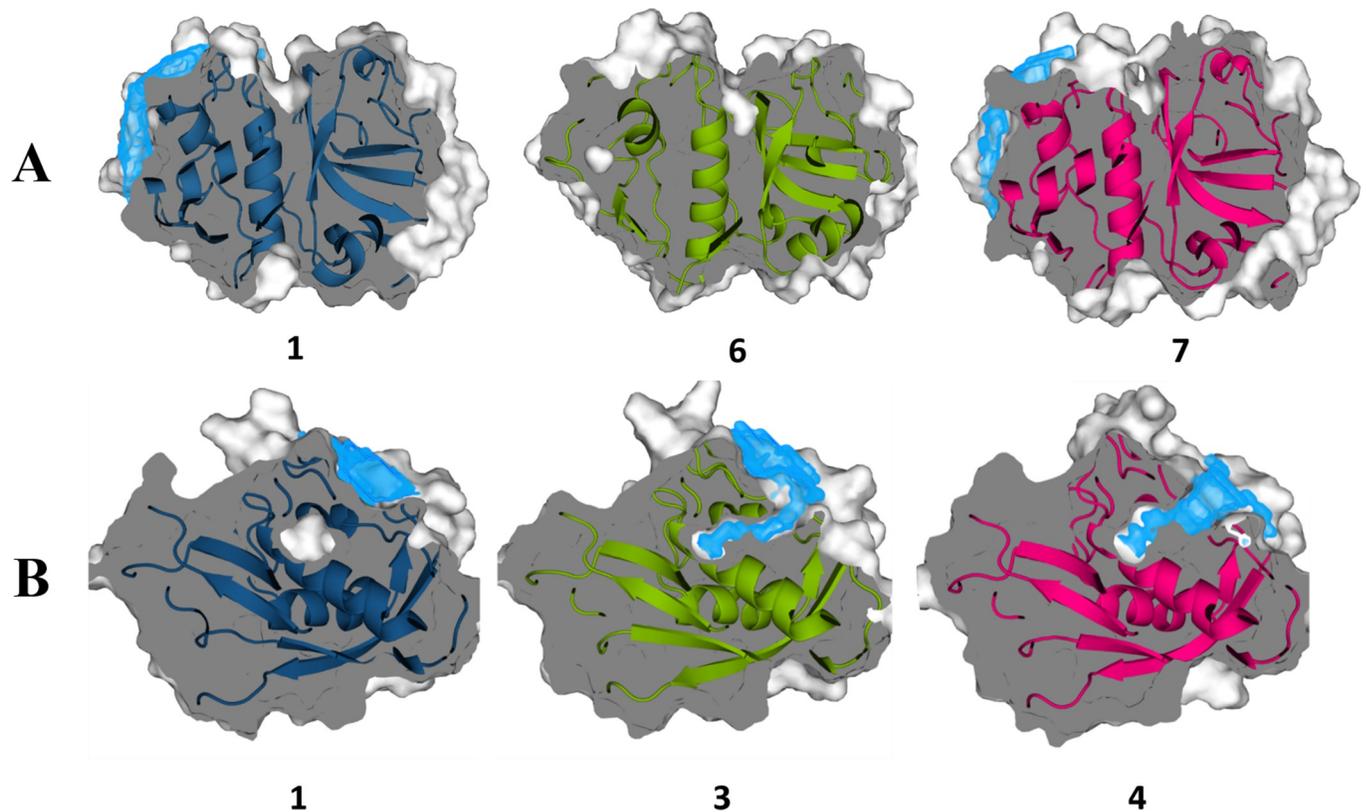


Fig 4. Representative structures of cruzain selected for ensemble VS against sites 1 and 3. Representative structures of (A) site 1 and (B) site 3. Volume densities of the three selected representative structures are depicted in blue. The corresponding cluster identifier is shown in each case.

<https://doi.org/10.1371/journal.pone.0211227.g004>

intermediate representative structure (cluster 3) that shared structural characteristics with the remaining two (Fig 4B). Our predictions revealed the occurrence of an opening/occlusion dynamics of site 3 internal cavity (cluster 1 in Fig 4B), in which the salt bridge established between conserved residues LYS17 and GLU35 (see sequence alignment of [11]) acts as “functional gate” of this groove (S5 Fig).

Additionally, the RMSF analysis showed the same fluctuation pattern in the five replicas of the apo-form (S6B and S6C Fig). Here, we observed the main flexibility in cruzain loops, mainly in the region corresponding to loop₈₄₋₁₀₉. These findings explained the huge volume variation of site 1, since several residues of the previous loop region are flanking this site (S4A Fig). Moreover, RMSD values calculated for heavy atoms showed different time evolution behavior between site 1 and site 3 (S6A Fig). In this sense, site 3 showed relatively stable RMSD value; whereas site 1 displayed structural fluctuations during simulation time, as expected for a site that displays volume instability.

Preferential binding of identified hits to allosteric site 3 of cruzain

The prediction of allosteric modulators of cruzain was based on two rounds of high-throughput selection using: (i) Autodock Vina scoring function, and (ii) combination of MD simulations with MM-GBSA calculations as the final post-processing step. The comparison of the compounds identified through the previous strategy showed non-overlapping scaffolds targeting each of the three selected clusters of the same cavity. This arises from the large structural differences between the three representative structures selected for the ensemble VS (Fig 4),

which reinforces the importance of using different receptor conformations in VS experiments to increase the structural diversity of the selected compounds.

In Fig 5, the top five compounds are listed and numbered in increasing order according to the MM-GBSA free energy values. These results showed that the ΔG_{eff} values of the compound list ranged from -41 to -23 kcal/mol (Fig 5). Interestingly, ligands having the lowest energy values among all sets were detected in VS against the representative structures that possessed the largest volumes (Figs 4 and 5); thus underlining the importance of exploring the protein conformational space to identify different pocket structures. In our case, despite having many cruzain crystal structures (25), it was necessary to perform MD simulations to sample appreciable conformational changes prior to conduct VS simulations.

Finally, we selected three top-scoring compounds (compounds 1, 2 and 4, see Fig 5) in complex with cruzain in order to assess the time stability of these complexes in longer MD simulations, which, in turn, would suggest the suitability of the selected compounds and/or the druggability of the potential allosteric sites. Note that our selection was based on the overall best hits identified for the studied pocket, in terms of ΔG_{eff} values. We also considered that two compounds targeting site 3 were, in principle, a handful subset to investigate the potential allosteric modulation of cruzain arising from the ligand binding to this site, which requires the usage of computationally-demanding calculations. The ensemble generated from MD simulations are useful to analyze the potential allosteric effects triggered by the binding of the selected compounds to cruzain sites 1 and 3. Note that within the list of best hits (Fig 5), only compound 4 is derived from VS against site 1. This is a consequence of large solvent accessible area present in site 1 (S4A Fig), where the interactions of ligands with solvent molecules are likely

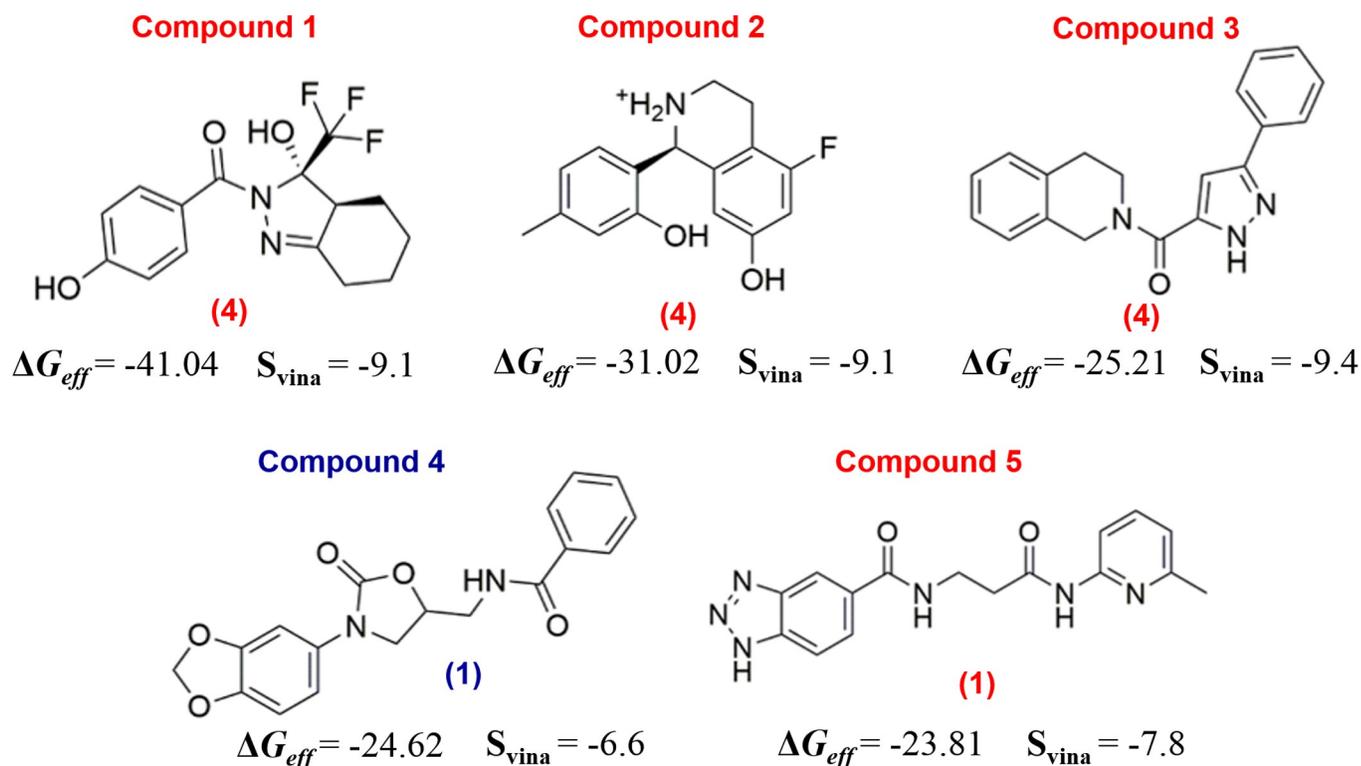


Fig 5. Chemical structures at pH = 5.5 of top five hits selected from the combination of VS and MM-GBSA calculations. Compound names are colored differently, according to the allosteric sites wherein they were docked, i.e., red for site 3 and blue for site 1. Below each chemical structure, the numerical IDs of cluster central structures that bind each ligand are depicted between parentheses. The S_{vina} and ΔG_{eff} values are expressed in kcal/mol units.

<https://doi.org/10.1371/journal.pone.0211227.g005>

to destabilize ligand-cavity interactions in short MD simulations (10 ns). Conversely, site 3 is a buried groove (S4B Fig), which can establish more stable interactions with ligands, but, again, at expense of large desolvation of the ligand.

In Fig 6, the top-scoring docking poses of these ligands are represented. Moreover, each pocket surface was colored according to the electrostatic potential and hydrophobicity. These results show that site 1 is made up of residues with dissimilar electrostatic and hydrophobicity properties (Fig 6C). Hence, compound scaffolds that possess polar and non-polar interacting groups, such as compound 4, are suitable to target this site. On the other hand, site 3 has a positive region at the entrance, in contrast with its internal area, which is highly negative. This is due to the presence of LYS17 in a more solvent-exposed region of site 3 and GLU35, GLU50 and GLU86 positioned in the pocket inner region. The two top-scoring ligands selected for this site (compounds 1 and 2) contain aromatic rings placed within the pocket hydrophobic region (Fig 6A and 6B). In addition, the polar substituents of these ligands may interact with polar regions by forming hydrogen bonds and/or salt bridges, as in the case of compound 2 (positively charged, see Fig 5).

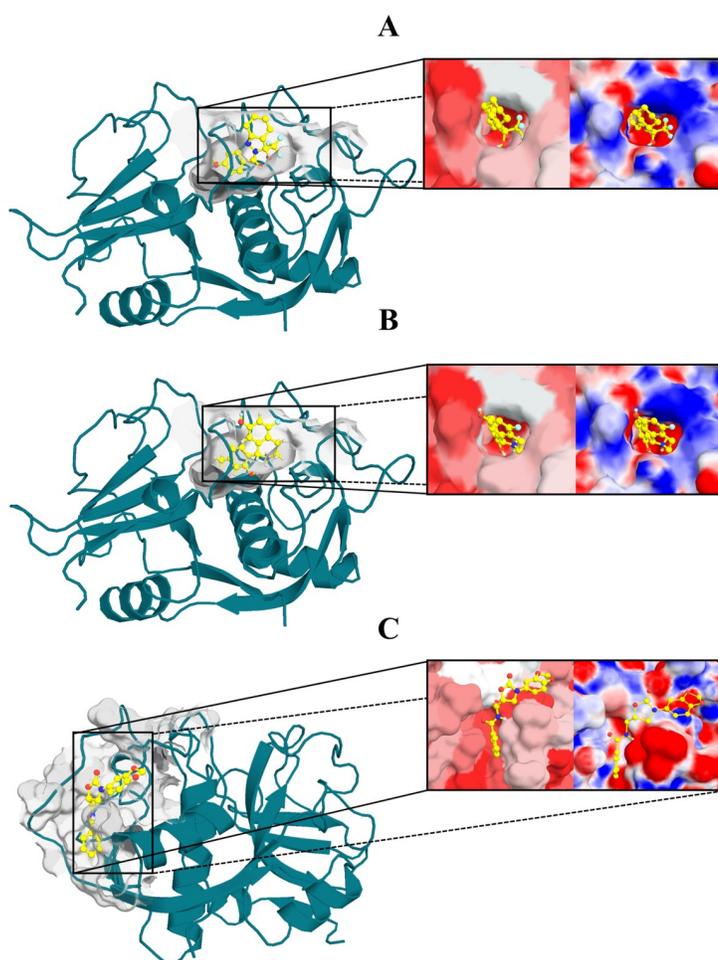


Fig 6. Selected docking poses for the best hits of each pocket. Top-scoring poses of (A) compound 1 and (B) 2 docked into site 3 pocket, and (C) top-scoring pose of compound 4 docked into site 1. Protein surface is colored according to hydrophobicity and electrostatic potential. The color scale for hydrophobicity ranges from red (hydrophobic) to white (hydrophilic), and for electrostatic potential, from red (electronegative potential) to blue (electropositive potential).

<https://doi.org/10.1371/journal.pone.0211227.g006>

Selected hits display stable interactions with site 3 during the long MD simulations

A comparative analysis of conformational dynamics and energetic profiles was performed, including the apo form, cruzain with a peptidic substrate, and the former systems bound to the allosteric ligands. The stability and convergence of long MD simulations (200 ns per replica) performed for cruzain complexes were monitored through per-frame ΔG_{eff} and ligand RMSDs (S7–S12 Figs). The accumulated average values of ΔG_{eff} became stable in most cases (S7–S11 Figs). The equilibration time of complexes was selected according to the behavior of ΔG_{eff} values across the simulation replicas. Accordingly, 60 ns was chosen as the start point for the rest of MD simulations analysis (S7–S11 Figs). On the other hand, in the MD simulation replicas of the cruzain-compound 4 system, the ligand dissociation was observed after 50 ns of simulation time. The latter points out the lack of a stability of the ligand in this site.

Finally, the ΔG_{eff} values of ligand binding ($\Delta G_{eff(ligs)}$) were calculated as the average of five replicas in order to establish the best hits (Table 1). The results show subtle differences between both compounds (~1 kcal/mol); therefore, both ligands may have the same priority in terms of enthalpic contribution for further experimental assays. In addition, moderate decrease in substrate affinity was observed when we compared the $\Delta G_{eff(peptide)}$ values in the presence of ligands ($\Delta\Delta G_{pep,compd1}^{pep} = 2.86 \pm 1.39$ and $\Delta\Delta G_{pep,compd2}^{pep} = 2.64 \pm 1.39$, see Table 1). These lines of evidence indicate a moderate destabilizing effect on the formation of enzyme-substrate complex (Fig 7). Moreover, the hydrogen bond analysis of cruzain-substrate complexes showed that the most distinctive differences caused by the ligand binding to the allosteric site 3 were mainly in those established at the S1'-P1' and S2'-P2' interfaces (Fig 7 and S1 Table). Note that these hydrogen bonds were also the most unstable in MD simulations corresponding to cruzain-substrate system (S1 Table). Interestingly, site 3 is close to the S1'-S2' subsites (Fig 7), and the presence of ligands in that cavity directly affects the movement of loop₁₀₋₂₅ and loop₁₈₂₋₁₉₂, and, thus, the dynamic of aforementioned subsites. As was previously discussed, GLN19 and TRP184 are conserved residues placed in the former loops [11], and it has also been reported that they are essential for the catalytic mechanism in papain-like cysteine proteases [102]. Although moderate variations in ΔG_{eff} values for substrate binding were observed, the variations in the motions of these protein regions may affect the cruzain catalytic rate constant as well.

Table 1. Effective free energies calculated for intermolecular interaction of each system.^a

System	cruzain-compound 1	cruzain-compound 2	cruzain-peptide	cruzain-peptide-compound 1	cruzain-peptide-compound 2		
Replicas	-28.83±0.07 ^b	-30.55±0.1	-64.59±0.05	-27.73±0.04 ^c	-65.26±0.06 ^d	-30.73±0.05	-63.37±0.05
	-31.11±0.11	-32.5±0.08	-65.29±0.05	-26.71±0.03	-61.45±0.05	-31.84±0.04	-58.64±0.05
	-27.05±0.1	-24.72±0.13	-63.84±0.05	-28.53±0.04	-59.16±0.05	-31.54±0.04	-65.03±0.05
	-26.52±0.09	-28.78±0.09	-64.08±0.05	-27.96±0.04	-63.46±0.06	-28.77±0.04	-64.83±0.05
	-29.49±0.12	-30.15±0.11	-63.58±0.05	-28.30±0.04	-56.79±0.05	-31.64±0.04	-58.25±0.05
			-64.67±0.05	-29.35±0.03	-62.78±0.05	-30.3±2.98	-60.05±0.09
Average	-28.6±0.83	-29.34±1.3	-64.34±0.28	-28.1±0.29	-61.48±1.37	-30.8±0.52	-61.7±1.36

^a All energies are in kcal/mol

^b Standard Errors of mean

^c Energy calculated for ligand

^d Energy calculated for peptide

<https://doi.org/10.1371/journal.pone.0211227.t001>

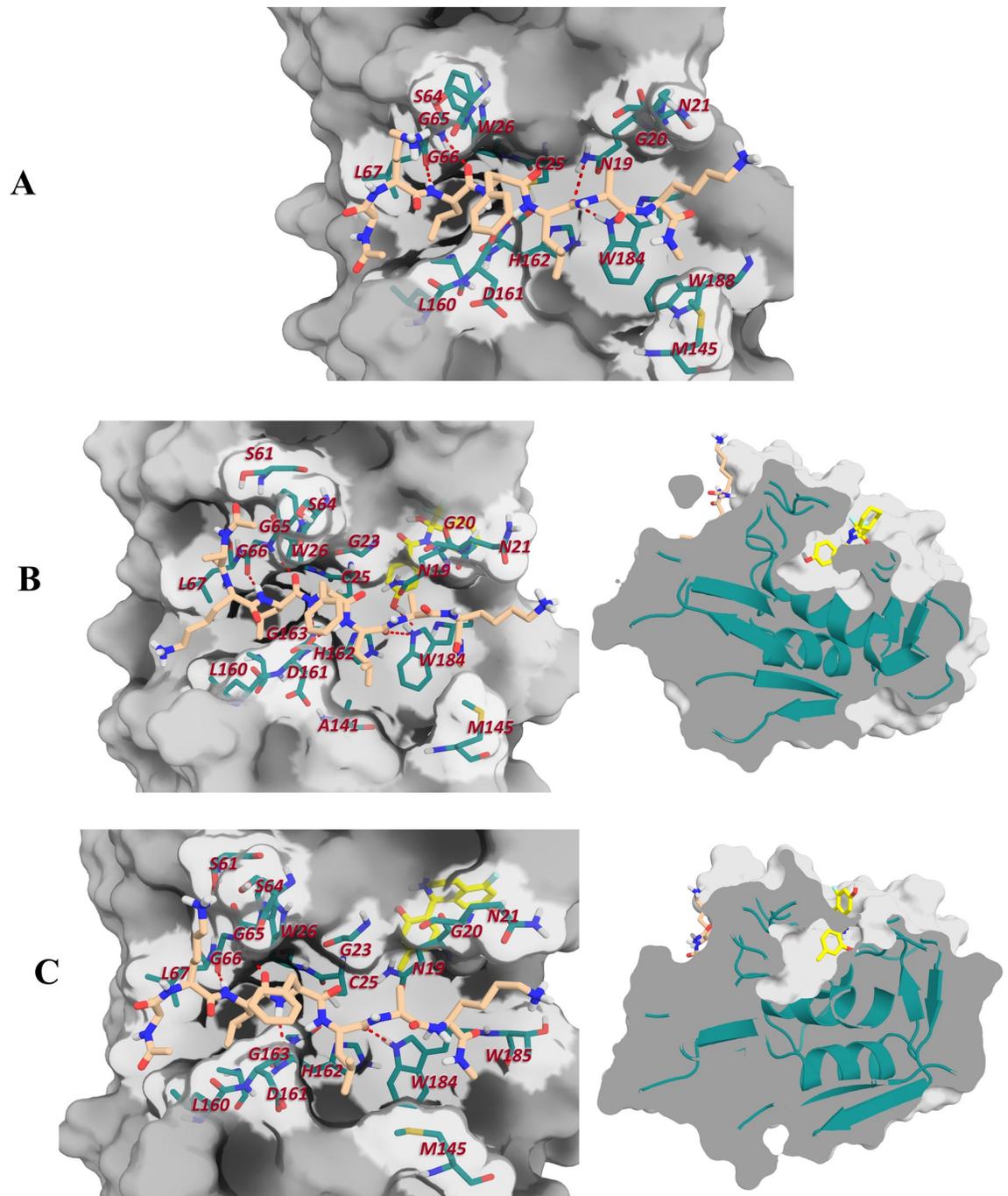


Fig 7. Main interactions established between substrate and cruzain binding site. (A) cruzain-substrate (B) cruzain-peptide-compound 1 and (C) cruzain-peptide-compound 2. The interacting residues are labeled in each case and the relative position of ligand (yellow) with respect to substrate (lightorange) is represented in the right panel.

<https://doi.org/10.1371/journal.pone.0211227.g007>

Aliphatic chains of site 3 display large energy contribution to ligand binding

In order to get insights into the ligand/peptide interactions with cruzain, the per-residue decomposition of ΔG_{eff} (ΔG_{res}) was employed. As it is shown in Fig 8, the non-polar interactions displayed the largest energy contribution to the binding process between cruzain and the substrate. Residues GLN19, CYS25, TRP26, LEU67, MET145, LEU160 and TRP184 were predicted as the main hot-spots of cruzain active site. Interestingly, there is a noticeable decrease

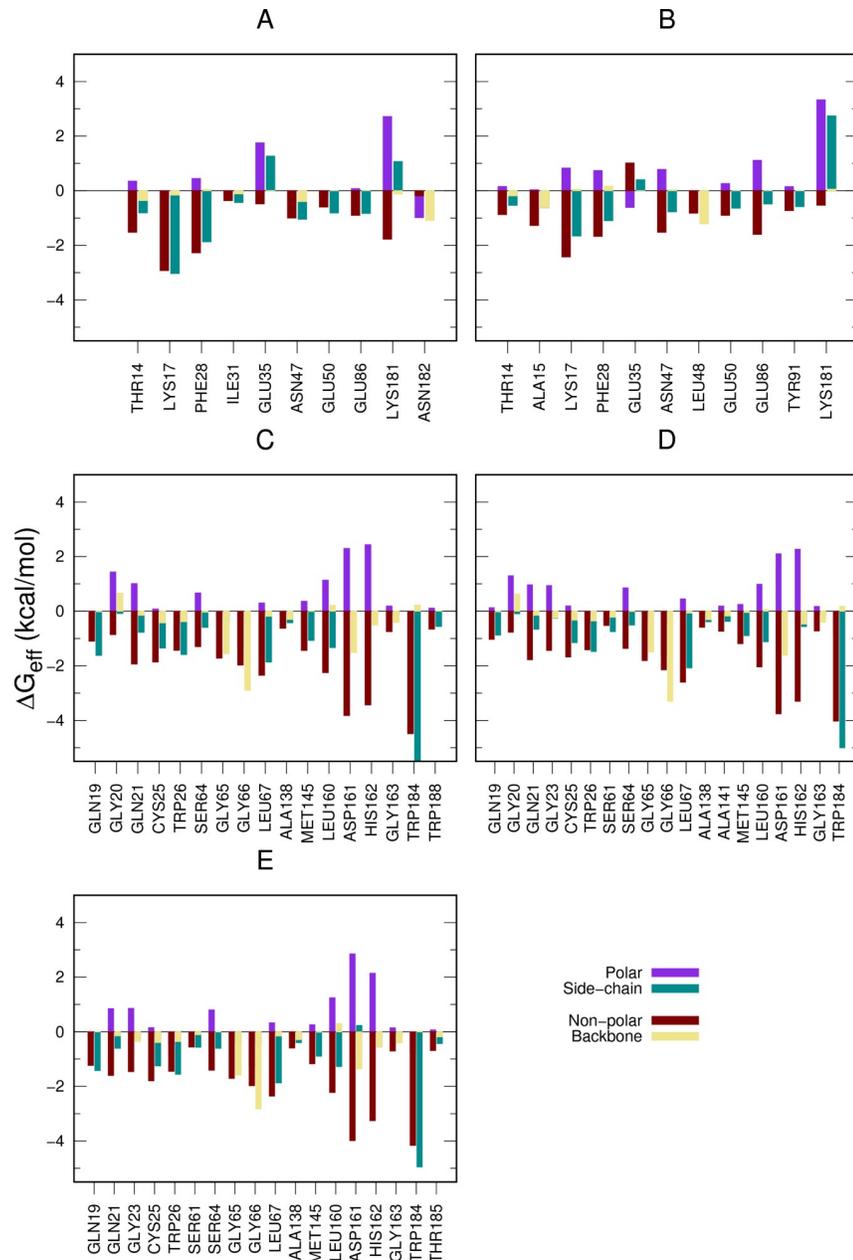


Fig 8. Per-residue free energy decomposition of cruzain complexes. The side chain, backbone, polar and non-polar contributions of each residue are displayed for all the analyzed complexes, i. e. (A) cruzain-compound 1 and (B) cruzain-compound 2, (C) cruzain-peptide, (D) cruzain-peptide-compound 1 and (E) cruzain-peptide-compound 2.

<https://doi.org/10.1371/journal.pone.0211227.g008>

in 1 kcal/mol of $\Delta G_{res(TRP184)}$ when ligands are bound to the allosteric site 3 (Fig 8C, 8D and 8E). This change partially accounts for the $\Delta\Delta G_{eff(peptide)}$ observed between cruzain-peptide and cruzain-peptide-ligand systems (Table 1). Additionally, $\Delta G_{res(GLN19)}$ decreased in 1 kcal/mol when compound 1 is positioned within site 3. These variations reinforce our previous conclusion that the binding to site 3 affects the dynamics of S1'-S3' subsites.

In general, we observed that despite site 3 surface being highly charged, van der Waals interactions are predominant in cruzain-ligand interfaces (Fig 8A and 8B). This is because the polar desolvation energy is more unfavorable in residues that establish polar interactions, i. e., GLU35, GLU86 and LYS181. The major hot spot residue at the interfaces of both cruzain compound complexes is LYS17, which interacts with the ligands through its aliphatic side chain, exposing its charged amine group to the solvent. Also, THR14, PHE28, ASN47 and GLU50 were predicted as critical interacting residues in studied cruzain-ligand systems.

The further decomposition of ΔG_{res} into side chain and backbone contributions led to the identification of energetically-relevant positions within site 3. For almost every residue represented in Fig 8A and 8B, the energy contribution of side chain is larger than that of the backbone, except for ALA15, LEU48 and ASN182. This result suggests the essential role of positions of residues conforming site 3 in the interaction with ligands. Remarkably, three energetically-relevant residues, i. e., THR14, LYS17 and PHE28, are conserved positions within papain-like cysteine proteases [11], and were also characterized through SCA method as residues belonging to functional sectors. Therefore, they are likely to participate in allosteric regulation of HCatK [36]. On the other hand, GLU50 is a semi-conserved position [11], also lying within the same sector of the former residues [36], that is able to form hydrogen bonds with both ligands (Fig 9 and S1 Table). In addition, ASN47 is a non-conserved residue that establishes hydrogen bonds with both ligands. Thus, it may be crucial to the selectivity of cruzain allosteric inhibitors targeting site 3 (Fig 9 and S1 Table). Finally, note the unfavorable overall energetic contribution of conserved GLU35 in both complexes, where the polar desolvation energy is more unfavorable than its hydrogen bond energetic contribution.

In the cruzain-compound 1 complex, ASN182 establishes the most stable hydrogen bond through its backbone carbonylic oxygen (Fig 9 and S1 Table). Interestingly, ASN182 is a highly conserved residue whose participation in the catalytic mechanism of cruzain has been proposed [102]. In the cruzain-compound 2 complex, GLU86 also forms a hydrogen bond through its carboxylic oxygen atoms, but its polar desolvation energy is more unfavorable because of the presence of positive charge on the compound N atom. In addition, compound 2 establishes important van der Waals contacts with THR14, ALA15, PHE28, GLU86, GLU50, LEU48 and TYR91. However, LYS181 displays a more negative contribution to cruzain-compound 2 binding process because of the high desolvation penalty of the amine group, which is facing the phenolic ring of this ligand (Fig 9B).

Compound 2 increases the flexibility of cruzain structure

Subsequently, we turned the attention to the characterization of the effects on cruzain internal dynamics triggered by ligand binding to site 3. The *CP* was calculated for each system in order to characterize the global rigidity/flexibility patterns. Here, we intended to seek for changes produced by ligand binding in the coordination of amino acids belonging to different protein domains. The matrices obtained from the *CP* differences between apo and holo forms are shown in Fig 10. Positive values in the matrices (red color) indicate higher flexibility in the apo form, while negative ones (blue color) denote higher flexibility in the holo form.

The results show that cruzain structure is less flexible when bound to the peptide than it is in the apo form (Fig 10A). Interestingly, the phenomenon of protein structure stabilization by

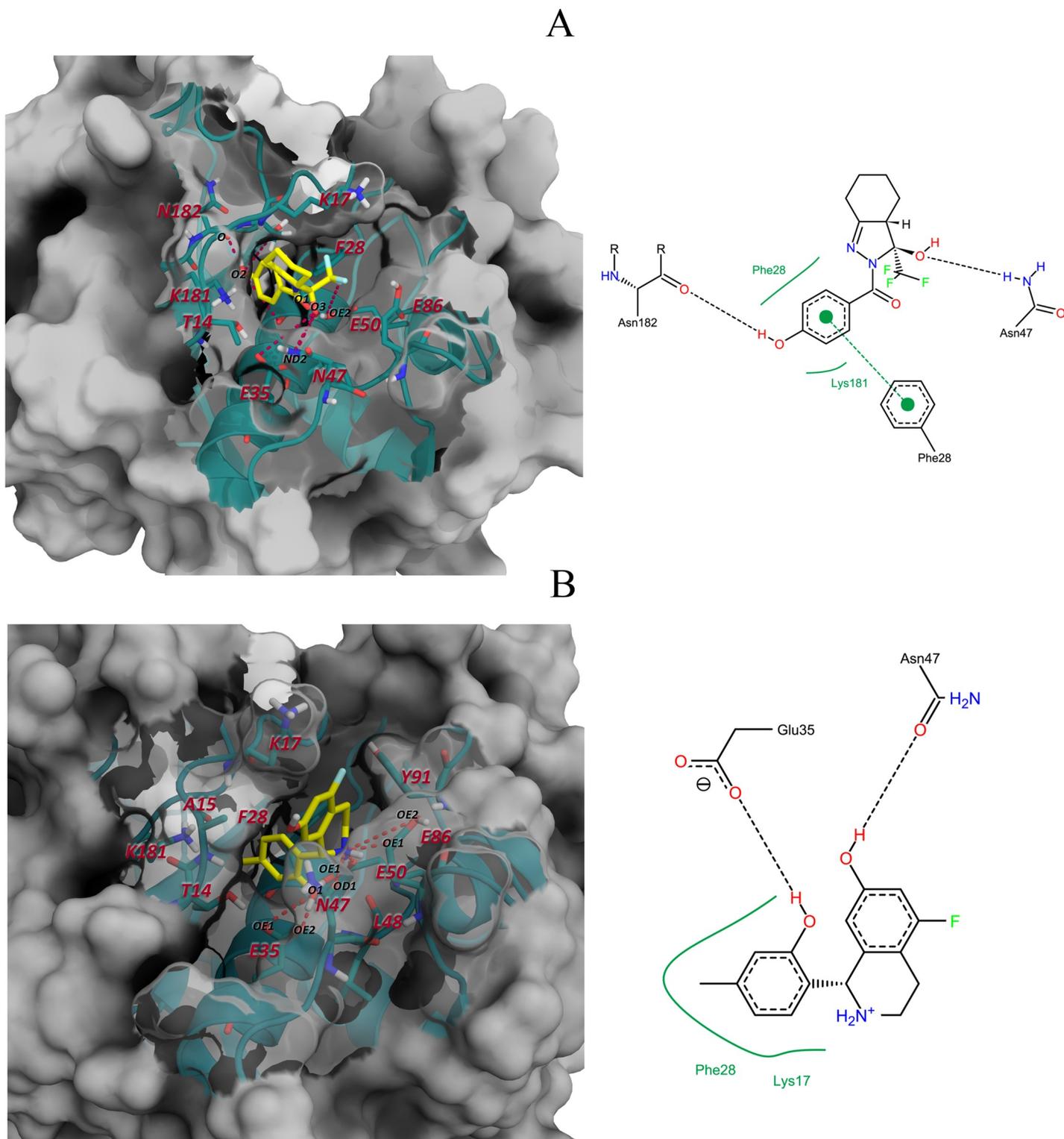


Fig 9. Structural representation of cruzain-ligand interfaces. 3D and 2D interfaces of (A) cruzain-compound 1 and (B) cruzain-compound 2. In each diagram, the residues considered energetically relevant are labeled together with the name of atoms forming the main inter-molecular hydrogen bonds (red dashed lines in left panel).

<https://doi.org/10.1371/journal.pone.0211227.g009>

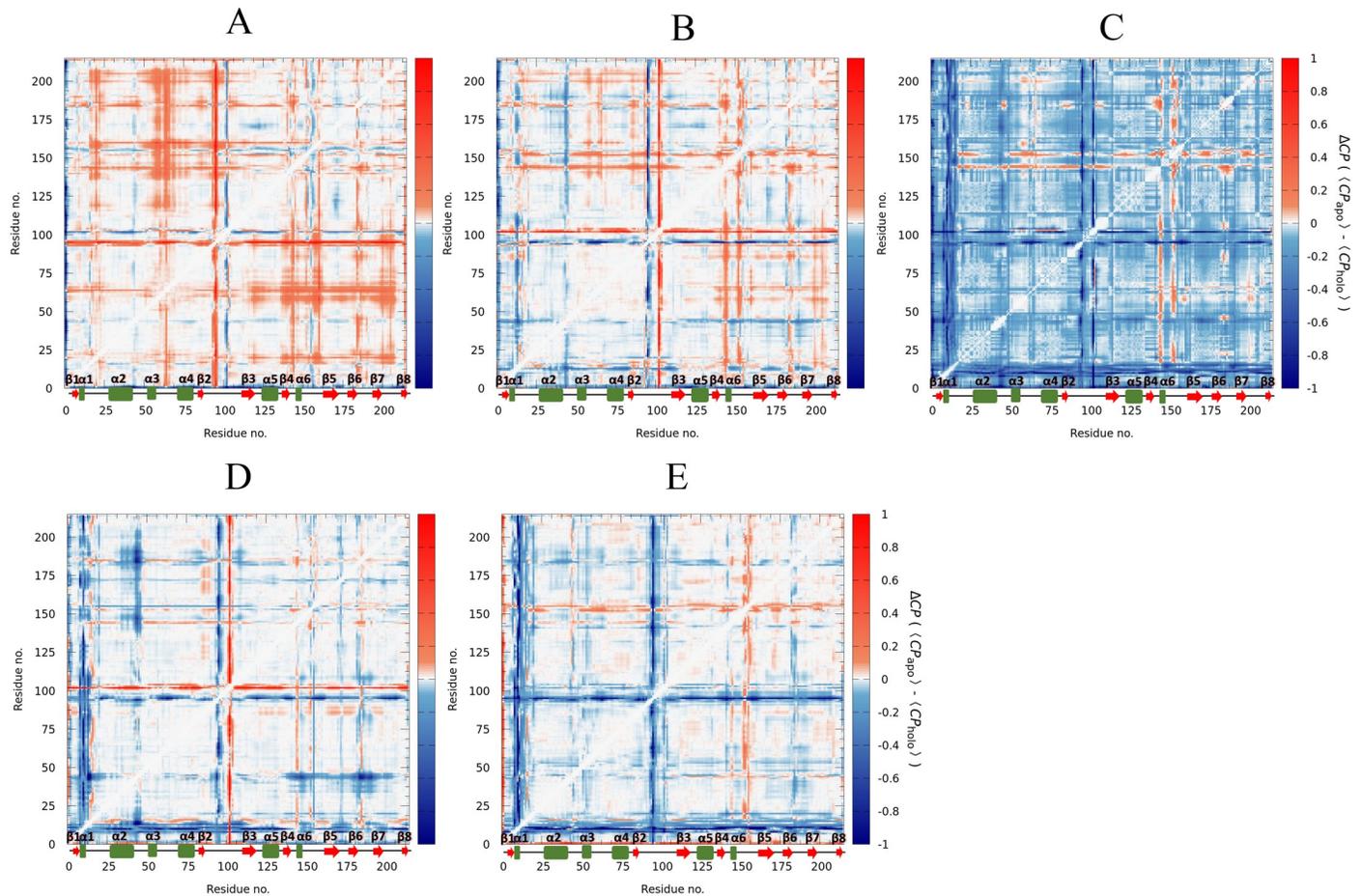


Fig 10. Analysis of motions of the apo and holo forms of cruzain. Graphic representation of matrices corresponding to CP differences (ΔCP). The CP values corresponding to (A) cruzain-peptide, (B) cruzain-compound 1 and (C) cruzain-compound 2 were subtracted from those of the apo form, and the CP values of (D) cruzain-peptide-compound 1 and (E) cruzain-peptide-compound 2 were subtracted from those corresponding to the cruzain-peptide system. The matrix scale is colored from blue (higher flexibility in holo form) to red (higher flexibility in apo form). The coloring scale was set between 1 and -1 in order to improve the visualization of ΔCP values.

<https://doi.org/10.1371/journal.pone.0211227.g010>

substrate binding is commonly described in biological systems [103]. In addition, the more flexible regions of cruzain were those corresponding to loops, which is in accordance with RMSF results (S6 Fig). On the other hand, the large white region of matrix corresponding to cruzain-compound 1 complex indicates similar patterns in flexibility/rigidity between apo and holo forms (Fig 10B). The main differences are also located in loop regions, while the same degree of internal coordination is maintained in residues belonging to secondary structures. In these systems, the greatest ΔCP values occurred at loop₈₄₋₁₀₉ and loop₁₃₉₋₁₆₁, where differential patterns in residues flexibility in both apo and holo forms were observed (Fig 10B). Conversely, the MD simulations of cruzain-compound 2 complex indicated a uniform increase in overall protein flexibility, favoring distortions in both, loops and well-defined secondary structures (Fig 10C).

Interestingly, the results of pairwise distance distributions of key residues of cruzain active site show, in most cases, that their respective maxima shifted towards lower values when compound 2 is bound to cruzain (S2 Text and S13 Fig). This suggested that the S3 and S2 subsites are slightly narrower when the former ligand is bound to site 3. Therefore, the substrate binding site is likely to undergo a small structural reorganization. Noteworthy, compound 2 is positively charged at the pH employed for MD simulations (pH = 5.5, see Fig 5), which could induce a

perturbation on protein movements and electrostatic balance. However, we observed no significant changes on protein surface electrostatic potential subsequent to this ligand binding (S15 Fig).

A parallel analysis of matrices corresponding to cruzain-peptide-ligand systems displayed low changes in *CP* values due to the ligand presence in site 3. It is worth noting the increase of flexibility in helix 1 (7–10 residues) and regions corresponding to positions 40–45 and 90–95. Interestingly, we observed a noticeable less perturbation in protein flexibility when we compared these systems with those lacking the peptide in the active site, notwithstanding the ligand binding to site 3.

Other significant changes in structure dynamics of cruzain were detected in RMSD and RMSF analyses (Fig 11). The RMSF results detected once again the cruzain loops as the most flexible regions of this enzyme (Fig 11A), and these predictions were also confirmed by the essential dynamics analysis (S3 Text). Note, that loop₁₀₋₂₅ increases its flexibility in presence of ligands, and some residues belonging to this loop were also defined as hot spots in the section of per-residue energy decomposition of cruzain-ligands complexes. The RMSD distributions of the analyzed systems showed partially overlapping populations between them (Fig 11B). Here, in agreement with previous results, we could observe that the substrate binding induced small dynamic changes in cruzain structure, leading to an increased stability of peptide-bound complexes, which had the narrowest range of RMSD values. Conversely, a notable shift of the RMSD distribution toward higher values was detected when cruzain is in complex with both ligands, indicating an increase of enzyme conformational motions. Interestingly, the highest RMSD values were observed for cruzain-compound 2 complex (an increase of ~1.0 Å, see Fig 11B), which agreed with previously-discussed results of *CP* analysis (Fig 10C). Therefore, the dynamic changes induced by compound 1 were smaller than the respective variations of cruzain complexed with compound 2.

Correlation network analysis reveals state-specific differences in residue couplings

The analysis of residue-residue couplings derived from MD simulations showed remarkable differences between apo and cruzain-peptide systems (Figs 12A and S18). Remarkably, the correlation decreases in regions corresponding to protein loops in cruzain-peptide system. The

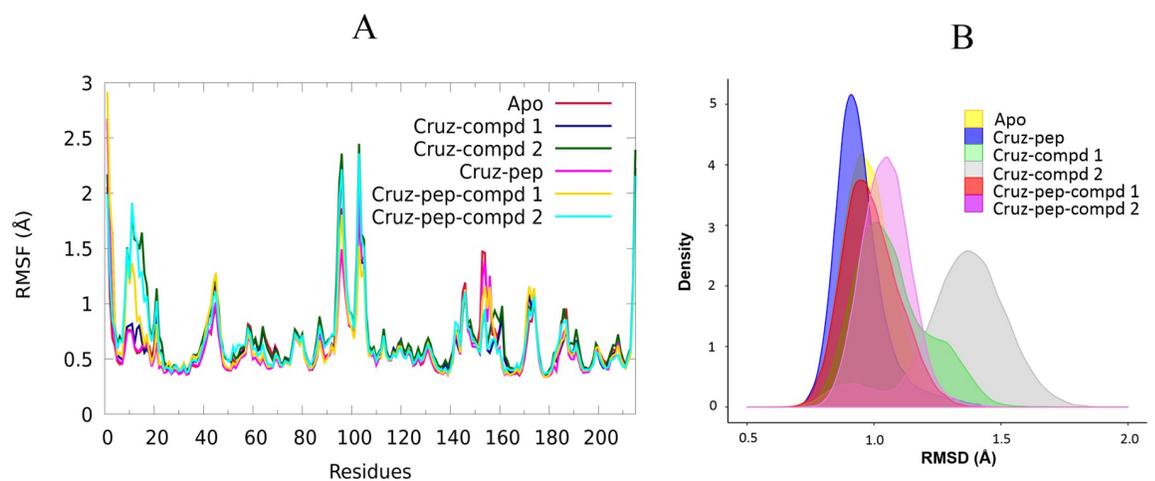


Fig 11. RMSF and RMSD values of cruzain analyzed complexes. (A) RMSF profiles and (B) RMSD distribution calculated for analyzed systems of cruzain. Each parameter was calculated for backbone atoms and taking as reference structure the starting frame of productive MD simulations.

<https://doi.org/10.1371/journal.pone.0211227.g011>

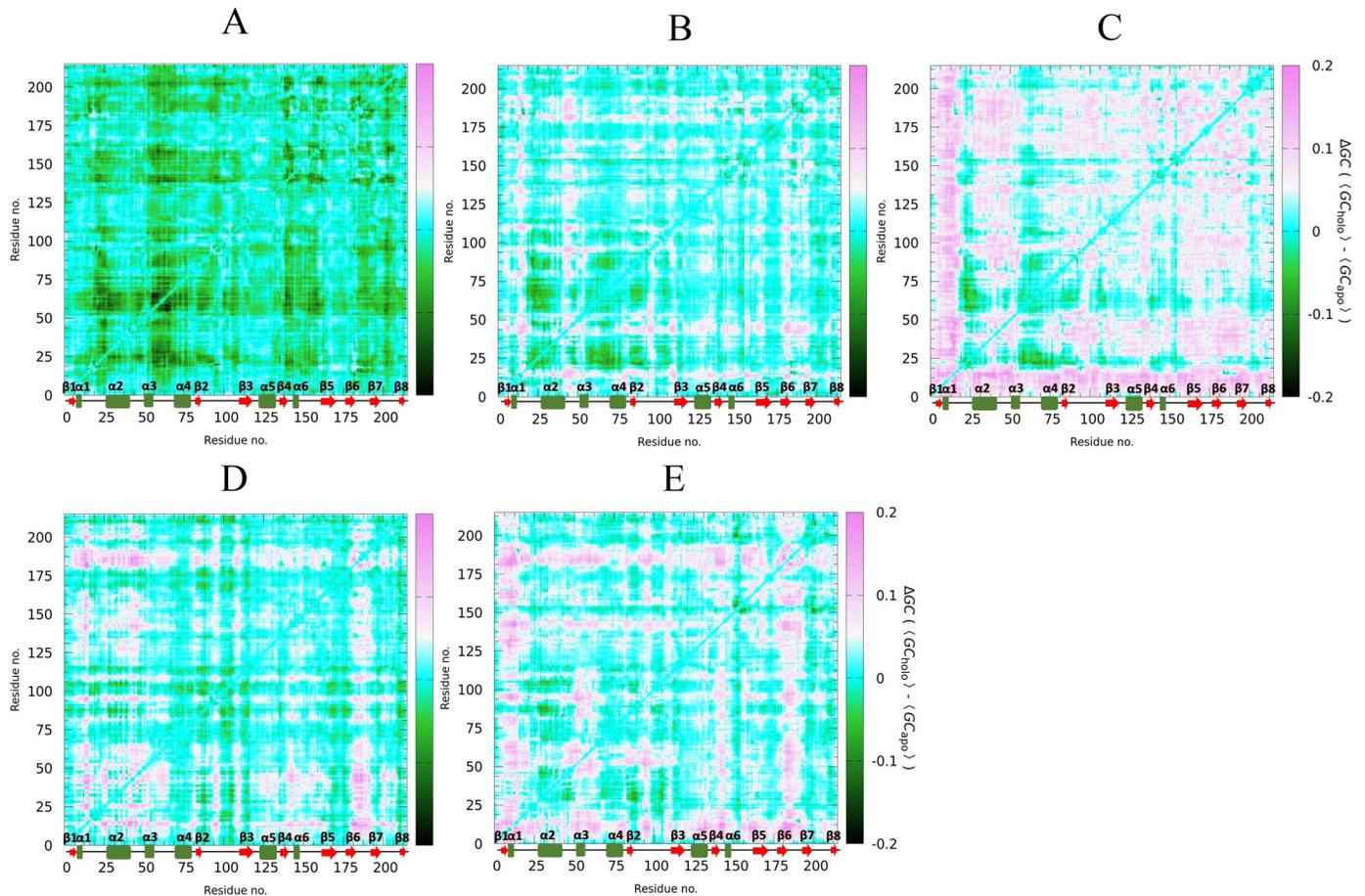


Fig 12. Comparison of generalized correlation coefficients of the analyzed cruzain systems. ΔGC matrices are represented as the subtraction of the apo form values from those corresponding to (A) cruzain-peptide, (B) cruzain-compound 1 and (C) cruzain-compound 2 systems, and as the subtraction of the cruzain-peptide from those calculated for (D) cruzain-peptide-compound 1 and (E) cruzain-peptide-compound 2 complexes.

<https://doi.org/10.1371/journal.pone.0211227.g012>

latter agrees with previous results that revealed the decrease in cruzain motions when the substrate lies within the enzyme active site. In parallel, the results obtained for apo and holo enzyme revealed an increase in correlations, triggered by ligand binding, mostly in cruzain-compound 2 complex (Fig 12C). This is demonstrated by the prevalence of positive values in the ΔGC matrix, i.e., holo minus apo correlations (pink spots). In the case of cruzain-compound 1 complex, an appreciable increment in correlation values was observed for residues lying in positions 14–19, 45–50, 90–100 and 180–190 (Fig 12B); while in the cruzain-compound 2 complex, many regions of the protein increased their coupled motions (Fig 12C). Besides, a decrease in correlation values (green points) of helix 2 (CYS25 localization) is observed in both systems (Fig 12B and 12C). Changes in coupled motions involving to interface residues of holo form are associated with the ligand-mediated “linkage” between loop₁₀₋₂₅ and loop₈₄₋₁₀₉. In addition, the comparison of correlated motions between cruzain-peptide and cruzain-peptide-ligand systems, revealed an intensification of couplings in regions comprising positions 10–55 and 182–192 (loop between $\beta 6$ and $\beta 7$) when compound 1 is bound to site 3 (Fig 12D). Likewise, when compound 2 is present, the correlated values increased in regions corresponding to N-terminal-20, $\alpha 3$, loop between $\beta 2$ and $\beta 3$, and loop between $\beta 4$ and $\alpha 6$ (Fig 12E). It is important to highlight that in both cases, several residues interacting

with aforementioned ligands lie in structural regions, which augmented their coupling motions.

The correlation network analysis further dissects residue couplings and reveals residue communities having the potential to facilitate long-range allosteric signal propagation. Also, the modification in the total number of communities and in the rearrangement of their residue composition provides a description of dynamical changes induced by effector binding during simulation time. In this sense, the study of intercommunity coupling strength in cruzain revealed the state-specific coupling paths and overall dynamical modules within the enzyme. The structure of papain-like proteases family are clearly partitioned into two main domains, and the analysis of time-averaged weighted networks adequately captured both of them in all the studied systems (Fig 13). In the network-based representation of cruzain apo form, its catalytic residues are distributed in three different communities, i. e., CYS25 in community 5, HIS162 in community 9 and ASN182 in community 2; while the two domains are linked by communities 7–9, 7–8 and 2–3. This highlights the indirect coupling of the catalytic residues. The analysis of intercommunity coupling strengths reveals state-specific allosteric coupling between site 3 and the active site through the edges between communities 6 and 5, and communities 2 and 9 (Fig 13A). On the other hand, the peptide binding induced changes in both

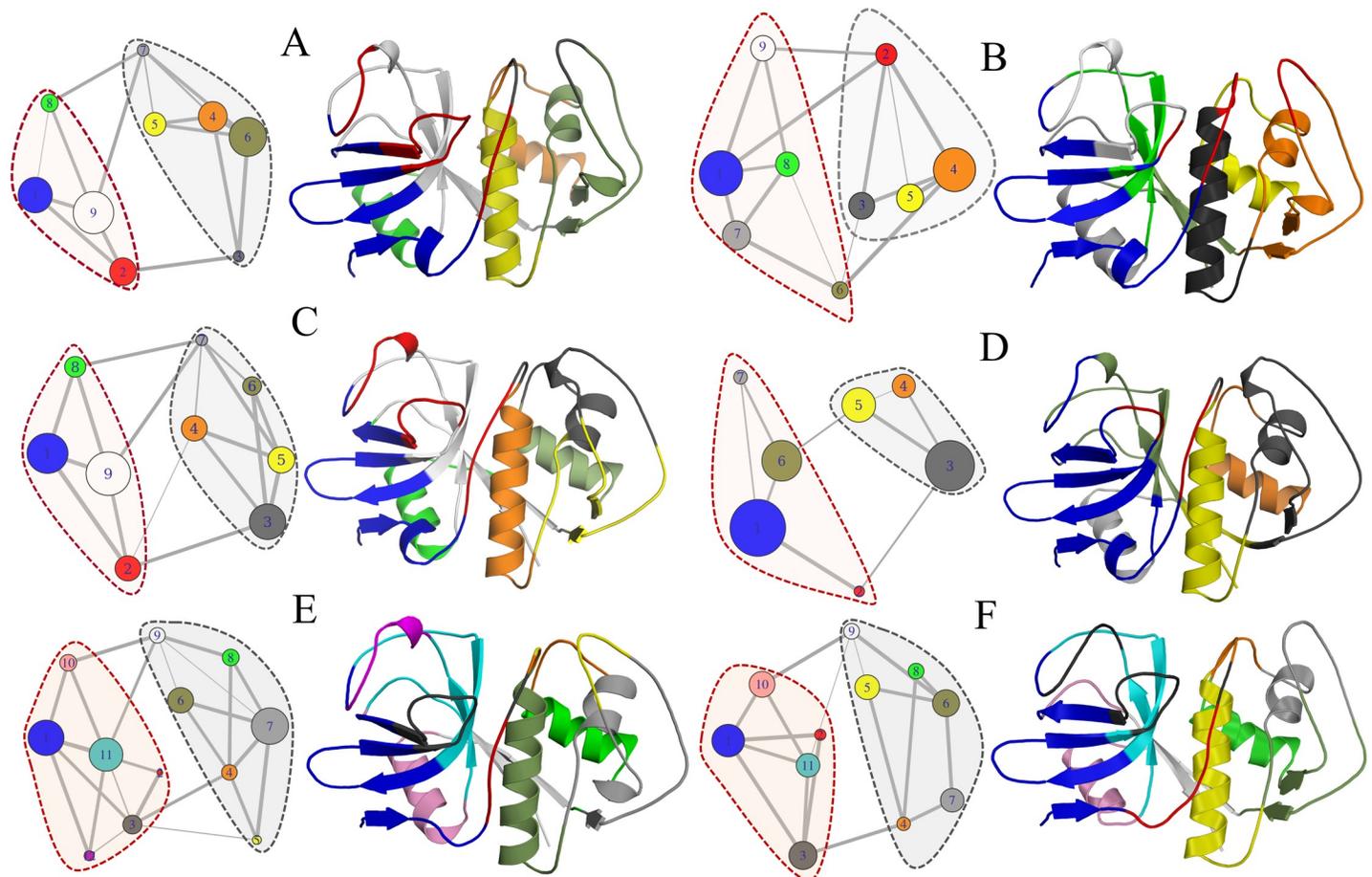


Fig 13. Community network analysis. Community structure corresponding to (A) cruzain apo form, (B) cruzain-peptide, (C) cruzain-compound 1, (D) cruzain-compound 2, (E) cruzain-peptide-compound 1 and (F) cruzain-peptide-compound 2. Dashed lines enclose the cruzain structural domains and edge thickness is proportional to the number of shortest path passing through those junctions. The circle diameter is proportional to the community size (i. e., number of residues).

<https://doi.org/10.1371/journal.pone.0211227.g013>

domains and residue rearrangement within communities. Note that β -sheets-rich domain (delimited by red dashed lines in Fig 13) increased its community number; while the α -helix-rich domain (delimited by gray dashed lines in Fig 13) reduced the number of communities by lumping some of them together (Fig 13B). More importantly, the peptide binding induced a significant reorganization within the communities forming the cruzain active site (communities 2, 3, and 8 in Fig 13B). In this system, the variation in correlated motions comprising loop₁₃₉₋₁₆₁, loop₁₈₁₋₁₉₃, β 3 and β 8 played the main role in the fragmentation of the community containing HIS162 (community 9 in Fig 13A, which was fragmented in communities 8 and 6 in Fig 13B). Moreover, in the α -helix rich domain, the communities reordering involved the “packing” of loop₁₀₋₂₅ and loop₈₄₋₁₀₉. It is also important to highlight the loss of a direct correlation edge between the community containing the HIS162 residue and those corresponding to the α -helix rich domain in the cruzain-peptide complex (Fig 13B).

The comparison of community structure between the apo and complexed systems revealed remarkable conformational changes induced by effector binding (Fig 13C and 13D, see also S2 Table), highlighting a rearrangement of community numbers and its residue members in holo systems. The cruzain-compound 1 complex displayed a change in community 3 size and a new weak coupling between the main domains appeared (edge between community 4–2 and 4–3). However, the strength of direct interactions remained unchanged with respect to the apo form, except for the interaction between communities 1 and 8, which enhanced their correlation. Note that community reordering took place in the α -helix domain, specifically in loop₄₁₋₅₀ and loop₈₄₋₁₀₉, which are located within site 3 and site 1, respectively (Fig 13C). Conversely, in the cruzain-compound 2 complex, a whole rearrangement in communities was observed in both protein domains (Fig 13D). Here, the domain containing the helix 2 lost two communities, inducing more concerted motions within its structure. On the other hand, the binding of compound 2 induced changes in β -sheets-rich domain, with reassignment of a few residues from communities 2, 8 and 9 in the newly formed community 1. The decrease in community numbers suggests that compound 2 triggers an overall structural reorganization of cruzain. The strength of inter- and intra-domain couplings increased considerably in this system, which could affect the overall enzyme activity. Indeed, the appearance of a new correlation between two domains through communities containing residues of active site was observed, indicating a perturbation in the correlations of these residues (edge between community 6 and 5 in Fig 13D).

The comparison between the community networks of cruzain-peptide-ligand systems with respect to the remaining complexes shows an increase in community number, which indicates a disruption in correlated motions because of ligand presence in site 3 (Fig 13E and 13F). Note that the α -helix rich domain was split into six communities in both ternary systems (compare Fig 13B with Fig 13E and 13F). In this case, the main changes were detected in correlations of residues lying in loop₈₄₋₁₀₉ and loop₁₀₋₂₅ (Fig 13E and 13F). Another key factor that drives the fragmentation of communities in these systems was the substrate presence, which also prevented the increase in correlated motions (compared Fig 13C and 13D with Fig 13E and 13F). Therefore, these findings indicate that there are less overall protein movement when both, the binding site and site 3 are occupied.

Network path analysis reveals couplings between cruzain orthosteric site and allosteric site 3

The residue centrality based on average betweenness can characterize and differentiate highly connected residues that mediate stable interaction networks and allosteric communications in protein structures. This parameter can be presented as profiles, being a guiding indicator for the identification of functional residues critical to allosteric regulation. The centrality analysis

highlighted remarkable differences in the way information is propagated through free cruzain versus the enzyme bound to allosteric modulators (Fig 14). These results showed changes in the centrality degree upon ligand association, indicating an alteration in dynamic couplings and a reduction in the diversity of the shortest pathways connecting all residues.

In the cruzain apo form, some residues with high-density network connections, i.e., ASP18, GLN19, ASN33, HIS115, ALA136, LEU165 and residues belonging to $\beta 8$ were detected (Fig 14A and 14B). Note that most of these residues lie at the interface of the two domains of cruzain (i.e., GLN19, ASN33, and residues from $\beta 3$, $\beta 4$, $\beta 5$ and $\beta 8$), thus suggesting the important role of interdomain signal transduction in the protein function. It is noteworthy the high centrality values of GLN19, ALA136 and LEU165, which are key positions for catalysis and substrate recognition in papain-like proteases [102].

In the cruzain-compound 1 complex, new highly connected residues arouse, i. e., loop₁₀₋₂₅, GLU35 and LEU48 (Fig 14A), many of which lie within site 3. These results point out that ligand presence increases the number of shortest paths crossing over site 3, making this groove a potential region for control of enzyme function. In the case of cruzain-compound 2 complex, residues LYS17, TRP26, SER29, GLU86, TYR91, VAL135, ASN182 and TRP188 increased their centrality (Fig 14B), being LYS17, GLU86, TYR91 and ASN182 part of site 3.

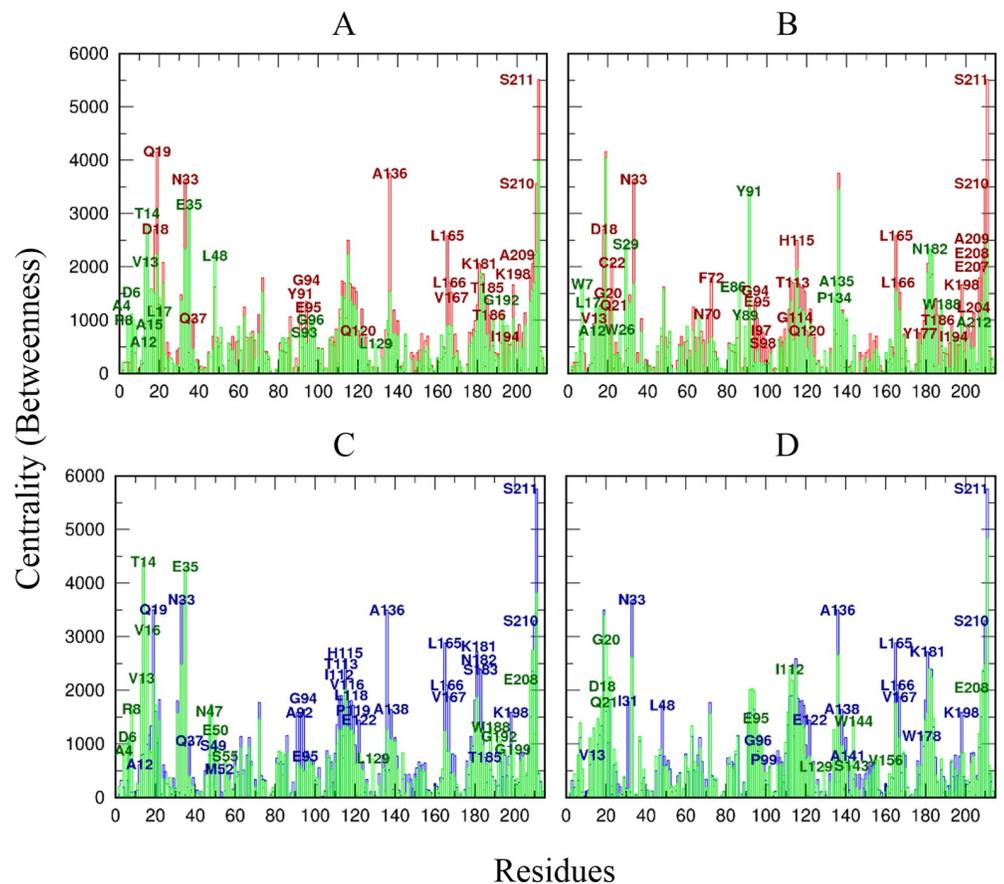


Fig 14. Residue centrality in cruzain analyzed systems. For convenience, the centrality profiles are represented as a comparison between ligand-bound and reference systems, i. e., cruzain apo form (red lines) and cruzain-peptide complex (blue lines), and cruzain-ligand complexes (green lines): (A) cruzain-compound 1, (B) cruzain-compound 2, (C) cruzain-peptide-compound 1 and (D) cruzain-peptide-compound 2. Residues with a Δ Centrality greater than 500 are labeled with the same color of their corresponding system.

<https://doi.org/10.1371/journal.pone.0211227.g014>

Interestingly, most residues mentioned above are conserved across the papain-like cysteine proteases [11]. This suggests that ligands designed against site 3 may generate an allosteric signal involving common residues of this protease family, which could trigger the inhibition of the enzyme catalytic function.

Similarly to the apo form, in cruzain-peptide system, the residues positioned at the interface between the two domains of the enzyme showed the highest values of centrality, i.e., GLN19, ASN33 and β 3, β 5 and β 8 (Fig 14C and 14D). In cruzain-peptide-compound 1 system, there was an increase in centrality values within the N-terminal region including loop₁₀₋₂₅ (Fig 14C). However, there was no appreciable differences between cruzain-peptide-compound 2 complex and the reference system (Fig 14D).

On the other hand, optimal and suboptimal paths were calculated between CYS25 and each hot spot previously identified for site 3. Among all the pair of residues analyzed here, CYS25-THR14 was the one showing the largest increase in correlation between site 3 and the catalytic core because of ligand binding. This is consistent with centrality analysis where THR14 was one of the main residues displaying significant changes in its connections as a consequence of ligand binding (Fig 14).

By examining the path lengths distribution, we found that the shortest path of cruzain-compound 1 complex do not differ substantially from that of the apo enzyme (3.12 for apo form; 2.88 for cruzain-compound 1) (Fig 15A). Nevertheless, the distribution derived from this system trajectory is slightly deviated toward shorter path lengths, indicating that motions of the residues linking the allosteric and catalytic sites are more tightly correlated when compound 1 is bound. The determination of residues critical to the allosteric signal propagation (residue with highest degeneracy values), revealed the occurrence of different paths between the free enzyme and cruzain-compound 1 complex (Figs 16A, 16C and S19). The shortest path between THR14 and CYS25 in the apo state was: THR14 → VAL16 → ASP18 → GLN19 → CYS22 → SER24 → CYS25. In contrast, the optimal path when compound 1 is bound comprise the following residues: THR14 → GLU35 → ILE31 → PHE28 → CYS25. Here, a considerable decrease in node degeneracy values and the increment in the number of new residues involved in suboptimal paths are observed when compound 1 is bound (S19 Fig). This analysis identifies two main routes for allosteric communication, i.e., one that goes through helix 2 and the other one through loop₁₀₋₂₅ (Fig 16A and 16C).

On the other hand, a shift of centrality values towards lower ranges was also observed for cruzain-peptide-compound 1 complex (3.41 for cruzain-peptide; 3.35 for cruzain-peptide-compound 1) (Fig 15C). In this case, the same shortest path was identified for cruzain-peptide-compound 1 and its reference system, which passed through loop₁₀₋₂₅ (i. e., THR14 → VAL16 → LYS17 → ASP18 → GLY20 → CYS22 → SER24 → CYS25) (Fig 16B and 16E). Overall, we conclude that the binding of compound 1 alters the distribution of cruzain internal pathways.

The suboptimal paths between THR14 from site 3 and catalytic cysteine transverse almost the same nodes in cruzain-compound 2 complex and apo form (Fig 16A and 16D). Interestingly, the signal propagates directly to the catalytic core through the residues of loop₁₀₋₂₅, being the optimal path for cruzain-compound 2 complex: THR14 → VAL16 → LYS17 → ASP18 → GLY20 → CYS22 → GLY23 → SER24 → CYS25. Even though the two systems share a similar optimal path, a strongest correlation between site 3 and the catalytic site in the holo form is detected. This becomes apparent by observing that the shortest path in cruzain-compound 2 system shifts towards lower length value (apo, 3.12; holo, 2.88) (Fig 15B). The previous shift is likely to arise from a more coherent signal propagation in the holo simulation, indicating a possible decrease in the entropy along the pathways due to compound 2 binding. In addition, an increase in the number of suboptimal paths with lower length values is observed in cruzain compound 2 complex with respect to the cruzain-compound 1 (Fig 15A and 15B).

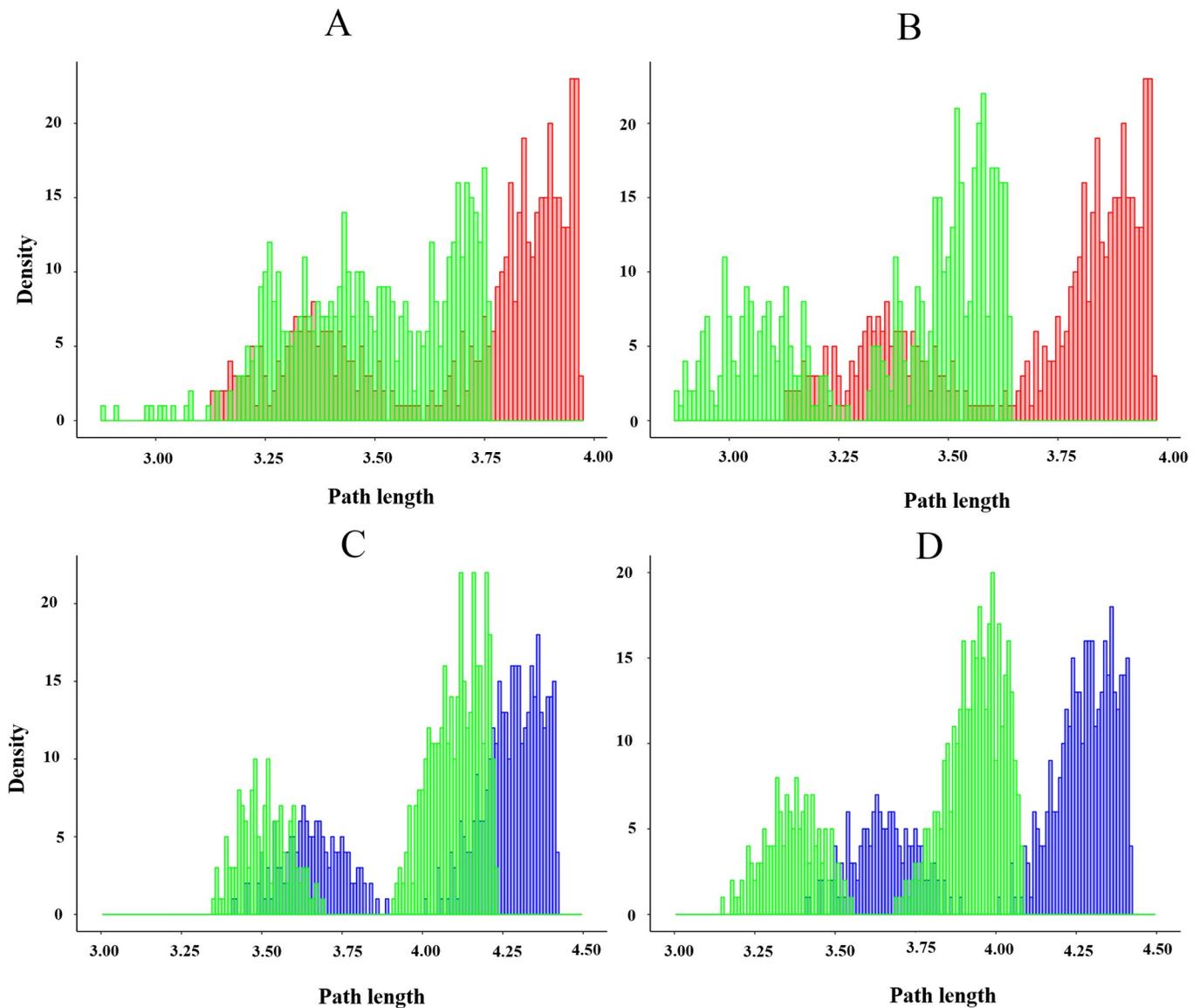


Fig 15. Statistical distribution of suboptimal paths. The histogram of the 500 path lengths associated with the apo and holo trajectories. Histograms corresponding to (A) cruzain-compound 1 and (B) cruzain-compound 2 were compared with apo distribution (red lines), while the histograms concerning to (C) cruzain-peptide-compound 1 and (D) cruzain-peptide-compound 2 complexes were compared with that of cruzain-peptide system (blue lines).

<https://doi.org/10.1371/journal.pone.0211227.g015>

Finally, the cruzain-peptide-compound 2 complex displayed an identical shortest path to that of the cruzain-peptide and cruzain-peptide-compound 1 systems (see explanation above). However, the distance of this path in terms of correlated motions was shorter than the other two (3.14) (Figs 15D and 16F). All these lines of evidence point out a dynamical tightening of the active site with site 3 in presence of compound 2, which might suggest, more potent allosteric influences of this compound in comparison to the other one.

Energy interactions networks of cruzain reveal significant changes in interactions of ionizable residues of site 3 in presence of allosteric ligands

The structural stability and function of proteins depend on intricate network of inter-residue interactions. In this sense, the PENs approach accounts for the structural plasticity, stabilizing

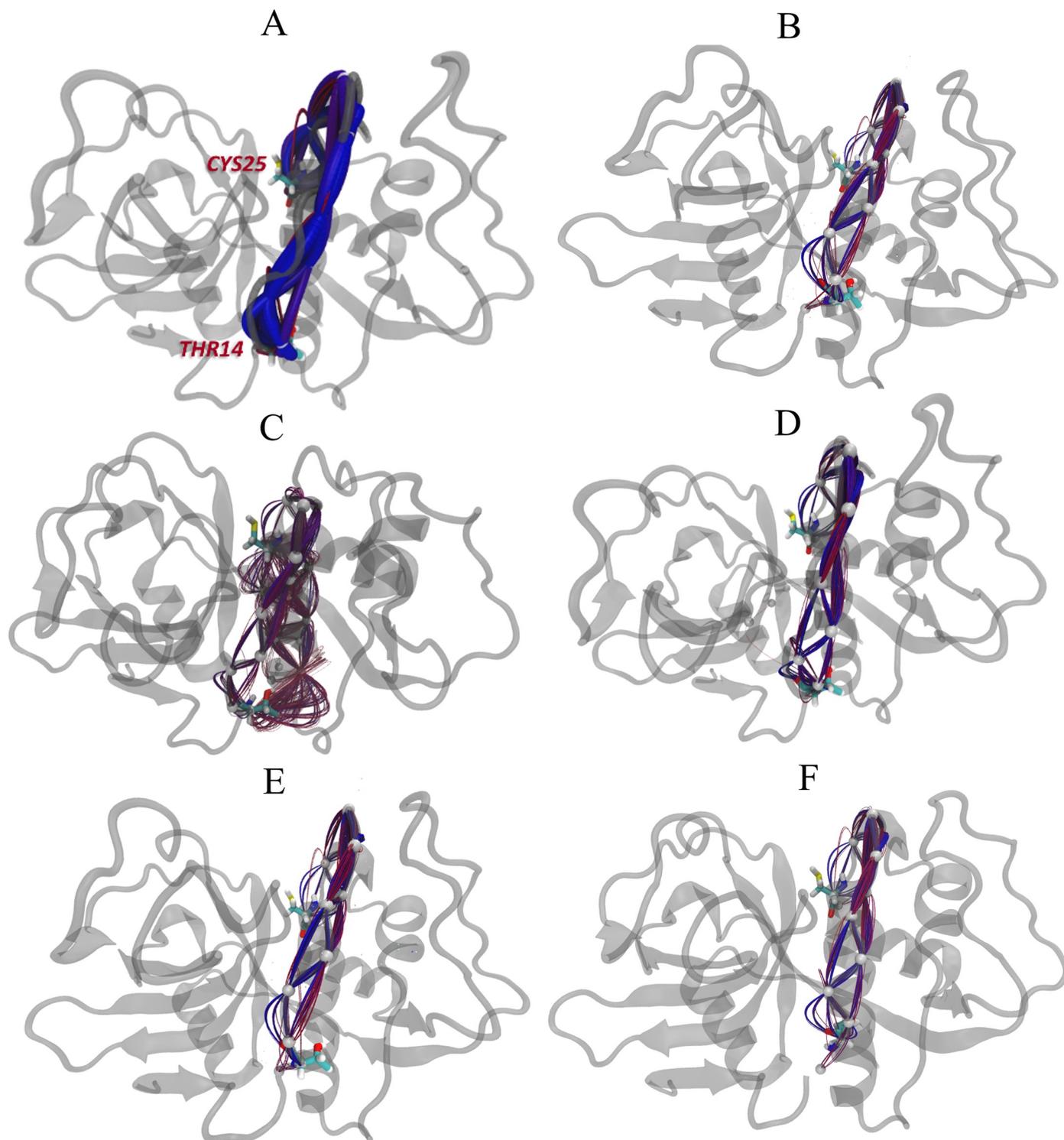


Fig 16. Suboptimal paths of cruzain analyzed systems. The 500 suboptimal paths between THR14 and CYS25 residues were calculated for (A) cruzain apo form, (B) cruzain-peptide, (C) cruzain-compound 1, (D) cruzain-compound 2, (E) cruzain-peptide-compound 1 and (F) cruzain-peptide-compound 2 trajectories. The shorter and longer paths are shown as blue and red splines, respectively.

<https://doi.org/10.1371/journal.pone.0211227.g016>

ligand binding to site 3 (Fig 17). The latter event sterically hinders the interaction between LYS17 and GLU35. In consequence, the role of LYS17 in the propagation of energetic paths across cruzain structure is affected, as well as that of some other surrounding residues, i.e., ASP18 and PHE28, which increased their centrality values in various analyzed systems. Moreover, there is a noticeable decrease in centrality values of GLU73 and residues lying within the $\alpha 1$ helix in ligand-bound systems, i. e., ASP6, ARG8 and ARG10 (Fig 17). Note that ASP6 and ARG8 are highly conserved residues from the N-termini of these cysteine proteases [11], and the previously-mentioned changes in their local interactions may affect the structural instability of this protein regions. However, residues like LYS181, GLU122, HIS106 and PHE28, increased their centrality values in almost all complexed systems. Interestingly, the aforementioned residues are semi-conserved and non-conserved within the C1A clan, except for PHE28. This suggests that changes in interaction energies involving the previous residues upon ligand binding may perturb the function of cruzain.

Discussion

Notwithstanding cruzain being an attractive therapeutic target for Chagas disease, few of the reported inhibitors of this protease have been selected for clinical trials in humans to date. This is probably due to their toxicity, low pharmacologic profiles and/or off-target effects. There are some aspects of this enzyme requiring further study that may foster the discovery of new drugs, such as the search for potential druggable sites to design non-competitive inhibitors by using conformational-sampling technics. So far, there are no reported studies on the identification of allosteric inhibitors of cruzain through structure-based virtual screening approaches. In this sense, our results constitute an *in silico* strategy for designing novel allosteric inhibitors of this enzyme. In the current work, several approaches of structure-based drug design were employed in order to rank the best hits that could contribute to a more specific inhibition of this protease.

The discovery of “hidden” and transient pockets, which only occur in some conformations of the protein and with no association with experimental structure, has vastly increased the number of druggable allosteric sites in these biomolecules [28, 75, 104]. Due to the difficulties of designing allosteric inhibitors targeting allosteric binding sites employing structure-based approaches, the combination between computational and experimental studies is required to explore this promising area [105]. By means of MD simulation techniques, it is possible to assess the protein conformational states, as well as the energetic penalty associated with the population shift toward the conformation required for the exposure of the hidden pocket [49, 106]. Therefore, MD simulations are particularly useful to the study of allostery, which inspired us to employ it in the present work [75, 105].

First, we performed a structural comparison of all available crystal structures of cruzain, and a few differences were found between them. Moreover, the dynamical survey of previously identified cavities showed the presence of a hidden pocket with a functional gate (site 3) and one transient pocket (site 1). Additionally, we found that the main residues of site 1 are non-conserved [11], thus making this site suitable for the design of specific allosteric inhibitors. Ultimately, the characterization of site 1 and site 3 would not have been feasible through the mere superposition of crystallographic structures of cruzain. Our study of these allosteric sites, which are potentially druggable and can modulate the enzymatic activity [37], provides a solid basis for further drug discovery.

A concrete evidence of allosteric modulation in papain-like protease family, mediated by ligand binding to site 1 and 3, is a previous study reporting several allosteric regulators of HCatK [37]. In that work, various allosteric inhibitors were predicted as compounds binding

the sites 1 and 3 of the human protease. However, none of the previously-identified compounds were predicted to bind a narrow groove conformation of site 3 equivalent to that observed during cruzain MD simulations [37]. On the other hand, the hypothesis that site 3 is involved in allosteric regulation and/or enzyme structural stability was previously proposed by Durrant *et al* [11]. In that work, the authors emphasized that the presence of two conserved residues lying outside site 3, i.e., SER49 and GLN51, may stabilize this groove within papain-like proteases. In addition, they underlined the role of conserved residues TYR89 and PRO90 in the rigidity of loop₈₄₋₁₀₉ and the occurrence of a hydrogen bond network involving residues 47, 86 and 91. All these findings point out the role of site 3 in the internal stabilization of cruzain structure [11]. It is important to highlight that the “closed form” of site 3 is the most prevalent in cruzain crystal structures reported so far, since, generally, LYS17 and GLU35 are at optimal distance to form a salt bridge (~3.5 Å in crystal structures). Noteworthy, two non-conserved glutamic residues, predicted in the protonated state at pH 5.5, are present in the vicinity of LYS17 and GLU35 (Fig 9). Interestingly, the protonation states of this patch of ionizable residues at the protein surface is likely to be highly susceptible to pH variations, which, in turn, may have important functional implications. This phenomenon cannot be observed through conventional MD simulations techniques, but, in theory, it can affect the “open-close” dynamical equilibrium of this groove.

Secondly, the combination of Autodock Vina score and MM-GBSA free energy calculations allowed us to propose some compounds with the potential to bind the predicted allosteric pockets. Our results suggest that site 3 is more druggable than site 1, as none of the selected compounds from the database seemed to form a stable complex with cruzain through the latter site, as shown through long MD simulations. Therefore, the *in silico* design of ligands displaying affinity for site 1 remains challenging and the usage of larger compound databases and other transient conformations of the pocket are likely to be required. On the other hand, compounds 1 and 2 were proposed here as promising scaffolds of cruzain allosteric modulators. We verified that the aliphatic chains of the site 3 residues mediated the main protein-ligand interactions within the groove. It is important to highlight that the internal part of this pocket is formed by many aliphatic side chains, which are facing to each other, enabling the stability of secondary structures like $\alpha 2$ and $\alpha 3$ through intra-molecular hydrogen bonds (S4 and S5 Figs). Moreover, several side chains of loop₁₀₋₂₅, loop₄₁₋₅₀ and loop₈₄₋₁₀₉ contribute to the stabilization of this internal structure of the groove. The disruption of the internal interaction pattern of the pocket due to the ligand binding might have a functional impact on cruzain activity.

The structural analysis of cruzain in complex with identified compounds showed slight conformational changes in enzyme binding site. However, through these structural analyses, we could also observed that compound 2 increased the flexibility of cruzain structure to a larger extent than compound 1. Remarkably, allostery can occur in the absence of well-defined conformational changes, indicating that subtle changes in protein dynamics can induce a population shift in conformational ensemble without substantially changing the mean conformation of the protein [107, 108]. The results shown here suggest that the proposed compounds would exert a modulation of the protein through the generalized allostery model, which involves the redistribution of the existing population of the conformations upon the ligand binding [26]. The modification of protein flexibility caused by the studied compounds can produce, in turn, a perturbation of enzyme processativity. Interestingly, a previous study on allosteric control in HCatK suggested that this enzyme possesses an allosteric inhibition without significant conformational changes, but through the alterations of protein motions [38]. On the other hand, the MM-GBSA calculations performed here indicated that ligand binding to site 3 moderately perturbed the substrate affinity, and that the slight differences in $\Delta G_{eff}(\text{peptide})$ values were mainly because of modifications in the complementarity at the interfaces

between S1'-S3' subsites and the peptide positions P1'-P3'. According to the results obtained from RMSD distributions, CP analysis and correlation motions, the substrate binding to the orthosteric site decreases the flexibility of cruzain, thereby making the allosteric effect less appreciable upon ligand binding to site 3.

Thirdly, a community analysis of cruzain was performed, which made possible the detection of topological changes in correlated motions of this enzyme that result from ligand binding. In this analysis, we mapped the domains of papain-like cysteine proteases based on coupled movements of cruzain. As observed from our results, an increase of global correlated motions is produced when compound 2 is bound to cruzain. However, the interaction with this ligand also leads to a decrease of the "signal transduction" between the two domains of the enzyme (e.g. edges between communities 5 and 6, and between 2 and 3 when compound 2 is into site 3). A significant repartitioning among the community network stemmed from the weaker interface edges connecting the β -sheet and α -helix rich domains upon the ligand binding. In particular, we observed reorganization of communities containing the active site residues due to the presence of the compounds. This could lead to a modulation of the enzymatic activity, which requires appropriate coupled motions of the residues involved in the catalytic process. In the ternary systems (cruzain-substrate-compounds), the ligand presence into site 3 also leads to remarkable community reorganization as well as changes in the inter-domain correlated motions. This suggest the capacity of the compounds to modify the overall motions of the enzyme even when the substrate is bound to the active site, a likely signature of non-competitive inhibition.

Moreover, in the two active cruzain forms analyzed here (apo and cruzain-substrate complex), specific residues across the two domain interfaces are characterized by high centrality values, thereby suggesting their involvement in the flow and transmission of dynamical information within cruzain. These results are useful to devise experimental procedures, such as point mutations, that would yield valuable information about structure-function relationship and structural organization of cruzain and other papain-like proteases. To our knowledge, this is first work that has performed this type of study for this cysteine protease superfamily based on MD simulations data, thus providing an insight into their regulation, function and a possible pharmaceutical intervention.

The allosteric paths that link site 3 to the orthosteric site were elucidated by combining MD simulations with correlation of protein motions based in network theory. Our results showed that compound 2 triggers the propagation of signals by creating shorter pathways and stronger correlations when compared to the apo form. Furthermore, compound 2 perturbs cruzain motions in a higher degree than compound 1, pointing out that structural dissimilarities between ligands may define different communication paths, even when binding the same allosteric site. Our study also underlines the critical role of loop₁₀₋₂₅ and helix 2 in allosteric communication between site 3 and catalytic site of cruzain. In this sense, previous studies on evolutionary conservation have shown that most residues of loop₁₀₋₂₅ and helix 2 are partially or strongly conserved across papain-like proteases [11]. The changes in dynamical correlations observed here indicate a possible regulation of key allosteric residues of this protein family through the site 3 narrow groove. All these results evidence an effective signal propagation, which, in turn, may enable the allosteric modulation. The design of drugs interacting with key residues involved in the allosteric signal propagation and/or modifying their existing couplings, e.g., THR14, LYS17, ASP18, GLN19, CYS22, SER24, constitutes a promising approach to the modulation of cruzain activity, extensible, in principle to other papain-like proteases.

We extended our analysis to the calculation of residue-residue interaction energies separately for apo form, enzyme-substrate and compound-bound systems. Here, we included the computation of the specific residue node-betweenness centralities based on the interactions

network. Therefore, we were able to identify the main residues involved in the transmission of energetic information in the active and bound states of cruzain. Indeed, our data suggest that in the cruzain-ligand forms the ionizable residues belonging to site 3 and its vicinity substantially change their pairwise energies. In this analysis, we corroborated the important role of the interaction between LYS17 and GLU35, which was the most affected one by the ligand presence. Moreover, we referred to the perturbations occurring in charged residues lying in $\alpha 1$, $\beta 1$ and loop₁₀₋₂₅ upon binding to site 3. We also underlined the role of three non-conserved histidine residues (i. e., HIS43, HIS106 and HIS115) in the connections between energetic paths within cruzain structure, as well as the structural implications of the interactions of conserved residues in modulation of papain-like cysteine proteases. Remarkably, a recent MD study proposes that allostery in the well-studied [PDZ domain](#) is driven by changes in [electrostatic](#) effects rather than solely changes in dynamics [99, 109]. Therefore, the predicted electrostatic rearrangement within the cruzain structure suggests its possible involvement in the enzyme allosteric modulation.

Remarkably, the *in silico* identification of non-competitive inhibitors is a challenging task because of various aspects. Firstly, the occurrence of transient and hidden pockets that are not observed in the available crystal structures of cruzain has to be characterized. This is a step that, when carried out *in silico*, requires experimental validation. Another major issue of the approach presented here is that the *in silico* selection of the most promising compounds was based only on their affinity for the putative allosteric pockets. However, until experimental assessment, we cannot confirm whether the selected compounds will act either as inhibitors or as simple binders unable to trigger an allosteric signal. In a worst scenario, some molecules targeting the allosteric sites might even act as activators of the enzymatic activity.

In order to understand and exploit allostery it is necessary to propose the allosteric sites, allosteric modulators and residues involved in propagating the allosteric signal. Although this work constitutes an entirely computational approximation to the discovery of allosteric inhibitors, there are various experimental techniques that can corroborate these predictions. Kinetic assays with fluorogenic substrates are usually performed in order to determine the inhibition mechanism, which, in turn, can confirm the non-competitive nature of the inhibition exerted by the proposed compounds [35, 37]. More insightful information can be obtained through the determination of the binding modes of the inhibitors using X-ray crystallography and NMR. The latter technique has also proven quite useful in the study of the allosteric mechanisms by tracking the fluctuations in ensemble populations in free and bound states, and changes in the motions of residues involved in allosteric pathways [106, 110–113]. Site-directed mutagenesis can also be used to validate the [binding site](#) and the predicted allosteric communication mechanisms [114–117].

Overall, we have characterized transient and hidden pockets in cruzain with the capacity of exerting allosteric modulation of the enzymatic activity. We also proposed some compounds displaying suitable affinity for site 3, which was the most likely druggable pocket predicted in this work. The *in silico* analyses performed here suggest that allosteric control in cruzain could be achieved due to ensemble redistribution and without major conformational changes, as has been described for other papain-like proteases. Our approach constitutes a promising way of designing novel allosteric inhibitors of cruzain.

Supporting information

S1 Table. Principal hydrogen bonds established along simulation time between cruzain allosteric site and hit compounds.

(PDF)

S2 Table. Distribution of cruzain residues within optimal community structures of apo and holo forms.

(PDF)

S1 Text. Comparison and selection of cruzain crystal structures.

(PDF)

S2 Text. Pairwise distance distributions identify differences in key residues of S1, S2 and S3 subsites between the apo and holo forms.

(PDF)

S3 Text. Small changes along PC2 detected for cruzain apo and holo forms.

(PDF)

S4 Text. Comparison of cross-correlation (CC) and generalized correlation results (GC).

(PDF)

S1 Fig. Representation of grid boxes employed in VS simulations into cruzain potential allosteric sites. (A) site 1 and (B) site 3 are highlighted in cruzain surface representation together with the corresponding docking boxes. The axis dimensions are labeled in each case and the residues contained in both sites are colored in magenta.

(TIF)

S2 Fig. Comparison of crystal structures reported for cruzain. (A) Superimposed structures of cruzain obtained from PDB database. Secondary structure elements are colored as follows: alpha helices (red), beta sheets (yellow) and loops/turn (green). All ligands are positioned in the active site. (B) Pairwise RMSD calculated for the backbone atoms of cruzain crystal structures (47 in total considering each cruzain copy solved within the same PDB file). PDBID of each analyzed structure is specified in alphabetical order on the right hand side.

(TIF)

S3 Fig. Comparison of the cluster central structures generated for each site by volume-based and pairwise RMSD cluster analysis. (A) Clusters (clust) calculated for site 1 (S1) and (B) site 3 (S3). Node size is proportional to average volume value and color gradient correspond to the frames number conforming each cluster. Edge color and thickness are proportional to pairwise RMSD values between the cluster central structures.

(TIF)

S4 Fig. Residue composition of cruzain site 1 and site 3. Surface representation of two opposite orientations and top view of cruzain (A) site 1 and (B) site 3. Protein surface is colored according to atom type and volume size is depicted as a grid of points.

(TIF)

S5 Fig. Dynamical opening/occlusion motions of site 3 gate. Snapshots of site 3 were taken at different times of cruzain apo simulations. The salt bridge formed between LYS17 and GLU35 represents the functional gate of this pocket. Cavities are colored in lightgrey and the distances between the former residues are labeled in each case.

(TIF)

S6 Fig. Allosteric site stability in apo form of cruzain. (A) Time evolution of RMSD values calculated for heavy atoms of site 1 (red) and site 3 (green). Each replica was labeled with the PDBID of the cruzain structure employed as starting conformation in each case. (B) RMSF calculated for each replica of cruzain apo form. The secondary structure of cruzain is represented along the protein sequence. (C) Tube representation of free enzyme average conformation.

Tube width is proportional to the per-residue atomic fluctuations computed for C-alpha atoms and the regions with high fluctuations are colored in red.

(TIFF)

S7 Fig. Time evolution of instantaneous ΔG_{eff} values for cruzain-compound 1 complex.

Effective binding free energies of compound 1 are shown together with accumulated mean values (black lines). Dashed lines indicate the equilibration time of the MD simulations. Every replica was labeled with the corresponding average ΔG_{eff} value and its standard error of mean.

(TIFF)

S8 Fig. Time evolution of instantaneous ΔG_{eff} values for cruzain-compound 2 complex.

Effective binding free energies of compound 2 are shown together with accumulated mean values (black lines). Dashed lines indicate the equilibration time of the MD simulations. Every replica was labeled with the corresponding average ΔG_{eff} value and its standard error of mean.

(TIFF)

S9 Fig. Time evolution of instantaneous ΔG_{eff} values for cruzain-peptide complex.

Effective binding free energies of peptide are shown together with accumulated mean values (black lines). Dashed lines indicate the equilibration time of the MD simulations. Every replica was labeled with the corresponding average ΔG_{eff} value and its standard error of mean.

(TIFF)

S10 Fig. Time evolution of instantaneous ΔG_{eff} values for cruzain-peptide-compound 1 complex.

Effective binding free energies of compound 1 are shown in grey color in order to highlight the deviation in ΔG_{eff} profiles of peptide (colored lines) in each replica. The accumulated mean values are shown as black lines. Dashed lines indicate the equilibration time of the MD simulations. Every replica was labeled with the corresponding average ΔG_{eff} value and its standard error of the mean.

(TIFF)

S11 Fig. Time evolution of instantaneous ΔG_{eff} values for cruzain-peptide-compound 2 complex.

Effective binding free energies of compound 2 are shown in grey color to highlight the deviation in ΔG_{eff} profiles of the peptide (colored graphs) in each replica. The accumulated mean values are shown as black lines. Dashed lines indicate the equilibration time of the MD simulations. Every replica was labeled with the corresponding average ΔG_{eff} value and its standard error of mean.

(TIFF)

S12 Fig. RMSDs for the ligands and the peptide along simulation time. RMSD time profiles calculated with respect to the initial frame of MD simulations. The RMSD values for the heavy atoms of ligands are represented for the following systems (A) cruzain-compound 1, (B) cruzain-compound 2, (C) cruzain-peptide-compound 1 and (D) cruzain-peptide-compound 2.

The RMSD values with respect to the peptide heavy atoms are represented for (E) cruzain-peptide, (F) cruzain-peptide-compound 1 and (G) cruzain-peptide-compound 2 systems. Different colors represent an individual replica in each case.

(TIFF)

S13 Fig. Distributions of pairwise interatomic distance of cruzain binding site. (A) S1, S2 and S3 region of cruzain with the interatomic distances of its principal residues. Selected residues are labeled and their side-chains are depicted as stick. (B) Distance distributions obtained from the MD simulations of three analyzed systems and from cruzain crystal structures.

Graphs are labeled with the atomic pair analyzed in each case.
(TIF)

S14 Fig. Comparison of contact maps of the analyzed systems. The contact map matrices are represented as a comparison of two simulated systems positioned in upper and lower triangles, respectively. Each system is identified by a different color, i. e. red for apo-form, green for cruzain-peptide, blue for cruzain-compound 1, cyan for cruzain-compound 2, magenta for cruzain-peptide-compound 1 and black for cruzain-peptide-compound 2.
(TIF)

S15 Fig. Comparison of electrostatic surface of cruzain site 3. (A) Electrostatic surface representation of cruzain apo-form and (B) cruzain-compound 2 systems. Compound 2 is displayed as orange stick. The electrostatic potential ranges from $-kbT/e$ (red) to $+kbT/e$ (blue), where kb is the Boltzmann constant, T , the temperature and e , the electron elementary charge.
(TIF)

S16 Fig. Principal component analysis (PCA) of cruzain apo and holo forms. (A) Amplitude of the first 12 eigenvectors calculated from the covariance matrix of $C\alpha$ coordinates from MD simulations. (B) Percentage of total variance accounted for each of the first 12 eigenvectors. Labels beside each point indicate the cumulative fluctuation of the selected eigenvectors. The trajectory was projected onto the principal planes defined by the first two principal components. The holo form (orange) was projected onto the eigenvectors of the apo form (blue) for (C) cruzain-compound 1 and (D) cruzain-compound 2 complexes. A color gradient was employed to represent the density of structures in each region of phase space. The projections of the MD motions were represented along the first two eigenvector for (E) the apo form, (F) cruzain-compound 1 and (G) cruzain-compound 2 systems. The black arrows show the direction of collective motions (from red to green) and the principal loops are labeled.
(TIF)

S17 Fig. Comparison between standard and generalized cross-correlation coefficients. Upper triangle corresponds to standard correlation and lower triangle the generalized correlation. The framed values are related to strong correlations. The violet rectangles enclose residues belonging to the catalytic core and to specificity-related subsites of cruzain. On the right side, the secondary structure elements of cruzain are numbered from N- to C-termini. This numbering scheme is also used in a linear sequence of cruzain at the bottom of the correlation matrix.
(TIF)

S18 Fig. Representation of correlation networks onto cruzain structure. Each edge represents the GC value between a residue pair. All networks were filtered by a cut-off of GC values (>0.5) and the protein is colored by element of secondary structure. The edges colored in dark possess tighter correlations. This representation is shown for all analyzed systems, i. e., (A) apo-form, (B) cruzain-peptide, (C) cruzain-compound 1, (D) cruzain-compound 2, (E) cruzain-peptide-compound 1 and (F) cruzain-peptide-compound 2.
(TIF)

S19 Fig. Node degeneracy in signaling pathways calculated for the analyzed systems. Each ligand-bounded complex (green lines) was compared to its reference system, respectively, i. e., cruzain apo-form (red lines) and cruzain-peptide (blue lines). The graphs are separated as follows: (A) cruzain-compound 1, (B) cruzain-compound 2, (C) cruzain-peptide-compound 1 and (D) cruzain-peptide-compound 2.
(TIFF)

S20 Fig. Residue centrality obtained from PENs analysis of cruzain complexes. For convenience, the centrality profiles are represented as a comparison between reference (red lines) systems (i. e., apo form and cruzain-peptide complex) and ligand-bounded systems (blue lines) in each case. (A) cruzain-compound 1, (B) cruzain-compound 2, (C) cruzain-peptide-compound 1 and (D) cruzain-peptide-compound 2. (TIF)

Acknowledgments

The authors are thankful to Bruno Fernandes for technical support and to Prof. Jorge Chahine for his helpful discussions. We thank to Prof. Leandro Cristante de Oliveira for kindly provide computer time at Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD) (grant 613). Also, we thank to mesocenter of Strasbourg for running time allocation. Finally, we thank to LMDM at UFRJ, and to the research group of Prof. Vitor Leite for their support with computational resources.

Author Contributions

Conceptualization: Lilian Hernández Alvarez, Pedro Geraldo Pascutti.

Data curation: Lilian Hernández Alvarez, Diego Enry Barreto Gomes, Jorge Enrique Hernández González.

Formal analysis: Lilian Hernández Alvarez.

Funding acquisition: Pedro Geraldo Pascutti.

Investigation: Lilian Hernández Alvarez, Jorge Enrique Hernández González.

Methodology: Lilian Hernández Alvarez, Diego Enry Barreto Gomes, Pedro Geraldo Pascutti.

Project administration: Pedro Geraldo Pascutti.

Resources: Pedro Geraldo Pascutti.

Supervision: Pedro Geraldo Pascutti.

Validation: Lilian Hernández Alvarez, Pedro Geraldo Pascutti.

Visualization: Lilian Hernández Alvarez, Jorge Enrique Hernández González.

Writing – original draft: Lilian Hernández Alvarez.

Writing – review & editing: Lilian Hernández Alvarez, Diego Enry Barreto Gomes, Jorge Enrique Hernández González, Pedro Geraldo Pascutti.

References

1. McGrath ME, Eakin AE, Engel JC, McKerrow JH, Craik CS, Fletterick RJ. The crystal structure of cruzain: a therapeutic target for Chagas' disease. *J Mol Biol.* 1995; 247(2):251–9. <https://doi.org/10.1006/jmbi.1994.0137> PMID: 7707373
2. Sajid M, Robertson SA, Brinen LS, McKerrow JH. Cruzain: the path from target validation to the clinic. *Adv Exp Med Biol.* 2011; 712:100–15. https://doi.org/10.1007/978-1-4419-8414-2_7 PMID: 21660661
3. Mark W. Robinson PhD JPDP. *Cysteine Proteases of Pathogenic Organisms*: Springer US; 2011.
4. Ribeiro I, Sevcsik AM, Alves F, Diap G, Don R, Harhay MO, et al. New, improved treatments for Chagas disease: from the R&D pipeline to the patients. *PLoS Negl Trop Dis.* 2009; 3(7):e484. <https://doi.org/10.1371/journal.pntd.0000484> PMID: 19582163

5. Wilkinson SR, Kelly JM. Trypanocidal drugs: mechanisms, resistance and new targets. *Expert Rev Mol Med*. 2009; 11:e31. <https://doi.org/10.1017/S1462399409001252> PMID: 19863838
6. Pena I, Pilar Manzano M, Cantizani J, Kessler A, Alonso-Padilla J, Bardera AI, et al. New compound sets identified from high throughput phenotypic screening against three kinetoplastid parasites: an open resource. *Sci Rep*. 2015; 5:8771. <https://doi.org/10.1038/srep08771> PMID: 25740547
7. Mott BT, Ferreira RS, Simeonov A, Jadhav A, Ang KK, Leister W, et al. Identification and optimization of inhibitors of Trypanosomal cysteine proteases: cruzain, rhodesain, and TbCatB. *J Med Chem*. 2010; 53(1):52–60. <https://doi.org/10.1021/jm901069a> PMID: 19908842
8. Rogers KE, Keranen H, Durrant JD, Ratnam J, Doak A, Arkin MR, et al. Novel cruzain inhibitors for the treatment of Chagas' disease. *Chem Biol Drug Des*. 2012; 80(3):398–405. <https://doi.org/10.1111/j.1747-0285.2012.01416.x> PMID: 22613098
9. Martinez-Mayorga K, Byler KG, Ramirez-Hernandez AI, Terrazas-Alvares DE. Cruzain inhibitors: efforts made, current leads and a structural outlook of new hits. *Drug Discov Today*. 2015; 20(7):890–8. <https://doi.org/10.1016/j.drudis.2015.02.004> PMID: 25697479
10. Ferreira RS, Simeonov A, Jadhav A, Eidam O, Mott BT, Keiser MJ, et al. Complementarity Between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors. *J Med Chem*. 2010; 53(13):4891–905. <https://doi.org/10.1021/jm100488w> PMID: 20540517
11. Durrant JD, Keranen H, Wilson BA, McCammon JA. Computational identification of uncharacterized cruzain binding sites. *PLoS Negl Trop Dis*. 2010; 4(5):e676. <https://doi.org/10.1371/journal.pntd.0000676> PMID: 20485483
12. Trossini GH, Guido RV, Oliva G, Ferreira EI, Andricopulo AD. Quantitative structure-activity relationships for a series of inhibitors of cruzain from *Trypanosoma cruzi*: molecular modeling, CoMFA and CoMSIA studies. *J Mol Graph Model*. 2009; 28(1):3–11. <https://doi.org/10.1016/j.jmgm.2009.03.001> PMID: 19376735
13. Hotez P, Bundy DAP. The PLOS Neglected Tropical Diseases decade. *PLoS Negl Trop Dis*. 2017; 11(4):e0005479. <https://doi.org/10.1371/journal.pntd.0005479> PMID: 28426658
14. Lu S, Li S, Zhang J. Harnessing allostery: a novel approach to drug discovery. *Med Res Rev*. 2014; 34(6):1242–85. <https://doi.org/10.1002/med.21317> PMID: 24827416
15. Liu J, Nussinov R. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput Biol*. 2016; 12(6):e1004966. <https://doi.org/10.1371/journal.pcbi.1004966> PMID: 27253437
16. Nussinov R. Introduction to Protein Ensembles and Allostery. *Chem Rev*. 2016; 116(11):6263–6. <https://doi.org/10.1021/acs.chemrev.6b00283> PMID: 27268255
17. Dokholyan NV. Controlling Allosteric Networks in Proteins. *Chem Rev*. 2016; 116(11):6463–87. <https://doi.org/10.1021/acs.chemrev.5b00544> PMID: 26894745
18. Monod J, Wyman J, Changeux JP. On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol*. 1965; 12:88–118. PMID: 14343300
19. Koshland DE Jr., Nemethy G, Filmer D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*. 1966; 5(1):365–85. PMID: 5938952
20. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? *Proteins*. 2004; 57(3):433–43. <https://doi.org/10.1002/prot.20232> PMID: 15382234
21. Kumar S, Ma B, Tsai CJ, Wolfson H, Nussinov R. Folding funnels and conformational transitions via hinge-bending motions. *Cell Biochem Biophys*. 1999; 31(2):141–64. <https://doi.org/10.1007/BF02738169> PMID: 10593256
22. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. *Protein Eng*. 1999; 12(9):713–20. PMID: 10506280
23. Tsai CJ, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. *Protein Sci*. 1999; 8(6):1181–90. <https://doi.org/10.1110/ps.8.6.1181> PMID: 10386868
24. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem Rev*. 2016; 116(11):6370–90. <https://doi.org/10.1021/acs.chemrev.5b00631> PMID: 27074285
25. Tsai CJ, Nussinov R. A unified view of "how allostery works". *PLoS Comput Biol*. 2014; 10(2):e1003394. <https://doi.org/10.1371/journal.pcbi.1003394> PMID: 24516370
26. Xiao J, Melvin RL, Salsbury FR. Mechanistic insights into thrombin's switch between "slow" and "fast" forms. *Phys Chem Chem Phys*. 2017; 19(36):24522–33. <https://doi.org/10.1039/c7cp03671j> PMID: 28849814
27. Nussinov R, Tsai CJ. Allostery in disease and in drug discovery. *Cell*. 2013; 153(2):293–305. <https://doi.org/10.1016/j.cell.2013.03.034> PMID: 23582321

28. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem*. 2016; 59(9):4035–61. <https://doi.org/10.1021/acs.jmedchem.5b01684> PMID: 26807648
29. Novinec M, Lenarcic B, Turk B. Cysteine cathepsin activity regulation by glycosaminoglycans. *Biomed Res Int*. 2014; 2014:309718. <https://doi.org/10.1155/2014/309718> PMID: 25587532
30. Costa TF, dos Reis FC, Lima AP. Substrate inhibition and allosteric regulation by heparan sulfate of *Trypanosoma brucei* cathepsin L. *Biochim Biophys Acta*. 2012; 1824(3):493–501. <https://doi.org/10.1016/j.bbapap.2011.12.006> PMID: 22234330
31. Jilkova A, Horn M, Rezacova P, Maresova L, Fajtova P, Brynda J, et al. Activation route of the *Schistosoma mansoni* cathepsin B1 drug target: structural map with a glycosaminoglycan switch. *Structure*. 2014; 22(12):1786–98. <https://doi.org/10.1016/j.str.2014.09.015> PMID: 25456815
32. Judice WA, Manfredi MA, Souza GP, Sansevero TM, Almeida PC, Shida CS, et al. Heparin modulates the endopeptidase activity of *Leishmania mexicana* cysteine protease cathepsin L-Like rCPB2.8. *PLoS One*. 2013; 8(11):e80153. <https://doi.org/10.1371/journal.pone.0080153> PMID: 24278253
33. Marques AF, Esser D, Rosenthal PJ, Kassack MU, Lima LM. Falcipain-2 inhibition by suramin and suramin analogues. *Bioorg Med Chem*. 2013; 21(13):3667–73. <https://doi.org/10.1016/j.bmc.2013.04.047> PMID: 23680445
34. Marques AF, Gomes PS, Oliveira PL, Rosenthal PJ, Pascutti PG, Lima LM. Allosteric regulation of the *Plasmodium falciparum* cysteine protease falcipain-2 by heme. *Arch Biochem Biophys*. 2015; 573:92–9. <https://doi.org/10.1016/j.abb.2015.03.007> PMID: 25791019
35. Bertoldo JB, Chiaradia-Delatorre LD, Mascarello A, Leal PC, Cordeiro MN, Nunes RJ, et al. Synthetic compounds from an in house library as inhibitors of falcipain-2 from *Plasmodium falciparum*. *J Enzyme Inhib Med Chem*. 2015; 30(2):299–307. <https://doi.org/10.3109/14756366.2014.920839> PMID: 24964346
36. Novinec M, Korenc M, Caflich A, Ranganathan R, Lenarcic B, Baici A. A novel allosteric mechanism in the cysteine peptidase cathepsin K discovered by computational methods. *Nat Commun*. 2014; 5:3287. <https://doi.org/10.1038/ncomms4287> PMID: 24518821
37. Novinec M, Lenarcic B, Baici A. Probing the activity modification space of the cysteine peptidase cathepsin K with novel allosteric modifiers. *PLoS One*. 2014; 9(9):e106642. <https://doi.org/10.1371/journal.pone.0106642> PMID: 25184245
38. Novinec M, Rebernik M, Lenarcic B. An allosteric site enables fine-tuning of cathepsin K by diverse effectors. *FEBS Lett*. 2016; 590(24):4507–18. <https://doi.org/10.1002/1873-3468.12495> PMID: 27859061
39. Almeida PC, Nantes IL, Rizzi CC, Judice WA, Chagas JR, Juliano L, et al. Cysteine proteinase activity regulation. A possible role of heparin and heparin-like glycosaminoglycans. *J Biol Chem*. 1999; 274(43):30433–8. PMID: 10521421
40. Almeida PC, Nantes IL, Chagas JR, Rizzi CC, Faljoni-Alario A, Carmona E, et al. Cathepsin B activity regulation. Heparin-like glycosaminoglycans protect human cathepsin B from alkaline pH-induced inactivation. *J Biol Chem*. 2001; 276(2):944–51. <https://doi.org/10.1074/jbc.M003820200> PMID: 11016923
41. Li Z, Yasuda Y, Li W, Bogyo M, Katz N, Gordon RE, et al. Regulation of collagenase activities of human cathepsins by glycosaminoglycans. *J Biol Chem*. 2004; 279(7):5470–9. <https://doi.org/10.1074/jbc.M310349200> PMID: 14645229
42. Lima AP, Almeida PC, Tersariol IL, Schmitz V, Schmaier AH, Juliano L, et al. Heparan sulfate modulates kinin release by *Trypanosoma cruzi* through the activity of cruzipain. *J Biol Chem*. 2002; 277(8):5875–81. <https://doi.org/10.1074/jbc.M108518200> PMID: 11726662
43. Costa MG, Batista PR, Shida CS, Robert CH, Bisch PM, Pascutti PG. How does heparin prevent the pH inactivation of cathepsin B? Allosteric mechanism elucidated by docking and molecular dynamics. *BMC Genomics*. 2010; 11 Suppl 5:S5.
44. Verkhivker GM, Dixit A, Morra G, Colombo G. Structural and computational biology of the molecular chaperone Hsp90: from understanding molecular mechanisms to computer-based inhibitor design. *Curr Top Med Chem*. 2009; 9(15):1369–85. PMID: 19860735
45. Rastelli G, Anighoro A, Chripkova M, Carrassa L, Brogginini M. Structure-based discovery of the first allosteric inhibitors of cyclin-dependent kinase 2. *Cell Cycle*. 2014; 13(14):2296–305. <https://doi.org/10.4161/cc.29295> PMID: 24911186
46. Dalafave DS, Dalafave RE. Computational Design of Allosteric Inhibitors of AKT and SGK Kinases. *Biophys J*. 108(2):320a.

47. Papaleo E, Saladino G, Lambrugh M, Lindorff-Larsen K, Gervasio FL, Nussinov R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem Rev.* 2016; 116(11):6391–423. <https://doi.org/10.1021/acs.chemrev.5b00623> PMID: 26889708
48. Ribeiro AAST, Ortiz V. A Chemical Perspective on Allostery. *Chem Rev.* 2016; 116(11):6488–502. <https://doi.org/10.1021/acs.chemrev.5b00543> PMID: 26741913
49. Hertig S, Latorraca NR, Dror RO. Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations. *PLoS Comput Biol.* 2016; 12(6):e1004746. <https://doi.org/10.1371/journal.pcbi.1004746> PMID: 27285999
50. Stanley N, De Fabritiis G. High throughput molecular dynamics for drug discovery. *In Silico Pharmacol.* 2015; 3:3. <https://doi.org/10.1186/s40203-015-0007-0> PMID: 25717426
51. Zhao H, Caffisch A. Molecular dynamics in drug design. *Eur J Med Chem.* 2015; 91:4–14. <https://doi.org/10.1016/j.ejmech.2014.08.004> PMID: 25108504
52. Hernandez-Rodriguez M, Rosales-Hernandez MC, Mendieta-Wejebe JE, Martinez-Archundia M, Basurto JC. Current tools and methods in Molecular Dynamics (MD) simulations for drug design. *Curr Med Chem.* 2016.
53. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A.* 2009; 106(16):6620–5. <https://doi.org/10.1073/pnas.0810961106> PMID: 19351898
54. Chemical Computing Group ULC. Molecular Operating Environment (MOE). 1010 Sherbooke St West, Suite #910, Montreal, QC, Canada, H3A 2R7. 2017.
55. Huey R, Morris GM. Using AutoDock 4 with AutoDocktools: a tutorial. The Scripps Research Institute, USA. 2008:54–6.
56. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010; 31(2):455–61. <https://doi.org/10.1002/jcc.21334> PMID: 19499576
57. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem.* 2000; 21(2):132–46.
58. Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc.* 1992; 114(25):10024–35.
59. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09. Gaussian Inc. 2009; Wallingford CT.
60. Besler BH, Merz KM, Kollman PA. Atomic charges derived from semiempirical methods *J Comput Chem.* 1990; 11:431–9.
61. Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, et al. Amber 14 2014.
62. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem.* 2004; 25(9):1157–74. <https://doi.org/10.1002/jcc.20035> PMID: 15116359
63. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000; 28(1):235–42. PMID: 10592235
64. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* 2004; 32(suppl 2):W665–W7.
65. Jorgensen WL, Jenson C. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *J Comput Chem.* 1998; 19(10):1179–86.
66. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput.* 2015; 11(8):3696–713. <https://doi.org/10.1021/acs.jctc.5b00255> PMID: 26574453
67. DeLano W. Use of PyMOL as a communications tool for molecular science. *Abst Pap Am Chem Soc.* 2004; 228:U313–U4.
68. Nery ED, Juliano MA, Meldal M, Svendsen I, Scharfstein J, Walmsley A, et al. Characterization of the substrate specificity of the major cysteine protease (cruzain) from *Trypanosoma cruzi* using a portion-mixing combinatorial library and fluorogenic peptides. *Biochem J.* 1997; 323 (Pt 2):427–33.
69. Judice WA, Cezari MH, Lima AP, Scharfstein J, Chagas JR, Tersariol IL, et al. Comparison of the specificity, stability and individual rate constants with respective activation parameters for the peptidase activity of cruzain and its recombinant form, cruzain, from *Trypanosoma cruzi*. *Eur J Biochem.* 2001; 268(24):6578–86. PMID: 11737212
70. Berendsen HJ, Postma Jv, van Gunsteren WF, DiNola A, Haak J. Molecular dynamics with coupling to an external bath. *J Chem Phys.* 1984; 81(8):3684–90.

71. Schneider T, Stoll E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys Rev B*. 1978; 17(3):1302.
72. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys*. 1993; 98(12):10089–92.
73. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys*. 1977; 23(3):327–41.
74. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput*. 2013; 9(9):3878–88. <https://doi.org/10.1021/ct400314y> PMID: 26592383
75. Stank A, Kokh DB, Fuller JC, Wade RC. Protein Binding Pocket Dynamics. *Acc Chem Res*. 2016; 49(5):809–15. <https://doi.org/10.1021/acs.accounts.5b00516> PMID: 27110726
76. Durrant JD, de Oliveira CA, McCammon JA. POVME: an algorithm for measuring binding-pocket volumes. *J Mol Graph Model*. 2011; 29(5):773–6. <https://doi.org/10.1016/j.jmgm.2010.10.007> PMID: 21147010
77. Swami A, Sadler B, editors. Modulation classification via hierarchical agglomerative cluster analysis. *Signal Processing Advances in Wireless Communications, First IEEE Signal Processing Workshop on*; 1997: IEEE.
78. Roe DR, Cheatham TE, 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput*. 2013; 9(7):3084–95. <https://doi.org/10.1021/ct400341p> PMID: 26583988
79. Bashford D, Case DA. Generalized born models of macromolecular solvation effects. *Annu Rev Phys Chem*. 2000; 51:129–52. <https://doi.org/10.1146/annurev.physchem.51.1.129> PMID: 11031278
80. Homeyer N, Gohlke H. Free energy calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area method. *Mol Inform*. 2012; 31(2):114–22. <https://doi.org/10.1002/minf.201100135> PMID: 27476956
81. Kleinjung J, Fraternali F. Design and application of implicit solvent models in biomolecular simulations. *Curr Opin Struct Biol*. 2014; 25:126–34. <https://doi.org/10.1016/j.sbi.2014.04.003> PMID: 24841242
82. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model*. 2011; 51(1):69–82. <https://doi.org/10.1021/ci100275a> PMID: 21117705
83. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput*. 2012; 8(9):3314–21. <https://doi.org/10.1021/ct300418h> PMID: 26605738
84. Zoete V, Michielin O. Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-pMHC complex. *Proteins*. 2007; 67(4):1026–47. <https://doi.org/10.1002/prot.21395> PMID: 17377991
85. Gohlke H, Kiel C, Case DA. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J Mol Biol*. 2003; 330(4):891–913. PMID: 12850155
86. Humphris EL, Kortemme T. Prediction of protein-protein interface sequence diversity using flexible backbone computational protein design. *Structure*. 2008; 16(12):1777–88. <https://doi.org/10.1016/j.str.2008.09.012> PMID: 19081054
87. Morra G, Potestio R, Micheletti C, Colombo G. Corresponding functional dynamics across the Hsp90 Chaperone family: insights from a multiscale analysis of MD simulations. *PLoS Comput Biol*. 2012; 8(3):e1002433. <https://doi.org/10.1371/journal.pcbi.1002433> PMID: 22457611
88. Vettoretti G, Moroni E, Sattin S, Tao J, Agard DA, Bernardi A, et al. Molecular Dynamics Simulations Reveal the Mechanisms of Allosteric Activation of Hsp90 by Designed Ligands. *Sci Rep*. 2016; 6:23830. <https://doi.org/10.1038/srep23830> PMID: 27032695
89. Lange OF, Grubmuller H. Generalized correlation for biomolecular dynamics. *Proteins*. 2006; 62(4):1053–61. <https://doi.org/10.1002/prot.20784> PMID: 16355416
90. Skjaerven L, Yao XQ, Scarabelli G, Grant BJ. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics*. 2014; 15:399. <https://doi.org/10.1186/s12859-014-0399-6> PMID: 25491031
91. Rivalta I, Sultan MM, Lee NS, Manley GA, Loria JP, Batista VS. Allosteric pathways in imidazole glycerol phosphate synthase. *Proc Natl Acad Sci U S A*. 2012; 109(22):E1428–36. <https://doi.org/10.1073/pnas.1120536109> PMID: 22586084
92. Girvan M, Newman ME. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*. 2002; 99(12):7821–6. <https://doi.org/10.1073/pnas.122653799> PMID: 12060727

93. Dijkstra EW. A note on two problems in connexion with graphs. *Numer Math.* 1959; 1(1):269–71.
94. Van Wart AT, Durrant J, Votapka L, Amaro RE. Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *J Chem Theory Comput.* 2014; 10(2):511–7. <https://doi.org/10.1021/ct4008603> PMID: 24803851
95. Vijayabaskar MS, Vishveshwara S. Interaction energy based protein structure networks. *Biophys J.* 2010; 99(11):3704–15. <https://doi.org/10.1016/j.bpj.2010.08.079> PMID: 21112295
96. Bhattacharyya M, Vishveshwara S. Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism. *Biochemistry.* 2011; 50(28):6225–36. <https://doi.org/10.1021/bi200306u> PMID: 21650159
97. Morra G, Genoni A, Colombo G. Mechanisms of Differential Allosteric Modulation in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains. *J Chem Theory Comput.* 2014; 10(12):5677–89. <https://doi.org/10.1021/ct500326g> PMID: 26583250
98. Ribeiro AA, Ortiz V. Energy propagation and network energetic coupling in proteins. *J Phys Chem B.* 2015; 119(5):1835–46. <https://doi.org/10.1021/jp509906m> PMID: 25569787
99. Kumawat A, Chakrabarty S. Hidden electrostatic basis of dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A.* 2017; 114(29):E5825–E34. <https://doi.org/10.1073/pnas.1705311114> PMID: 28634294
100. Sercinoglu O, Ozbek P. gRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations. *Nucleic Acids Res.* 2018.
101. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems.* 2006; 1695(5):1–9.
102. Arafet K, Ferrer S, Moliner V. Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease. *ACS Catalysis.* 2017; 7(2):1207–15.
103. Alber T, Gilbert WA, Ponzi DR, Petsko GA. The role of mobility in the substrate binding and catalytic machinery of enzymes. *Ciba Found Symp.* 1983; 93:4–24. PMID: 6551232
104. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol.* 2009; 5(11):789–96. <https://doi.org/10.1038/nchembio.232> PMID: 19841628
105. Greener JG, Sternberg MJ. Structure-based prediction of protein allostery. *Curr Opin Struct Biol.* 2017; 50:1–8. <https://doi.org/10.1016/j.sbi.2017.10.002> PMID: 29080471
106. Guo J, Zhou HX. Protein Allostery and Conformational Dynamics. *Chem Rev.* 2016; 116(11):6503–15. <https://doi.org/10.1021/acs.chemrev.5b00590> PMID: 26876046
107. Tsai CJ, del Sol A, Nussinov R. Allostery: absence of a change in shape does not imply that allostery is not at play. *J Mol Biol.* 2008; 378(1):1–11. <https://doi.org/10.1016/j.jmb.2008.02.034> PMID: 18353365
108. Nussinov R, Tsai CJ. Allostery without a conformational change? Revisiting the paradigm. *Curr Opin Struct Biol.* 2015; 30:17–24. <https://doi.org/10.1016/j.sbi.2014.11.005> PMID: 25500675
109. Liu J, Nussinov R. Energetic redistribution in allostery to execute protein function. *Proc Natl Acad Sci U S A.* 2017; 114(29):7480–2. <https://doi.org/10.1073/pnas.1709071114> PMID: 28696318
110. Boulton S, Melacini G. Advances in NMR Methods To Map Allosteric Sites: From Models to Translation. *Chem Rev.* 2016; 116(11):6267–304. <https://doi.org/10.1021/acs.chemrev.5b00718> PMID: 27111288
111. Grutsch S, Bruschweiler S, Tollinger M. NMR Methods to Study Dynamic Allostery. *PLoS Comput Biol.* 2016; 12(3):e1004620. <https://doi.org/10.1371/journal.pcbi.1004620> PMID: 26964042
112. Manley G, Loria JP. NMR insights into protein allostery. *Arch Biochem Biophys.* 2012; 519(2):223–31. <https://doi.org/10.1016/j.abb.2011.10.023> PMID: 22198279
113. Muchmore SW, Hajduk PJ. Crystallography, NMR and virtual screening: integrated tools for drug discovery. *Curr Opin Drug Discov Devel.* 2003; 6(4):544–9. PMID: 12951817
114. Naganathan AN. Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Curr Opin Struct Biol.* 2018; 54:1–9. <https://doi.org/10.1016/j.sbi.2018.09.004> PMID: 30268910
115. Weinkam P, Chen YC, Pons J, Sali A. Impact of mutations on the allosteric conformational equilibrium. *J Mol Biol.* 2013; 425(3):647–61. <https://doi.org/10.1016/j.jmb.2012.11.041> PMID: 23228330
116. Tang Q, Fenton AW. Whole-protein alanine-scanning mutagenesis of allostery: A large percentage of a protein can contribute to mechanism. *Hum Mutat.* 2017; 38(9):1132–43. <https://doi.org/10.1002/humu.23231> PMID: 28407397
117. Chopra N, Wales TE, Joseph RE, Boyken SE, Engen JR, Jernigan RL, et al. Dynamic Allostery Mediated by a Conserved Tryptophan in the Tec Family Kinases. *PLoS Comput Biol.* 2016; 12(3):e1004826. <https://doi.org/10.1371/journal.pcbi.1004826> PMID: 27010561