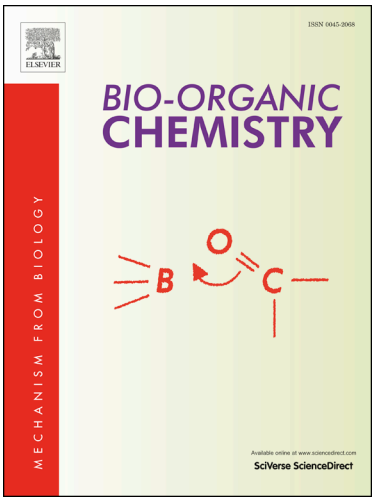


Leveraging the Cruzain S3 Subsite to Increase Affinity for Reversible Covalent Inhibitors

Lorenzo Cianni, Geraldo Sartori, Fabiana Rosini, Daniela De Vita, Gabriel Pires, Bianca Rebelo Lopes, Andrei Leitão, Antonio C.B. Burtoloso, Carlos A. Montanari

PII: S0045-2068(17)30781-2
DOI: https://doi.org/10.1016/j.bioorg.2018.04.006
Reference: YBIOO 2329

To appear in: *Bioorganic Chemistry*

Received Date: 13 October 2017
Revised Date: 12 April 2018
Accepted Date: 12 April 2018

Please cite this article as: L. Cianni, G. Sartori, F. Rosini, D. De Vita, G. Pires, B. Rebelo Lopes, A. Leitão, A.C.B. Burtoloso, C.A. Montanari, Leveraging the Cruzain S3 Subsite to Increase Affinity for Reversible Covalent Inhibitors, *Bioorganic Chemistry* (2018), doi: https://doi.org/10.1016/j.bioorg.2018.04.006

Leveraging the Cruzain S3 Subsite to Increase Affinity for Reversible Covalent Inhibitors

Lorenzo Cianni[†*], Geraldo Sartori[†], Fabiana Rosini[†], Daniela De Vita[†], Gabriel Pires, Bianca Rebelo Lopes[†], Andrei Leitão[†*], Antonio C. B. Burtoloso[††], Carlos A. Montanari[†]

[†]*Grupo de Química Medicinal do IQSC/USP, Instituto de Química de São Carlos, Universidade de São Paulo, Av. Trab. Sancarlense, 400 - Parque Arnold Schimidt, São Carlos - SP, 13566-590, São Paulo, Brazil*

[††]*Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brasil*

*corresponding authors: andleitao@iqsc.usp.br (Andrei Leitão). +55 (16) 3373-9943 / +55 (16) 3373-9985. lorenzo.cianni@usp.br (Lorenzo Cianni). +55 (16) 3373-8060

Leveraging the Cruzain S3 Subsite to Increase Affinity for Reversible Covalent Inhibitors

Cruzain is the major cysteine protease of *Trypanosoma cruzi*, the etiological agent of Chagas disease. Reversible covalent cruzain inhibitors can block the steps of cell differentiation in the parasite and kill the organism. To this end, the description of how inhibitors modified at the P2/P3 positions lead to analogs with greater cruzain affinity to the S2/S3 subsites is of fundamental importance. Albeit many efforts are being employed in the characterization of the interaction processes with S2 subsite, little is known about the cruzain S3 subsite. In this work, we show a brief but consistent study to identify favorable substitutions in P3 of dipeptidyl nitriles that increase cruzain affinity. Using molecular dynamics simulations, we have identified some dipeptidyl nitrile analogs with modifications at P3 position that had higher cruzain inhibition than the original unsubstituted compound. A matched molecular pair analysis shows the importance of including a chlorine atom in the P3-*meta* position. The modifications implemented in P3 are confirmed when profiling the thermodynamic parameters via isothermal titration calorimetry. The classical enthalpy-entropy compensation phenomenon, in which enthalpy changes are counterbalanced by entropy results in a small modification of ΔG. The inclusion of the chlorine atom in the P3-*meta* position results in the highest reduction of the detrimental entropic contribution observed in P3.

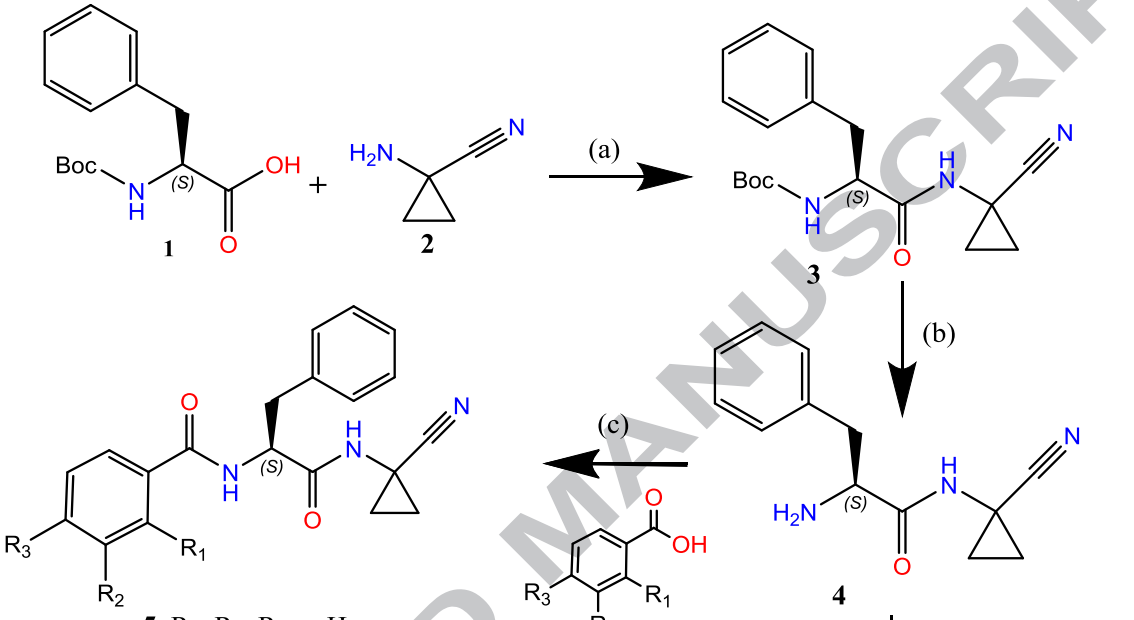
Keywords: Cruzain inhibitors; Dipeptidyl nitriles; Molecular Dynamics Simulation; Matched Molecular Pair Analysis; Isothermal Titration Calorimetry

1. Introduction

Chagas disease, whose etiological agent is the protozoan parasite *Trypanosoma cruzi*, is a serious health and social problem for people living in Latin America and areas previously considered non-endemic such as Japan, East Europe and North America [1]. More than 300,000 new cases are reported every year in 21 countries around the world with an average of one million people currently infected with *T. cruzi* [2-4]. At present, there is no vaccine available to prevent Chagas disease or effective chemotherapies. The only two existing drugs in the market, Benznidazole and Nifurtimox, show high side toxic effects and inefficiency in the chronic stage of the disease [5,6]. Beyond doubt, new drugs that are safe and efficacious are therefore critically needed. One approach consists in the discovery and development of cruzain (Cz) inhibitors, the major *T. cruzi* cysteine protease, responsible for the survival and propagation of the protozoan parasite [7]. Consequently, its inhibition is an important strategy for the treatment of this disease. Three-dimensional (3D) structures of the enzyme with a variety of ligands have already been resolved [8], thereby prompting us to apply target based molecular designing with the aim of finding new high-affinity Cz inhibitors. In particular, our recent studies [9] showed key interactions of P1, P2 and P3 portions of dipeptidyl nitrile ligands with the respective subsites of the enzyme (S1, S2, S3). Specifically, it is already known the importance of Cz inhibition via hydrogen bond between the ligand and residues Asp16 and Gly66 located at S2; and His162 in S1 subsite [10].

The importance of the S3 region still needs to be better understood due to its open-flat water accessible surface area. However, relevant data and information are available for the hydrogen bonding interaction of the Ser61 residue with the corresponding P3 region of different inhibitors co-crystallized with Cz (for instance, see

PDB code 1ME3). Herein, we report the molecular design and synthesis of a new series of high Cz affinity dipeptidyl nitrile inhibitors, based on the structure of substance **5** (Scheme 1), to study and evaluate the regioselective interaction with residue Ser61 via *in silico* studies, kinetics and thermodynamic data.



Scheme 1. Synthetic route for dipeptidyl-nitrile derivatives. Conditions for the reaction steps: (a) HATU, DIPEA, DMF, Argon atmosphere, 16 h, RT; (b) HCOOH, 16 h, RT; (c, d) HATU, DIPEA, DMF, Argon atmosphere, 16 h, RT.

2. Methods

2.1 *Molecular Dynamics Simulation*

All ligands in this work were considered covalently bonded to the catalytic Cys25 residue from Cz. Initially, we generated the modified cysteine amino acid cross-linked with the inhibitor and capped at N and C terminal with ACE (acetyl) and NME (methylamine) groups, respectively. The formal charge of the compound was predicted by LigPrep [11] considering a pH of 5.8. Electrostatic potential of each modified cysteine was calculated at HF/6-31G+ level without geometry optimization in Gaussian09 to a subsequent RESP [12] partial atomic charge derivation using RED-vIII.4 [13]. In this step, we attributed the partial charge of main chain N, H, C and O atoms as -0.4157, 0.2719, 0.5973 and -0.5679, respectively (as per standard AMBER amino acids) and defined total charge of ACE and NME groups as zero. Then, the library and topological parameters of each nonstandard cysteine were generated using Antechamber package implemented in AmberTools14 considering FF14SB [14] and gaff (general amber force field) [15] to treat the original cysteine residue and ligand atoms, respectively [16]. The missing parameters at the interface of gaff-FF14SB atoms were estimated based on comparable connectivity in gaff force field.

Cruzain structure was retrieved from Protein Data Bank (PDB code 1ME3) [17]. All water molecules were removed, histidine residues were renamed according to their protonation and tautomeric states, and Asp60 side chain was considered in its protonated state. Cross-linked adducts prepared in the previous section replaced the standard Cys25 amino acid. Complexes were parametrized with FF14SB force field to amino acid atoms and gaff force field for ligand atoms in tleap. Sodium atoms were added to neutralize the system, and a truncated octahedral box of TIP3P water molecules [18] was created extending at least 10 Å beyond any protein atom.

Molecular dynamics simulations were carried out using pmemd.cuda script from Amber14 package. All simulation steps were done using SHAKE method[19] to constrain bonds containing hydrogen atoms and Langevin approach to control temperature (frequency of 2 s^{-1}). A cutoff of 10 Å was defined for nonbonded interaction considering a timestep of 2 fs. After, we performed the equilibration protocol consisting of (i) minimization with positional constraint of 500 kcal.mol^{-1} for solute atoms; (ii) minimization without constraints; (iii) system heating from 0 to 310 K in NVT ensemble and constraint of 10 kcal.mol^{-1} for solute atoms during 50 ps; (iv) 5 ns equilibration at NPT ensemble. Then, the whole system was simulated in the NPT ensemble using three replicates of 100 ns production simulations that were subdivided into 5 ns sequential simulations. Each simulation is restarted with a new seed for Langevin simulation and coordinates were saved every 50 ps.

Data regarding the RMSD of the ligand and the distance between the ligand and protein were calculated for production simulations using the cpptraj as implemented within the AmberTools14 package. Graphs were generated using Origin software to perform the analyses.

2.2 Chemistry

All chemicals were purchased as reagent grade and used without further purification unless otherwise noted. *N*,*N*-Dimethylformamide (DMF) was dried over 3 Å activated molecular sieves for 72 hours. All non-aqueous reactions were carried out under argon atmosphere in oven-dried glassware. Solvents used in high-performance chromatography (HPLC) were supplied by Tedia and used without further purification

Thin layer chromatography was performed on Fluka Analytical Sigma-Aldrich silica gel matrix, pre-coated plates with fluorescent indicator 254 nm and/or staining

solutions. Flash column chromatography was performed on silica gel (pore size 60 Å, 70-230 mesh).

^{1}H and ^{13}C NMR spectra were recorded on HP – 400 and 500 MHz instruments in $CDCl_3$ or DMSO-d_6. Chemical shifts are referenced to the residual solvent peak, and J values are given in Hz. The following multiplicity abbreviations are used: (s) singlet, (d) doublet, (dd) doublet of doublets, (ddd) doublet of doublet of doublets, (dt) doublet of triplet, (t) triplet, (q) quartet, (m) multiplet, and (br) broad.

Characterization and separation of substances were carried out with a HPLC system. The analytical HPLC system consisted of a Shimadzu LC (Kyoto, Japan) equipped with a LC-20AT pump, a LC-20AD pump, a SIL-20A HT autosampler, a DGU-20A5 degasser, a CBM-20A, SPD-M20A DAD detector and a FRC-10A fraction collector. Data acquisition was performed using LCsolution software version 1.26 SP5. The LC system was coupled to an AmaZon SL ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with electrospray ionization (ESI) interface. Data acquisition was performed with Bruker Daltonics Data Analysis software (version 4.2.383.1). Spectra for all compounds and further details can be found in the supplementary information.

Solvents were filtered through a 0.45 μm Merck-Millipore membrane before use and degassed in an ultrasonic bath. In the established HPLC protocol, chiral analysis and separation were carried out at 32 °C (column oven) where not otherwise specified, using analytical and semi-preparative cellulose-2 Phenomenex column (Analytical: 5 μm, 250 mm x 4.6 mm I.D, semi-preparative: 5 μm, 250 mm x 10 mm I.D) or Diacel column (IC-chiralpak: 5 μm, 250 mm x 4.6 mm) via isocratic elution with a flow rate of 0.5 (analytical) and 2.36 mL min^{-1} (semi-preparative). The most common mobile phase composition was acetonitrile-water (50:50) (v/v). Volumes of 10 μL (analytical) and

1000 μL (semi-preparative) were injected. Quantification was carried out at 200–800 nm and the chromatographic run time varied according to the sample.

Specific rotations ($[\alpha]^T = 100\alpha/lc$, in deg mL g^{-1} dm^{-1}, but reported herein in degrees) were observed at the wavelength 589 nm, the D line of a sodium lamp. The temperature was set to be 25 °C. Samples were weighted using a precision balance (Sartorius, Model CPA26P) and dissolved in methanol (HPLC grade, Panreac). Rotations were measured using a Digital Polarimeter (P2000, Jasco): α = observed rotation in degrees; l = cell path length of 0.1 decimeter; c = concentration in g 100 mL^{-1}. Values were calculated using 5 measurements for each substance. Melting points were determined by a Quimica Micro MQAPF-302 apparatus and are uncorrected.

The synthetic route was developed to optimize the set of substituents to be placed in P1, P2, and P3 that have been defined after the planning and design studies. All syntheses were started with enantiopure amino acids. Due to the diversity of building blocks, it was necessary to evaluate different coupling reagents, aiming at the best yield and lower racemization [9].

The peptide coupling reaction was performed twice; first to bind the amino acid (enantiomeric pure, amine-Boc-protected) to the nitrile moiety, and secondly (after removing Boc group) to the wanted benzoyl acids (Scheme 1).

*Synthesis of (S)- tert-butyl (1-((cyanomethyl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**3**)*

N, N-diisopropylethylamine (DIPEA, 2.6 Eq) was added to a solution of Boc-L-phenylalanine (**1**) (3.77 mmol, 1 Eq), HATU (1.3 Eq) and 1-Amino-1-cyclopropanecarbonitrile hydrochloride (**2**) (1.3 Eq.) in DMF (5 mL), under argon atmosphere. The resulting solution was stirred at room temperature for 16 hours. The reaction mixture was diluted with ethyl acetate (10 mL) and washed with a saturated

$NaHCO_3$ solution (3 x 20 mL) and brine (3 x 20 mL). The organic phase was dried over $MgSO_4$ and evaporated to give a crude residue that was purified by chromatographic column on Flash Silica using ethyl acetate: *n*-hexane (7:3) as the mobile phase to give a white solid with a good yield.

Yield 92%. White solid. R_f= 0.9 (Ethyl Acetate: *n*-Hexane; 7:3). MP. 146-147 °C. ^{1}H NMR (500 MHz,$CDCl_3$) δ 7.27- 7.04 (m, 5H), 4.58 (m, 1H), 3.19 (dd, *J* =13.8, 5.0 Hz, 1H), 2.88 (dd, *J* = 9.5, 5.0 Hz, 1H), 1.31 (s, 9H), 1.25 (m, 2H), 1.04 (m, 2H).

^{13}C NMR. (125 MHz, $CDCl_3$) δ 142.88, 131.04, 127.69, 126.21, 120.81, 107.59, 37.36.

*Synthesis of (S)-2-amino-N-(1-cyanocyclopropyl)-3-phenylpropanamide (**4**)*

Substance **3** (500 mg) was treated with formic acid (2 mL) at room temperature and stirred for 18 h. The reaction mixture was evaporated under vacuum to get yellowish oil. NaOH 1 M was added dropwise to reach pH 9. The product was extracted with ethyl acetate (4 times) and then washed twice with brine. Then, the organic phase was evaporated to obtain the colorless oil. The product formation was confirmed by TLC (ethyl acetate) with ~ 90% yield, and the product was used for the next step without any further purification.

*General procedure for synthesis of substances: **5 -15***

Substance (**4**) (1.3 Eq) was added to a solution with the respective benzoic acid derivative (1 mmol, 1 Eq), HATU (1.3 Eq) and *N*,*N*-diisopropylethylamine (DIPEA, 2.6 Eq) in DMF (5 mL), under argon atmosphere. The resulting solution was stirred at room temperature for 16 h. This reaction mixture was diluted with ethyl acetate (10 mL) and washed with a saturated solution of $NaHCO_3$ (3 x 20 mL) and brine (3 x 20 mL). The organic phase was dried over $MgSO_4$ and evaporated to give a crude residue that was

purified by chromatographic column on Flash Silica using as mobile phase ethyl acetate: hexane (7:3) to give a white solid with 65-80% yield.

*(2S)-N-(1-cyanocyclopropyl)-3-phenyl-2-(phenylformamido)propanamide (**5**)*

70% yield. White solid. MP 189-190°C. NMR ^{1}H (500 MHz, DMSO-d_6) δ 9.02 (s, 1H), 8.64 (d, *J* = 8.1 Hz, 1H), 7.82 (d, *J* = 7.2 Hz, 2H), 7.52 (t, *J* = 7.3 Hz, 1H), 7.44 (t, *J* = 7.5 Hz, 2H), 7.30 (d, *J* = 7.2 Hz, 2H), 7.25 (t, *J* = 7.5 Hz, 2H), 7.17 (t, *J* = 7.2 Hz, 1H), 4.60 (m, 1H), 3.03 (dd, *J* = 23.4, 13.6, 7.6 Hz, 2H), 1.46 (d, *J* = 2.7 Hz, 2H), 1.04 (d, *J* = 9.7 Hz, 2H). NMR ^{13}C (125 MHz, DMSO-d_6) δ 173.18, 166.71, 138.29, 134.26, 131.81, 129.61, 128.61, 128.55, 127.93, 126.80, 121.18, 55.05, 37.39, 20.19, 16.15, 16.13. HRMS (ESI (+)) Calcd. [$C_{20}H_{20}N_3O_2$] 334.1550, Observed: 334.1550 [M+H]. HPLC (50:50 ACN:Water): t_R(min)= 12.01. Purity: 97.73 %.

*(S)-2-chloro-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)benzamide (**6**)*

80% yield. White solid. MP 175- 176 °C $[\alpha]_D^5$ = -45.3° ± 0.5 (CH_3OH, C=0.1). FT-IR (KBr, cm^{-1}) 3300, 3284, 3030, 2247, 1663, 1650, 1525, 1502, 1329, 1053, 744, 698. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.95 (s, 1H), 8.71 (d, *J*= 10.0 Hz, 1H), 7.42 (m, 2H), 7.35 (td, *J*= 9.5, 3.0 Hz, 1H), 7.28-7.17 (m, 5H), 4.57 (m, 1H), 3.06 (dd, *J*= 13.5, 5.0 Hz, 1H), 2.98 (dd, *J*= 19.0, 13.5 Hz, 1H), 1.44 (m, 2H), 1.02 (m, 2H).^{13}C NMR (125 MHz, DMSO-d_6) δ 172.21, 166.27, 137.53, 136.38, 131.05, 130.21, 129.75, 129.42, 129.17, 128.28, 127.07, 126.62, 120.81, 54.30, 39.52, 37.22, 19.89, 15.89. ESI-MS (-) Calc. for [$C_{20}H_{18}ClN_3O_2$] 367.83, found: 366.25 [M-H^+]$^-$ .HPLC (50:50 ACN:Water): t_R(min)= 12.01. Purity: 96.39 %.

*((S)-3-chloro-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)benzamide (**7**)*

75% yield. White solid. MP 214-215 °C $[\alpha]_D^5$ = -47.3 ° ± 0.2 (CH_3OH, C=0.1). FT-IR (KBr, cm^{-1}) 3302, 2924, 2243, 1672, 1649, 1537, 1525, 1319, 1076, 750, 696, 686. ^1H NMR (500 MHz, DMSO-d_6) δ 9.03 (s, 1H), 8.83 (d, *J*= 10.0 Hz, 1H), 7.87 (t, *J*= 2 Hz, 1H), 7.77 (dt, *J*= 7.5, 1.5 Hz, 1H), 7.59 (ddd, *J*= 8.0, 2.0, 1.0 Hz, 1H), 7.48 (t, *J*= 8 Hz, 1H), 7.29-7.23 (m, 4H), 7.17 (td, *J*= 7.5 , 1.5 Hz, 1H), 4.59 (m, 1H), 3.06 (dd, *J*= 13.5, 5.0 Hz, 1H), 2.98 (dd, *J*= 19.0, 13.5 Hz, 1H), 1.24 (m, 2H), 1.03 (m, 2H).^{13}C NMR (125MHz, DMSO-d_6) δ 172.95, 165.31, 138.21, 136.21, 133.49, 131.66, 130.68, 129.57, 128.56, 127.73, 126.84, 126.71, 121.16, 55.12, 37.32, 20.20, 16.14, 16.12. ESI-MS (-) Calc. for [$C_{20}H_{18}ClN_3O_2$] 367.83, found: 366.25 [M-H^+]$^-$. HPLC (50:50 ACN:Water): t_R (min)= 14.7. Purity: 94.57 %.

*(S)-4-chloro-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)benzamide (**8**)*

75% yield. White solid. MP 218 -219 °C. $[\alpha]_D^5$ = -48.0° ± 1 (CH_3OH, C=0.1). FT-IR (KBr, cm^{-1}) 3300, 3284, 3030, 2247, 1663, 1649, 1525, 1502, 1329, 1053, 744, 698. ^1H NMR (500 MHz, DMSO-d_6) δ 9.03 (s, 1H), 8.78 (d, *J*= 10.0 Hz, 1H), 7.83 (dt, *J* = 10.5, 2.5 Hz, 2H), 7.52 (dt, *J*= 10.5, 2.5 Hz, 2H), 7.28-7.22 (m, 4H), 7.16 (tt, *J* = 7, 1.5 Hz, 1H), 4.58 (m, 1H), 3.05 (dd, *J*= 13.5, 5.0 Hz, 1H), 2.96 (dd, *J*= 19.0, 13.5 Hz, 1H), 1.45 (m, 2H), 1.01 (m, 2H). ^{13}C NMR (125 MHz, DMSO-d_6) δ 172.74, 165.38, 137.92, 136.35, 132.69, 129.58, 129.29, 128.43, 128.26, 126.53, 120.87, 54.81, 39.52, 37.06, 19.89, 15.85, 15.82. ESI-MS (-) Calc. for [$C_{20}H_{18}ClN_3O_2$] 367.83, found: 402.22 [M+Cl^-]$^-$. HPLC (50:50 ACN:Water): t_R (min)= 16.11. Purity: 99.99 %.

*(S)-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)-3-methylbenzamide (**9**)*

73% yield. White solid. MP 140 – 141 °C $[\alpha]_D^5$ = -51.3° ± 0.7 (CH_3OH, C=0.1). FT-IR (KBr, cm^{-1}) 3303, 3285, 3030, 2247, 1663, 1650, 1525, 1502, 1329, 1053, 744, 698. ^1H NMR (500 MHz, DMSO-d_6) δ 8.87 (s, 1H), 8.38 (d, *J*= 10.0 Hz, 1H), 7.61 (m, 2H), 7.29 – 7.12 (m, 7H), 4.62 (m, 1H), 3.06 (dd, *J*= 13.5, 5.0 Hz, 1H), 2.99 (dd, *J*= 19.0, 13.5 Hz, 1H), 2.33 (s, 3H) 1.37 (m, 2H), 1.04 (m, 2H). ^{13}C NMR (125 MHz, DMSO-d_6) δ 173.00, 166.93, 137.90, 137.67, 134.24, 133.47, 132.09, 130.08, 129.49, 128.38, 128.31, 126.78, 124.92, 120.71, 79.34, 78.68, 54.77, 37.56, 21.36, 20.13, 16.13. ESI-MS (-) Calc. for [$C_{21}H_{21}N_3O_2$] 347.42, found: 346.31 [$M-H^+$]$^-$. HPLC (50:50 ACN:Water): t_R (min)= 13.5. Purity: 95.78 %.

*(S)-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)-3-hydroxybenzamide (**10**)*

65% yield. White solid. MP 228 - 229 °C. $[\alpha]_D^5$ = -46.7° ± 0.6 (CH_3OH, C=0.1). FT-IR (KBr, cm^{-1}) 3416, 3324, 3305, 2246, 1673, 1649, 1525, 1284, 1027, 706. ^1H NMR (500 MHz, DMSO-d_6) δ 9.62 (s, 1H), 8.98 (s, 1H), 8.52 (d, *J*= 10 Hz, 1H), 8.12 (m, 1H), 7.66 (m, 2H), 7.26-7.13(m, 5H), 6.91 (td, *J*= 7.5, 1.5 Hz, 1H), 4.57 (m, 1H), 3.02 (dd, *J*= 13.5, 5.0 Hz, 1H), 2.69 (dd, *J*= 19.0, 13.5 Hz, 1H),1.46 (m, 2H), 1.03 (m, 2H). ^{13}C NMR (125 MHz, DMSO-d_6) δ 175.47, 167.64, 156.77, 137.59, 135.53, 129.87, 129.42, 129.42, 129.07, 129.07, 127.17, 121.62, 121.57, 120.20, 114.53, 56.87, 38.46, 19.09, 19.09. ESI-MS (-) Calc. for [$C_{20}H_{19}N_3O_3$] 349.39. found: 348.23 [$M-H^+$]$^-$.The substance **10** was subjected to semi-preparative chromatography using the following conditions to afford 20 mg: C18 column (250x 10 mm; 10 μm; 100 A; LUNA®); isocratic elution with ACN/water, 30:70; flow rate 2.5 mL/min; injection volume 100 μL; stock solution

of substance **10** was 10.0 mg/mL in mobile phase. Analyses were carried out on a mass spectrometer operating in the scan mode (enhanced resolution; 100 - 1200 m/z) under positive ionization with detection of $[M+H]^+$ 350.16 m/z or $[M+Na]^+$ 372.15 m/z. The ESI parameters were as follows: capillary voltage – 4500 V; end plate voltage – 500 V; dry gas – 9 L/min; nebulizer – 40 psi; dry temperature – 300 °C. HPLC (Analytic) (50:50 ACN:Water): t_R (min)= 14.53. Purity: 98.61 %.

*(S)-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)nicotinamide (**11**)*

65% yield. White solid. MP 180 - 181 °C $[\alpha]_D^5$ = -63.2 ° ± 0.4 (CH_3OH). FT-IR (KBr, cm^{-1}) 3561.30, 3332.46, 2246.20, 1674.28, 1649.50, 1535.40, 1524.40, 1484.95, 1027.14, 705.0. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.11 (s, 1H), 8.97- 8.94 (m, 2H), 8.70(dd, *J*= 10.5, 5.5 Hz, 1H), 8.15 (dt, *J*= 10.5, 5.5 Hz , 2H), 7.49 (ddd, *J*= 15.5, 10,5 5.0 Hz, 1H), 7.29 (m, 4H), 7.18 (tt, *J*= 10.0, 3.2 Hz, 1H), 7.48 (m, 1H), 4.62 (m, 1H), 3.11 (dd, *J*= 13.5, 5.0 Hz, 1H), 3.10 (dd, *J*= 19.0, 13.5 Hz, 1H), 1.47 (m, 2H), 1.05 (m, 2H).^{13}C NMR (125 MHz, DMSO-d_6) δ 172.91, 165.31, 152.40, 149.04, 138.11, 135.61, 129.80, 129.60, 128.55, 126.83, 123.77, 121.16, 55.07, 39.77, 37.40, 20.19, 16.14, 16.09. ESI-MS (-) Calc. for [$C_{19}H_{18}N_4O_2$] 334.38, found: 333.21 $[M-H^+]^-$. HPLC (50:50 ACN:Water): t_R (min)= 8.27. Purity: 98.03 %.

*(S)-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)-3-fluorobenzamide (**12**)*

75% yield. White solid. MP 219 - 220 °C $[\alpha]_D^5$ = - 41.4 ° ± 0.6 (CH_3OH). FT-IR (KBr, cm^{-1}) 3271, 3064, 3026, 2297, 2237, 1676, 1664, 1587, 1541, 1523, 1485, 1435, 1377, 1325, 1300, 1288, 1271, 1224, 1215, 1180, 750, 702. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.02 (s, 1H), 8.76 (d, J = 6.5 Hz, 1H), 7.68 - 7.67 (m, 1H), 7.63 - 7.59 (m, 1H), 7.50 (dt,

J = 10.0, 7.5 Hz, 1H), 7.40 – 7.35 (m, 1H), 7.30 – 7.23 (m, 4H), 7.16 (tt, J = 6.0, 1.5 Hz, 1H), 4.62 – 4.56 (m, 1H), 3.08 (dd, , J = 17.0, 16.5 Hz, 1H), 2.98 (dd, , J = 17.0, 12.5 Hz, 1H), 1.47 - 1.44 (m, 2H), 1.04 – 1.01 (m, 2H).^{13}C NMR (125 MHz, DMSO-d_6) δ 172.95, 165.38, 165.36, 163.49, 161.06, 138.17, 136.59, 136.53, 130.86, 130.78, 129.57, 128.55, 126.83, 124.12, 124.09, 121.15, 118.82, 118.61, 114.80, 114.57, 55.11, 40.58, 40.38, 40.17, 39.96, 39.75, 39.54, 39.33, 37.34, 20.17, 16.13. ESI-MS (-) Calc. for [$C_{20}H_{18}FN_3O_2$] 351.14, found: 352.03 $[M+H]^+$. HPLC (50:50 ACN:Water): t_R (min)= 12.41. Purity: 96.8 %.

(S)-3-bromo-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)benzamide ***(13)***

72% yield. White solid. MP 229 - 230 °C $[\alpha]_D^5$ = - 66.9 ° ± 0.6 (CH_3OH). FT-IR (KBr, cm^{-1}) 3302, 3029, 3024, 2243, 1672, 1649, 1539, 1525, 1377, 1319, 1303, 1219, 1064, 896, 848, 746, 700, 684. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.05 (s, 1H), 8.86 (d, J = 8.0 Hz, 1H), 8.04 (t, J = 1.5 Hz, 1H), 7.84 - 7.82 (m, 1H), 7.76 – 7.74 (m, 1H), 7.46 (t, J = 8.0 Hz, 1H), 7.32 – 7.26 (m, 4H), 7.21 (tt, J = 6.0, 1.5 Hz, 1H), 4.64 – 4.60 (m, 1H), 3.15 (dd, , J = 17.0, 16.5 Hz, 1H), 3.01 (dd, , J = 17.0, 12.5 Hz, 1H), 1.50 - 1.47 (m, 2H), 1.08 – 1.03 (m, 2H).^{13}C NMR (125 MHz, DMSO-d_6) δ 172.95, 165.23, 138.19, 136.41, 134.54, 130.94, 130.60, 129.57, 128.56, 128.01, 127.09, 126.84, 121.97, 121.16, 55.11, 40.58, 40.49, 40.41, 40.32, 40.24, 40.15, 40.08, 39.99, 39.91, 39.82, 39.65, 39.48, 37.32, 20.18, 16.12. ESI-MS (-) Calc. for [$C_{20}H_{18}BrN_3O_2$] 412.29, found: 413.50 $[M+H]^+$. HPLC (50:50 ACN:Water): t_R (min)= 18.33. Purity: 99.5 %.

(S)-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)-3-iodobenzamide ***(14)***

86% yield. White solid. MP 205 - 207 °C $[\alpha]_D^5$ = - 44.3 ° ± 0.5 (CH_3OH). FT-IR (KBr, cm^{-1}) 3286, 3029, 3022, 2241, 1672, 1649, 1522, 1317, 1284, 1299, 1219, 1066, 883, 792, 744, 684. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.00 (s, 1H), 8.78 (d, J = 8.0 Hz, 1H), 8.17 (t, J = 1.5 Hz, 1H), 7.88 - 7.86 (m, 1H), 7.28 – 7.22 (m, 5H), 7.16 (tt, J = 7.0, 2.0 Hz, 1H), 4.60 – 4.55 (m, 1H), 3.05 (dd, J = 15.0, 5.0 Hz, 1H), 2.96 (dd, , J = 14.0, 10.0 Hz, 1H), 1.47 - 1.42 (m, 2H), 1.04 – 1.00 (m, 2H).^{13}C NMR (125 MHz, DMSO-d_6) δ 172.97, 165.23, 140.32, 138.19, 136.34, 136.26, 130.85, 129.57, 128.55, 127.41, 126.83, 121.16, 94.96, 55.06, 40.58, 40.49, 40.41, 40.32, 40.24, 40.15, 40.08, 39.99, 39.91, 39.82, 39.65, 39.49, 37.32, 20.18, 16.12.. ESI-MS (-) Calc. for [$C_{20}H_{18}IN_3O_2$] 460.39, found: 461.32 $[M+H]^+$. HPLC (50:50 ACN:Water): t_R (min)= 22.26 Purity: 99.1 %.

*(S)-N-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)-3-(trifluoromethyl)benzamide (**15**)*

62% yield. White solid. MP 193 - 195 °C $[\alpha]_D^5$ = - 65.7 ° ± 0.2 (CH_3OH). FT-IR (KBr, cm^{-1}) 3273, 3025, 2296, 2243, 1683, 1674, 1645, 1539, 1489, 1440, 1325, 1303, 1282, 1170, 1124, 1097, 1072, 748, 696, 684. ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.04 (s, 1H), 8.98 (d, J = 8.0 Hz, 1H), 8.15 (s, 1H), 8.10 (d, J = 8.0 Hz, 1H), 7.90 (d, J = 7.5 Hz, 1H), 7.70 (t, J = 8.0 Hz, 1H), 7.29 - 7.22 (m, 4H), 7.19 (t, J = 7.5, Hz, 1H), 4.64 – 4.60 (m, 1H), 3.09 (dd, J = 13.5, 5.0 Hz, 1H), 2.98 (dd, J = 13.5, 10.0 Hz, 1H), 1.47 - 1.44 (m, 2H), 1.04 – 1.03 (m, 2H).^{13}C NMR (125 MHz, DMSO-d_6) δ 172.91, 165.32, 138.13, 135.13, 132.06, 130.01, 129.58, 129.33, 128.56, 128.39, 126.86, 125.51, 124.57, 121.15, 55.12, 40.55, 40.46, 40.39, 40.30, 40.22, 40.13, 40.05, 39.96, 39.89, 39.79, 39.63, 39.46, 37.37, 20.19, 16.14. ESI-MS (-) Calc. for [$C_{21}H_{18}F_3N_3O_2$] 401.14, found: 402.15 $[M+H]^+$. HPLC (50:50 ACN:Water): t_R (min)= 14.75 Purity: 97.9 %.

2.3 ***Enzymatic inhibition studies***

The recombinant cruzain, consisting of the catalytic domain of cruzipain but excluding the carboxy-terminal extension, was expressed and purified as previously described. [10] The enzymatic activity was evaluated by fluorometric assays (Biotek Synergy HT) monitoring the hydrolysis rate of the fluorogenic substrate Z-Phe-Arg-7-amido-4-methylcoumarin (Z-FR-MCA, Sigma-Aldrich) with fluorescence emission at 460 nm (excitation at 355 nm) at 37 °C. The reaction was followed over 5 minutes. Enzyme kinetic assays were carried out in a Corning 96-well black flat bottom microplates containing 200 μL of a solution constituted by 100 mM acetate buffer pH 5.5, 300 mM NaCl, 5 mM DTT (dithiothreitol), 5% v/v DMSO (dimethyl sulfoxide), 0.01 % v/v Triton X-100 and 0.15 nM Cz. The enzyme stock aliquot was rapidly thawed at 37 °C and kept on ice until activation, in which it was incubated for 20 min in the assay buffer (100 mM acetate pH 5.5 and 5 mM DTT) followed by additional 2 min with inhibitors before the addition of the substrate to trigger the reaction.

Visual inspection and a pre-reading of plate wells were performed to check for possible precipitation and background fluorescence, respectively. None of the substances displayed a significant fluorescence signal around 460 nm; the emission wavelength used to monitor reaction kinetics. Thus, potential inner-filter effects did not have to be taken into account in our experiments. Analysis and manipulation of the data were performed with Sigma Plot 10.

Each experiment was performed in triplicate for each substance. Initial velocities of the substrate hydrolysis under the first-order reaction were calculated using Gen5^{TM} Biotek software. The apparent inhibition constant $K_i^{'}$ was determined by non-linear regression using equation $V_s = V_o/(1+[I]/K_i^{'})$, where V_s is the steady-state rate, V_o is the rate in the absence of inhibitor, and [I] is the inhibitor concentration. The true inhibition

constant *K*i was calculated by the correction of $K_i^{'}$ according to $K_i = K_i^{'}/(1+[S]/K_M)$, where [S] is the substrate concentration and K_m is the Michaelis constant. A control measurement with the covalent reversible fast-binding inhibitor Neq0570 was done for each setup plate. All inhibitors were evaluated at seven different concentrations that were chosen based on a previous screening (percent inhibition) at 2.5 µM substrate concentration (~K_M) [20]. Further details can be found in the supplementary information.

2.4 Isothermal Titration Calorimetry (ITC)

An aliquot of the protein was taken, and a new dialysis was performed using the last buffer employed in this process during the purification of the protein. Dialysis and concentration were performed over 30 to 60 min period on Amicon ultra 10 kDa membranes (Merck Millipore). The inhibitor was placed in the syringe and the protein in the cell. 0.001% Triton-X 100 was added to the solution of the syringe and the cell to prevent the protein from aggregation. All thermodynamic parameters were calculated with Microcal PEAQ-ITC Analysis Software, and experiments were performed at least in duplicate. All graphs and experimental conditions can be found in the supplementary information.

3. Results and Discussion

To describe the influence of P3 substitution in the phenyl ring of the known dipeptidyl nitrile inhibitor (**5**) acting on Cz [9], we performed three replicates of 100 ns molecular dynamics (MD) simulations for the covalent nitrile-enzyme complexes. We observed a better interaction pattern between the Ser61 hydroxyl group and the chlorine atom at the

meta-position (**7**) in relation to *orto* (**6**) or *para* (**8**) derivatives (described below). We considered the unsubstituted phenyl ring in P3 (compound **5**) for comparative purpose (Figure 1). The MD for substance **7** showed two well-defined distributions for the distance between H3 (bound to carbon at the *meta* position of the aromatic ring) and the hydroxyl group of Ser61 ($d_{m\text{-Phe\_Ser61}}$), centered at 3 and 6.8 Å (Figure S2). The *para*-chlorophenyl compound (**8**) did not bring any further stability, showing free rotation of the phenyl ring in P3 and no preference between parallel or perpendicular conformation in respect to Gly65 and Gly66 main chains. On the other hand, we observed that contacts between *orto*-chlorophenyl substance and protein prevent the ring rotation during MD. This was confirmed by the high stability of the initial conformation throughout the simulation, with chlorine pointed respectively to Gly65 (5.1 Å) and Leu67 (8.3 Å). Regarding the flexibility, the molecule containing *meta*-chlorophenyl in P3 lays in the midrange when compared with *orto*-chlorophenyl derivative (highly constrained) and *para*-chlorophenyl (highly flexible), once the ring is free to rotate but stay during 75 % of simulation time with $d_{m\text{-Phe\_Ser61}}$ less than 5.5 Å (Figure 1). From these frames, we observed that the angle and distance of chlorine atom from the ligand and Ser61 hydroxyl group fluctuate between 90 and 170 degrees and 3 and 5.5 Å, respectively. This pattern is in agreement with what Ibrahim [11] has proposed for a halogen bond in MD simulations using the Amber force field without extra-point correction. Finally, for substance **11**, it was not observed any interaction pattern with residue Ser61 (not shown) as for substance **5.**

It is known that the Cz Ser61 residue interacts with pyridine and triazole groups via hydrogen bond, as shown in structures of PDB code 1ME3 [17] and 4KLB [8]. To understand the possibility of halogen and hydrogen bond interaction of *meta*-substituted phenyl derivatives with Ser61, we also carried out MD simulations for *m*-methyl and *m*-

hydroxyl derivatives (Figure 1). The analysis of the P3 subsite interactions reveals the ability of *meta*-hydroxyphenyl moiety (**10**) to perform weak hydrogen bond with Ser61, once the interaction amid hydroxyl groups of ligand and Ser61 could be retrieved from the simulation timeframes leading to a frequency of 14 % (not shown). Also, we observed that *meta*-methyl phenyl substituent in (**9**) has a higher tendency to move out from the initial position, reaching values as higher as 4 Å, while the other two *meta*-substituted molecules present a maximum deviation of 3 Å. The $d_{m\text{-Phe\_Ser61}}$ of 3.75 Å is similar to 3.5 Å observed to *meta*-chlorophenyl derivative (**7**) that could be explained by similar van der Waals radius. Both values are considerably higher than the $d_{m\text{-Phe\_Ser61}}$ of 2.9 Å observed for *meta*-hydroxyphenyl analog (**10**). Simulations showed the preference of *meta*-position substitution on the phenyl ring at the P3 portion of the substance as well as indicate a possible hydrogen bond interaction between hydroxyl and chlorophenyl derivatives with Ser61 residue. To better understand this hypothesis we also evaluated the influence of different halogen atoms in *meta*-position of P3 ring. The fluorine derivative molecule (compound **12**) showed high rotational flexibility of ring, similar to that found for the molecule **5**. It is an indication of the interaction lost between S3 subsite and P3 ring and thus, between the fluorine atom and Ser61 side chain. On the other hand, for bromine-derivative molecule (compound **13**), we observed more sampling for smaller $d_{m\text{-phe\_Ser61}}$ described by 75% of simulation frames with $d_{m\text{-Phe\_Ser61}}$ below 5.5 Å and maximum peak at 3.5 Å, which are similar with those found for compound **7**. Based on these simulations, probably the molecules **7**, **10** and **13** can perform a hydrogen bond interaction with residue Ser61 of the cruzain S3 cavity.

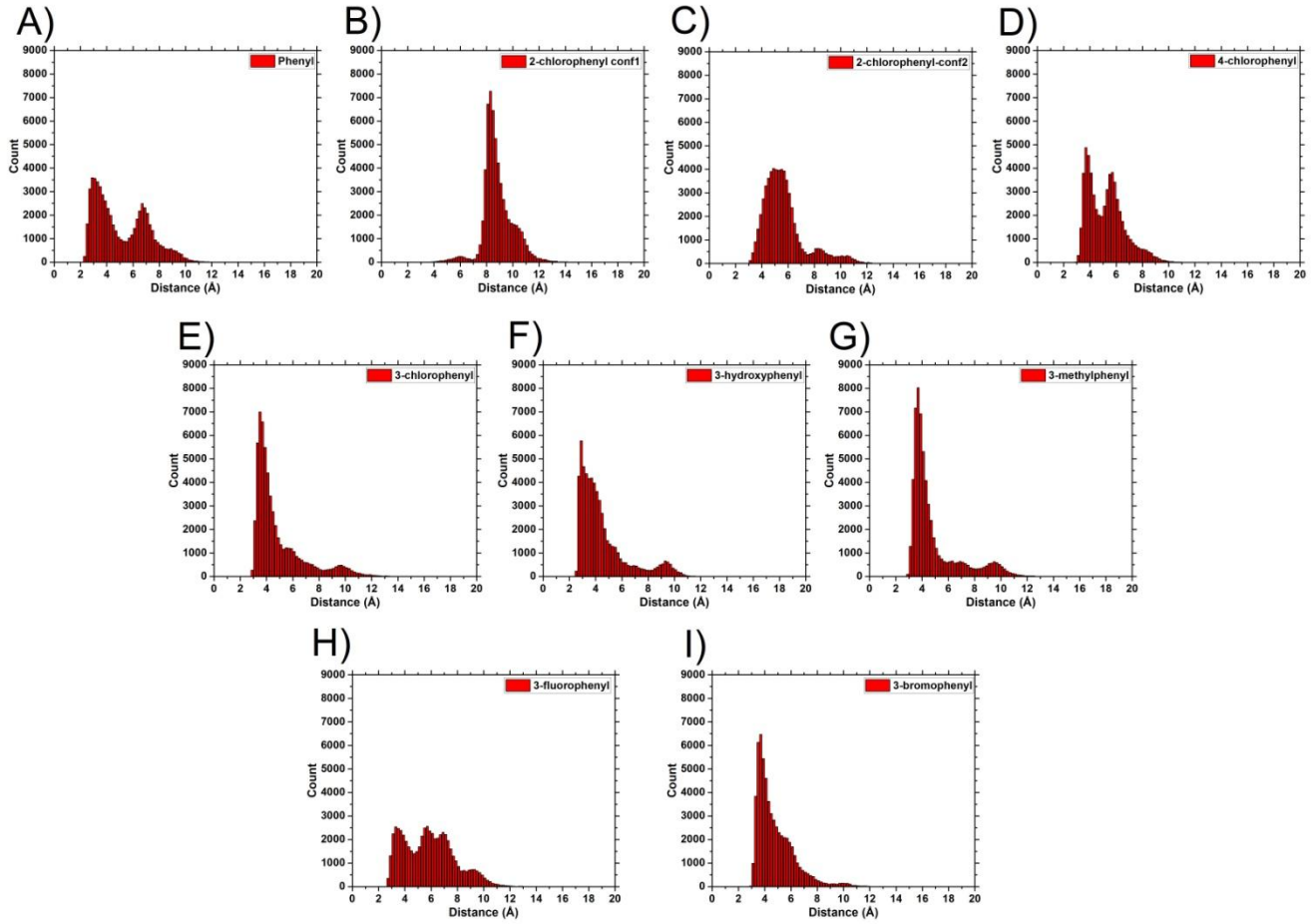


Figure 1. Histograms (A-I) with the occurrence of molecular dynamics simulation frames binned by the distance between HO-Ser61 and the atom in *meta* position of the ring in P3 subsite and (H) labeling of the $d_{m\text{-Phe\_Ser61}}$ distance using PDB ID 4QH6. Derivatives are displayed according to the substituents at the P3 phenyl ring: A) Phenyl (**5**), B) *o*-chlorophenyl-conf1 (**6**), C) *o*-chlorophenyl-conf2 (**6**), D) *p*-chlorophenyl (**8**), E) *m*-chlorophenyl (**7**), F) *m*-hydroxyphenyl (**10**), G) *m*-methyl phenyl (**9**), H) *m*-

fluorophenyl (**12**), I) *m*-iodophenyl (**14**) groups in complex with cruzain. Conf1: conformation 1. Conf2: conformation 2.

To better evaluate such a preference and to confirm our hypothesis of possible interaction with Ser61 in S3 subsite, we carried out an in-depth Cz inhibition study for substances **5-15**. Using pK_i affinity values [21] from Table 1, we performed a Matched Molecular Pair Analysis (MMPA) that provides the opportunity to evaluate molecular modifications for significant structure-activity relationships (SAR). The analysis was performed in molecular pairs that differ only in positions where there is a promising chemical transformation. In this approach, transformations occur by identifying the change from one functional group to another in which its regiochemistry is also taken into account. MMPA may be useful in interpreting the bimolecular recognition process and the individual contributions of each transformation in that process. Our major interest was focused on P3 modifications considering that substitutions at this position impact the S3 subsite of the cruzain in a manner that is not well elucidated. This analysis shows that the pK_i median of 6.6 spanned in the range 5.9-7.2, and it displays some significant transformations occurring among the following pairs: *p*-Cl/*m*-Cl, *m*-H/*m*-Cl, *m*-Cl/*m*-OH, *m*-H/*m*-OH, *m*-Me/*m*-Cl, *m*-Me/*m*-OH/, *m*-H/*m*-Br, *m*-Cl/*m*-Br, *m*-H/*m*-I, *m*-Cl/*m*-CF_3, and *p*-Cl/*o*-Cl.[7] The most substantial difference in the observed pK_i values is found in the *m*-H/*m*-I substitution (**5** *vs.* **14**) with the value of 0.8 log units in favor of the iodine atom in *meta*-position. Moreover, the inclusion of the chlorine atom in the *meta*-position resulted in the second major difference observed in the series with affinity gain about 0.7 log units (**5** *vs.* **7**). This result is corroborated by the average pose obtained after the molecular modeling, where the P3 phenyl ring occupies the S3 subsite of the cruzain in which the meta position is pointing toward the hydroxyl group

of Ser61 (Figure 2). Instead, the exchange of the hydrogen atom by a chlorine atom in *para* position (**5** *vs.* **8**) resulted in affinity loss of only 0.07 log units, which is not significant due to experimental uncertainty. Quite surprisingly but not completely unexpected, the *m*-Cl/*m*-OH substitution (**7** *vs.* **10**) resulted in an affinity loss of 0.10 log units, while the inclusion of the hydroxyl group in *meta* showed a gain of 0.61 log units (**10** *vs.* **5**). The same gain in affinity was observed for the *m*-Me/*m*-Cl derivatives (**7** *vs.* **9**) or *m*-H/*m*-Br derivatives (**5** vs **13**). A minor gain of 0.2 log units was observed for transformation *o*-Cl/*p*-Cl (**6** *vs.* **8**). As we expected, the replacement of chlorine, bromine or iodine atom by a trifluoromethyl substituent resulted in a loss of affinity. In particular, this effect is observed for the pair *m*-Br/*m*-CF_3 with a loss of 0.6 log units; while we pictured a small loss in affinity (0.1 log units) for the substitution *m*-H/*m*-F (**5** *vs*, **12**). The most considerable loss in affinity (-1.3 log units) occurred by the transformation observed between substances **7** and **11**, where the inclusion of the nitrogen heteroatom for the formation of the pyridine moiety did not represent a good approach.

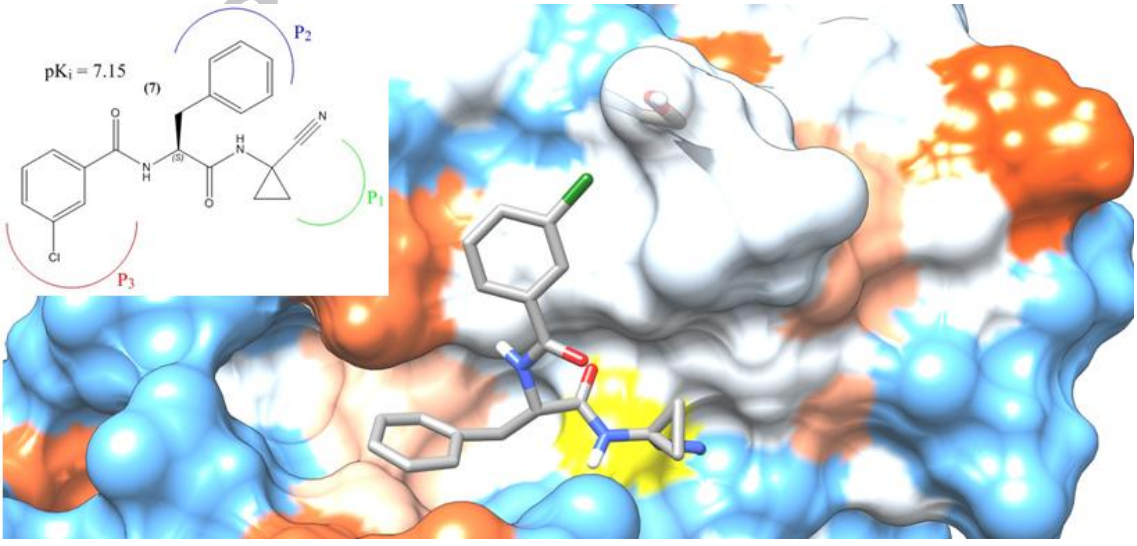


Figure 2. Average pose obtained after the molecular dynamic simulation with substance **7** covalently bond with cruzain (PDB code 1ME3). Colour for substance **7**: Carbon chain is gray; Chlorine is green; Oxygen is red. Protein is represented as a polar surface;

from white to red there is an increment of polarity. The catalytic residue of Cysteine is spotted with yellow color. P1, P2 and P3 represent the substituents of the compound that interact with their respective subsites (S1, S2 and S3) at the enzyme.

In summary, these results point out the direct influence of the Cl, Br and I regiochemistry in the P3 phenyl ring as an adequate strategy to gather the halogen bond with the hydroxy group from Ser61. Known hydrogen bond donors and acceptors (*m*-phenol and 3-pyridine) in comparison with the *m*-Cl derivative led to less potent cruzain inhibitors. It corroborates the existence of the halogen bond intermediating the bimolecular recognition process.

The inclusion of halogen atoms in ligands with the aim of improving interaction with targets is considered an adequate strategy in molecular design [22]. The attractive interaction between the electrophilic region of the chlorine, bromine or iodine (the so-called σ-hole) and that nucleophile of another molecular fragment may represent a good approximation for the inclusion of one of these atoms in the phenyl *meta*-position of P3. Furthermore, even if bromine and iodine have a greater ability to form halogen bond than chlorine, their atomic radius could influence their ability to fit properly in the S3 region next to Ser61 residue. To better evaluate this trend we employed isothermal titration calorimetry to determine thermodynamic parameters of cruzain–dipeptidyl nitrile interactions (Table 1).

Table 1. Affinity (pK_i) values for substances **5-15** against cruzain.

ID	pK_i[a]	K_d (nM)	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	-TΔS (kcal mol^{-1})
5	6.4 (0.026)	252 (24.4)	-9.00	-14.6	5.60
6	6.6 (0.058)	ND	ND	ND	ND
7	7.2 (0.020)	121 (17.5)	-9.43	-10.7	1.32
8	6.4 (0.018)	ND	ND	ND	ND
9	6.6 (0.0099)	155 (12.6)	-9.29	-12.8	3.55
10	7.0 (0.019)	175 (9.51)	-9.22	-11.9	2.64
11	5.9 (0.017)	338 (91.7)	-8.83	-12.1	3.30
12	6.3 (0.023)	211 (2.50)	-9.11	-13.9	5.96
13	6.9 (0.032)	93.1 (1.30)	-9.58	-11.3	1.77
14	7.2 (0.02)	70.2 (4.50)	-9.76	-12.0	2.28
15	6.6 (0.02)	125 (4.0)	-9.45	-12.7	3.29

[a] pK_i = -log(K_i/M); Uncertainty = desvpad/(K_i*ln(10))

ND – Not Determined. Standard deviations are reported in brackets. Deviations for ΔG are lower than 0.4%; ΔH < 4.5%; -TΔS < 15% (except for compound **13**, with 22%).

As already described in our recent study, [23] isothermal titration calorimetry (ITC) is a technique used for the thermodynamic characterization of interactions between biomolecules and small molecules, which is vital for understanding the process of molecular recognition. Hence, we studied the thermodynamic profiles of substances **5**, **7**, and **9-15** to better estimate our halogen-bonding hypothesis. The values of K_d (Table 1) showed the same trend as depicted by pK_i, with substance **14** having the highest affinity for Cz, followed by substances **13**, **7** and **10**. This ultimately corroborates the idea of the *meta*-halogen derivative (**7, 14** and **13**) having the required positioning for a putative interaction with Ser61.

The thermodynamic signatures gathered from ITC data reveal that binding of cruzain inhibitors was driven by favorable enthalpy contributions, with detrimental entropy for all study substances.

Figure 3 shows the thermodynamic signatures of the interactions that have, mostly, the same free energy (ΔG) of interaction. The interaction energy is related to the affinity, which is a combination between enthalpy (ΔH) and entropy (ΔS). Classically, enthalpy reflects the strength of the interactions due to the formation of hydrogen bonds and van der Waals, while the binding entropy is a combination of the change in the desolvation and conformational changes after the formation of the bimolecular ligand-cruzain complex [25].

A plot (not shown) of enthalpy (ΔH), versus the entropy contribution (-TΔS) unveils a *quasi*-perfect omnipresent phenomenon of enthalpy-entropy compensation ($\Delta H_{intrinsic}$ = -9.4(± 0.25), -TΔS = 0.9(± 0.071), R^2: 0.982, Adjusted R^2: 0.976, F-statistic: 167.2, *p*-value: 0.001). Interestingly, the effect of the *m*-chlorophenyl moiety on binding affinity decreases both binding enthalpy and entropy contributions (see compounds **5** and **7** in Fig. 3). For substance **7**, the decrease in entropy is more efficient for the enthalpy-entropy compensation process than that observed for the decrease in enthalpy change, which leads to a lower Gibb's free energy than **5**.

As mentioned, ΔG values are quite steady over the series. This effect could be due to an enthalpy-entropy compensation effect [24], where solvent reorganization could play a central role as reported [25]. The substitution of *m*-H (-TΔS = 5.6 kcal mol^{-1}) by *m*-CH_3 (-TΔS = 3.5 kcal mol^{-1}), *m*-OH (-TΔS = 2.6 kcal mol^{-1}) or *m*-Cl (-TΔS = 1.3 kcal mol^{-1}) groups results in a decrease of both enthalpy and entropy (-TΔS) from 5.6 to 1.3 kcal mol^{-1} (except for substance **12**). This yields an entropic difference (-TΔΔS) gain of 4.3 kcal mol^{-1} (Table 1, Figure 3). This trend is also followed for *m*-Br (-

$T\Delta S$ = 1.8 kcal mol^{-1}) and *m*-I ($-T\Delta S$ = 2.3 kcal mol^{-1}). In particular, there is a small but consistent gain in enthalpy from *m*-Cl to *m*-Br ($\Delta\Delta H$ = -0.6 kcal mol^{-1}) as well as from *m*-Cl to *m*-Br ($\Delta\Delta H$ = -1.3 kcal mol^{-1}). Furthermore, the enthalpy-entropy compensation effect is visible also for *m*-F and *m*-CF_3, where both improve the enthalpy as from the *m*-Br, but with a stronger loss in entropy.

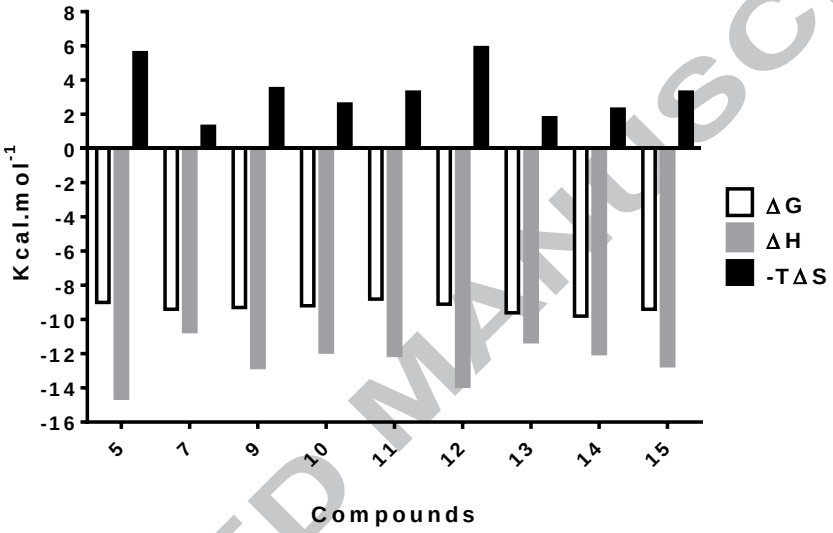


Figure 3. Thermodynamic profiling signatures for the set of inhibitors interacting with cruzain.

4. Conclusions

Previous studies have shown that dipeptidyl nitriles are potent cruzain inhibitors and may be as potent as benznidazole [9] in *in vitro* parasite killing. To better understand and characterize cruzain inhibition, we synthesized key derivatives that allowed us to establish critical structure-activity relationships at the P3 ring consistent with an increase in cruzain affinity. From our studies of molecular dynamics simulation, we have shown that the *m*-chlorophenyl, *m*-bromophenyl, and *m*-hydroxyphenyl moieties have the right orientation inside P3 to interact with the oxygen of Ser61 residue. So we

synthesized 11 new cruzain inhibitors with different substituents in P3 to better evaluate this trend. The results are in good agreement with the observed MD simulations, once the best compounds are precisely those with a lower fluctuation of P3 ring. Indeed the substitution of a hydrogen atom to an iodine atom in *meta* position in P3 increased the pK_i of 0.8 log unit. Moreover, our ITC studies pictured that the mechanism of binding of cruzain inhibitors is driven by favorable enthalpy contributions with *quasi*-perfect omnipresent phenomenon of enthalpy-entropy compensation. This tendency is well visible when we screened the 17th group of the periodic table as possible substituents of the hydrogen atom in *meta* position.

Overall, our studies provide further details of the SAR regarding the physical-chemical profiling of cruzain inhibitors that could also be extended to other cysteine proteases. This may ultimately aid advancing molecular designing efforts directed to reversible covalent inhibitors.

5. Acknowledgments

We are grateful to São Paulo Research Foundation (FAPESP) for funding this research (grant number: 2013/18009-4). Authors declare no conflict of interest.

References

[1] WHO - Chagas disease (American trypanosomiasis). http://www.who.int/mediacentre/factsheets/fs340/en; (Accessed 08 August 2017).

[2] Y. Jackson, A. Angheben, B. Carrilero Fernandez, J.M. Jansa i Lopez del Vallado, J.G. Jannin, P. Albajar-Viñas, Management of Chagas disease in Europe. Experiences and challenges in Spain, Switzerland and Italy, Bull. Soc. Pathol. Exot. 102 (1990) 326-329.

[3] C. Chagas, Nova tripanozomiaze humana: estudos sobre a morfolojia e o ciclo evolutivo do Schizotrypanum cruzi n. gen., n. sp., ajente etiolojico de nova entidade morbida do homem, Mem. Inst. Oswaldo Cruz 1 (1909) 159-218.

[4] A. Rassi Jr., A. Rassi, J.A. Marin-Neto, Chagas disease, Lancet 375 (2010) 1388-1402.

[5] J.R. Coura, P.A. Vinas, Chagas disease: a new worldwide challenge, Nature 465 (2010) S6-S7.

[6] J.A. Castro, M.M. de Mecca, L.C. Bartel, Toxic side effects of drugs used to treat Chagas disease (American trypanosomiasis), Hum. Exp. Toxicol. 25 (2006) 471-479.

[7] M. Sajid, S.A. Robertson, L.S. Brinen, J.H. McKerrow, Cruzain: the path from target validation to the clinic. Adv. Exp. Med. Biol. 712 (2011) 100-115.

[8] H.J. Wiggers, J.R. Rocha, W.B. Fernandes, R. Sesti-Costa, Z.A. Carneiro, J. Cheleski, A.B. da Silva, L. Juliano, M.H. Cezari, J.S. Silva, J.H. McKerrow, C.A. Montanari, Non-peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay, PLoS Negl. Trop. Dis. 7 (2013) e2370.

[9] L.A. Avelar, C.D. Camilo, S. de Albuquerque, W.B. Fernandes, C. Gonçalez, P.W. Kenny, A. Leitão, J.H. McKerrow, C.A. Montanari, E.V. Orozco, J.F. Ribeiro, J.R. Rocha, F. Rosini, M.E. Saidel, Molecular Design, Synthesis and Trypanocidal Activity of Dipeptidyl Nitriles as Cruzain Inhibitors, PLoS Negl. Trop. Dis. 9 (2015) e0003916.

[10] H.J. Wiggers, J.R. Rocha, W.B. Fernandes, R. Sesti-Costa, Z.A. Carneiro, J. Cheleski, A.B. da Silva, L. Juliano, M.H. Cezari, J.S. Silva, J.H. McKerrow, C.A. Montanari, Non-peptidic cruzain inhibitors with trypanocidal activity discovered by virtual screening and in vitro assay, PLoS Negl. Trop. Dis. 7 (2013) e2370.

[11] M.A. Ibrahim, Molecular mechanical study of halogen bonding in drug discovery, J. Comput. Chem. 32 (2011) 2564-2574.

[12] W.D. Cornell, P. Cieplak, C.I. Bayly, P.A. Kollman, Application of resp charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation, J. Am. Chem. Soc. 115 (1993) 9620-9631.

[13] F.Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak, The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building, Phys. Chem. 12 (2010) 7821-7839.

[14] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99sb, J. Chem. Theory Comput. 11 (2015) 3696-3713.

[15] J.M. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157-1174.

[16] D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, P.A. Kollman, AMBER 14, University of California, San Francisco, 2014.

[17] L. Huang, L.S. Brinen, J.A. Ellman, Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain, Bioorg. Med. Chem. 11 (2003) 21-29.

[18] W.D. Cornell, P. Cieplak, C.I. Bayly, P.A. Kollman, Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation, J. Am. Chem. Soc. 115 (1993) 9620-9631.

[19] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327-341.

[20] M. Frizler, F. Lohr, N. Furtmann, J. Kläs, M. Gütschow, Structural optimization of azadipeptide nitriles strongly increases association rates and allows the development of selective cathepsin inhibitors, J. Med. Chem. 54 (2010) 396-400.

[21] P.W. Kenny, J. Sadowski, Structure Modification in Chemical Databases, in: T.I. Oprea(Ed.), Chemoinformatics in Drug Discovery, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2005, pp. 271-285.

[22] S. Sirimulla, J.B. Bailey, R. Vegesna, M. Narayan, Halogen interactions in protein-ligand complexes: implications of halogen bonding for rational drug design, J. Chem. Inf. Model. 53 (2013) 2781-2791.

[23] I.M. Prokopczyk, J.F. Ribeiro, G.R. Sartori, R. Sesti-Costa, J.S. Silva, R.F. Freitas, A. Leitão, C.A. Montanari, Integration of methods in cheminformatics and biocalorimetry for the design of trypanosomatid enzyme inhibitors, Future Med. Chem. 6 (2014) 17-33.

[24] J.D. Chodera, D.L. Mobley, Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular, Annu. Rev. Biophys. 42 (2013) 121-142.

[25] N.J. Deng, P. Zhang, P. Cieplak, L. Lai, Elucidating the energetics of entropically driven protein-ligand association: calculations of absolute binding free energy and entropy, J. Phys. Chem. B 115 (2011) 11902-11910.

Leveraging the Cruzain S3 Subsite to Increase Affinity for Reversible Covalent Inhibitors

Lorenzo Cianni[†*], Geraldo Sartori[†], Fabiana Rosini[†], Daniela De Vita[†], Gabriel Pires, Bianca Rebelo Lopes[†], Andrei Leitão[†*], Antonio C. B. Burtoloso[††], Carlos A. Montanari[†]

[†]*Grupo de Química Medicinal do IQSC/USP, Instituto de Química de São Carlos, Universidade de São Paulo, Av. Trab. Sancarlense, 400 - Parque Arnold Schimidt, São Carlos - SP, 13566-590, São Paulo, Brazil*

[††]*Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brasil*

- Little is known about the S3 subsite interaction between inhibitors and cruzain
- New cruzain inhibitors were designed to probe the S3 subsite of cruzain
- Halogen bonding was observed with OH-Ser61
- pKi and thermodynamics data were in agreement for the whole set
- The molecular dynamics provided the putative mode of binding

Leveraging the Cruzain S3 Subsite to Increase Affinity for Reversible Covalent Inhibitors

Lorenzo Cianni[†*], Geraldo Sartori[†], Fabiana Rosini[†], Daniela De Vita[†], Gabriel Pires, Bianca Rebelo Lopes[†], Andrei Leitão[†*], Antonio C. B. Burtoloso[††], Carlos A. Montanari[†]

[†]*Grupo de Química Medicinal do IQSC/USP, Instituto de Química de São Carlos, Universidade de São Paulo, Av. Trab. Sancarlense, 400 - Parque Arnold Schimidt, São Carlos - SP, 13566-590, São Paulo, Brazil*

[††]*Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brasil*