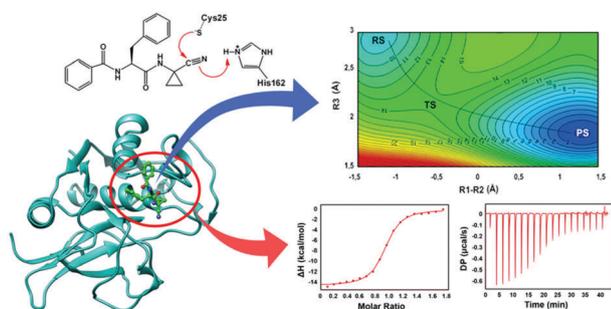We have presented the Graphical Abstract text and image for your article below. This brief summary of your work will appear in the contents pages of the issue in which your article appears.

Q1

Experimental study and computational modelling of cruzain cysteine protease inhibition by dipeptidyl nitriles

Alberto Monteiro Dos Santos, Lorenzo Cianni, Daniela De Vita, Fabiana Rosini, Andrei Leitão, Charles A. Laughton, Jerônimo Lameira* and Carlos A. Montanari*

A combined computational and experimental study aimed to gain insights into the reaction inhibition mechanism of cruzain by dipeptidyl nitriles.

Please check this proof carefully. Our staff will not read it in detail after you have returned it.

Please send your corrections either as a copy of the proof PDF with electronic notes attached or as a list of corrections. **Do not edit the text within the PDF or send a revised manuscript** as we will not be able to apply your corrections. Corrections at this stage should be minor and not involve extensive changes.

Proof corrections must be returned as a single set of corrections, approved by all co-authors. No further corrections can be made after you have submitted your proof corrections as we will publish your article online as soon as possible after they are received.

Please ensure that:

- The spelling and format of all author names and affiliations are checked carefully. You can check how we have identified the authors' first and last names in the researcher information table on the next page. **Names will be indexed and cited as shown on the proof, so these must be correct.**
- Any funding bodies have been acknowledged appropriately and included both in the paper and in the funder information table on the next page.
- All of the editor's queries are answered.
- Any necessary attachments, such as updated images or ESI files, are provided.

Translation errors can occur during conversion to typesetting systems so you need to read the whole proof. In particular please check tables, equations, numerical data, figures and graphics, and references carefully.

Please return your **final** corrections, where possible within **48 hours** of receipt, by e-mail to: pccp@rsc.org. If you require more time, please notify us by email.

Funding information

Providing accurate funding information will enable us to help you comply with your funders' reporting mandates. Clear acknowledgement of funder support is an important consideration in funding evaluation and can increase your chances of securing funding in the future.

We work closely with Crossref to make your research discoverable through the Funding Data search tool (http://search.crossref.org/funding). Funding Data provides a reliable way to track the impact of the work that funders support. Accurate funder information will also help us (i) identify articles that are mandated to be deposited in **PubMed Central (PMC)** and deposit these on your behalf, and (ii) identify articles funded as part of the **CHORUS** initiative and display the Accepted Manuscript on our web site after an embargo period of 12 months.

Further information can be found on our webpage (http://rsc.li/funding-info).

What we do with funding information

We have combined the information you gave us on submission with the information in your acknowledgements. This will help ensure the funding information is as complete as possible and matches funders listed in the Crossref Funder Registry.

If a funding organisation you included in your acknowledgements or on submission of your article is not currently listed in the registry it will not appear in the table on this page. We can only deposit data if funders are already listed in the Crossref Funder Registry, but we will pass all funding information on to Crossref so that additional funders can be included in future.

Please check your funding information

The table below contains the information we will share with Crossref so that your article can be found *via* the Funding Data search tool. **Please check that the funder names and grant numbers in the table are correct and indicate if any changes are necessary to the Acknowledgements text.**

Funder name	Funder's main country of origin	Funder ID (for RSC use only)	Award/grant number
Conselho Nacional de Desenvolvimento Científico e Tecnológico	Brazil	501100003593	#400658-2014-3
Fundação de Amparo à Pesquisa do Estado de São Paulo	Brazil	501100001807	#2013/18009-4

Researcher information

Please check that the researcher information in the table below is correct, including the spelling and formatting of all author names, and that the authors' first, middle and last names have been correctly identified. **Names will be indexed and cited as shown on the proof, so these must be correct.**

If any authors have ORCID or ResearcherID details that are not listed below, please provide these with your proof corrections. Please ensure that the ORCID and ResearcherID details listed below have been assigned to the correct author. Authors should have their own unique ORCID iD and should not use another researcher's, as errors will delay publication.

Please also update your account on our online manuscript submission system to add your ORCID details, which will then be automatically included in all future submissions. See here for step-by-step instructions and more information on author identifiers.

First (given) and middle name(s)	Last (family) name(s)	ResearcherID	ORCID iD
Alberto Monteiro	Dos Santos		
Lorenzo	Cianni		
Daniela	De Vita		
Fabiana	Rosini		
Andrei	Leitão	B-7942-2012	0000-0002-6601-6609
Charles A.	Laughton		0000-0003-4090-3960
Jerônimo	Lameira		0000-0001-7270-1517
Carlos A.	Montanari	C-2799-2012	0000-0002-4963-0316

Queries for the attention of the authors

For your information: You can cite this article before you receive notification of the page numbers by using the following format: (authors), Phys. Chem. Chem. Phys., (year), DOI: 10.1039/c8cp03320j.

Editor's queries are marked on your proof like this Q1, Q2, etc. and for your convenience line numbers are indicated like this 5, 10, 15, ...

Please ensure that all queries are answered when returning your proof corrections so that publication of your article is not delayed.

Query reference	Query	Remarks
Q1	Please confirm that the spelling and format of all author names is correct. Names will be indexed and cited as shown on the proof, so these must be correct. No late corrections can be made.	
Q2	Text has been provided for footnote a in Table 3, but there does not appear to be a corresponding citation in the table. Please indicate a suitable location for the footnote citation.	
Q3	A citation to Fig. 5 has been added here, please check that the placement of this citation is suitable. If the location is not suitable, please indicate where in the text the citation should be inserted.	
Q4	Ref. 22 and 23: Can these references be updated? If so, please provide the relevant information such as year, volume and page or article numbers as appropriate.	
Q5	Ref. 26–28: Please provide the journal title.	

PAPER

Experimental study and computational modelling of cruzain cysteine protease inhibition by dipeptidyl nitriles

Q1

Alberto Monteiro Dos Santos,[a] Lorenzo Cianni,[b] Daniela De Vita,[b] Fabiana Rosini,[b] Andrei Leitão, [ID] [b] Charles A. Laughton, [ID] [c] Jerônimo Lameira [ID] *[a] and Carlos A. Montanari [ID] *[b]

Chagas disease affects millions of people in Latin America. This disease is caused by the protozoan parasite *Trypanossoma cruzi*. The cysteine protease cruzain is a key enzyme for the survival and propagation of this parasite lifecycle. Nitrile-based inhibitors are efficient inhibitors of cruzain that bind by forming a covalent bond with this enzyme. Here, three nitrile-based inhibitors dubbed Neq0409, Neq0410 and Neq0570 were synthesized, and the thermodynamic profile of the bimolecular interaction with cruzain was determined using isothermal titration calorimetry (ITC). The result suggests the inhibition process is enthalpy driven, with a detrimental contribution of entropy. In addition, we have used hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) and Molecular Dynamics (MD) simulations to investigate the reaction mechanism of reversible covalent modification of cruzain by Neq0409, Neq0410 and Neq0570. The computed free energy profile shows that the nucleophilic attack of Cys25 on the carbon C1 of inhibitiors and the proton transfer from His162 to N1 of the dipeptidyl nitrile inhibitor take place in a single step. The calculated free energy of the inhibiton reaction is in agreement with covalent experimental binding. Altogether, the results reported here suggests that nitrile-based inhibitors are good candidates for the development of reversible covalent inhibitors of cruzain and other cysteine proteases.

1 Introduction

Human American trypanosomiasis (Chagas disease) is a protozoan infection caused by the flagellate *Trypanosoma cruzi*.[1] Chagas disease occurs mainly in Latin America, where it is estimated that the illness affects about 6 million to 7 million people.[2] Although discovered in 1909 by the Brazilian sanitary doctor Carlos Chagas, the disease remains a serious public health problem in Latin America.[3] The drugs most commonly used to combat the disease, nifurtimox and benznidazole, are not very efficient and cause strong side effects.[4] Therefore, there is an urgent need for developing an effective therapy against this disease. The cysteine protease cruzain is considered an attractive target for therapeutic intervention in the treatment of Chagas disease.[5,6] Recently, a series of compounds based on the dipeptidyl nitrile scaffold were synthesized and assayed for their inhibitory activity against the *T. cruzi* cysteine protease cruzain.[7] It is important to point out that the presence of the electrophilic nitrile[8] in a molecular structure is not necessarily incompatible with achieving pharmacokinetic and toxicological profiles that are compatible with dosing in humans. For example, the cathepsin K inhibitor odanacatib[9] has been evaluated extensively as a treatment for osteoporosis.[10] This inhibitor has a nitrile 'warhead' to form a covalent bond with the catalytic cysteine in a reversible manner.[9,11,12]

Experimental data demonstrated that inhibition of cruzain by nitrile-based is both competitive and reversible.[7] Cruzain react *via* a nucleophilic reaction mechanism that involves a cysteine residue and the proton of a histidine. Recently, Moliner and coworkers have studied the catalytic mechanism of cruzain cysteine protease with computational techniques.[13] Using computational approaches, Moliner and coworkers also investigated the inhibition mechanism of cruzain by two different irreversible peptidyl halomethyl ketone (PHK) inhibitors.[14] Hitherto, the molecular details of the inhibition reaction mechanism of cruzain by nitrile-based inhibitors are still poorly understood.

[a] *Laboratório de Planejamento e Desenvolvimento de Fármacos, Universidade Federal do Pará, Cidade Universitária Prof. José da Silveira Netto, Rua Augusto Correa S/N, Belém-PA, Brazil. E-mail: lameira@ufpa.br*

[b] *Grupo de Química Medicinal do Instituto de Química de São Carlos da Universidade de São Paulo, NEQUIMED/IQSC/USP, 13566-590, São Carlos/SP, Brazil. E-mail: carlos.montanari@usp.br*

[c] *School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

This journal is © the Owner Societies 2018

Phys. Chem. Chem. Phys., 2018, **00**, 1–12 | 1
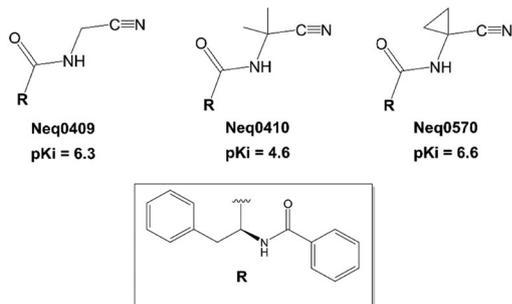
Scheme 1 Dipeptidyl nitriles cruzain inhibitors.

The catalytic mechanism of the cysteine proteases depends on a Cys and His residue. It has been proposed that the imidazole group of His polarizes the SH group of the Cys and enables deprotonation of the SH group, thus forming a highly nucleophilic CysS⁻/HisH⁺ ion pair.[15] The existence of the ion pair was experimentally proven by different studies[16–18] and confirmed by computational modelling.[19] Inhibitors that contain an electrophilic functional group, such as nitriles, covalently bind to cruzain via nucleophilic attack of the active site CysS⁻.[20] Experimental data suggest that dipeptidyl nitriles inhibitors bind tightly within the active site of cruzain and inhibit via the reversible formation of a covalent bond.[7] In this contribution, isothermal titration calorimetry (ITC) was used to assess experimental binding of cruzain to Neq0409, Neq0410 and Neq0570 inhibitors. To gain further insight into the inhibition of cruzain by these molecules, we also modelled the reaction of cruzain with three inhibitors (Scheme 1).

Scheme 2 shows two possible mechanisms for the inhibition reaction. In the first proposed mechanism, the protonation of N1 occurs concertedly with the nucleophilic attack of the thiolate anion (see Scheme 2). The second proposed mechanism is the nucleophilic addition and proton transfer in a stepwise mechanism, where the thiolate group of Cys25 attacks the carbon atom of the nitrile group first to form a covalent bond.

Recently, we have used free energy surface and quantum mechanics/molecular mechanics (QM/MM) approach,[21] to investigate biomolecular systems with an emphasis on enzymatic reactions.[22–24] In the QM/MM approach a small part of the system (ligand/substrate species) is described by quantum mechanics, while MM force fields represent the protein and solvent environment.[22] Herein, we are reporting the first QM/MM free energy surface study of the inhibition reaction of cruzain by dipeptidyl nitriles. Overall, we present a combined computational and experimental study aimed to gain insights into the mechanism whereby nitriles inhibitors bound to a cysteine protease.

2 Methods

2.1 Expression and purification of cruzain enzyme

2.1.1 Cell transformation procedure. 3 μL of the plasmid (80 ng μL⁻¹) Epoch, were added to 100 μL of Artic Express DE3-RIL (Agilent) *E. coli* cells. The solution was placed in an ice bath for 30 minutes. Subsequently, the heat shock was effected at 42 °C in a water bath for 30 seconds and was again glazed in an ice bath for another 3 minutes. Then 250 μL of SOC medium (Sigma-Aldrich) was added and the solution was placed in a shaker incubator at 37 °C at 250 rpm for 60 minutes. Meanwhile, one plate with LB-Agar medium was prepared containing 20 μg mL⁻¹ gentamicin and 100 μg mL⁻¹ ampicillin. After 1 hour of stirring, the solution of the bacterium with the plasmid was placed on the plate and left in an oven (New Ethics) at 37 °C overnight. The next day, bacterial colonies containing the plasmid appeared on the plaque. Expression was attained with medium of high induction, which was prepared using the following reagents for 1 L of culture medium. Autoclaving the mixture: N-Z amine (Sigma Aldrich), 10 g; yeast extract (Sigma Aldrich), 5 g; glycerol (Vetec), 5 mL; deionized water (MilliQ-Millipore, Direct Q3), 900 mL. Added salts (also pre-autoclaved): Na₂HPO₄ (Sigma Aldrich), 25 mL of 1 M (stock); KH₂PO₄ (Alfa Aesar) 25 mL of 1 M; NH₄Cl (Vetec) 50 mL of 1 M; Na₂SO₄ (Sigma Aldrich) 5 mL of 1 M; MgSO₄ (Sigma Aldrich) 2 mL of 1 M. Added antibiotics: ampicillin (Sigma-Aldrich), 1 mL from 100 mg mL⁻¹ stock solution (final concentration 100 μg mL⁻¹); gentamicin (Sigma Aldrich), 400 μL from 50 mg mL⁻¹ stock solution (final concentration 20 μg mL⁻¹). Added sugars: ᴅ-glucose (Vetec), 10 mL 5% (w/v) stock or 5 g in 100 mL of deionized water. α-Lactose (monohydrate, Sigma-Aldrich),
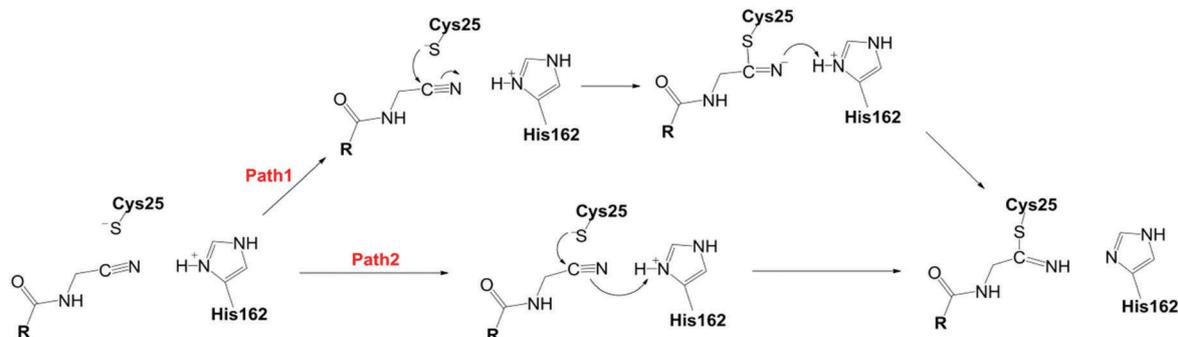
Scheme 2 Possible reaction mechanisms catalyzed by cruzain in complex with dipeptidyl nitrile derivatives inhibitors. Path 1 corresponds to a concerted mechanism and Path 2 illustrates the stepwise mechanism.

2 | *Phys. Chem. Chem. Phys.*, 2018, **00**, 1–12

This journal is © the Owner Societies 2018

40 mL of 5% (w/v) or 10 g in 200 mL of deionized water. A 0.2% lactose solution was used to induce the bacteria to express the desired protein.

2.1.2 Pre-inoculum. Gentamicin (20 µg mL^{-1}) and ampicillin (100 µg mL^{-1}) were placed in autoclaved three flasks of 150 mL culture medium (LB, Sigma Aldrich). In the laminar flow hood, a colony (the most isolated) was collected and added to each flask containing the culture medium and then placed in the shaker (Marconi, refrigerated incubator) at 37 °C overnight at 250 rpm. On the following day, the high induction medium was prepared by adding to the first autoclaved mixture (yeast extract, NZ amine, *etc.*), the previously autoclaved salts, sugars, antibiotics and the volume of solution with the bacterium sufficient to reach the optical density (OD) of 0.16. This preparation should be done in a sterile environment using a laminar flow hood or a Bunsen nozzle to prevent contamination. The mixture was placed in a shaker (Lab Companion Incubated Shaker model SIF600R) at 18 °C at 200 rpm for 72 hours. Subsequently, this solution was centrifuged at 5000 rpm (Hitachi Centrifuge model CR21GIII) at 4 °C and the pellet was frozen at −80 °C.

2.1.3 Purification. The following buffers are required for purification: buffer A: 50 mM Tris (Sigma-Aldrich), 300 mM NaCl (Synth), 10 mM imidazole (Sigma Aldrich) pH = 10. Buffer B: 50 mM Tris (Sigma Aldrich), 300 mM NaCl (Sigma-Aldrich), 10 mM imidazole (Sigma Aldrich) pH = 10. Lysis solution: for a 500 mL pellet of the culture medium, 60 mL of buffer A; 0.5 mM CaCl$_2$ (Sigma-Aldrich); 0.5 mM MgSO$_4$ (Sigma Aldrich), 20 µL DNase (Promega), a few milligrams of lysozyme (Sigma-Aldrich) and 100 µM PMSF (phenylmethylsulfonyl fluoride, Sigma-Aldrich). The pellet was resuspended in the lysis solution and allowed to stand for one hour. Soon thereafter, it is sonicated using a Fisher Scientific Sonic Dismembrator (model 500) with 30 s pulse on and 30 s pulse off. When the sonication was completed, the solution was centrifuged at 19 000 rpm for 30 minutes (Hitachi centrifuge model CR21GIII). The supernatant solution was mixed with 5 mL nickel resin (Ni, Sepharose 6 fast flow, GE Healthcare) and allowed to stir for approximately 3 hours at 4 °C. For this procedure, a cold chamber was used in the laboratory. At the end of this time, the solution was placed on a bench column containing Ni resin (Ni Sepharose™ 6 Fast Flow – GE Healthcare) then the resin was washed with 50 mL of buffer A and shortly thereafter, the protein was eluted with 60 mL of buffer B. The eluted solution containing the protein was then placed on a cellulose dialysis tubing (Sigma Aldrich), dialysis membrane and left overnight in activation buffer. The composition of the activation buffer was: 100 mM NaAC (Sigma Aldrich), 5 mM NaAc (Sigma Aldrich) and 300 mM NaCl (Synth), pH = 5.5. The next day, the protein was placed in another liter of activation buffer for a few more hours. At the end, the pH (Qualxtron model 8010 pH) was measured, which was adjusted to 5.2 or 5.0 if the protein is precipitated. The protein concentration should be less than or equal to 0.5 mg mL to avoid precipitation during the activation step. If the concentration is higher in this step, the dilution should be done using the same buffer used in the dialysis. For the

determination of total protein concentration, the Denovix DS-11 + Spectrophotometer was used.

2.1.4 Activation step. In this step, 1 mM of β-mercaptoethanol (Sigma-Aldrich), which is used with cysteine reducing agent, was added and left in the Shaker (Marconi) at 37 °C under slow stirring for approximately one hour (the color of the solution passed from cloudy to milky white and then to colorless during this process). This is an important step where the enzyme loses its pro-domain. At the end of activation, the protein was concentrated using amicon Ultra, Merck-Millipore (10 kDa) and a bench centrifuge (Eppendorf centrifuge model 5804R) at 4500 RCF for 30 minutes. At the end of the process, it was aliquoted, frozen in liquid nitrogen and stored in a freezer at −80 °C (Ultra-Low, Sanyo).

2.2 PEAQ-ITC experimental procedure

2.2.1 Preparation of the sample. An aliquot of the protein (frozen at −80 °C in activation buffer pH = 5.5) was taken and a new dialysis was performed using the last buffer employed in the dialysis process during the purification of the protein. Dialysis and concentration were performed over the 30 to 60 minute period on Amicon ultra 10 kDa membranes (Merck/Millipore) using a 4500 RCF benchtop centrifuge (Eppendorf centrifuge model 5804R) at 4 °C. At the end, solutions from the cell and the syringe were prepared using the dialysis buffer at the concentrations desired for the assay. In that case, the inhibitor was placed in the syringe and the protein in the cell. The inhibitor solution was prepared at the concentration of 170 µM and the protein solution was 20 to 30 µM and was adjusted by the equipment software according to the actual concentration of protein that was interacting with the inhibitor. 0.001% Triton-X 100 (Sigma-Aldrich) was added to the protein solution to prevent aggregation formation. The same surfactant was added to the inhibitor solution to avoid unbalance between the syringe and the cell.

2.2.2 PEAQ-ITC run. A microcalorimeter Microcal PEAQ-ITC (Malvern Intruments) was used for carrying out the experiments. The parameters used in all tests are: temperature (°C), 25; reference power (µcal s^{-1}), 5; feed back, high; stirring speed (rpm), 750; initial delay (s), 90; injections number, 19; injection volume (µL), 2; duration (s), 4; spacing (s), 120; total sample cell volume, 200 µL; total syringe volume, 40 µL. The thermodynamic parameters were calculated by Microcal PEAQ ITC Analysis Software. All experiments were performed at least in duplicate.

2.3 Synthesis of compounds

2.3.1 General considerations. All reagents were purchased from Sigma-Aldrich or Combi-blocks and used as received. Solvents such as hexane, ethyl acetate, chloroform, dimethylformamide, methanol, were all purchased from Synth and VETEC. *N,N*-Diisopropylethylamine was distilled over CaH$_2$ before used. *N,N*-Dimethylformamide (DMF) was dried over 3 Å activated molecular sieves for 72 hours. Solvents used in high-performance chromatography (HPLC) were supplied by Tedia and used without further purification. All non-aqueous

This journal is © the Owner Societies 2018

Phys. Chem. Chem. Phys., 2018, **00**, 1–12 | 3

reactions were carried out under argon atmosphere in oven-dried glassware.

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 the constant coupling values J 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 purification of compounds 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).

The solvents were filtered through a 0.45 μm Merck-Millipore filter before the use and degassed in an ultrasonic bath. In the established HPLC protocol, chiral analysis was carried out at 32 °C (column oven) where not otherwise specified, using analytical Diacel column (IC-chiralpak: 5 μm, 250 mm × 4.6 mm), by isocratic elution with a flow rate of 0.5 (analytical) and 2.36 mL min^{-1} (semi-preparative). The mobile phase composition was acetonitrile–water (50 : 50) (v/v). Volumes of 10 μL (analytical) was injected. Quantification was carried out at 200–800 nm and the chromatographic run time varied according to the sample.

Specific rotations ([α]^T = 100α/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. T was set to be 25 °C. Samples were weight using a precision balance (Sartorius, Model CPA26P) and were fully dissolved in methanol (HPLC grade, Panreac). The rotations were measured using a Digital Polarimeter (P2000, Jasco): α = observed rotation in degrees; l = cell path length of 0.1 decimeter in length; c = concentration in g 100 mL^{-1}. Values were calculated using 5 measurements for each compound. Melting points were determined by a Quimica Micro MQAPF-302 apparatus and are uncorrected. Infrared spectra were obtained from FT-IR Bomen Hartman & Braun mod MB-102.

2.4 Procedure and spectral data

2.4.1 General synthesis of Boc-(L)phenylalanine-NC(R$_1$R$_2$)CN (4, 5, 6).
The synthesis of the dipeptidyl nitriles are depicted in Scheme 3. Aminoacetonitrile hydrochloride (299 mg, 2.45 mmol, 1.3 eq.) was added to a solution of Boc-(L)phenylalanine (500 mg,
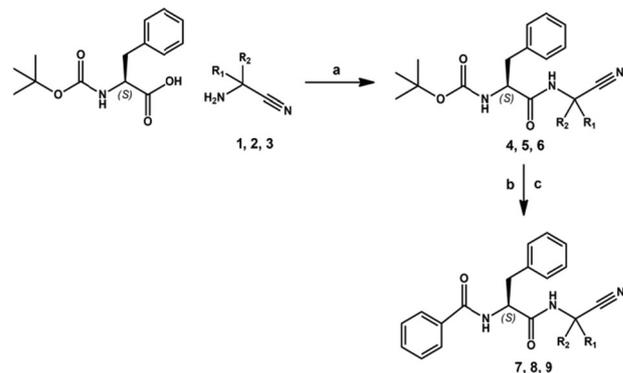
Scheme 3 (a) HATU, DIPEA, DMF, 16 h, RT, Ar; (b) HCOOH, 16 h; (c) benzoic acid, HATU, DIPEA, DMF, 16 h, RT. (1) R$_1$, R$_2$ = −H; (2) R$_1$, R$_2$ = −CH$_2$−; (3) R$_1$, R$_2$ = −CH$_3$; (4) R$_1$, R$_2$ = −H; (5) R$_1$, R$_2$ = −CH$_2$−; (6) R$_1$, R$_2$ = −CH$_3$; (7) R$_1$, R$_2$ = −H; (8) R$_1$, R$_2$ = −CH$_2$−; (9) R$_1$, R$_2$ = −CH$_3$.

1.88 mmol, 1 eq.), HATU (932 mg, 2.45 mmol, 1.3 eq.) and N,N-diisopropylethylamine (0.904 mL, 4.9 mmol, 2.6 eq.) in DMF (10 mL), under argon atmosphere. The resulting solution was stirred at room temperature for 16 hours. The reaction mixture was diluted with ethyl acetate (50 mL) and washed with a saturated NaHCO$_3$ solution (3 × 20 mL), HCl 1 N (2 × 20 mL) and brine (3 × 20 mL). The organic phase was dried over MgSO$_4$ and evaporated to give a crude residue that was used for the next step without further purification.

2.4.2 tert-Butyl-(1-((cyanomethyl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (4).
Yield 92%. White solid. R_f = 0.7 (ethyl acetate : n-hexane; 7 : 3). M.P. 133–136 °C. ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.27 (dd, J = 12.9, 6.0 Hz; 3H), 7.19 (t, J = 6.8 Hz, 2H), 4.87 (m, 1H), 4.24 (d, J = 5.5 Hz, 2H), 3.19 (dd, J = 13.8, 5.0 Hz, 1H), 2.88 (dd, J = 9.5, 5.0 Hz, 1H), 1.43 (s, 9H).

2.4.3 tert-Butyl-(1-((1-cyanocyclopropyl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (5).
Yield 90%. White solid. R_f = 0.9 (ethyl acetate : n-hexane; 7 : 3). M.P. 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).

2.4.4 tert-Butyl-(S)-(1-((2-cyanopropan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (6).
Yield 94%. White solid. R_f = 0.9 (ethyl acetate : n-hexane; 7 : 3). M.P. 120–123 °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), 1.68 (s, 3H), 1.67 (s, 3H), 1.31 (s, 9H).

2.5 General procedure for removal of Boc group

Compounds (4, 5, 6) were treated with formic acid (2 mL) at room temperature. The solution was stirred for 18 hours. The reaction mixture was evaporated under vacuum to get a yellowish oil. It was treated with 1 M NaOH until pH 9. The product was extracted with ethyl acetate (4 × 20 mL). The organic phase was evaporated to obtain a colorless oil. The formation of the product was confirmed by TLC (ethyl acetate). The product was used for the next step without any further purification. Yield ∼90%.

4 | Phys. Chem. Chem. Phys., 2018, 00, 1–12

This journal is © the Owner Societies 2018

2.6 General procedure for synthesis of compound 7, 8, 9

Compound **4** or **5** or **6** (2.45 mmol, 1.3 eq.) was added to a solution of benzoic acid (229.6 mg, 1.88 mmol, 1 eq.), HATU (932 mg, 2.45 mmol, 1.3 eq.) and *N*,*N*-diisopropylethylamine (0.904 mL, 4.9 mmol, 2.6 eq.) in DMF (10 mL), under argon atmosphere. The resulting solution was stirred at room temperature for 16 hours. The reaction mixture was diluted with ethyl acetate (50 mL) and washed with a saturated NaHCO$_3$ solution (3 × 20 mL), HCl 1 N (2 × 20 mL) and brine (3 × 20 mL). The organic phase was dried over MgSO$_4$ and evaporated to give a crude residue that was used for the next step without further purification.

2.6.1 (2*S*)-*N*-(Cyanomethyl)-3-phenyl-2-(phenylformamido)-propanamide (7) Neq0409.
Yield 85%. White solid. M.P. = 179–180 °C. ^{1}H NMR (400 MHz, DMSO-d_6) δ 3.02 (dd, 13.6, 10.8 Hz, 1H), 3.14 (dd, 13.7, 4.3 Hz, 1H), 4.17 (d, 5.6 Hz, 2H), 4.70 (ddd, 10.7, 8.4, 4.4 Hz, 1H), 7.17 (t, 7.3 Hz, 1H), 7.26 (t, 7.5 Hz, 2H), 7.34 (d, 7.3 Hz, 2H), 7.45 (t, 7.4 Hz, 2H), 7.52 (t, 7.3 Hz, 1H), 7.84–7.78 (m, 2H), 8.69 (d, 8.3 Hz, 1H), 8.79 (t, 5.5 Hz, 1H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 27.62, 37.25, 55.22, 117.32, 127.11, 127.49, 128.75, 128.87, 129.35, 132.30, 133.53, 137.24, 168.21, 172.88. IR (KBr) 3299, 3060, 2236, 1673, 1636, 1532, 1326, 693 cm^{-1}. HRMS (ESI(+)) calcd for [C$_{18}$H$_{17}$N$_3$O$_2$] 307.1320, observed: 308.1387 [M + H]$^+$. [α]$_D^{10}$ = −54.50°. HPLC (ACN : H$_2$O/50 : 50), t_R = 12.8 min. Purity 98.2% (220 nm).

2.6.2 (2*S*)-*N*-(1-Cyanocyclopropyl)-3-phenyl-2-(phenylforma-mido)propanamide (8) Neq0570.
Yield 82%. M.P. 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 (ddd, *J* = 9.6, 8.2, 5.4 Hz, 1H), 3.03 (ddd, *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 (126 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$_3$O$_2$] 334.1550, observed: 334.1550 [M + H]$^+$. [α]$_D^{10}$ = −35.05°. HPLC (ACN : H$_2$O/50 : 50), t_R = 10.1 min. Purity 99.2% (220 nm).

2.6.3 (2*S*)-*N*-(1-Cyano-1-methylethyl)-3-phenyl-2-(phenyl-formamido)propanamide (9) Neq0410.
Yield 91%. M.P. 114–117 °C. NMR ^{1}H (400 MHz, DMSO-d_6) δ 8.66 (s, H), 8.60 (d, *J* = 8.3 Hz, H), 7.86–7.78 (m, 2H), 7.51 (d, *J* = 7.3 Hz, H), 7.45 (t, *J* = 7.4 Hz, 2H), 7.36 (d, *J* = 7.1 Hz, 2H), 7.27 (t, *J* = 7.5 Hz, 2H), 7.18 (t, *J* = 7.3 Hz, H), 4.74 (td, *J* = 8.9, 5.9 Hz, H), 3.04 (dd, *J* = 11.5, 6.9 Hz, H), 1.58 (d, *J* = 15.1 Hz, 6H). NMR ^{13}C (126 MHz, DMSO-d_6) δ 171.74, 167.34, 137.56, 133.72, 132.14, 129.57, 128.82, 128.58, 127.61, 126.97, 121.89, 54.98, 49.00, 46.07, 37.62, 26.64, 26.35. HRMS (ESI(+)) calcd [C$_{20}$H$_{22}$N$_3$O$_2$] 336.1706. observed: 336.1689 [M + H]$^+$. HPLC (ACN : H$_2$O/ 50 : 50), t_R = 10.8 min. Purity 98.6% (220 nm).

2.7 Computational model

In this report, the computational model for the QM/MM MD calculations was based on the crystal structure of the cysteine protease cruzain (PDB code: 1ME4),[25] where we have replaced a ketone-based inhibitor by the dipeptidyl nitrile inhibitors studied here (see Scheme 1). Before starting the molecular dynamic simulations an assignment of the protonation states of the residues at pH = 7 was carried out using the empirical propKa program.[26] After adding the hydrogen atoms to the full structure, series of optimization algorithms[27] were applied. All the heavy atoms of the protein and substrate were restrained by means of a Cartesian harmonic umbrella with force constant of 500 kJ mol^{-1} Å^{-2}. Afterward, the system was fully relaxed, but the peptide backbone was restrained with a lower constant of 100 kJ mol^{-1} Å^{-2}. Then, the optimized protein was placed in an octahedral box of pre-equilibrated waters (80 Å side), using the center of mass of the complex as the geometrical center. Afterward, 100 ps of hybrid QM/MM Langevin–Verlet MD at 300 K and in a canonical thermodynamic ensemble (*NVT*) were used to equilibrate the system. For the hybrid QM/MM calculations, the atoms of the inhibitor and the side chain of Cys25 and His162 residues were selected to be treated by QM, using a semiempirical AM1/d-phot hamiltonian.[28] The other atoms of the system, protein and water molecules, were described using the CHARMM/TIP3P[29] force fields, respectively. The number of QM atoms resulted then to be 25 for Neq0409, 31 for Neq0410 and 29 for Neq0570 while the final system contains 33 098, 33 104 and 33 102 atoms, respectively.

Due to the number of degrees of freedom, any residue 20 Å apart from any of the atoms of the inhibitors was selected to be kept frozen in the remaining calculations. Cut-offs for the non-bonding interactions were applied using a switching scheme, within a range radius from 14.5 to 16 Å. Afterward, the system was equilibrated by means of 2.0 ns of QM/MM MD at temperature of 298 K. The computed RMSD for the protein during the last 1 ns renders a value always below 0.9 Å. Furthermore, the RMS of the temperature along the different equilibration steps was always lower than 2.5 K and the variation coefficient of the potential energy during the dynamics simulations was never higher than 0.3%.

2.7.1 Free energy surface.
Recently, the increased computer power, as well as parallel computing and advances in electronic structures calculations, have opened new possibilities for including *ab initio* QM/MM simulations upon understanding chemical reactions in a polar environment on a molecular level.[30] However, it is still extremely challenging to obtain computationally converging sampling of *ab initio* QM/MM (QM(*aî*)/MM) free energy surfaces in condensed phases (due to large number of gradient values evaluation). In the present case, we have described quantum region using AM1/d-PhoT hamiltonian[28] to compute the activation free energies at a significantly reduced computational cost, and that incorporates d-extension for the phosphorus atom and modified AM1 parameters for oxygen and hydrogen atoms. AM1/d-PhoT potential has been designed to accurately reproduce high-level DFT results such as geometries, dipole moments, proton affinities, and relative energies for a broad set of molecules, complexes, and chemical reactions relevant to biological transfers.[28]

To study the mechanism, we constructed 2D-PMF using the combination of R1–R2 antisymmetric distance that corresponds to the proton transfer from His162 to N1 of the
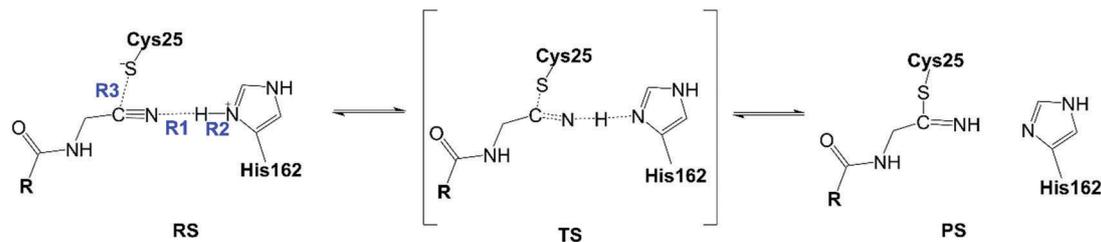
Scheme 4 Reaction coordinates and reaction inhibition of cruzain.

dipeptidyl nitrile inhibitor and R3 distance that corresponds to the nucleophilic attack of the negatively charged Cys25 on the carbon C1 of inhibitors (see Scheme 4). It is important to mention that the initial structures for 2D-PMF calculations were obtained from the last point of 2 ns QM/MM MD calculations for each system.

All PMFs have been calculated using the weighted histogram analysis method (WHAM) combined with the umbrella sampling approach[31,32] as implemented in the pDynamo program.[33] The PMF calculation requires series of molecular dynamics simulations in which the distinguished reaction coordinate variable, ξ, is constrained around particular values.[34] The values of the variables sampled during the simulations are then pieced together to construct a distribution function from which the PMF is obtained as a function of the distinguished reaction coordinate ($W(\xi)$). The PMF is related to the normalized probability of finding the system at a particular value of the chosen coordinate by eqn (1):

$$W(\xi) = C - kT \ln \int \rho(r^N)\delta(\xi(r^N) - \xi)\mathrm{d}r^{N-1} \qquad (1)$$

The activation free energy can be then expressed as:[34]

$$\Delta G^{\ddagger}(\xi) = W(\xi^{\ddagger}) - [W(\xi^{R}) + G_{\xi}(\xi^{R})] \qquad (2)$$

where the superscripts indicate the value of the reaction coordinate at the reactants and transition state and $G_{\xi}(\xi^{R})$ are the free energy associated with setting the reaction coordinate to a specific value at the reactant state. Usually, this last term makes a small contribution and the activation free energy is directly estimated from the PMF change between the maximum of the profile and the reactant's minimum:

$$\Delta G^{\ddagger}(\xi) \approx W(\xi^{\ddagger}) - W(\xi^{R}) = \Delta W^{\ddagger}(\xi) \qquad (3)$$

A total of 30 simulations was performed at different values of R1–R2 antisymmetric combination of distances (ranging from −1.5 Å to 1.5 Å), with an umbrella force constant of 500 kJ mol^{-1} Å^{-2} applied to this distinguished reaction coordinate. In addition, 36 simulations were performed at different values of R3 (ranging from 1.2 Å to 3.0 Å), also with an umbrella force constant of 500 kJ mol^{-1} Å^{-2} on this combination of distances. Consequently, 1080 simulation windows were needed to obtain the 2D-PMF for the Neq0409 inhibition mechanism. The values of the variables sampled during the simulations were then pieced together to construct a full distribution function from which the 2D-PMF was obtained. In each window, 10 ps of

relaxation were followed by 15 ps of production with a time step of 0.5 fs due to the nature of the chemical step involving a hydrogen transfer. The Verlet algorithm was used to update the velocities. The same protocol of the simulation was used to describe the 2D-PMF of the Neq0410 and Neq0570 inhibition mechanism, using the R2–R1 and R3 distances, where also 1080 simulation windows were needed to obtain the 2D-PMFs.

2.7.2 Activation free energy at M06-2X/6-31+G(d)/MM level. In the study, we have used AM1/d-PhoT/MM as the reference potential and M06-2X/6-31+G(d)/MM as target potential in order to explore the challenge of evaluation QM(ai)/MM activation free energies of a reaction catalyzed by cysteine protease cruzain. We have used the same protocol previously,[22] where we have obtained single-point at M062X/6-31+G(d)/MM potential using representative snapshots from RS and TS ensembles from 2D-PMF QM/MM simulations computed at AM1/d-PhoT potential, where a total of 2000 representatives' snapshots for each state were used in the simulations. In this way, the extensive sampling required to calculate the QM(ai)/MM free energy barrier can be done on a computationally inexpensive reference potential rather than directly on the high target potential.[35]

3 Results and discussion

3.1 Thermodynamic signature of ligand binding to cruzain

We started the work with isothermal titration calorimetry ITC to assess binding of cruzain to Neq0409, Neq0410 and Neq0570. Following experimental data, the binding enthalpy was exothermic for all ligands, Fig. 1. However, the three different cruzain inhibitors are not equipotent: Neq0570 (ΔG_{bind} = −9.0 kcal mol^{-1}) and Neq0409 (ΔG_{bind} = −8.9 kcal mol^{-1}) showed better binding affinities than Neq0410 (ΔG_{bind} = −7.5 kcal mol^{-1}). The structural variation of the inhibitors studied is shown in Scheme 1. The affinities of cruzain inhibitors are factored differently in their enthalpic and entropic contributions. The detrimental entropic contribution of Neq0409 as a whole is more effective than that observed for Neq0570, being Neq0410 the worst case. The introduction of the geminal methyl group in Neq0410 (see Scheme 1) causes the most significant detrimental entropic contribution to the binding free energy (Fig. 1 and Table 1). Fig. 2 shows the thermodynamic signatures of cruzain–ligand complexes as histograms (expressed in kcal mol^{-1}). Although there is no solution to the enthalpy–entropy compensation, there is a
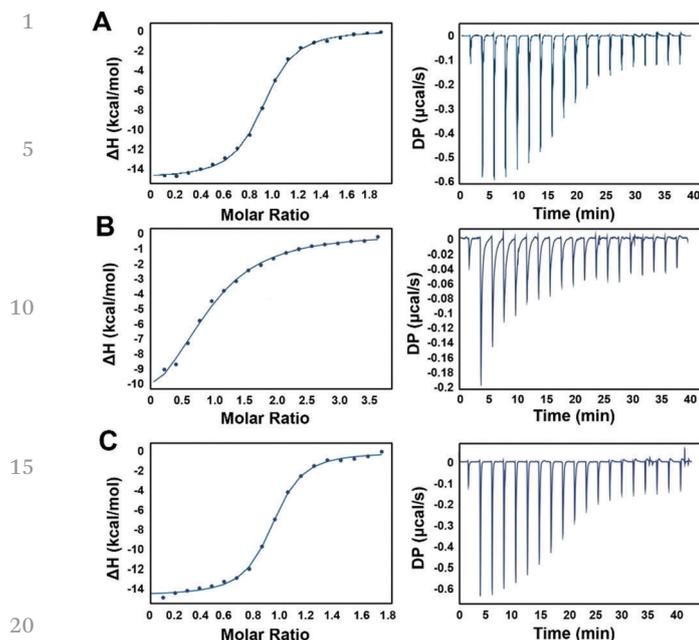
Fig. 1   (A) Neq0409, $\chi^2 = 2.9 \times 10^{-2}$ kcal mol^{-1}. (B) Neq0410, $\chi^2 = 3.7 \times 10^{-2}$ kcal mol^{-1}. (C) Neq0570, $\chi^2 = 4.5 \times 10^{-2}$ kcal mol^{-1}.
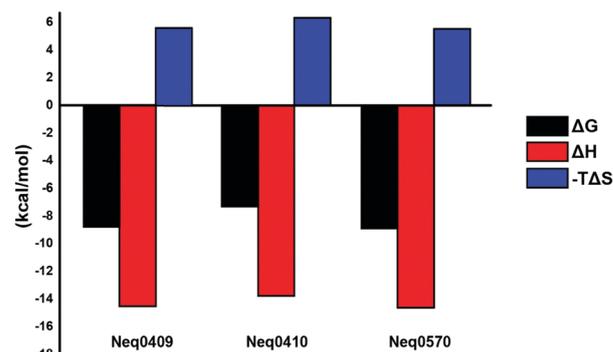

Fig. 2   Thermodynamic signatures of cruzain−inhibitor complexes Neq0409, Neq0410 and Neq0570. The Gibbs free energy binding values (ΔG; black), enthalpy (ΔH; red) and entropy ($-T\Delta S$; blue) are shown in the form of histograms (all data in kcal mol^{-1}).
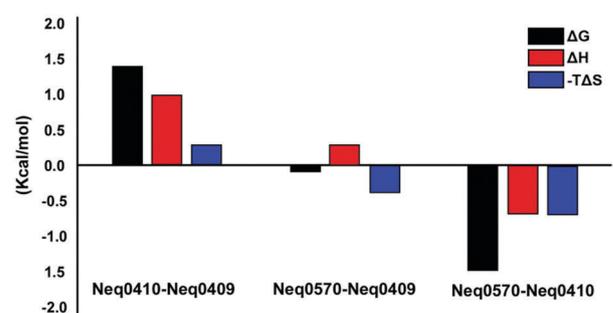

Fig. 3   Net thermodynamic profile for cruzain inhibitors. Values of dΔG, dΔH and $-T$dΔS are relative to unsubustitued Neq0409 (pairs Neq0410–Neq0409 and Neq0570–Neq0409) whereas that for pair Neq0570–Neq0410 is related to the exchange of the geminal methylene by the cyclopropyl moiety.

direct reflection on the affinity of Neq0410 which is advocated by the introduction of the geminal dimethyl moiety.

In addition, there is the danger that irreversibly modified proteins could cause an immune response. The precise effects of irreversible protein modifications are difficult to predict, and the risks are most relevant for the long-term treatment of chronic diseases. Therefore, the development of targeted covalent inhibitors would benefit from strategies that minimize the occurrence of such protein modifications.

When comparing and contrasting molecules of similar structure to the same target, the differences in ΔG (dΔG), ΔH (dΔH) and $-T\Delta S$ ($-T$dΔS) are more informative because the effect of removing solvent from the internal surfaces during complex formation will be canceled.

Matched Molecular Pair Analysis (MMPa) provides the opportunity to evaluate molecular modifications in which the thermodynamic contributions are rationalized according to the individual contributions of their unique elements. The analysis is performed in molecular pairs that differ only in positions where there is a modification in molecular structure. Transformations occur by identifying the change from one functional group to another in which its regiochemistry is also taken into account. Herein, an example of transformation refers to the exchange of a functional group (a fragment) by another whose

Table 1   TD parameters for Neq0409, Neq0570 and Neq0570

| Compound | TD parameters | | |
	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$-T\Delta S$ (kcal mol^{-1})
Neq0409	−8.9	−14.9	6.0
Neq0410	−7.5	−13.9	6.3
Neq0570	−9.0	−14.6	5.6

significant fragmental conversion reflects the change in the thermodynamic properties. MMPa may be useful in interpreting the bimolecular recognition process and the individual contributions of each alteration in that process. One of our objectives was to identify possible small molecular modifications that included simple fragments, but that resulted in some significant modification in the properties of interest. The subtle changes in the thermodynamic signature upon binding can be observed in Fig. 3.

The modifications implemented in the inhibitors (see Scheme 1) impact in the binding affinity and the thermodynamic profile of the ligands when co-complexed with cruzain. For comparisons between the pairs to be significant, the thermodynamic parameters were evaluated concerning the Neq0409. The pair Neq0410–Neq0409 presents a thermodynamic signature in which the introduction of the geminal methylene results in a total detrimental contribution towards binding and the three thermodynamic parameters are unfavorable to the spontaneity of the bimolecular interaction process. On the other hand, the introduction of the cyclopropyl moiety in Neq0409 (see Scheme 1) thereby yielding Neq0570 contributes favorably to the decrease in the detrimental entropic contribution that makes the spontaneous process to happen

This journal is © the Owner Societies 2018

Phys. Chem. Chem. Phys., 2018, 00, 1–12 | 7

even at the expense of the detrimental contribution of its net enthalpy. Nonetheless, values result in a small difference between dΔG, dΔH and $-T$dΔS.

The thermodynamic profile of the pair Neq0570–Neq0410 has the same signal for the contributions of dΔH and $-T$dΔS, both favorable to the change in the spontaneity, dΔG, of such bimolecular interaction.

At this point, it is important to clarify that in the processes of discovery of new bioactive chemical entities, it is accepted that molecules that occupy the same (or similar) chemical space within a set of properties exhibit potencies in similar magnitudes. Yet, it is quite common to find structurally similar molecules with high capacity prediction of the potency within such a chemical space, which, however, present great differences in potency. These discontinuities are known as "activity cliffs",[36] and it is exactly this phenomenon that we are describing in this work, where the compounds selected for this study contain in their structures a fundamental characteristic that is related to their peculiarly similar structure with discrete modifications. They range from a methylene group to cyclopropyl through the geminal methylene group (see Scheme 1). The change of affinity in the magnitude improves from 25.1 μM for Neq0410 to 0.25 μM (or 251 nM) for Neq0570.[7] In other words, Neq0570 is the most potent compound in the series with 2 log units in relation to the Neq0409. This observation is in agreement with our ITC results described above which show that inhibition of cruzain by Neq0570 is thermodynamically more favourable than Neq0410 (Fig. 3).

3.2 Computational modelling

Computational descriptions of chemical reactions in enzymatic environments are usually based on the selection of a distinguished reaction coordinates.[37] Recently, Moliner and coworkers have performed a computational study of the

inhibition of cruzain by two irreversible peptidyl halomethyl ketones inhibitors using hybrid AM1d/MM and M06-2X/MM potentials.[38] More recently, Chatterjee and coworkers[38] proposed an approach that can utilize relative FEP calculations to minimize the amount of QM/MM calculation needed to estimate the thermodynamics and kinetics of reversible covalent inhibitors. In this study, we are using free energy surface simulations to explore and test the cruzain inhibition reaction by dipeptidyl nitrile inhibitors Neq0409, Neq0410 and Neq0570.

The covalent reversible inhibition mechanism of a dipeptidyl nitrile inhibitor in cruzain requires the covalent link of Cys25 on the carbon C1 of the inhibitor, as shown by Avelar and coworkers[7] (Scheme 2). We have traced the 2D-PMFs using the combination of R1–R2 antisymmetric distance and R3 distance (see Scheme 4), where R1–R2 corresponds to the proton transfer from His162 to N1 of the dipeptidyl nitrile inhibitor, while R3 corresponds to the nucleophilic attack of negatively charged Cys25 on the carbon C1 of Neq0409, Neq0410 or Neq0570. Fig. 4a shows the 2D-PMF obtained for reaction inhibition of cruzain by Neq0409. The mechanism is described at AM1/d-PhoT/MM level, as a function of the combination of the distances (R1–R2, and R3, see Scheme 4) defining the breaking and forming bonds. In this PMF, the RS corresponds to the non-covalent cruzain–Neq0409 complex, a structure where the inhibitor binds tightly within the active site of the protein. In the TS (the transition state for both proton transfer and nucleophilic attack), the bond distance for R1 corresponds to 1.23 Å, whereas R2 corresponds to 1.33 Å, indicating a very advanced stage for the proton transfer, concomitantly the bond distance for R3 is verified as 2.21 Å, indicating the C–O bond formation (see Table 2 and Fig. 4).

Simulation of the reaction inhibition of cruzain by Neq0409 reveals that reaction proceeds *via* a concerted mechanism, where the proton transfer from His162 to N1 of Neq0409 co-
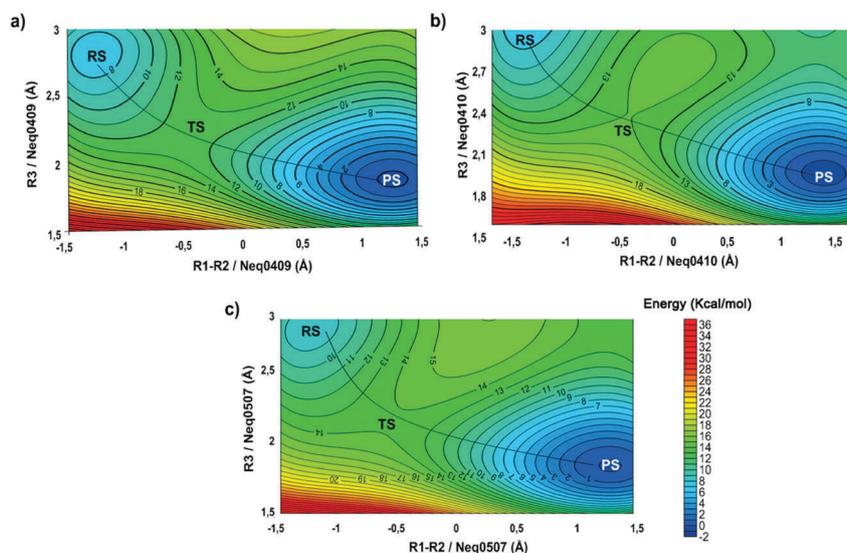
Fig. 4 2D-PMF (in kcal mol^{-1}) calculated at AM1/d-PhoT/MM level. R1–R2 and R3 antisymmetric combinations of distances describing the cruzain inhibition mechanism of dipeptidyl nitrile derivatives (a) Neq0409 (b) Neq0410 (c) Neq0570. Values of contour potential energy lines are reported in kcal mol^{-1} and coordinates in Å. Values on isoenergetic lines are reported for each 1 kcal mol^{-1}.

8 | *Phys. Chem. Chem. Phys.*, 2018, **00**, 1–12

This journal is © the Owner Societies 2018

Table 2 Average key distances (Å) for structures obtained from 2D-FESs[a]

Distance	RS	TS	PS
Neq0409			
R1	1.05	1.23	2.39
R2	2.30	1.33	1.05
R3	2.75	2.21	1.86
Neq0410			
R1	1.13	1.12	2.47
R2	2.43	1.51	1.06
R3	3.04	2.32	1.88
Neq0570			
R1	1.06	1.14	2.45
R2	2.54	1.76	1.05
R3	2.96	2.19	1.78

[a] Reaction coordinates R1–R3 defined in Scheme 4.

Table 3 Activation free energy calculations of cruzain inhibition at AM1/d-PhoT/MM and M06-2X/6-31+G(d)/MM levels (in kcal mol^{-1})

Method	Neq0409		Neq0410		Neq0570	
	$\Delta G^{\ddagger}$	ΔG	$\Delta G^{\ddagger}$	ΔG	$\Delta G^{\ddagger}$	ΔG
AM1/d 2D PMF	5.1	−7.5	6.9	−7.6	5.2	−8.3
M06-2X/6-31G(d)	14.8	−17.3	17.9	−18.9	16.0	−21.5

Q2

[a] The values on the table were obtained considering the perturbation from AM1/d-PhoT/MM reference potential to M06-2X/6-31G(d)/MM target potential.

occurs with a nucleophilic attack from negatively charged Cys25 at C1 of Neq0409. The calculated activation free energy at AM1/d-PhoT level is 5.1 kcal mol^{-1} (see Table 3), and the calculated reaction free energy is moderately exergonic (−7.5 kcal mol^{-1}). It is known from experiment that this reaction is reversible, where the covalent adduct can revert to the non-covalent protein–inhibitor complex, which means that the computed values at the AM1/d-PhoT level are in very good agreement with experimental data.

As commented in the Method section, we have used AM1/d-PhoT/MM as the reference potential and M06-2X/6-31+G(d,p)/MM as target potential to determine the higher level energy corrections for QM/MM activation free energies of cruzain inhibition mechanism by Neq0409. The representative snapshots from RS, TS and PS ensembles from AM1/d-PhoT 2D-PMF
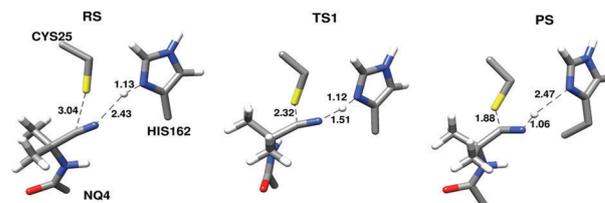

Fig. 6 Representative structures along the 2D-PMF obtained from the QM/MM free energy simulations for Neq0410. Structures named according to the locations along the 2D PMF.

were used as start point for evaluating the activation free energy of cruzain at M06-2X/6-31+G(d,p)/MM level following the approach described in methods (see ref. 22 for more details). It is worth to mention that estimating the average difference between target potential and reference potential allows us to compute the activation free energies at a significantly reduced computational cost, using 2 trajectories: one at the reactants and one at the transition state. Therefore, the reference potential idea seems to be a feasible approach regarding computational cost in calculating the M06-2X/6-31+G(d,p)/MM free-energy barrier for an enzymatic reaction. After M06-2X/6-31+G(d,p)/MM corrections, the resulting activation barrier at M06-2X/6-31+G(d,p)/MM level is 14.8 kcal mol^{-1} for the transfer of a proton from His162 to N1 and nucleophilic attack from negatively charged Cys25 at C1 of Neq0409 (concerted step). Computational work on the reference solution reaction for studies of cysteine protease found a barrier of 24 kcal mol^{-1} for nucleophilic attack of thiomethanol S^{-} on amide,[39] which suggest that for the corresponding reaction in water, the barrier can be about higher by 10 kcal mol^{-1}. It is also important to point out that there is a considerable difference between values computed at AM1d and M06-2X level (see Table 3). The convergence of one-step perturbation can be problematic when the difference between the reference potential (RP) and target potential (TP) is significant.[40] The full FEP treatment should be used to reduce the source of the error. Olsson and Ryde have proposed the use of reference-potential approach, converting the ligands with FEP at the molecular mechanics (MM) level and then perform also MM → QM/MM FEP for each ligand.[41] However, the one-step perturbation approach drastically reduces the computational cost of implementation of the perturbation from reference to target potential. The values reported in Table 3
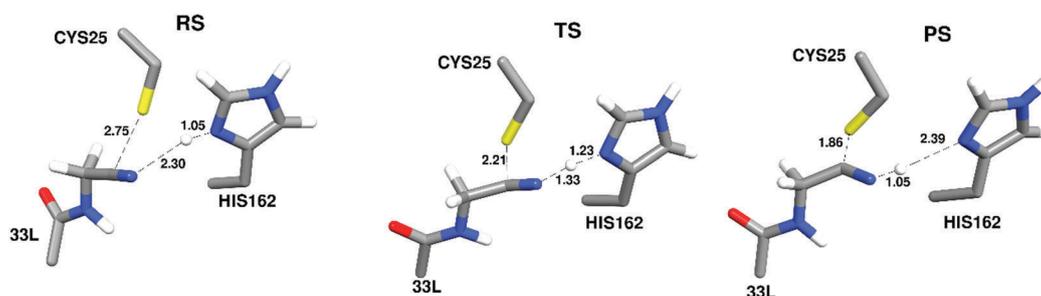

Fig. 5 Representative structures along the 2D-PMF obtained from the QM/MM free energy simulations for Neq0409. Structures named according to the locations along the 2D PMF.
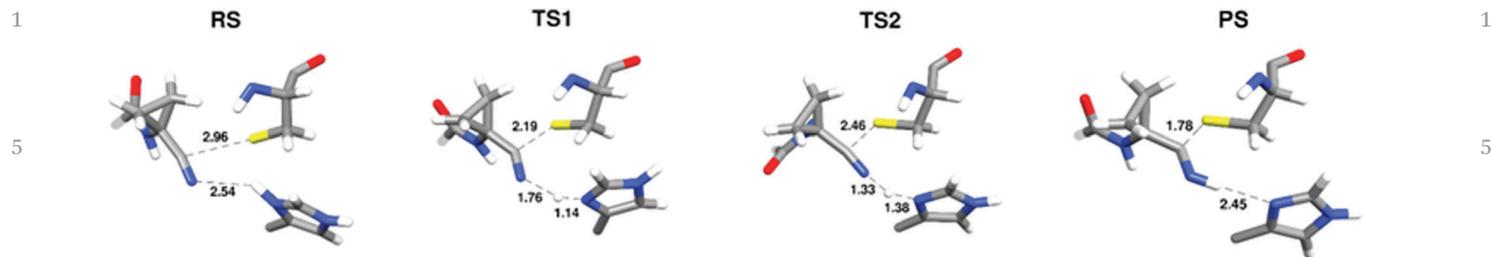
Fig. 7   Representative structures along the 2D-PMF obtained from the QM/MM free energy simulations for Neq0570. Structures named according to the locations along the 2D PMF shown in Fig. 4c for the Neq0570 simulation.

shows that energies have a consistent trend for each inhibitor of reaction which suggests that the perturbation was performed in the right direction. Here, we are assuming that AM1-d reproduces DFT geometry accurately and consequently we may be matching the correct stationary points. It is important to note that the nucleophilic addition barrier calculated at the M06-2X/6-31+G(d,p)/MM level is in reasonable agreement with that found in previous studies for inhibition of cathepsin K by nitrile-based inhibitors, for which a barrier of 19.9 kcal mol^{-1} was obtained from DFT/MM calculations (Fig. 5).[42]

In the present study, we have also explored the molecular mechanism of the cruzain inhibition by dipeptidyl nitrile inhibitors Neq0410 and Neq0570, where we have traced a 2D-FES using the same reaction coordinates described for Neq0409 (R1–R2, and R3, see Scheme 4). The results show a concerted mechanism for both Neq0410 and Neq0570 (Fig. 4). Furthermore, FES results suggest that the activation barrier for Neq0410 is more significant than the barrier observed for Neq0409 (see Fig. 4 and Table 3). Fig. 4b shows the 2D Free energy surface (2D-FES) for Neq0410 using the combination of distances R1–R2 and R3, as described above. The TS1 observed on the FES for inhibition by Neq0410 corresponds to the transition state for both proton transfer and nucleophilic attack. The bond distance for R1 and R2 corresponds to 1.12 and 1.51 Å, respectively (see Table 2 and Fig. 6), and the R3 is 2.32 Å, indicating a concomitant step of both proton transfer and nucleophilic attack.

The FES for Neq0570 (Fig. 4c) suggests that its activation barrier is also lower than that barrier for Neq0410 (Table 3). Fig. 4c shows the 2D Free energy surface (2D-FES) for Neq0570 using the combination of distances R1–R2 and R3, as described. The TS1 observed on the FES corresponds to the transition state for both proton transfer and nucleophilic attack, where bond distance for R1 and R2 corresponds to 1.38 and 1.33 Å, respectively (see Table 1 and Fig. 7), and the R3 is 2.46 Å, also indicating a concerted process.

The reactions of irreversible inhibitors were proposed to be more exothermic than −22 kcal mol^{-1} (see ref. 43). Analysis of the reaction energies obtained from QM/MM FES and ITC data reported here for the three inhibitors studied shows that the process is exergonic and reversible since the ΔG obtained from ITC and computational calculations are less exothermic than −22 kcal mol^{-1} (see Table 3), which is agree with previous experimental work[7] that have shown that inhibition process is reversible for inhibition reactions of cruzain by nitrile-based

inhibitors. Therefore, the Neq0570 and Neq0409 and Neq0410 inhibitors binding covalently and in a reversible fashion. It is important to point out that our QM/MM result is consistent with the ITC data that demonstrates Neq0570 and Neq0409 are more potent inhibitors of cruzain from *Trypanosoma cruzi* than Neq0410. Finally, we hope this protocol used here may be helpful for the development of reversible covalent inhibitors of cruzain.

4 Conclusions

In the present study, ITC analysis shows the cruzain inhibition by Neq0570 and Neq0409 and Neq0410 is enthalpy driven with a clear detrimental contribution of the entropy difference to the free binding energy. Furthermore, an important conclusion from the hybrid QM/MM approach and FESs obtained is that the inhibition reaction of cruzain by Neq0409, Neq0410 and Neq0570 follows a concerted mechanism. A proton transfer (PT) from His162 to N1 co-occurs with the nucleophilic attack of the negatively charged Cys25 on the C1 atom of inhibitors. The barrier for a forward reaction involving Neq0409 and Neq0570 inhibitors corresponds to 5.1 and 5.2 kcal mol^{-1}, respectively at AM1-d/MM level. The barrier for Neq0409 and Neq0570 is lower than the barrier for Neq0410 (6.9 kcal mol^{-1}). It is also worth to mention that results from the AM1-d/MM calculated activation free energy are in agreement with activation free energy derived from the M06-2X/6-31(d)/MM estimations. Our experimental and computational results also demonstrate that inhibition of cruzain by Neq0570 and Neq0409 is thermodynamically more favourable than Neq0410. Finally, we hope these results reported here may provide valuable information for designing new inhibitors of cruzain from *Trypanosoma cruzi* with known mode of action (MoA).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 C. A. Hoare and F. G. Wallace, Developmental Stages of Trypanosomatid Flagellates: a New Terminology, *Nature*, 1966, **212**, 1385–1386.

2 WHO, Chagas disease (American trypanosomiasis), http://www.who.int/chagas/en/.

3 A. Moncayo and A. C. Silveira, Current epidemiological trends of Chagas disease in Latin America and future challenges: epidemiology, surveillance, and health policies, *Mem. Inst. Oswaldo Cruz*, 2009, **104**, 17–30.

4 L. V. Kirchhoff, American Trypanosomiasis (Chagas' Disease) – A Tropical Disease now in the United States, *N. Engl. J. Med.*, 1993, **329**, 639–644.

5 P. S. Doyle, Y. M. Zhou, J. C. Engel and J. H. McKerrow, A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection, *Antimicrob. Agents Chemother.*, 2007, **51**, 3932–3939.

6 J. C. Engel, P. S. Doyle, I. Hsieh and J. H. McKerrow, Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection, *J. Exp. Med.*, 1998, **188**, 725–734.

7 L. A. 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. M. Orozco, J. F. R. Ribeiro, J. R. Rocha, F. Rosini and M. E. Saidel, Molecular design, synthesis and trypanocidal activity of dipeptidyl nitriles as cruzain inhibitors, *PLoS Neglected Trop. Dis.*, 2015, **9**, 1–24.

8 F. F. Fleming, L. Yao, P. C. Ravikumar, L. Funk and B. C. Shook, Nitrile-Containing Pharmaceuticals: Efficacious Roles of the Nitrile Pharmacophore, *J. Med. Chem.*, 2010, **53**, 7902–7917.

9 J. Y. Gauthier, N. Chauret, W. Cromlish, S. Desmarais, L. T. Duong, J. P. Falgueyret, D. B. Kimmel, S. Lamontagne, S. Léger, T. LeRiche, C. S. Li, F. Massé, D. J. McKay, D. A. Nicoll-Griffith, R. M. Oballa, J. T. Palmer, M. D. Percival, D. Riendeau, J. Robichaud, G. A. Rodan, S. B. Rodan, C. Seto, M. Thérien, V. L. Truong, M. C. Venuti, G. Wesolowski, R. N. Young, R. Zamboni and W. C. Black, The discovery of odanacatib (MK-0822), a selective inhibitor of cathepsin K, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 923–928.

10 A. G. Costa, N. E. Cusano, B. C. Silva, S. Cremers and J. P. Bilezikian, *Nat. Rev. Rheumatol.*, 2011, **7**, 447–456.

11 N. Asaad, P. A. Bethel, M. D. Coulson, J. E. Dawson, S. J. Ford, S. Gerhardt, M. Grist, G. A. Hamlin, M. J. James, E. V. Jones, G. I. Karoutchi, P. W. Kenny, A. D. Morley, K. Oldham, N. Rankine, D. Ryan, S. L. Wells, L. Wood, M. Augustin, S. Krapp, H. Simader and S. Steinbacher, Dipeptidyl nitrile inhibitors of Cathepsin L, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 4280–4283.

12 A. G. Dossetter, H. Beeley, J. Bowyer, C. R. Cook, J. J. Crawford, J. E. Finlayson, N. M. Heron, C. Heyes,

A. J. Highton, J. A. Hudson, A. Jestel, P. W. Kenny, S. Krapp, S. Martin, P. A. MacFaul, T. M. McGuire, P. M. Gutierrez, A. D. Morley, J. J. Morris, K. M. Page, L. R. Ribeiro, H. Sawney, S. Steinbacher, C. Smith and M. Vickers, (1R,2R)-N-(1-Cyanocyclopropyl)-2-(6-methoxy-1,3,4,5-tetrahydropyrido[4,3-b]indole-2-carbonyl)cyclohexanecarboxamide (AZD4996): A Potent and Highly Selective Cathepsin K Inhibitor for the Treatment of Osteoarthritis, *J. Med. Chem.*, 2012, **55**, 6363–6374.

13 K. Arafet, S. Ferrer and V. Moliner, Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease, *ACS Catal.*, 2017, **7**, 1207–1215.

14 K. Arafet, S. Ferrer and V. Moliner, First Quantum Mechanics/Molecular Mechanics Studies of the Inhibition Mechanism of Cruzain by Peptidyl Halomethyl Ketones, *Biochemistry*, 2015, **54**, 3381–3391.

15 J. W. Keillor and R. S. Brown, Attack of Zwitterionic Ammonium Thiolates on a Distorted Anilide as a Model for the Acylation of Papain by Amides. A Simple Demonstration of a Bell-Shaped pH/Rate Profile, *J. Am. Chem. Soc.*, 1992, **114**, 7983–7989.

16 D. J. Creighton, M. S. Gessouroun and J. M. Heapes, Is the Thiolate-Imidazolium ion pair the Catalytically Important Form of Papain?, *FEBS Lett.*, 1980, **110**, 3–6.

17 S. D. Lewis, F. A. Johnson and J. A. Shafer, Potentiometric Determination of Ionizations at the Active Site of Papain, *Biochemistry*, 1976, **15**, 5009–5017.

18 S. D. Lewis, F. A. Johnson and J. A. Shafer, Effect of Cysteine-25 on the Ionization of Histidine-159 in Papain As Determined by Proton Nuclear Magnetic Resonance Spectroscopy. Evidence for a His-159–Cys-25 Ion Pair and Its Possible Role in Catalysis, *Biochemistry*, 1981, **20**, 48–51.

19 D. Wei, X. Huang, M. Tang and C.-G. Zhan, Reaction pathway and free energy profile for papain-catalysed hydrolysis of N-acetyl-phe-gly 4-nitroanilide, *Biochemistry*, 2013, **52**, 5145–5154.

20 W. R. Roush, J. Cheng, B. Knapp-Reed, A. Alvarez-Hernandez, J. H. McKerrow, E. Hansell and J. C. Engel, Potent second generation vinyl sulfonamide inhibitors of the trypanosomal cysteine protease cruzain, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 2759–2762.

21 A. Warshel and M. Levitt, Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *J. Mol. Biol.*, 1976, **103**, 227–249.

22 A. M. dos Santos, A. H. Lima, C. N. Alves and J. Lameira, Unraveling the Addition–Elimination Mechanism of EPSP Synthase through Computer Modeling, *J. Phys. Chem. B*, 2017, DOI: 10.1021/acs.jpcb.7b05063. Q4

23 E. Araújo, A. H. Lima and J. Lameira, Catalysis by solvation rather than desolvation effect: exploring the catalytic efficiency of SAM-dependent chlorinase, *Phys. Chem. Chem. Phys.*, DOI: 10.1039/C7CP02811C.

24 M. Reis, C. N. Alves, J. Lameira, I. Tuñón, S. Martí and V. Moliner, The catalytic mechanism of glyceraldehyde 3-phosphate dehydrogenase from *Trypanosoma cruzi*

elucidated via the QM/MM approach, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3772.

25 L. Huang, L. S. Brinen and J. A. Ellman, Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain, *Bioorg. Med. Chem.*, 2003, **11**, 21–29.

26 C. R. Sondergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values, , 2011, 2284–2295.

27 N. Shirish Keskar and A. Wächter, A Limited-Memory Quasi-Newton Algorithm for Bound-Constrained Nonsmooth Optimization, , 2016, 1–20.

28 K. Nam, Q. Cui, J. Gao and D. M. York, Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms, , 2007, 486–504.

29 J. Huang and A. D. Mackerell, CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, *J. Comput. Chem.*, 2013, **34**, 2135–2145.

30 S. C. L. Kamerlin and A. Warshel, Multiscale modeling of biological functions, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10401–10411.

31 G. M. Torrie and J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, *J. Comput. Phys.*, 1977, **23**, 187–199.

32 B. Roux, The calculation of the potential of mean force using computer simulations, *Comput. Phys. Commun.*, 1995, **91**, 275–282.

33 M. J. Field, The pDynamo program for molecular simulations using hybrid quantum chemical and molecular mechanical potentials, *J. Chem. Theory Comput.*, 2008, **4**, 1151–1161.

34 G. K. Schenter, B. C. Garrett and D. G. Truhlar, Generalized transition state theory in terms of the potential of mean force, *J. Chem. Phys.*, 2003, **119**, 5828.

35 N. V. Plotnikov and A. Warshel, *J. Phys. Chem. B*, 2012, **116**, 10342–10356.

36 J. Husby, G. Bottegoni, I. Kufareva, R. Abagyan and A. Cavalli, Structure-based predictions of activity cliffs, *J. Chem. Inf. Model.*, 2015, **55**, 1062–1076.

37 J. Lameira, C. N. Alves, I. Tuñón, S. Martí and V. Moliner, Enzyme molecular mechanism as a starting point to design new inhibitors: a theoretical study of O-GlcNAcase, *J. Phys. Chem. B*, 2011, **115**, 6764–6775.

38 P. Chatterjee, W. M. Botello-Smith, H. Zhang, L. Qian, A. Alsamarah, D. Kent, J. J. Lacroix, M. Baudry and Y. Luo, Can Relative Binding Free Energy Predict Selectivity of Reversible Covalent Inhibitors?, *J. Am. Chem. Soc.*, 2017, **139**, 17945–17952.

39 M. Ŝtrajbl, J. Florián and A. Warshel, *Ab initio* evaluation of the free energy surfaces for the general base/acid catalyzed thiolysis of formamide and the hydrolysis of methyl thiolformate: a reference solution reaction for studies of cysteine proteases, *J. Phys. Chem. B*, 2001, **105**, 4471–4484.

40 J. Lameira, I. Kupchencko and A. Warshel, Enhancing Paradynamics for QM/MM Sampling of Enzymatic Reactions, *J. Phys. Chem. B*, 2016, **120**, 2155–2164.

41 M. A. Olsson and U. Ryde, Comparison of QM/MM Methods To Obtain Ligand-Binding Free Energies, *J. Chem. Theory Comput.*, 2017, **13**, 2245–2253.

42 M. G. Quesne, R. a Ward and S. P. de Visser, Cysteine protease inhibition by nitrile-based inhibitors: a computational study, *Front. Chem.*, 2013, **1**, 1–10.

43 M. Mladenovic, K. Ansorg, R. F. Fink, W. Thiel, T. Schirmeister and B. Engels, Atomistic insights into the inhibition of cysteine proteases: first QM/MM calculations clarifying the stereoselectivity of epoxide-based inhibitors, *J. Phys. Chem. B*, 2008, **112**, 11798–11808.