Electronic Supplementary Material (ESI) for RSC Advances.
This journal is © The Royal Society of Chemistry 2014

Targeting the Leishmania mexicana cysteine protease CPB2.8ΔCTE by decorated fused benzo[b]thiophene scaffold

A. Scala,[a] N. Micale,[a] A. Piperno,[a] A. Rescifina,[b] T. Schirmeister,[c] J. Kesselring,[c] and G . Grassi[a*]

[a] Dipartimento di Scienze Chimiche, Biologiche, Farmaceutiche ed Ambientali, Università di Messina, V.le F. Stagno D'Alcontres 31, 98166 Messina, Italy. E-mail: ggrassi@unime.it

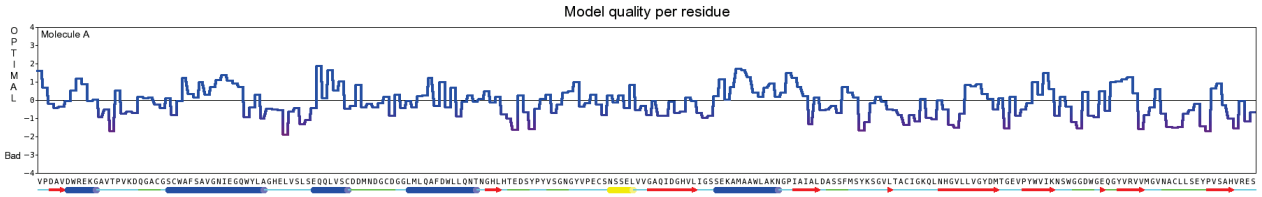
[b] Dipartimento di Scienze del Farmaco, Università degli Studi di Catania, V.le A. Doria, 95125 Catania, Italy.

[c] Institute of Pharmacy and Biochemistry, University of Mainz, Staudinger Weg 5, D 55099 Mainz, Germany.

Table of Contents

Assessment of Homology Model Quality

a) ModelQuality by YASARA[1]



Check type	Quality Z-score	Comment
Dihedrals	0.647	Optimal
Packing 1D	0.487	Optimal
Packing 3D	-1.551	Satisfactory
Overall	-0.437	Good

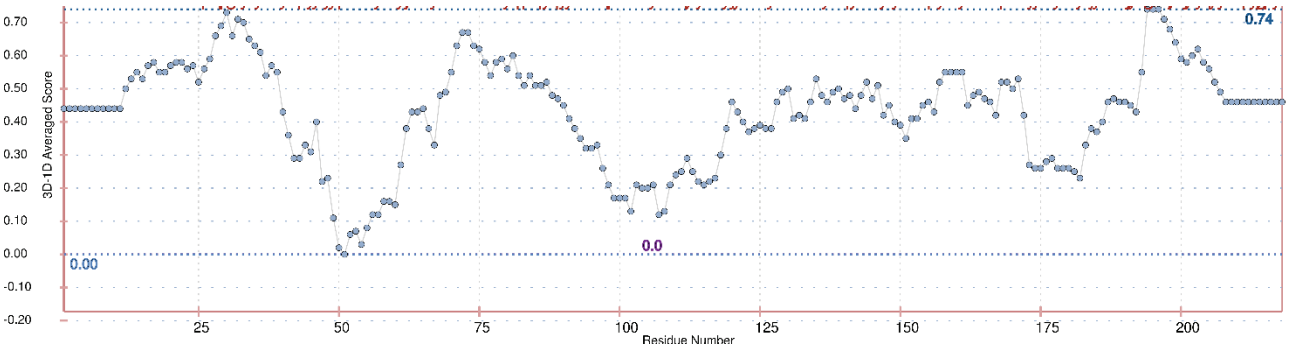
A Z-score describes how many standard deviations the model quality is away from the average high-resolution X-ray structure. Negative values indicate that the homology model looks worse than a high-resolution X-ray structure. The overall Z-scores for all models have been calculated as the weighted averages of the individual Z-scores using the formula Overall = 0.145*Dihedrals + 0.390*Packing1D + 0.465*Packing3D. The overall score thus captures the correctness of backbone-(Ramachandran plot) and side-chain dihedrals, as well as packing interactions.

b) Verify 3D Results plot[2, 3]

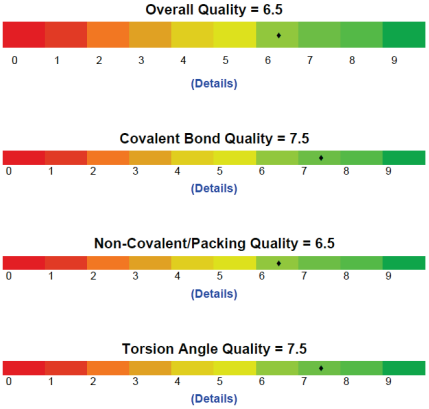
91.74% of the residues had an averaged 3D-1D score >= 0.2

Pass

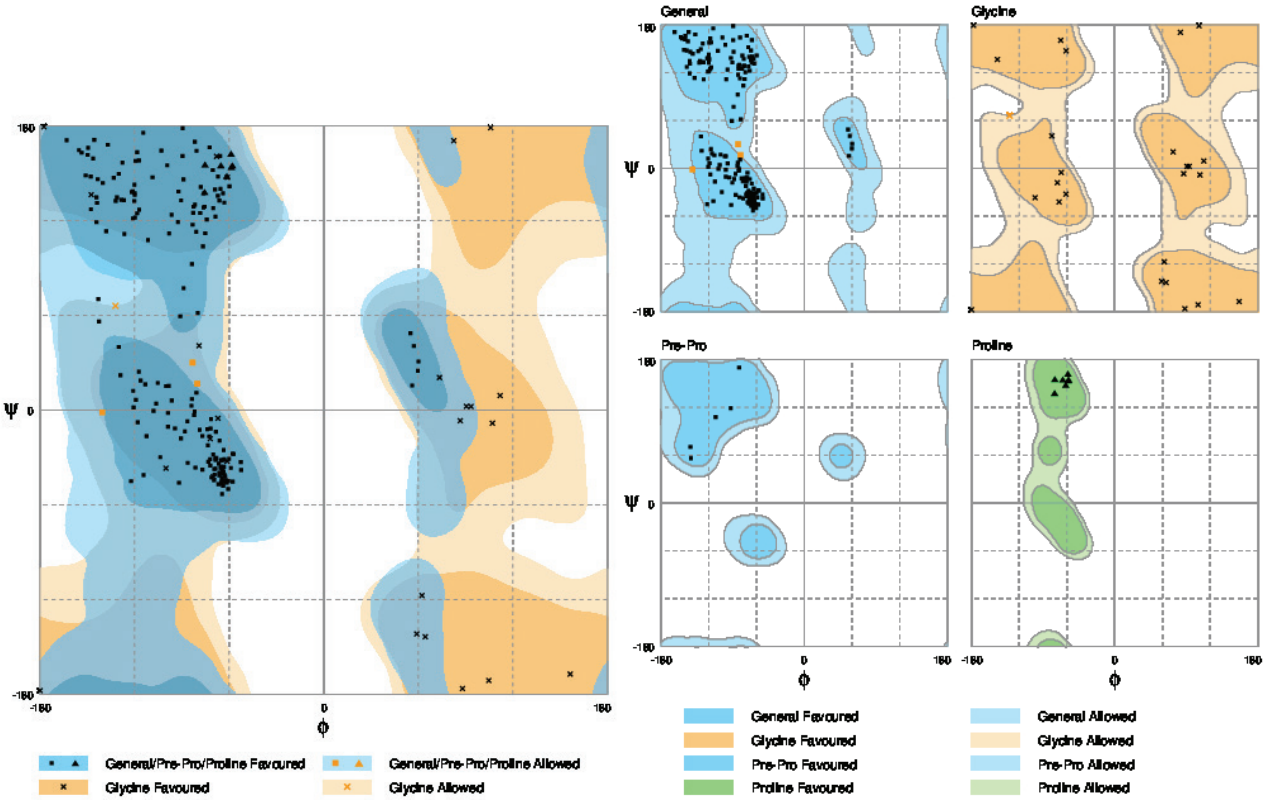
At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.



c) PROSESS Global Structure Assessment[4]



d) RAMPAGE: Assessment of the Ramachandran Plot[5]



Evaluation of residues
Residue [A 21 :ALA] (-83.14, 30.11) in Allowed region
Residue [A 24 :SER] (-80.12, 16.64) in Allowed region
Residue [A 158 :GLY] (-132.05, 66.00) in Allowed region
Residue [A 162 :ASN] (-140.40, -1.40) in Allowed region
Number of residues in favored region (~98.0% expected) : 212 (98.1%)
Number of residues in allowed region (~2.0% expected) : 4 (1.9%)
Number of residues in outlier region : 0 (0.0%)

e) Protein Structure Validation Suite Report

Structure Quality Analysis for CBP2.8DCTE

Analyses performed for order residues.

Procheck analysis, RMSD calculation and structure superimposition are based on Dihedral angle order parameter, with S(phi)+S(psi)>=1.8

Length (a.a.): 218
weight: 23367

Secondary Structure Elements:

alpha helices: 25A-40A, 50A-56A, 68A-78A, 123A-133A

beta strands: 4A-6A, 163A-171A, 136A-140A, 211A-214A, 115A-118A, 152A-153A, 178A-182A, 194A-198A, 82A-84A, 109A-112A

e1) Ramachandran Plot Summary from Procheck

Most favoured regions Additionally allowed regions Generously allowed regions Disallowed regions
90.7% 9.3% 0.0% 0.0%

e2) Ramachandran Plot Summary for from Richardson Lab's Molprobity

Most favoured regions	Allowed regions	Disallowed regions
98.1%	1.9%	0%

e3) Global quality scores

Program	Verify3D	ProsaII (–ve)	Procheck (phi-psi)	Procheck (all)	MolProbity Clashscore
Raw score	0.44	0.62	–0.25	–0.12	1.89
e4) Z-score[1]	–0.32	–0.12	–0.67	–0.71	1.20

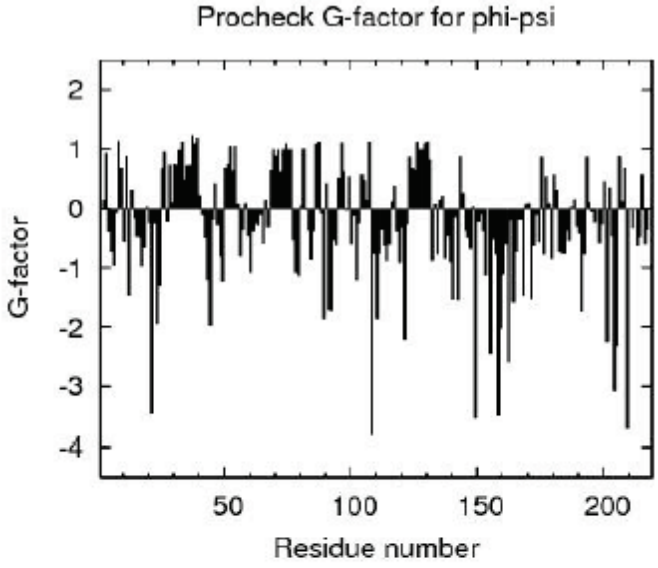
[1]With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score

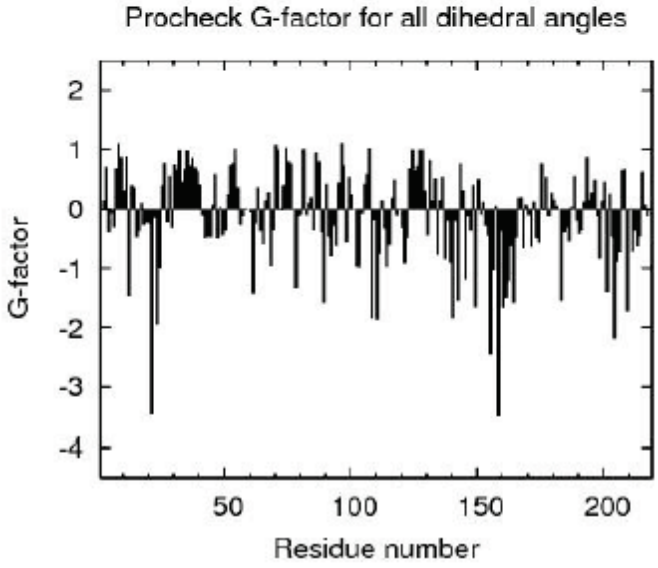
f) Close Contacts and Deviations from Ideal Geometry (from PDB validation software)

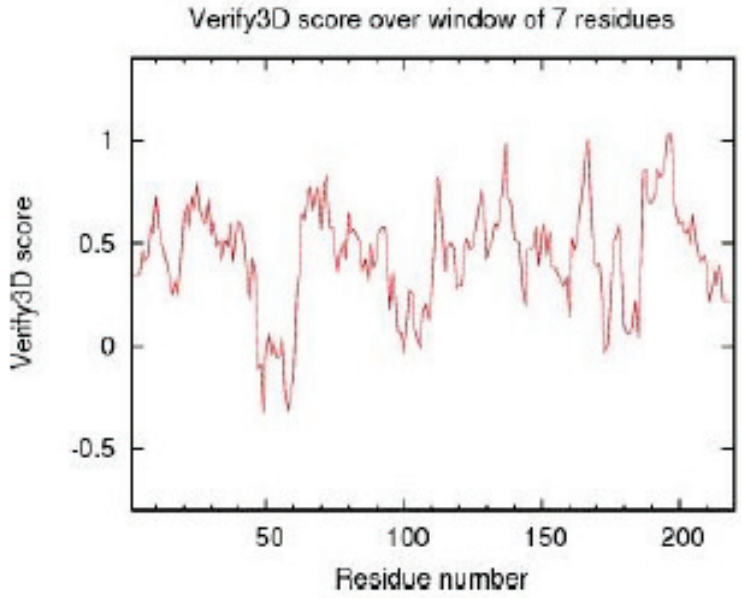
Number of close contacts (within 2.2 Å): 0

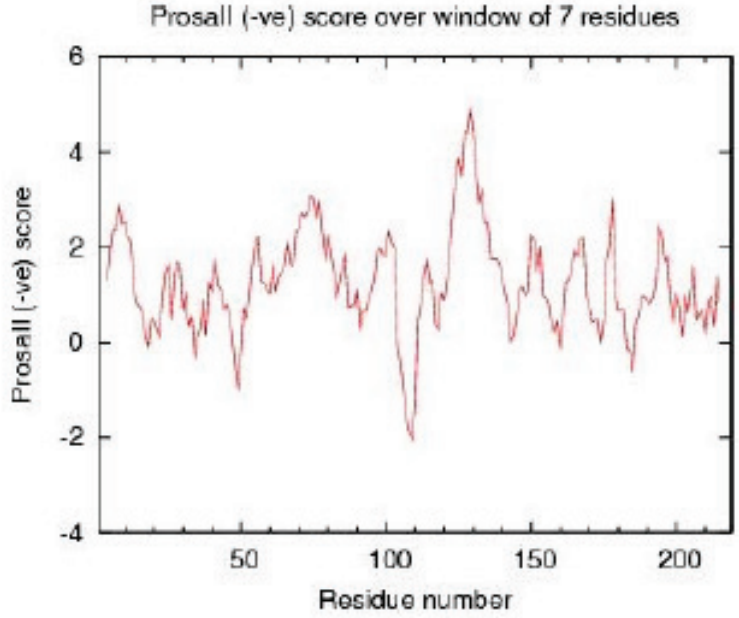
RMS deviation for bond angles: 1.7°

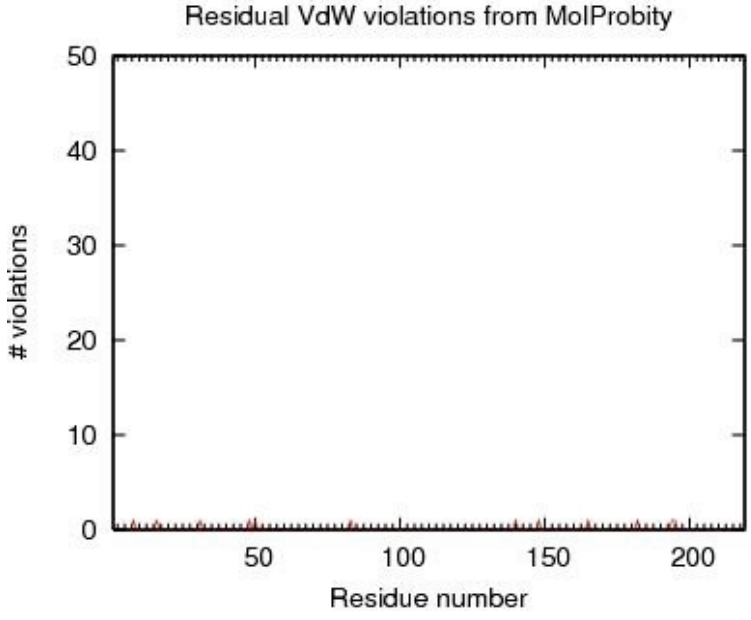
RMS deviation for bond lengths: 0.010 Å






Verify3D score over window of 7 residues
Verify3D score
1
0.5
0
-0.5
50
100
150
200
Residue number


Prosall (-ve) score over window of 7 residues
Prosall (-ve) score
6
4
2
0
-2
-4
50
100
150
200
Residue number

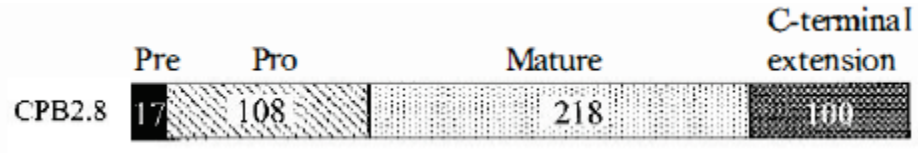

Residual VdW violations from MolProbity
violations
50
40
30
20
10
0
50
100
150
200
Residue number

Summary of structure quality factors

Structure Quality Factors - overall statistics	Mean score	SD	Z-score
Procheck G-factor (phi/psi only)	−0.25	N/A	−0.67
Procheck G-factor (all dihedral angles)	−0.12	N/A	−0.71
Verify3D	0.44	0.0000	−0.32
ProsaII (−ve)	0.62	0.0000	−0.12
MolProbity clashscore	1.89	0.0000	1.20
Ramachandran Plot Summary from Procheck			
Most favoured regions	90.7%		
Additionally allowed regions	9.3%		
Generously allowed regions	0.0%		
Disallowed regions	0.0%		
Ramachandran Plot Statistics from Richardson's			
Most favoured regions	98.1%		
Allowed regions	1.9%		
Disallowed regions	0%		

Considerations upon MD calculations

The cysteine proteinases of parasites in general, and the Type I of L. Mexicana in particular, encoded by cpb, are polymorphic and highly stage-regulated, with the highest level of activity occurring in the mature amastigote form.[6] Moreover, these Type I enzymes are distinguished from other cysteine proteinases by the possession of an unusual *C*-terminal extension, approximately 100 amino acids in length, whereas shared a pre- and a pro-region (Figure below).[7] L. mexicana CPB isoenzymes are expressed as inactive zymogens comprising an 17 amino acid pre-region, that is thought to be rapidly removed by a signal peptidase upon transfer into the endoplasmic reticulum, a 108 amino acid pro-region, a 218 amino acid mature domain that includes the active site, and a *C*-terminal domain of either 16 or 100 amino acids. The deletion of long *C*-terminal extension does not prevent in vitro processing of recombinant enzyme or abolish activity of mature enzyme. Conversion of zymogen into mature active CPB requires processing of the pro-region and possibly part of the *C*-terminal domain and results in mature enzymes.[8]



The cysteine proteinase enzyme used by us for MTT assays, CPB2.8, was expressed as an inactive pro-form lacking the characteristic *C*-terminal extension (CPB2.8ΔCTE) that by removal of the entire pro-region, give the full active "mature CPB2.8ΔCTE". This mature form is the most active and stable and all enzyme-ligand complexes of this typology, crystallized and studied by X-ray spectroscopy, are generally referred to the mature form and then lacking the first 125 amino acids and the last 100 ones. So, the 3D structure located at PDB 4PI3 (the mature form of cysteine protease of Trypanosoma

Cruzi), and used for the homology model, has residues 122–337, corresponding to the 216 residues of mature form. Also Paul M. Selzer et Al. built an homology model of mature CPB2.8ΔCTE using as template the PDB 1EWP containing residues 123–337.[9]

Since the in vitro assays are performed on the mature form it is reasonable to expect that the 3D structure is stable at physiological conditions and therefore even MD simulations are reliable, as confirmed by data on equilibrium (see below).

Validation of Molecular Dynamics Stability

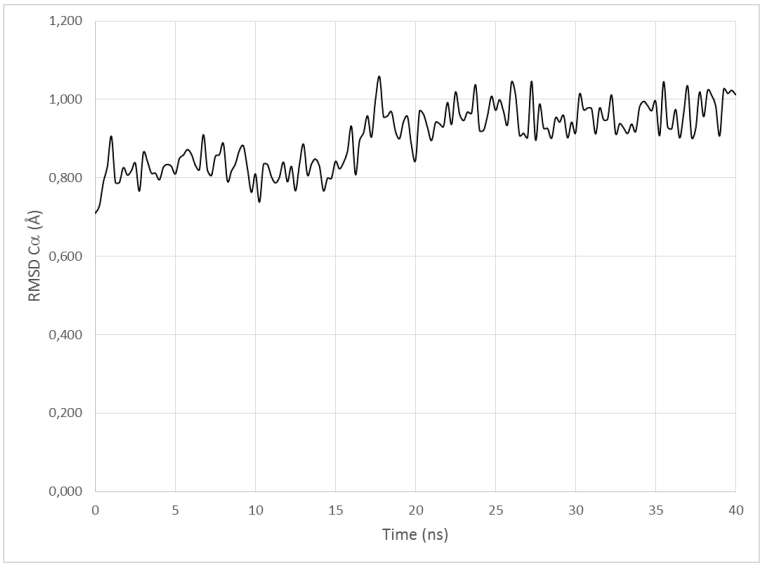


Fig. S1 RMSD Cα values during non-covalent **5**-H-CPB2.8ΔCTE complex MD simulations.

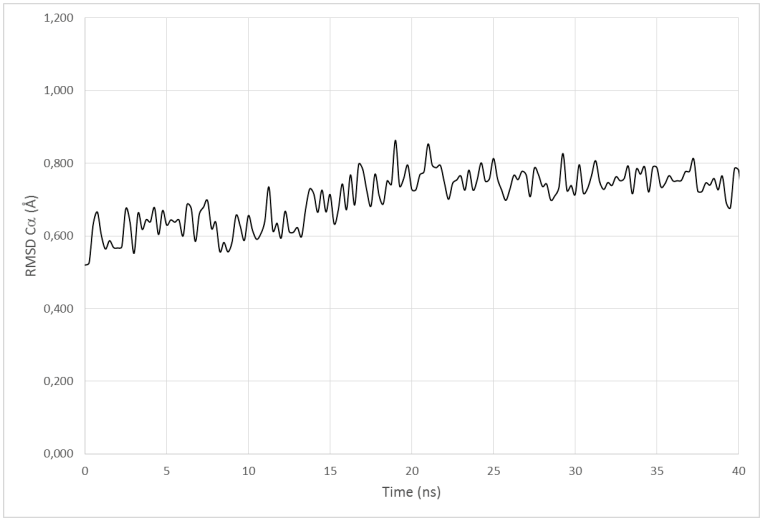


Fig. S2 RMSD Cα values during covalent **5**-H-CPB2.8ΔCTE complex MD simulations.

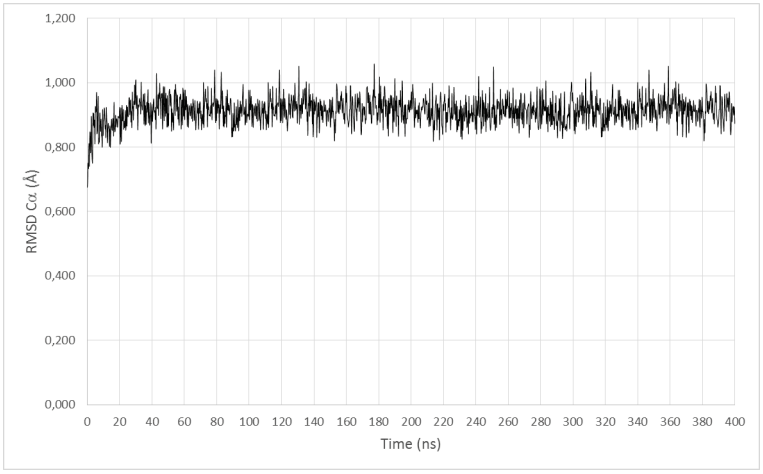
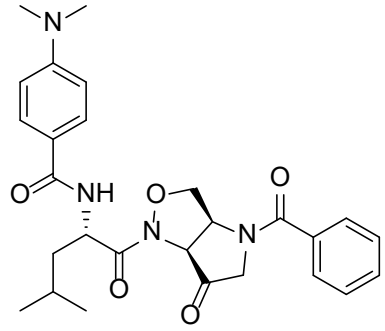
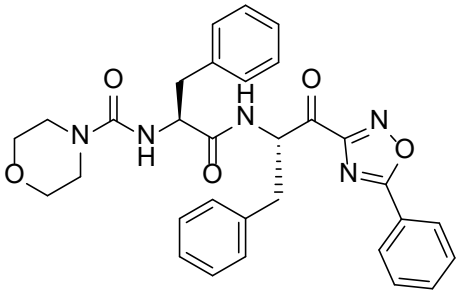


Fig. S3 RMSD Cα values during non-covalent re-docked **5**-H-CPB2.8ΔCTE complex MD simulations.

Validation of the homology model in performing a suitable level of docking accuracy

The validation was performed according to points i–iv of docking sequence, as described in the text, employing as ligands two molecules, **inhib_1**[10] and **inhib_2**,[11] that have been recognized as inhibitors of CPB2.8 and possessing inhibition constants that differ by one order of magnitude. The calculated K_i are reported in Table S1 and the interactions in Figure S4.

Table S1. Experimental and calculated K_i for **inhib_1** and **inhib_2** compounds.

Compound	Experimental K_i (μM)	Calculated ΔG_B (kcal/mol)	Calculated K_i (μM)
 **inhib_1**	1.0000	–7.5	3.1000
 **inhib_2**	0.0016	–10.9	0.0101

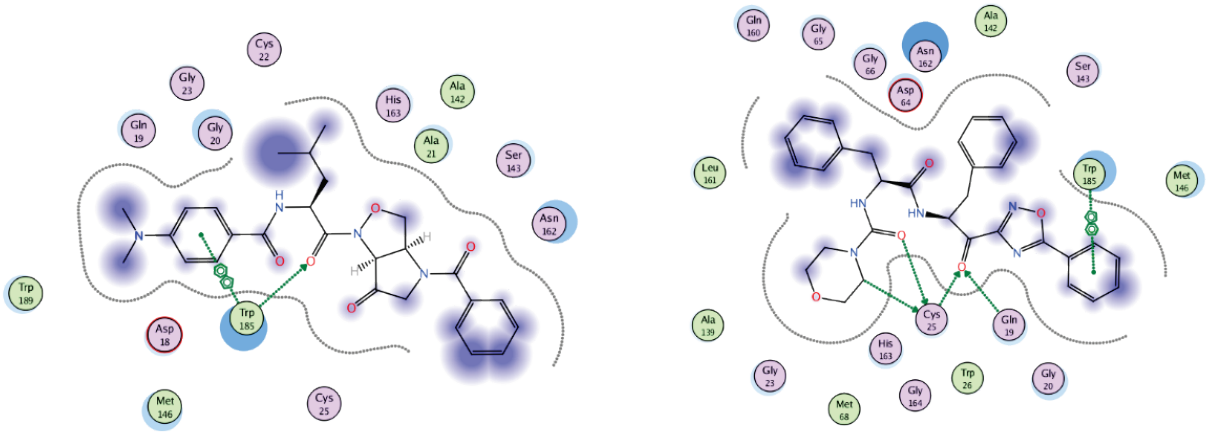


Fig. S4 2D sketch interactions of non-covalent docked pose of **inhib_1** (left) and **inhib_2** (right).

Table S2. Hydrogen bond interactions and their energies in non-covalent docked **5**-H-CPB2.8ΔCTE complex.

Ligand		Receptor		Interaction	Distance (Å)	E (kcal/mol)
S	13	SD	MET 146	H-donor	4.05	–0.5
O	1	NE2	GLN 19	H-acceptor	3.21	–1.0
O1	8	NE2	GLN 19	H-acceptor	2.98	–1.4
O1	8	NE1	TRP 185	H-acceptor	2.91	–2.0

Movie of ONIOM minimization in avi format

This movie evidence the nucleophilic attach of CYS 25 sulfur atom to C_a of ligand anhydride moiety and the formation of the intermediate.

References

1. Krieger, E.; 15.3.8 ed.; YASARA Biosciences GmbH: Vienna, Austria, 2013.
2. J. U. Bowie, R. Luthy and D. Eisenberg, *Science*, 1991, **253**, 164-170.
3. R. Luthy, J. U. Bowie and D. Eisenberg, *Nature*, 1992, **356**, 83-85.
4. M. Berjanskii, Y. J. Liang, J. J. Zhou, P. Tang, P. Stothard, Y. Zhou, J. Cruz, C. MacDonell, G. H. Lin, P. Lu and D. S. Wishart, *Nucleic Acids Res.*, 2010, **38**, W633-W640.
5. S. C. Lovell, I. W. Davis, W. B. Adrendall, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D. C. Richardson, *Proteins-Structure Function and Genetics*, 2003, **50**, 437-450.
6. J. C. Mottram, M. J. Frame, D. R. Brooks, L. Tetley, J. E. Hutchison, A. E. Souza and G. H. Coombs, *J. Biol. Chem.*, 1997, **272**, 14285-14293.
7. G. H. Coombs and J. C. Mottram, *Parasitology*, 1997, **114**, S61-S80.
8. S. J. Sanderson, K. G. J. Pollock, J. D. Hilley, M. Meldal, P. St Hilaire, M. A. Juliano, L. Juliano, J. C. Mottram and G. H. Coombs, *Biochem. J.*, 2000, **347**, 383-388.
9. J. Schröder, S. Noack, R. J. Marhöfer, J. C. Mottram, G. H. Coombs and P. M. Selzer, *PLoS One*, 2013, **8**.
10. Y. K. Wang, A. Benn, N. Flinn, T. Monk, M. Ramjee, J. Watts and M. Quibell, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 1327-1331.
11. K. Steert, M. Berg, J. C. Mottram, G. D. Westrop, G. H. Coombs, P. Cos, L. Maes, J. Joossens,