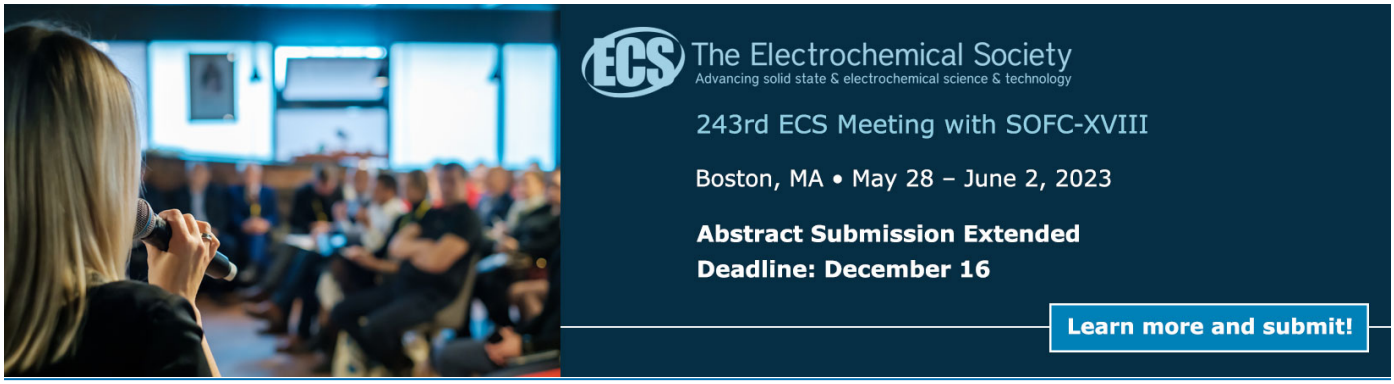
Mercapto-Benzamide Inhibitors effects on HIV NCp7 Protein: a parameter-free DFT based structural study



Mercapto-Benzamide Inhibitors effects on HIV NCp7 Protein: a parameter-free DFT based structural study

G Cappellini[1,2], R Cardia[1,3], M Valentini[3], E Pieroni[3]
[1]Dipartimento di Fisica, Università di Cagliari, 09042, Cagliari, Italy
[2]ETSF and Dipartimento di Fisica, Università di Cagliari, 09042, Cagliari, Italy
[3]CRS4, Modeling and Simulation Group, Bioscience Department, 09010, Pula, Italy

Abstract. The action of the Mercapto-Benzamide (MB) class of molecules on the HIV Nucleocapsid protein (HIV NCp7), a zinc finger protein, is an issue of relatively recent research interest, relevant to develop a new class of effective and well tolerated HIV antivirals, able to overcome virus escape strategies. MB molecules are easily and cheaply synthesized, and show the ability to unfold the HIV Zinc-finger region, thus avoiding effective viral replication. This effect is not still fully understood, and moreover is highly influenced by the precise composition of MB aromatic ring and chain. Our approach to this biological problem is to adopt a quantum parameter-free (ab-initio: AI) geometrical scheme based on density functional theory (DFT) for the treatment of the electronic degrees of freedom to study with atomistic resolution the action mechanism of MB molecules on NCp7. In particular with respect to the role played by each MB functional group.We report and discuss the outcomes of the here proposed DFT simulations with respect to the different final configurational structures obtained.

1. Introduction

HIV is still a major public health issue with more than 37 million people affected globally [1,2]. The therapeutical schemes being used to attack HIV infection inhibit the essential viral enzymes reverse transcriptase, protease, orintegrase [3].

The principal healing treatment for HIV is the highly active antiretroviral therapy (HAART), in which combinations of drugs simultaneously can target these viral enzymes [4]. Even if infected individuals who adhere to HAART can expect a normal life span, HAART requires regular and lifelong access to costly medication that often impedes many HIV-infected people from receiving the correct treatment. In addition HAART is affected by viral resistance [5] and may also lead patients to cardiovascular and neurological diseases [6,7].

As a consequence of the above points, it is important to continue the study and development of affordable and mutation-resistant small-molecule inhibitors of HIV. Nucleocapsid protein 7(NCp7, (see Figure 1), an 55-residue protein containing two highly conserved zinc-knuckle motifs,[8] results as an attractive target for antiretroviral drugs due to its essential role in viral replication and maturation [9].

NCp7 facilitates essential replication steps, such as strand transfers during reverse tran-scription of viral RNA and genome packaging during virion assembly [10].

In general, NCp7 has proven to be a difficult drug target [11], as there is no distinct three-dimensional structure for the protein other than the zinc-coordinated regions.

Furthermore, traditional medicinal chemistry approaches to target the active sites of enzymes are not applicable to a protein like NCp7. As a result there are currently no NCp7 inhibitors that are used clinically to treat HIV infection.

Some resarch groups are interested in the capacity of mercaptobenzamide(MB)-based molecules to

inhibit HIV-1 through the inactivation of NCp7 [12].
In this work we perform an AI structural minimization, based on DFT for the treatment of the electronic degrees of freedom, in order to study the action of the Mercapto-Benzamide (MB) class of molecules on the HIV Nucleocapsid [12] protein NCp7, a zinc finger protein [13].
More specifically, stimulated by some studies on animal models, we were interested in the impact of the binding of these molecules to the structure of the C-terminal HIV NCp7 "zinc-knuckle" motif [14,15,16]. This motif binds and stabilizes the viral RNA and is thus essential for viral replication and maturation [12,17,18]. Due to its relevance for virus reproduction, the zinc-knuckle is little or nothing muted across HIV strands and undergoes relatively little mutation upon host infection. Therefore, inhibition of NCp7 is a very attractive antiretroviral mechanism, able to escape viral mutation strategies even upon host infection [12,19,20,21]. Unfortunately, NCp7 is a difficult biological target and traditional medicinal chemistry approaches proved to be ineffective [12-15].
On the other hand, MB molecules are easily and cheaply synthesized, well tolerated and, as mentioned, some recent studies found they could be effective inhibitors of the Zinc-finger region, more precisely with a range of IC50 values ranging from 1 to 100 μM, strongly depending on very subtle changes to the aromatic ring and the MB molecular chain [12].
The precise mechanism of action of MB molecules is not fully understood, and the strong dependence on little changes of the atomic structure of the molecules remains definitely unclear. Moreover some MB forms are active in vitro while they lose their effectiveness in vivo and this fact is too yet poorly understood [15]. Fortunately, a 3D crystal structure of the two coordinated zinc-knuckles regions exists [22], that can be used to build the Zinc-finger motif model.
We adopted a modeling hypothesis, explained in the Numerical Methods section, that allowed us to chop the problem in simpler steps, in such a way to mimic complex biological mechanisms and such that each step can be simulated with the appropriate method.
The expected results will be fundamentals to design the molecule pharmacological activity.According to the existing literature, our hypothesis is that in the body biological environment, the MB2 molecules are partly acetylated to the form AcMB2, where the aromatic ring sulfhydryl group (SH) is replaced by the acetyl group ($SOCH_3$) (Figure 2). The acetylation of MB2 (a particular important form of MB molecules[12]) guarantees survival and migration of AcMB2 inside the cell, where it will eventually meet the NCp7 protein. Then, when the AcMB2 molecule interacts with the NCp7 protein, the acetyl group is transferred to NCp7 CYS413 (Figure 2), exposing the AcMB2 S ion, that is now free to interact with the motif coordinating Zn and eventually disrupts Zn coordination of the remaining CYS416, CYS426 and HIS421, leading to protein misfolding or even unfolding. Therefore we believe the active form of the MB2 molecule is precisely the AcMB2 molecule without the acetyl group and with the S ion exposed, indicated in the following as MB2act.

2. Modeling Protocol

According to our work hypothesis, we developed the following simplified modeling protocol:

- Protein model: we selected model number 1 from PDB NMR structure 2L44 [22] of the C-terminal zinc knuckle of the HIV NCp7. The structure is composed of 18 aminoacids representing the zinc knuckle motif Cys3-X2-Cys6-X4-His11-X4-Cys16 corresponding to NCP7 protein [Uniprot Q9YP46] aminoacids 411-429.
- AI DFT-based quantum simulation to optimize the MB2act molecular structure, and its electronic characterization (orbitals, ground state, ...).
- AI DFT-based quantum simulation of MB2act interacting with the Zn-finger motif in vacuum and with implicit water. The MB2act sulphur atom has been placed close to the zinc coordination atom.
- Analysis of the distance variations between the cysteine sidechain sulphur atoms, histidine nitrogen atom and the central coordinating zinc atom during the DFT runs is performed. To evidence structural changes, particularly motif unfolding, we also analysed the variations of distances between the NCp7 backbone Calphas relative to the histidine and cysteines that define the motif.

3. AI DFT Structural Simulations

We perform an AI DFT based [23,24] structural relaxation using the Berny algorithm [25-b] (in order to find the true local minimum on the potential energy surface of the system) as implemented in the Gaussian16 simulation package [25]. Geometry optimizations have been obtained using the Becke three-parameter Lee–Yang–Parr (B3LYP) hybrid electronic exchange–correlation (XC) functional [26,27], in combination with the 6–31 G* basis-set, a valence double-ζ set augmented with d polarization functions. The B3LYP electronic exchange-correlation (XC) functional has stable behavior, with some well-documented limitations [28]. The structural optimization of the system was performed without imposing symmetry constraints. The procedure here followed has been successfully tested for different systems, ranging from organic to inorganic clusters of different dimension and grade of complexity [29-33]. The calculation of the exact ground state is also of fundamental importance for systems of biological interest as the first step that should be accomplished for the calculation of electronic excitations and optical properties [34,35]. Notice that, being mainly interested to the protein structure evolution, we perform simulations with a total number of atoms quite larger than that used for a standard structural optimization. Moreover we performed all the main simulations using the SMD implicit solvent model [25-c] in order to take in account at least the main effects of the typical water environment in the living cells. In conclusion in this project the motivation to use the AI DFT scheme described above instead of a Molecular Dynamics (MD) one has been mainly related to the dimension of the studied protein-related problem which could be tackled safely within a parameter-free method, to the possibility to follow the creation and breaking of chemical bonds during the dynamics and to the possibility to find the real conformation of minimum energy of the studied system.

4. Results and Discussion

In the Figure 1 is reported within two different perspectives the Nucleocapsid protein 7(NCp7) which is the target of the present study. Reported in the right panel are also the main groups forming the protein, namely CYS426, HYA421,CYS416,CYS413.

In Fig.2 on the right panel the Mercaptobenzamide inhibitor is reported and in the left panel the final structure of unfolded NCp7 + MB2 inhibitor (considered in the implicit water simulation) obtained by the quantum relaxation sheme here used. In the left panel is also highlighted the final distance of Zn-CYS 426 after the structural relaxation**.**

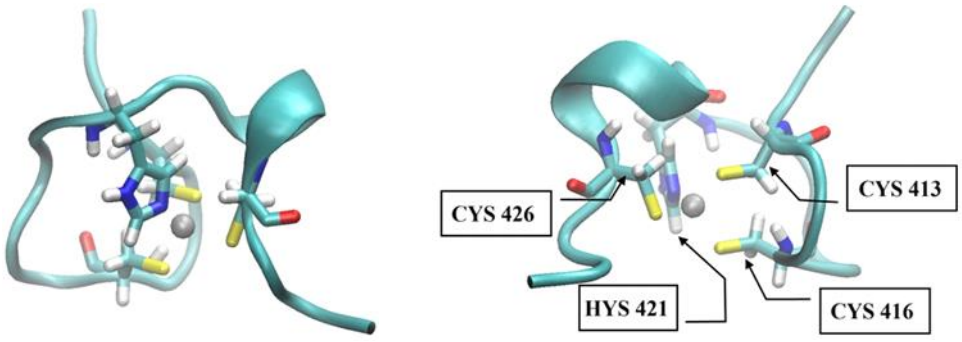
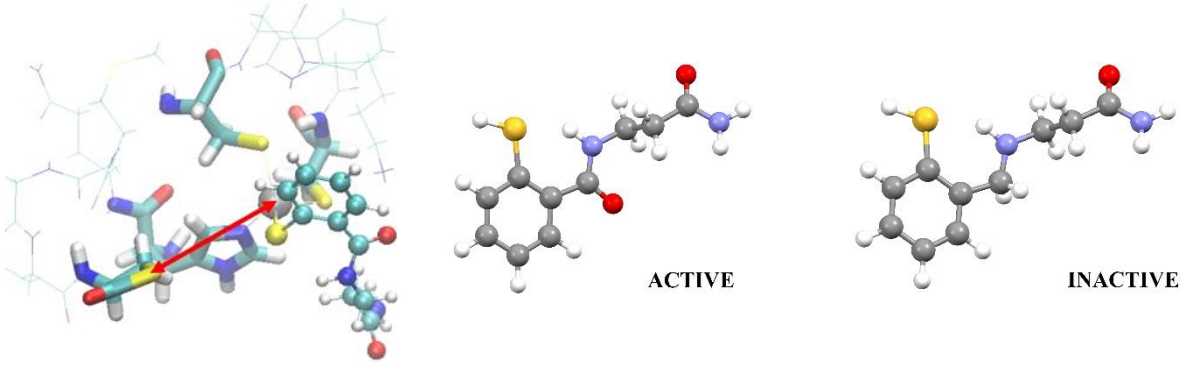


Fig. 1 Left panel Ncp7 secondary and tertiary structure. The 4 residues highlighted on the right panel are those forming the C-terminal Zinc-knuckle. Data taken from PDB file [22].

To make a more complete investigation we have also performed relaxation in case of a non active MB2 inhibitor. In Fig.(3) the outcome of that simulation are reported in the right panel. In this case the effect of the depontiated MB2 inhibitor is clearly visible in relation to the reduced opening of the zinc-

knuckle.
To have a quantitative insight in the effects of the MB2 inhibitor we report in Tab.(1) the distances between the Zn atom the groups of the NCp7 protein in the different simulation schemes. In the first column the experimental data after Ref. (22), in the second column the results of the simulation performed aftyer the optimization in vacuum. In paretheses the percentage of deviations with respect to the experiment.
The distance between the Zn atom and the CYS 426 group shows a consistent enlargement (more than 30%). In the third is present also the MB2 inhibitor. In this particular case the above distance is enlarged by more a factor 5, demonstrating the active action of the MB2 inhibitor and the unfolding of the Zn-knuckle. The consequence of the presence of real water molecules in the simulatioin is clearly highlighted in the values of the distances reported in the fourth column of this table. The distance between Zn atom and the CYS 426 group shows an enlargement with respect to the experiment. Interestingly, this enlargement is strongly hampered in case the simulation is started from the use of an MB2 inactivated inhibitor. The Zn-Cys 426 distance retourns to a value comparable with



that obtained for the free NCp7 protein in vacuum (see Fig.(3)).

Fig. 2 On the left the final structure of unfolded NCp7 + MB2 inhibitor (implicit water simulation). Highlighted the final distance of Zn-CYS 426 after the structural relaxation. In the center and right panels the Mercaptobenzamide inhibitor (Active form in center, Inactive form on the right).

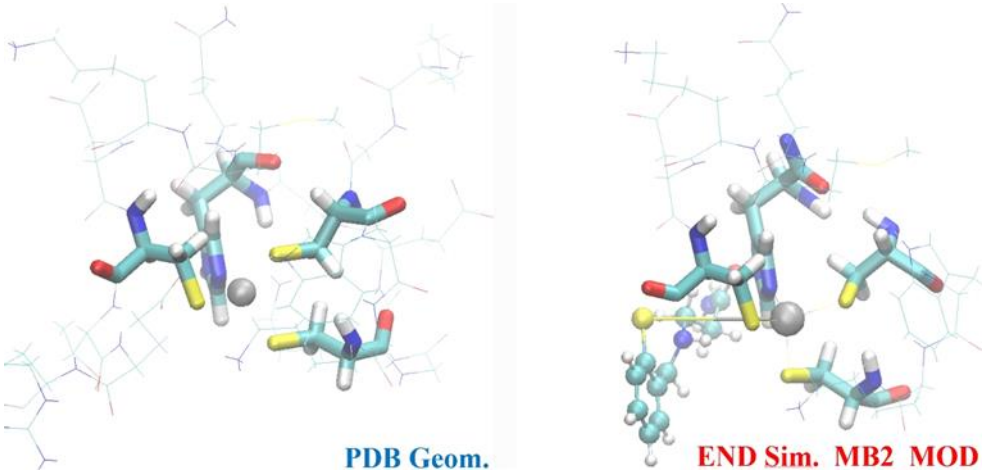


Fig. 3 Comparison between the experimental geometry of NCp7 (left panel) and final simulation frame in the case of non active MB2 inhibitor form (right panel, implicit water simulation).

Table I.The distance between the Zn atom the groups of the NCp7 protein in the different simulation schemes. In the first column the experimental data after reference Ref. (22), in the second column the results of the simulation performed after the optimization performed in vacuum. In the third is present also the MB2 inhibitor. In the fourth the MB2 inhibitor with implicit water (H_2O). In the fifth the MB2 inhibitor modified in the inactive form with implicit water (H_2O).

	NCp7 [Å] (Exp.)	NCp7 [Å] (Opt.Vacuum)	NCp7+MB2 [Å](Vacuum)	NCp7+MB2 [Å](H_2O)	NCp7 + MB2[Å] (Inactive MB2 - H_2O)
d $_{Zn\_Cys\ 413}$	*2.30*	*2.34* (+ 1.2 %)	*2.42* (+ 5.2 %)	*2.38* (+ 3.5 %)	*2.35* (+ 1.2 %)
d $_{Zn\_Cys\ 416}$	*2.27*	*2.25* (<1 %)	*2.44* (+ 7.5%)	*2.38* (+ 4 %)	*2.40* (+ 6 %)
d $_{Zn\_Cys\ 426}$	*2.37*	*3.17* (+ 34 %)	*13.03* (~ x 5.5)	*8.45* (~ x 3.6)	*3.24* (+ 37 %)
d $_{Zn\_Hys\ 421}$	*2.05*	*1.88* (+8.3 %)	*2.03* (+ 1 %)	*2.03* (+1 %)	*2.04* (<+1 %)

5. Conclusions

The effect of MB molecules upon HIV Zinc-finger region has been analysed within an “ab-initio” DFT based full quantum framework. Our simulations demonstrate that the main effects of the MB2 inhibitor is to unfold the Zn-knuckle of the on NCp7 protein, even in the presence of water. This result can be hampered by small structural changes in the MB2 inhibitor.

We believe the results presented here and their future development will be useful for rational drug design, including discovering interesting targets and understanding molecular action mechanisms.

Due to recent emerging results we are planning to continue and deepen this line of research by explicitly including the acetyl group and by MD simulation techniques to explicitly solvate the system and elucidate distinct protein equilibrium states, with potentially distinct roles on antiviral action mechanism.

6. Acknowledgements

The authors acknowledge the use of the ISCRA-CINECA facility within the ISCRA-C 2021 project HP10CDZJIB. G.C. acknowledges collaboration with "Progetto biennale d’Ateneo UniCa finanziato dalla Fondazione di Sardegna annualità 2020 : Molecular simulations and machine learning in bio- and medical physics: from antimicrobials to radiopharmaceuticals" and partial financial support from IDEA-AISBL-Bruxelles.

7. Bibliography

[1] AIDS by the numbers 2015,UNAIDS, November24, 2015:www.unaid-s.org/en/resources/documents/2015/AIDS_by_the_numbers_2015.
[2] Eriksen J et al. 2016 *Infect. Dis.* 1–34
[3] a) de Bethune M P 2010 *Antiviral Res.* **85** 75–90
b) Prokofjev M M et al. 2016 *Acta Naturae* **8** 23–32
c) Palomba M et al. 2016 *Org. Biomol. Chem.* **14**, 2015 –2024
d) Corona A et al. 2016 *ChemMedChem* **11** 1709 –1720
[4] Moore R D, Chaisson R E 1999 *AIDS* **13** 1933–1942
[5] Cortez K J, Maldarelli F, 2011 *Viruses* **3**, 347 –378
[6] Broder S 2010 *Antiviral Res.* **85**, 1–18
[7] Wensing A M et al. 2015 *Top. Antiviral Med.* **23**,132–141
[8] a) Berg J M 1986 *Science* **232** 485–487
b) Cornille F et al. 1990 *Int. J. Pept. Protein Res.* **36** 551–558
c) Morellet N et al. 1992 *EMBO J.* **11** 3059 –3065

[9] a) Rice G R et al. 1995 *Science* **270** 1194
b) Domagala J M et al. 1997 *Bioorg. Med. Chem.* **5** 569
c) Elslager E F et al. 1995 Oct 31 (Warner- Lambert Co.), US Pat. No. US5463122 A
d) Bolton G L et al. 1997 April 15 (Warner-Lambert Co.), US Pat. No. US5620997 A
e) Goldschmidt V et al. 2010 *HIV Ther.* **4** 179–198
[10] Muriaux D, Darlix J L 2010 *RNA Biol.* **7** 744 –753
[11] Spearman P 2016 *Curr. Top. Med. Chem.* **16** 1154–1166
[12] Saha M et al. 2017 *ChemMedChem* **12** 714 – 721
[13] Ren W et al. 2018 *PLoS ONE* 13 (5) e0196662
[14] Berg J M 1986 *Science* **232** 485-487
[15] Cornille F et al. 1990 *Int. J. Pept. Protein Res.* **36** 551-558
[16] Morellet N et al. 1992 *EMBO J.* **11** 3059-65
[17] Rice G R et al. 1995 *Science* **270** 1194
[18] Domagala J M et al. 1997 *Bioorg. Med. Chem.* **5** 569
[19] Iraci N et al. 2018 *Drug Discovery Today* **23-3** 687-695
[20] Sancineto L et al. 2018 *Drug Discovery Today*, **23-2,** 260-271
[21] Shvadchak V et al. 2018 *Biochemistry* **57-30** 4562-4573
[22] Quintal S et al. 2012 *Biochemistry* **51 (8)** 1752-1761
[23] Kohn W 1999 *Rev. Mod. Phys.* **71** 1253–1266
[24] Jones R.O., Gunnarsson O 1989 *Rev. Mod. Phys.* **61** 689–74
[25] a) Gaussian 16, Revision C.01, Frisch M J et al. 2016 Gaussian, Inc., Wallingford CT
b) Schlegel H B 1982 *J. Comp. Chem.* **3** 214-18
c) Marenich A V et al. 2009 *J. Phys. Chem. B* **113** 6378-6396
[26] Becke A D 1993 *J. Chem. Phys.* **98** 5648–5652
[27] Lee C et al. 1998 *Phys. Rev.* B **37** 785–789
[28] Grimme S, Parac M 2003 *ChemPhysChem* **4** 292–295
[29] Malloci G et al. 2004 *Phys. Rev.* B *Condens. Matter Mater. Phys.* **70** 1–6
[30] a) Cardia R et al. 2014 *J. Phys. Chem.* A **118** 5170–5177
b) Cardia R et al. 2016 *Phys. Rev.* B **93** 235132
c) Dardenne N et al. 2017 *J. Phys. Chem* C **121 (44)** 24480–24488
d) Antidormi A et al. 2018 *J. Phys. Chem* C **122 (49)** 28405–28415
[31] Cardia R et al. 2016 *Opt. Photonics J.* **6** 41–47
[32] a) Mocci P et al. 2018 *Journal of Physics: Conference Series* **956 (1)** 012020
b) Mocci P et al. 2018 *New Journal of Physics* **20 (11)** 113008
[33] Cappellini G et al. 2020 *ACS Omega* **5 (22)** 13268-13277
[34] Kumar A et al. 2018 Cellulose **25 (4)** 2191-2203
[35] Kumar A et al. 2019 Cellulose **26 (3)** 1489-1501