BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(14)

Insilico analysis and molecular docking of resuscitation promoting factor B (RpfB) protein of *Mycobacterium tuberculosis*

Baphilinia Jones Mylliemngap[1], Angshuman Borthakur[1], Devadasan Velmurugan[2] & Atanu Bhattacharjee[1]*

[1]Department of Biotechnology and Bioinformatics, North Eastern Hill University, Shillong-793022, Meghalaya; [2]Centre of Advanced Study in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai-600025, Tamil Nadu; Atanu Bhattacharjee – Email: atanubioinfo@gmail.com; *Corresponding author

Received June 23, 2012; Accepted June 26, 2012; Published July 21, 2012

Abstract:
Invulnerability of *Mycobacterium tuberculosis* to various drugs and its persistency has stood as a hurdle in the race against eradication of the pathogenecity of the bacteria. Identification of novel antituberculosis compounds is highly demanding as the available drugs are resistant. The ability of the bacteria to surpass the body's defenses and adapt itself to survive for disease reactivation is contributed by secreted proteins called resuscitating promoting factors (Rpfs). These factors aid in virulence and resuscitation from dormancy of the bacteria. Sequence analysis of RpfB was performed and compounds were first screened for toxicity and high-throughput virtual screening eliminating the toxic compounds. To understand the mechanism of ligand binding and interaction, molecular docking was performed for the compounds passing through the filter resulting with better docking studies predicting the possible binding mode of the inhibitors to the protein. Of all the active residues the binding conformation shows that residues Arg194, Arg196, Glu242, and Asn244 of the RpfB protein play vital role in the enzyme activity and interacts with the ligands. Promising compounds have been identified in the current study, thus holding promise for design of anti-tuberculosis drugs.

Key words: Resuscitating promoting factor B, *Mycobacterium tuberculosis*, molecular docking

Background:
Tuberculosis has been a challenge for human for ages claiming more lives than any other bacterial disease. With the availability of short-course chemotherapy (DOTS) and Bacille Calmette-Guerin (BCG) vaccine, the tubercle bacillus continues to claim more lives each year. The emergence of drug resistance-multidrug resistance (MDR) and extensive drug resistance (XDR) in tuberculosis is due to the extensive period of treatment in which patients fails to complete the therapy. Various drugs have been developed continuously targeting various proteins and other components of the microbe. In the end 60's, Rifampicin (RIF) was introduced as a combination therapy which succeeded in a declining the drug resistance and drug susceptibility of tuberculosis. However due to the arrival of HIV/AIDS in the 80's resulted in increase transmission of TB associated with outbreaks of multi drug resistant tuberculosis (MDR-TB) **[1, 2]** that are still resistant to most drugs including Rifampicin (RIF).

The potential threat of *Mycobacterium tuberculosis* is due to its ability to generate a dormant infection which evades host responses. The enigma of its dormancy and capability of infection in this phase is the prime reason for which most of the treatments have failed against it as a result of which one third of

the world population is infected **[3]** claiming two million deaths each year **[4].** Mycobacterium tuberculosis can persist in the host for decades after infection, non replicative, before reactivating to cause disease **[5]**. Persistency of the infection is due to the characteristic feature of the bacteria to reside inside the mononuclear phagocytes by exhibiting specific cellular equilibrium for the phagocytes, inferring about dynamic interactions between mycobacterial virulence factors and the human immune system **[6-9]**. The bacteria resides inside the alveolar macrophage vesicular compartment **[10, 11]** and inhibits phagosome- lysosome fusion which helps the organism to get away with direct anti microbial activity of the innate immune system as well as effective antigen presenting and overcoming adaptive immunity **[9, 12-14]**. The bacterium then replicates inside the macrophages and induces the release of cytokines that cause inflammatory response in lungs, to which macrophages and lymphocytes migrate to form a granuloma **[6].**The microbe can persist in this granuloma for years **[15, 16]** and this is the latent or the dormant phase which is clinically inactive.

The ability of the bacteria to adapt itself to survive for disease reactivation is contributed by secreted proteins called resuscitation promoting factors (Rpfs) these factors aid in virulence and resuscitating from dormancy of the bacteria, and helping in the growth of the microbe. Five such Rpfs were identified RpfA – E of which RpfB is the largest and most complex protein and is devoted to bacterial reactivation from the dormant state **[17]**. These proteins act on the bacterial cell wall causing hydrolysis of the peptidoglycan in association with other helping proteins. Resuscitation-promoting factor B (rpfB) is required for resuscitation of *M. tuberculosis* in a reactivation mouse model **[18]** and deletion of several combinations of three rpf genes results in viable bacteria that are unable to resuscitate from *in vitro* and *in vivo* resuscitation assays **[19].** RpfB have previously been shown to interact with the peptidoglycan-hydrolyzing endopeptidase, Rpf-interacting protein A (ripA) regulating its activity **[20].**

The present study is aimed to understand the molecular interaction of the protein resuscitation-promoting factor B and formulating inhibitors against the enzyme which would also help in eliminating the microbe before it attains resistance.

Methodology:

The structure of the RpfB protein was retrieved from the Protein Data Bank (PDB) having an identification number 3EO5. Sequence analysis of the protein was done using ProtParam and GOR **[21].** CATH and SCOP was performed for the classification of the protein structure **[22-23]**. The active residues of the protein were predicted using CastP server **[24]**. Ligands for study were retrieved from ZINC database containing about 2.7 million compounds **[25]** including compounds from other databases like PubChem, ACB blocks, NCI diversity II, Maybridge, Drugbank, etc. The compounds from Zinc database were first screened by selecting only the drug-like molecules. The compounds after ligand screening were then screened for AdmeTox (poor absorption, distribution, metabolism, elimination or toxicity) using FAF-Drugs2, a free ADME/tox filtering tool **[26]**. The compounds passing the AdmeTox filter were considered for high-throughput virtual screening with the target protein. Compounds showing an interaction with the protein were then selected for calculation of molecular properties using Molinspiration and calculating the drug-relevant properties using Osiris following the Lipinski rule of Five **[27]**. Molecular docking of the filtered compounds with the protein was performed using Gold suite 5.0.1.

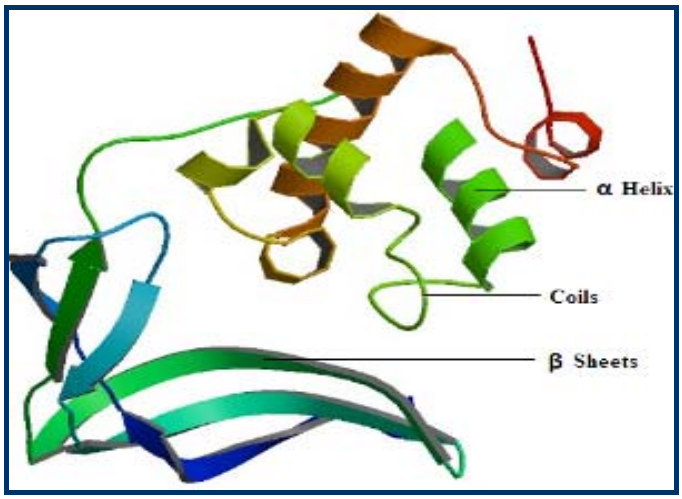


Figure 1: Three-dimensional structure of RpfB protein of *Mycobacterium tuberculosis*

Results and discussion:

The three-dimensional structure of the RpfB protein was retrieved from PDB **(Figure 1).**

Sequence analysis

The sequence analysis of the RpfB protein shows a theoretical PI of 5.36 with extinction co-efficient of 40575M^{-1} cm^{-1} and a stability index of 34.81 classifying the protein as stable. The protein sequence shows to have a more contribution of random colis of about 54.14% and a lesser contribution of alpha helix and extended strand of about 24.03% and 21.82% respectively **(Figure 2)**. CATH and SCOP results shows that rpfB belongs to class of mainly alpha and beta proteins, architecture which is an orthogonal bundle, topology lysozyme-like and a family of RPF-like.

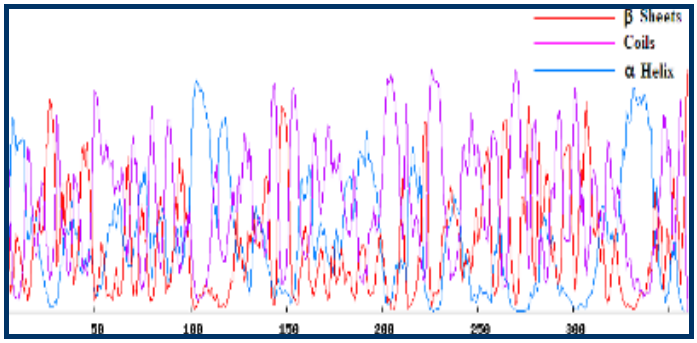


Figure 2: Graphical representation of secondary structure as predicted by GOR

Active residues

The active sites of the protein were predicted showing the amino acid sequence likely to be the binding site of the protein. The active sites targeted ranges from residue Arg194 – Gly 245 **(Figure 3).**

Ligand screening

The ligands were retrieved from ZINC database containing about 2.7 million ligands and only 25000 compounds were obtained after screening the drug-like compounds. The 25000 compounds were screened for AdmeTox and 5767 drugs were accepted by AdmeTox screening. Intermediate and rejected compounds were not considered for further study. The compounds accepted after AdmeTox were then virtual screened with the protein RpfB out of which 2982 compounds showed an interaction with the protein. Molecular properties of the compounds after virtual screening were calculated following Lipinski rule of five resulting in about 2526 compounds following the rule. **Table 1 (see supplementary material)** shows the calculated molecular properties of selected compounds.

The drug-relevant properties of the compounds were then screened by Osiris and 294 compounds showed to be non-toxic with low risk of side effects **Table 2 (see supplementary material)**. IUPAC name and structures of selected compounds are shown in **Table 3 (see supplementary material).**

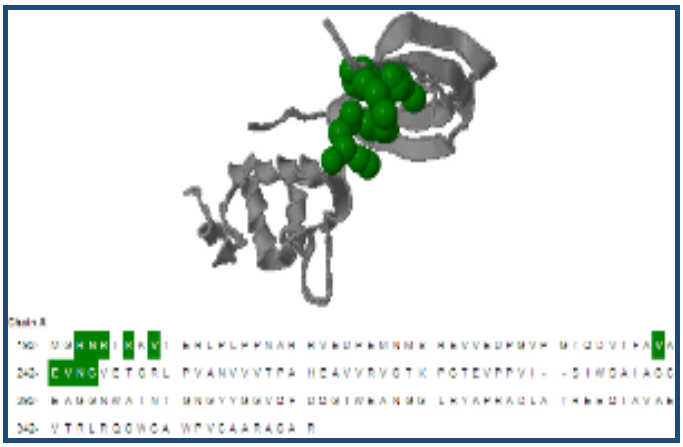


Figure 3: Active residues (space filled) of the RpfB protein

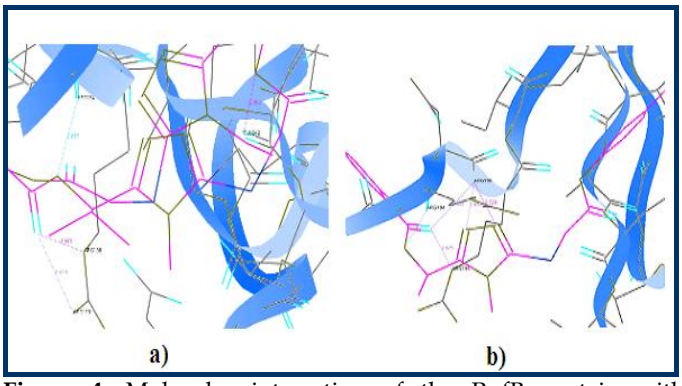


Figure 4: Molecular interaction of the RpfB protein with compounds **(a)** 3-methyl-N-[(1R)-2-methyl-1-[4-methyl-5-[2-oxo-2(phenethylamino)ethyl]sulfanyl-1,2,4-triazol-3-yl]propyl]benzamide and **(b)** 4-[[2-[[5-[(1R)-1-[(2-chlorobenzoyl)amino]ethyl]-4-methyl-1,2,4-triazol-3-yl]sulfanyl]acetyl]amino]benzoate.

Molecular Docking

The compounds passing through the filter were docked with the protein resulting with better docking studies predicting the possible binding mode of the inhibitors to the protein. The docking results show the compounds with ZincID ZINC01124772 and ZINC00687361 to have a high binding score of 70.1470 and 69.2838 with 4 and 1 H-bonds, respectively **(Figure 4)**. The compounds ZINC002146172, ZINC00687359 and ZINC00633743 have a comparatively good binding score of 68.6334, 68.5269 and 68.8880 with H-bonds of 5, 3 and 3, respectively showing a better interaction with the protein **Table 4 (see supplementary material).** Electrostatic interactions of the docked proteins are ubiquitous affecting the protein structure and stability with the ligand molecules inside the cavity **(Figure 5).** Residues Arg194, Arg196, Glu242 and Asn244 of the RpfB protein interact more with the compounds and may be the key residues to inhibit the protein activity.

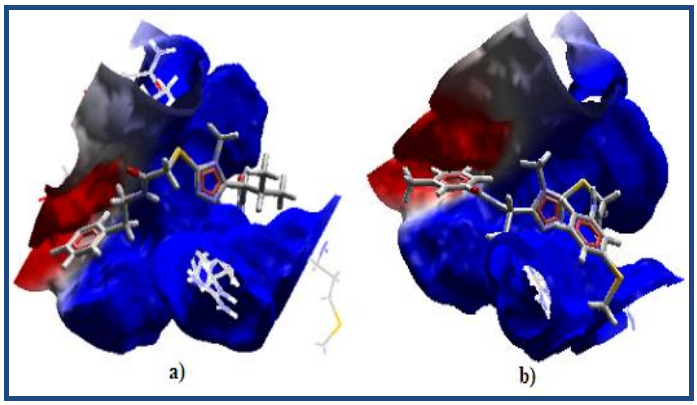


Figure 5: Electrostatic interaction of **(a)** 3-methyl-N-[(1R)-2-methyl-1-[4-methyl-5-[2-oxo-2(phenethylamino)ethyl]sulfanyl-1,2,4-triazol-3-yl]propyl]benzamide and **(b)** 3-methyl-N-[(1S)-1-[4-methyl-5-[2-(3-methylsulfanylanilino)-2-oxoethyl]sulfanyl-1,2,4-triazol-3-yl]ethyl]benzamide.

Conclusion:

Based on the results, the sequence analysis indicates that the protein is stable belonging to a class of mainly alpha and beta proteins. The docking results shows that the best compounds interacting with the protein are compounds with zinc Id ZINC01124772 (IUPAC Name: 3-methyl-N-[(1R)-2-methyl-1-[4-methyl-5-[2-oxo-2(phenethylamino)ethyl]sulfanyl-1,2,4-triazol-3-yl]propyl]benzamide), ZINC00687361 (IUPAC Name: 3-methyl-N-[(1S)-1-[4-methyl-5-[2-(3-methylsulfanylanilino)-2-oxoethyl]sulfanyl-1,2,4-triazol-3-yl]ethyl]benzamide) and ZINC00633743 (IUPAC Name: 2-({5-[(2,6-dimethylphenoxy)methyl]-4-methyl-4H-1,2,4-triazol-3-yl}sulfanyl)-N-[3-(4-morpholinylcarbonyl)phenyl]acetamide) having a good docking energy with an equivalent number of hydrogen bonds interaction which will act effectively against the protein interacting with residues Arg194, Arg196, Glu242 and Asn244 which may be the key residues to inhibit the protein activity. These compounds identified, thus holds promise for design of new anti-tuberculosis drugs and can be further validated by wet-lab studies its proper function in vivo with the target protein.

References:

[1] Edlin BR *et al. N Engl J Med.* 1992 **326**: 1514 [PMID: 1304721]
[2] Fischl MA *et al. Ann Internal Med.* 1992 **117**: 177 [PMID: 1616211]
[3] Dye C *et al. Bull World Health Organ.* 2002 **80**: 437 [PMID: PMC2567525]
[4] Corbett E L *et al. Arch Intern Med.* 2003 **163**: 1009 [PMID: 12742798]

[5] Stewart GR *et al*. *Nat Rev Microbiol*. 2003 **1:** 97 [PMID: 15035039]
[6] Dannenberg Jr AM *et al*. *ASM, Washington, DC* 1994, pp. 459–484
[7] Ellner JJ, *J Infect Dis*. 1997 **176**: 1351 [PMID: 9359738]
[8] Hingley-Wilson SM *et al*. *Nat Immunol*. 2003 **4**: 949 [PMID: 14515128]
[9] Russell DG, *Nat Rev Mol Cell Biol*. 2001 **2**: 569 [PMID: 11483990]
[10] Armstrong JA & Hart PD, *J Exp Med*. 1971 **134**: 713 [PMID: PMC2139093]
[11] Armstrong JA & Hart PD, *J Exp Med*. 1975 **152**: 1 [PMID: PMC2189870]
[12] Kusner DJ, *Horizon Scientific Press, New York, NY, 2004*, pp. 77–101
[13] Ramachandra L *et al*. *J Exp Med*. 2001 **194**: 1421 [PMID: 11714749]
[14] Vergne I *et al*. *Traffic*. 2003 **4**: 600 [PMID: 12911814]
[15] Noss EH *et al*. *Cell Immunol*. 2002 **201**: 63 [PMID: 10805975]
[16] Opie EL & Aronson JD, *Arch Path Lab Med*. 1927 **4**: 1
[17] Ruggiero A *et al*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2007 **63:** 870 [PMID: 17909292].
[18] Tufariello JM *et al*. *Infect Immun*. 2004 **72**: 515 [PMID: 14688133]
[19] Downing KJ *et al*. *Infect Immun*. 2005 **73**: 3038 [PMID: 15845511]
[20] Hett EC *et al*. *PLoS Pathog*. 2008 **4**: e1000001 [PMID: PMC2262848]
[21] Garnier J *et al*. *Methods in Enzymology* 1996 **266**: 540.
[22] Orengo CA *et al*. *Structure*. 1997 **5**: 1093 [PMID: 9309224].
[23] Murzin AG *et al*. *J Mol Biol*. 1995 **247**: 536 [PMID: 7723011].
[24] Dundas J *et al*. *Nucleic Acid Research*. 2006 **34**: W116 [PMID: 16844972]
[25] Irwin JJ & Shoichet BK, *J Chem Inf Model*. 2005 **45:** 177 [PMID: 15667143]
[26] Lagorce D *et al*. *BMC Bioinformatics*. 2008 **9**: 396 [PMID: 18816385]
[27] Lipinski CA *et al*. *Adv Drug Delivery*. 1997 **23**: 4

Edited by P Kangueane
Citation: Mylliemngap *et al*. Bioinformation 8(14): 646-651 (2012)

BIOINFORMATION

open access

Supplementary material:

Table 1: Validation of selected compound with Molinspiration

Identifier	miLogP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	volume
ZINC00633743	2.528	98.593	35	495.605	9	1	0	8	442.818
ZINC00687359	3.112	88.914	31	455.609	7	2	0	9	403.032
ZINC00687361	3.271	88.914	31	455.609	7	2	0	8	402.817
ZINC01124772	3.715	88.914	33	465.623	7	2	0	10	435.12
ZINC02146172	-0.431	129.042	32	472.934	9	2	0	8	389.361

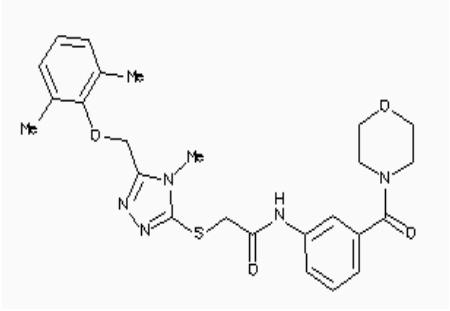
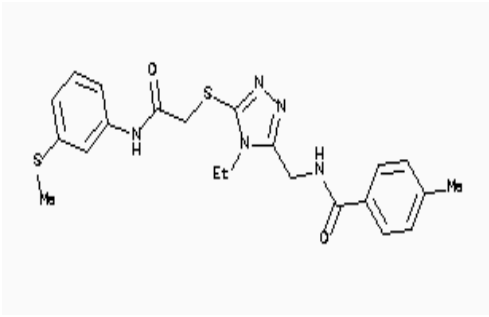
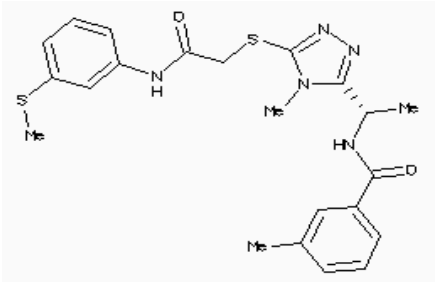
Were
miLogP: LogP (partition coefficient); TPSA: Molecular Polar Surface Area; natoms: number of atoms; MW: molecular weight nON: hydrogen bond acceptor; nOHNH: hydrogen bond donor; nviolations: number of violations; nrotb: number of rotatable bonds

Table 2: Toxicity risk of selected compounds as predicted by Osiris

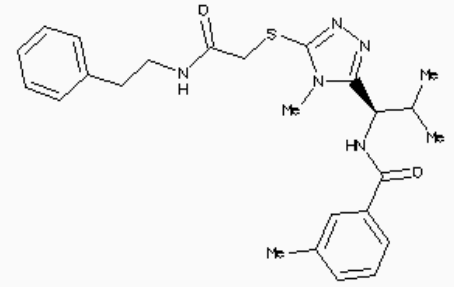
	Mutagenic	Tumorigenic	Irritant	Reproductive effect	cLogP	Solubility	Druglikeness	Drug-score
ZINC00633743	No risk	No risk	No risk	No risk	2.58	-3.08	3.31	0.67
ZINC00687359	No risk	No risk	No risk	No risk	3.06	-3.86	1.44	0.61
ZINC00687361	No risk	No risk	No risk	No risk	3.19	-3.93	2.12	0.63
ZINC01124772	No risk	No risk	No risk	No risk	3.62	-3.44	2.18	0.63
ZINC02146172	No risk	No risk	No risk	No risk	2.51	-3.48	3.28	0.68

Were
cLogP: logarithm of its partition coefficient between n-octanol and water $\log(c_{octanol}/c_{water})$

Table 3: IUPAC Name and structures of the selected lead molecules

ZINC ID	IUPAC Name	Structure
ZINC00633743	2-({5-[(2,6-dimethylphenoxy)methyl]-4-methyl-4H-1,2,4-triazol-3-yl}sulfanyl)-N-[3-(4-morpholinylcarbonyl)phenyl]acetamide	
ZINC00687359	N-({4-ethyl-5-[(2-{[3-(methylthio)phenyl]amino}-2-oxoethyl)thio]-4H-1,2,4-triazol-3-yl}methyl)-4-methylbenzamide	
ZINC00687361	3-methyl-N-[(1S)-1-[4-methyl-5-[2-(3-methylsulfanylanilino)-2-oxoethyl]sulfanyl-1,2,4-triazol-3-yl]ethyl]benzamide	

ZINC01124772 3-methyl-N-[(1R)-2-methyl-1-[4-methyl-5-[2-oxo-2(phenethylamino)ethyl]sulfanyl-1,2,4-triazol-3-yl]propyl]benzamide



ZINC02146172 4-[[2-[[5-[(1R)-1-[(2-chlorobenzoyl)amino]ethyl]-4-methyl-1,2,4-triazol-3-yl]sulfanyl]acetyl]amino]benzoate

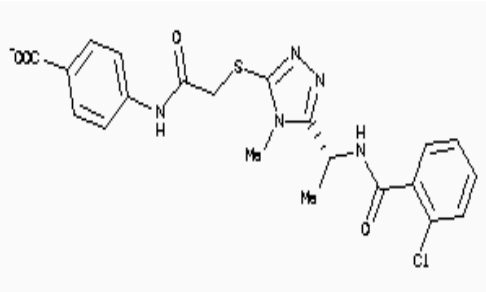


Table 4: Selected lead molecules docked with resuscitation-promoting factor B

Compounds	No. of H-bonds	Interaction (D…H–A)	H-bond length (Å)	Goldscore
		O(Arg194)….N4	2.964	
ZINC00633743	3	NE(Arg196)…N1	2.607	
		NE(Arg196)….N2	3.028	68.8880
		O(Arg194)….N4	2.509	
ZINC00687359	3	N(Arg196)….N3	2.742	
		N(Arg196)….O1	2.547	68.5269
ZINC00687361	1	O2(Arg194)…N	2.654	69.2838
		O(Arg194)….N1	2.08	
ZINC01124772	4	NH2(Arg196)….O1	2.833	
		NE(Arg196)….O1	2.473	
		O(Glu242)….N5	2.853	70.147
		O(Arg194)….N5	2.338	
		NE(Arg196)….O4	2.697	
ZINC02146172	5	NE(Arg196)….O4	2.675	
		N(Arg196)….N1	2.652	
		N(Arg196)….N2	2.729	68.6334