IN-SILICO IDENTIFICATION OF POTENTIAL PROTEIN ARGININE DEIMINASE IV (PAD4) INHIBITORS

(Pengenalpastian *In-Silico* yang Berpotensi Sebagai Perencat Protin Arginina Deiminase IV (PAD4))

Zalikha Ibrahim[1,4], Bimo Ario Tejo[2], Muhammad Alif Mohammad Latif[1], Roghayeh Abedi Karjiban[1], Abu Bakar Salleh[3], Mohd Basyaruddin Abdul Rahman[1,3]*

[1]*Macromolecular Simulation Laboratory, Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[2]*Faculty of Applied Science, UCSI University, No. 1, Jalan Menara Gading, UCSI Heights, 56000 Cheras, Kuala Lumpur, Malaysia*
[3]*Enzyme Microbial and Technology (EMTech), Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.*
[4]*Department of Pharmaceutical Chemistry, Kulliyyah of Pharmacy, International Islamic University Malaysia, Bandar Indera Mahkota Campus, 25200 Kuantan, Pahang, Malaysia*

*Corresponding author: basya@upm.edu.my

Abstract

Protein Arginine Deiminase IV (PAD4) is a promising target for treating rheumatoid arthritis. Here, an *in-silico* screening was performed using PAD4 crystal structure against National Cancer Institute Diversity Set III compounds. Results obtained from the docking studies showed that the compounds have high affinity towards the protein. Visual inspections of the top compounds indicated that they preferred to bind at the front door of the catalytic pocket instead of the back door. The current results from this screening could provide a basis for the development of new PAD4 inhibitors.

Keywords: protein arginine deiminase IV, PAD4, virtual screening, NCI diversity set, structure-based

Abstrak

Protin arginina deiminase IV (PAD4) adalah sasaran yang berharapan untuk merawat artritis reumatoid. Di sini, satu saringan *in-silico* telah dilakukan menggunakan struktur kristal PAD4 terhadap molekul dari Set Beraneka Institut Kanser Nasional III. Keputusan dari pengdokan molekular menunjukkan molekul-molekul tersebut mempunyai daya tarikan yang tinggi terhadap protin. Pemeriksaan secara visual ke atas molekul teratas menunjukkan mereka mempunyai kecenderungan untuk terikat di pintu hadapan poket katalisis daripada di pintu belakang. Keputusan daripada pencarian ini boleh menjadi asas kepada penemuan perencat PAD4 yang baru.

Kata kunci: protin arginina deiminase IV, PAD4, pencarian maya, set beraneka NCI, struktur berasas

Introduction

Protein citrullination is a post-translational modification process, which converts peptidyl-bound arginine into peptidyl-bound citrulline (see Figure 1) [1]. This process is catalyzed by calcium-dependent enzyme called protein arginine deiminase (PAD) [2]. To date, there are five PAD isozymes reported, namely PAD1, PAD2, PAD3, PAD4 and PAD6, where each of them has specific expression place in human body [3]. Among all isozymes, PAD4 has

been linked to an inflammatory disease called rheumatoid arthritis [4]. Overexpression of PAD4 and citrullinated products was found to break the immune system tolerance and produce autoantibody responses against them, which consequently leads to the disease onset. Therefore, modulation of PAD4 activities is considered as an alternative therapeutic strategy against the disease.
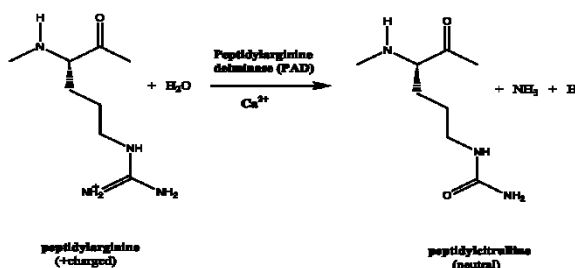

Figure 1. Citrullination process to convert peptidylarginine to peptidylcitrulline

Many researches have been conducted to design and discover PAD4 inhibitors. Based on the crystal structure of PAD4 in complex with small substrate, benzoyl-arginine amide [5], the citrullination process was catalyzed by four key residues, namely D350, H471, D473 and C645. These residues are arranged in a close proximity causing the overall shape of the pocket to be considerably narrow and deep. This significantly influenced PAD4 drug designs and studies were aimed at discovering an improved binding to one or more of the key residues [6-8]. Recent reports on PAD4 inhibitors [9] showed that the GSK199 and GSK147 compounds were binding the protein via another entrance of the catalytic pocket. The anchoring by residue N588 helps the GSK compounds to interact with H471 and distracts the protein activity. The knowledge about the entrance or the 'back door' has expanded the overall view of the pocket (see Figure 2).
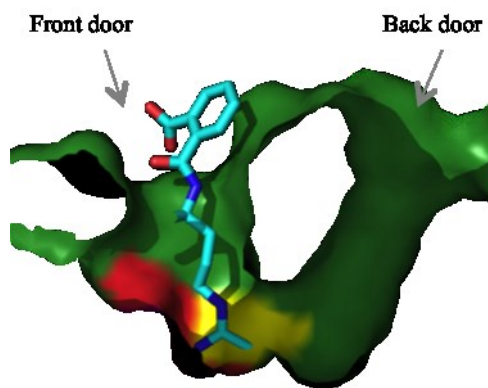

Figure 2. Two accessible doors of PAD4 catalytic pocket. Ligand such as *o*-F-amidine (rendered as blue stick) is
        bound at the front door. The main catalytic residue C645 is colored in yellow and H471 is colored in red.

Structure-based virtual screening (SBVS) has become an essential tool for drug pre-screening [10]. This approach is dependent on three-dimensional structures of target of interest i.e. protein, in which specific location on the protein can be aimed. The underlying concept in SBVS is the use of molecular docking algorithm to predict ligand poses corresponds to target of interest [11]. The method to evaluate the poses may include binding energies/affinities and reasonable binding interactions. SBVS has been applied in the previous PAD4 inhibitor screening [12]; however, the study was limited to exploring the PAD4 front door. Considering the potential of the PAD4 back door, it is

important that both doors should participate during the screening process. Here, we report SBVS performed using PAD4 crystal structure. The subjects of this study are from the National Cancer Institute Diversity Set 3 (NCIDS3), which is a compilation of pharmacophores with diverse skeleton and functional groups [13]. The poses of the compounds were predicted and evaluated computationally. It is hoped that the findings from this study can provide insight for searching and improvising PAD4 inhibitors.

Materials and Methods

Molecular docking

All molecular docking procedures were performed using AutoDock Vina program [14]. The crystal structure of PAD4 in complex with *o*-F-amidine inhibitors (PDB ID: 3B1U) [8] was used as receptor and control docking. Compounds from the previous PAD4 inhibitor screening report were used as reference compounds [15]. The NCIDS3 containing a total of 1880 ligands were downloaded in mol format from the ZINC database [16]. AutoDockTools version 1.5.6 [17] was used to add polar charges, assign Gasteiger charge [18] and prepare the input files.

Covalent docking protocol was applied exclusively for the control since the inhibitor was bound to PAD4 irreversibly [19-20]. This was achieved by docking a water molecule while setting the covalently bonded co-crystallized ligand, *o*-F-amidine to the thiol group of C645 as flexible residue. The volume of the grid box was 27 Å by 31 Å by 36 Å with a 1.0 Å spacing. The grid box was placed on the coordinate of -26.69, 14.74 and -28.8 centered on H640 to encompass both possible binding sites, namely the front door and the back door. The exhaustiveness was set to 8, while the number of run was set to 9. For the reference compounds and NCIDS3 ligands, conventional rigid protein and flexible ligand docking were performed while maintaining all the docking parameters.

In this work, the NCIDS3 ligands are referred to by their NCI index number (also known as the NSC number). For each ligand, the highest-affinity conformation of the docking result was selected as the docked pose for the ligand. The complexes were sorted by lowest binding affinity values. Complexes with predicted binding affinity equal to or less than -8.0 kcal/mol were filtered based on Lipinski's rule of 5 [21], here describes as: 1) molecular weight less than 500 g/mol, 2) number of hydrogen bond donor less than 5, 3) number of hydrogen bond acceptor less than 10 and 4) log P value less than 5. Only compounds that fulfilled all the criteria were further inspected. Analysis of the docking results was performed using AutoDock Vina shell script. The two-dimensional ligand structures were retrieved from chemicalize.org [22]. Analysis of binding interaction was performed using LigPlot[+] [23]. Visualization and rendering routines for the docking complexes were done using Pymol version 1.4.1 [24].

Results and Discussion

Validation of docking parameters

The *o*-F-amidine ligand was redocked in consistent manner as seen in the crystal structure as shown in Figure 3. The all-atom RMSD of the predicted pose with respect to its crystallized pose was 1.06 Å. The main difference between the docked pose and the crystallized pose was observed at the benzene carboxylate group. The group was interacting with inconsistent hydrophobic contact. Autodock 4 [25] was also tested as alternative docking tool, however higher RMSD value was found as compared to the one obtained using Vina. The redocking was successful, indicating that dependability of AutoDock Vina in predicting ligand experimental pose.

In conventional virtual screening routine, the accuracy of any computational docking protocol is assessed using inhibitors with known activity against the target. It is a vital step in order to justify the current parameters used. Here in our case, we look forward to ensure that the protocol was suitable and reliable to distinguish and rank PAD4 inhibitors. To achieve the goal, a set of compounds with known PAD4 inhibitions (here labeled as S1 to S12) reported by Bozdag et al. was used [15].
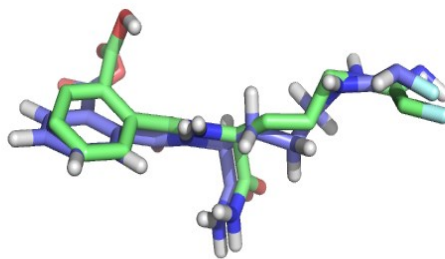
Figure 3. Redocking of *o*-F-amidine. The docked pose was rendered in green, while the crystallized structure was
in purple

Table 1 showed the experimental inhibition, experimental rank, predicted binding affinity and predicted rank for
compounds S1-S12. As can be seen from the table, it was obvious that Vina fairly predict the ranking of each
compound. There was undoubted correlation between the compounds which having the high inhibition and its
predicted binding affinity. Visual inspection of the docked poses showed that 11 out of 12 compounds have
preference to bind at the front door; only compound S5 was seen to occupy the back door. Although the compounds
with low inhibition were not in the exact rank, but it did pick the overall binding trend. Vina was able to distinguish
between the high-ranked compounds (i.e. compound S11) and the low-ranked compounds (i.e. compound S4).
These results suggested that predicted ranking reflect the experimental ranking with moderate degree of accuracy.

Table 1. Predicted binding affinity and ranks for compounds by Bozdag et al. [15]. The experimental inhibitions
were listed for reference purposes.

Compound	% Inhibition at 10 μM	Experimental Rank	Predicted Binding affinity (kcal/mol)	Predicted Rank
S1	6	7	-5.7	7
S2	3	9	-6.0	6
S3	5	8	-6.2	5
S4	1	10	-5.2	8
S5	7	6	-6.0	6
S6	5	8	-4.8	9
S7	3	9	-6.2	5
S8	8	5	-6.8	4
S9	10	3	-6.9	3
S10	34	2	-7.3	2
S11	36	1	-7.4	1
S12	9	4	-5.5	8

Structure-based virtual screening of NCIDS3

It is interesting to observe that ligands from NCIDS3 are in similar predicted binding affinity window as the known PAD4 inhibitors [15]. The plot of number of ligands with respect to their binding affinity range as shown in Figure 4 followed a bell-shaped trend, with majority population lies within the range of -6.0 kcal/mol. In prior to this study, focus was given to compounds that were predicted to have binding affinity equal to or less than -8.0 kcal/mol. This cut-off criterion of -8.0 kcal/mol was imposed as the first filter to yield compounds with predicted binding affinity lower than the predicted binding affinity of the known PAD4 inhibitors (see Table 1) [15]. Even though this NCIDS3 screening have not yet been validated experimentally, the work here is promising, given that half of the shortlisted compounds were predicted to be binding the protein via similar interacting residues with that seen in the crystal structure [8]. All shortlisted compounds were observed to dock at the traditional front door area instead of the back door. The top 11 compounds together with their predicted binding affinity were listed in Table 2.
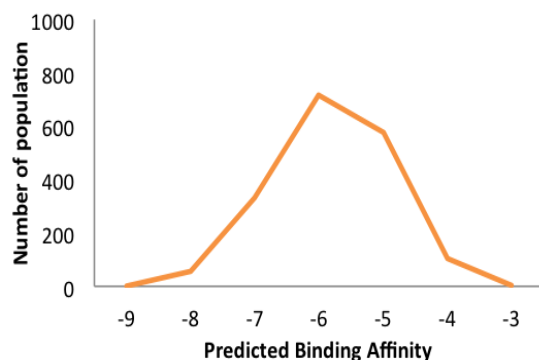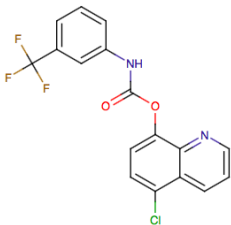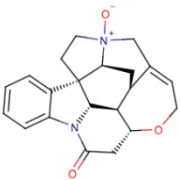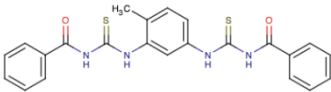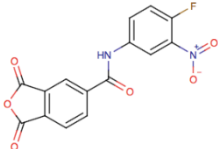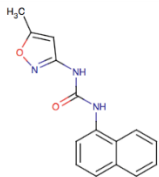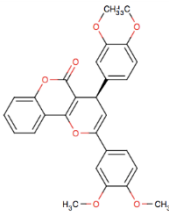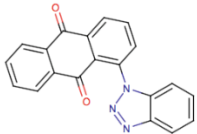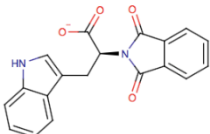

Figure 4. Binding affinities distribution of all NCIDS3 compounds towards PAD4 crystal structure (PDB ID: 3B1U)

Table 2. List of potential PAD4 inhibitors

Number	NSC number	Structure	Predicted Binding Affinity (kcal/mol)
1	127133		-8.8
2	670283		-8.6
3	58052		-8.4

Table 2 (cont.). List of potential PAD4 inhibitors

Number	NSC number	Structure	Predicted Binding Affinity (kcal/mol)
4	377852		-8.4
5	24951		-8.3
6	5856		-8.2
7	215629		-8.2
8	203837		-8.1
9	211490		-8.1
10	80313		-8.0
11	401077		-8.0

The ligands were found to linger around the front door region. None of the compounds were able to approach the main catalytic residue C645, possibly because of ring skeleton possesses by each compound. An example of the docked poses was depicted in Figure 5a. There was no specific trend in the type of ligand; but it was observed that each of them contains carbonyl group to assist interactions. For example, Compound 6. As can be seen in Figure 5b, Compound 6 made four hydrogen bonds and six hydrophobic contacts at the front door region. The first two hydrogen bonds were formed between the indole nitrogen of W347 and guanidinium nitrogen of R374 with the carbonyl group of Compound 6. The third hydrogen bond was established between the carbonyl group of R639 backbone and nitrogen atom of Compound 6. Another hydrogen bond was formed between the hydroxyl group of S468 and nitrogen atom of Compound 6. The compound also formed non-hydrogen-bonding interactions with G403, S406, V469, L578 and H640.
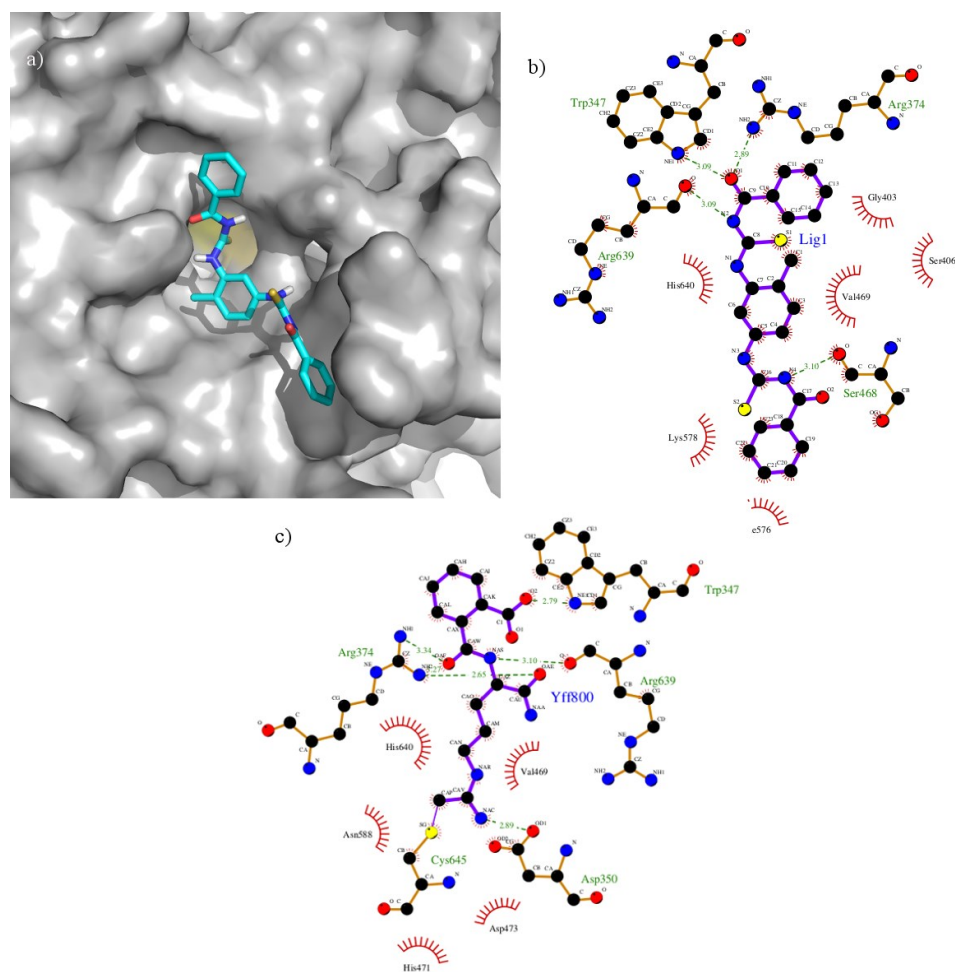

Figure 5. a) Docked pose of Compound 6 (rendered in cyan stick). PAD4 was shown as grey surface. C645 was colored in yellow. Other parts of the protein were discarded for visualization purpose, b) Interaction between PAD4 and Compound 6, c) Interaction of PAD in complex with *o*-F-amidine as reference [8].

Although Compound 6 did not interact with the main catalytic residue, C645, it interacted with important residues at the mouth of the pocket such as R374 and W347. Such interactions were also seen in PAD4-substrate/inhibitor complexes, as illustrated in Figure 5c. Arginine has high tendency to facilitate hydrogen bonds, since it is possible for the guanidinium side chain to donate five hydrogen bonds. Previous mutagenesis study revealed that the residue

R374 was important in PAD4 substrate recognition [26]. The role of W347 was also been studied [8,26]. Located at the line of the active site, the residue was observed to assist interaction via its indole nitrogen. Based on the above information, it is anticipated that the shortlisted compounds in theory would interfere with the binding of substrate to the catalytic residues. Undoubtedly, these compounds will need further optimization because of their current binding interaction. For improved binding, elongation of the compounds towards the active site tunnel may be required.

It is noticeable that the shortlisted compounds preferred to occupy the front door rather than the back door. This may be due to the use of bound front door protein structure. Since the back door did not contain any ligand, the shape of this entrance was not visibly made to occupy a ligand. To exclude the bias, it is suggested to use multiple protein structures to accommodate binding site dynamics in future screening.

Conclusion

SBVS has been successfully performed on PAD4 crystal structure using ligands from NCIDS3. This work indicates that NCIDS3 could serve as a good dataset for searching potential PAD4 inhibitors. All shortlisted compounds have preference to occupy the front door with good binding affinities. The compounds are ready to be tested experimentally. It may not be adequate to use only single protein structure in virtual screening, as binding site dynamics is unattainable. Future works suggested include flexible protein-flexible ligand docking and ensemble-based virtual screening.

Acknowledgement

References

1. van Venrooij, W. J. and Pruijn, G. J. M. (2000). Citrullination: a small change for a protein with great consequences for rheumatoid arthritis. *Arthritis Research*, 2: 249 – 251.
2. Jones, J., Causey, C., Knuckley, B., Slack-Noyes, J. L. and Thompson, P. R. (2009). Protein arginine deiminase 4 (PAD4): Current understanding and future therapeutic potential. *Current Opinion in Drug Discovery and Development*, 12(5): 616 – 627.
3. Vossenaar, E. R., Zendman, A. J., van Venrooij, W. J. and Pruijn, G. J. (2003). PAD, a growing family of citrullinating enzymes: Genes, features and involvement in disease. *BioEssays*, 25(11): 1106 – 1118.
4. van Boekel, M. A. Vossenaar, E. R., van den Hoogen, F. H. and van Venrooij, W. J. (2002). Autoantibody systems in rheumatoid arthritis: Specificity, sensitivity and diagnostic value. *Arthritis Research*, 4(2): 1 –7.
5. Arita, K., Hashimoto, H., Shimizu, T., Nakashima, K., Yamada, M. and Sato, M. S. (2004). Structural basis for Ca^{2+}-induced activation of human PAD4. *Nature Structure Molecular Biology*, 11(8): 777 – 783.
6. Luo, Y., Knuckley, B., Lee, Y. H., Stallcup, M. R. and Thompson P. R. (2006). A fluoroacetamidine-based inactivator of protein arginine deiminase 4: design, synthesis, and in vitro and in vivo evaluation. *Journal of the American Chemical Society*, 128(4): 1092 – 1093.
7. Luo, Y., Arita, K., Bhatia, M., Knuckley, B., Lee, Y. H., Stallcup, M. R. and Thompson P. R. (2006). Inhibitors and Inactivators of Protein Arginine Deiminase 4: Functional and Structural Characterization. *Biochemistry*, 45(39): 11727 – 11736.
8. Causey, C. P., Jones, J. E., Slack, J. L., Kamei, D., Jones, L. E., Subramanian, V., Knuckley, B., Ebrahimi, P., Chumanevich, A. A., Luo, Y., Hashimoto, H., Sato, M., Hofseth, L. J. and Thompson, P. R. (2011). The development of *N-α*-(2-carboxyl)benzoyl-*N*-5-(2-fluoro-1-iminoethyl)-L-ornithine amide (*o*-F-amidine) and *N-α*-(2-carboxyl)benzoyl-*N*-5-(2-chloro-1-iminoethyl)-L-ornithine amide (*o*-Cl-amidine) as second generation protein arginine deiminase (PAD) inhibit. *Journal of Medicinal Chemistry*, 54(19): 6919 – 6935.
9. Lewis, H. D., Liddle, J., Coote, J. E., Atkinson, S. J., Barker, M. D., Bax, B. D., Bicker, K. L., Bingham, R. P., Campbell, M., Chen, Y. H., Chung, C. W., Craggs, P. D., Davis, R. P., Eberhard, D., Joberty, G., Lind, K. E., Locke, K., Maller, C., Martinod, K., Patten, C., Polyakova, O., Rise, C. E., Rudiger, M., Sheppard, R. J., Slade, D. J., Thomas, P., Thorpe, J., Yao, G., Drewes, G., Wagner, D. D., Thompson, P. R., Prinjha, R. K. and Wilson,

D. M. (2015). Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. *Nature Chemical Biology*, 11(3): 189 – 191.

10. Kalyaanamoorthy, S. and Chen, Y. P. P. (2011). Structure-based drug design to augment hit discovery. *Drug Discovery Today*, 16(17-18): 831 – 839.

11. Kubinyi, H. (1998). Structure-based design of enzyme inhibitors and receptor ligands. *Current Opinion. Drug Discovery and Development*, 1(1): 4 – 15.

12. Teo, C. Y., Shave, S., Chor, A. L. T., Salleh, A. B., Rahman, M. B. B. A., Walkinshaw, M. D., and Tejo, B. A. (2012). Discovery of a new class of inhibitors for the protein arginine deiminase type 4 (PAD4) by structure-based virtual screening. *BMC bioinformatics*, 13(Supplementary 17): S4.

13. National Cancer Institute Diversity Set III (NCIDS3). Access online from *https://dtp.cancer.gov*

14. Trott, O. and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2): 455 – 461.

15. Bozdag, M., Dreker, T., Henry, C., Tosco, P., Vallaro, M., Fruttero, R., Scozzafava, A., Carta, F. and Supuran, C. T. (2013). Novel small molecule protein arginine deiminase 4 (PAD4) inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 23(3): 715 – 719.

16. Irwin, J. J. and Shoichet, B. K. (2004). ZINC- a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, 45(1): 177 – 182.

17. Sanner, M. F. (1999). Python: a programming language for software integration and development. *Journal of Molecular Graphics & Modelling*, 17(1): 57 – 61.

18. Gasteiger, J. and Marsili, M. (1980). Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron*, 36(22): 3219 – 3228.

19. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16): 2785 – 2791.

20. Ortar, G., Morera, E., De Petrocellis, L., Ligresti, A., Moriello, A.S., Morera, L., Nalli, M., Ragno, R., Pirolli, A. and Di Marzo, V. (2013). Biaryl tetrazolyl ureas as inhibitors of endocannabinoid metabolism: modulation at the N-portion and distal phenyl ring. *European Journal of Medicinal Chemistry*, 63: 118 –132.

21. Lipinski, C. A., Lombardo, F., Dominy, B.W. and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 46(1): 3 – 26.

22. Instant Cheminformatics Solution (2015). Access online from *www.chemicalize.org* (accessed on June 2015).

23. Laskowski, R. A. and Swindells, M. B. (2011). LigPlot[+]: Multiple ligand–protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*, 51(10): 2778 – 2786.

24. DeLano, W. L. (2002). The PyMOL Molecular Graphics System.

25. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14): 1639 – 1662.

26. Arita, K., Shimizu, T., Hashimoto, H., Hidaka, Y., Yamada, M. and Sato, M. (2006). Structural basis for histone N-terminal recognition by human peptidylarginine deiminase 4. *Proceedings of the National Academy of Sciences*, 103(14): 5291 – 5296.