# Affinity comparison of different THCA synthase to CBGA using modeling computational approaches

**Moulay Abdelaziz El Alaoui[1,2], Azeddine Ibrahimi[3], Oussama Semlali[3], Zineb Tarhda[3], Melloul Marouane[1, 2], Alaoui Najwa[3], Abdelmajid Soulaymani[2] & Elmostafa El Fahime[1]\***

[1]Functional Genomic Platform, UATRS, Center for Scientific and Technical Research [CNRST], Rabat, Morocco; [2]Laboratory of Genetics and Biometry, Faculty of Sciences, University Ibn Tofail; [3]MedBiotech Lab, Faculté de Médecine et de Pharmacie de Rabat, Université Mohammed V Souissi; Elmostafa El Fahime - Email: elfahime@cnrst.ma; Phone: +212 (0)537712983, Fax: +212 (0) 37713205; *Corresponding author

**Abstract:**
The Δ⁹-Tetrahydrocannabinol (THCA) is the primary psychoactive compound of *Cannabis Sativa*. It is produced by Δ¹-Tetrahydrocannabinolic acid synthase (THCA) which catalyzes the oxidative cyclization of cannabigerolic acid (CBGA) the precursor of the THCA. In this study, we were interested by the three dimensional structure of THCA synthase protein. Generation of models were done by MODELLER v9.11 and homology modeling with Δ¹-tetrahydrocannabinolic acid (THCA) synthase X ray structure (PDB code 3VTE) on the basis of sequences retrieved from GenBank. Procheck, Errat, and Verify 3D tools were used to verify the reliability of the six 3D models obtained, the overall quality factor and the Prosa Z-score were also used to check the quality of the six modeled proteins. The RMSDs for C-alpha atoms, main-chain atoms, side-chain atoms and all atoms between the modeled structures and the corresponding template ranged between 0.290 A°-1.252 A°, reflecting the good quality of the obtained models. Our study of the CBGA-THCA synthase docking demonstrated that the active site pocket was successfully recognized using computational approach. The interaction energy of CBGA computed in 'fiber types' proteins ranged between -4.1 95 kcal/mol and -5.95 kcal/mol whereas in the 'drug type' was about -7.02 kcal/mol to -7.16 kcal/mol, which maybe indicate the important role played by the interaction energy of CBGA in the determination of the THCA level in *Cannabis Sativa L.* varieties. Finally, we have proposed an experimental design in order to explore the binding energy source of ligand-enzyme in *Cannabis Sativa* and the production level of the THCA in the absence of any information regarding the correlation between the enzyme affinity and THCA level production. This report opens the doors to more studies predicting the binding site pocket with accuracy from the perspective of the protein affinity and THCA level produced in *Cannabis Sativa.*

**Keywords:** *Cannabis Sativa*, Homology modeling, Tetrahydrocannabinolic acid synthase potency, THC synthase binding energy, dynamic study, experimental design.

**Background:**
*Cannabis Sativa* is one of the most used drugs in the word. However currents researches are focused on its use in medicinal application as a medication in order to relieve specific symptoms such as : AIDS, cancer, multiple sclerosis, and neurological disorders **[1]**. The THCA synthase belongs to the oxygen-dependent FAD-linked oxidoreductase family containing 1 FAD-binding domain and has homology to several oxido-reductases including the berberine bridge enzyme (BBE).

This enzyme catalyzes the oxidative cyclization of the monoterpene moiety in cannabigerolic acid (CBGA) to produce Δ⁹- tetrahydrocannabinol (THCA) which is the major cannabioid in drug-type cannabis plants **[2]**. The THCA synthase is an aggregation of two domains (domains I and II) consisting of eight α -helices in addition of eight β -strands and two subdomains (Ia and Ib). Subdomain Ia comprises three α -helices surrounding three β –strands and subdomain Ib comprising five antiparallel β –strands surrounding five α –

helices. Whereas domain II contains eight antiparallel β -strands surrounded by six α –helices **(Figure 1)** **[2]**. A recent study **[2]** based on the mutational analysis of THCA synthase enzyme showed the implication of several residues in the binding to the ligand (CBGA/FAD), as well as residues which are essential for the enzyme activity: His114, His 292, Tyr 484. The His 292 is hypothesized to be involved in essential substrate binding of CBGA. The Tyr 417 might also contribute to the substrate binding but not to be essential for catalysis, the residues His114 and Cys 176 are involved to bind covalently the FAD **[2]**.

Most studies performed until to date were focused only on the determination of the kinetic parameters in recombinant THCA synthase **[3]** or the determination of kinetic parameters in wild-type and mutant THCA synthase enzymes **[4]**. However no studies have been carried out to correlate between the affinity of the THCA synthase to CBGA or FAD and the production level of the THCA. Therefore, it is interesting to use the computational approach to study the affinity of the THCA synthase to the substrate. The computation of the THCA synthase affinity to their substrates CBGA and /FAD will allow us to determine how strongly the CBGA and FAD binds to the active site pocket and correlate the enzyme biological response to THCA production levels. In 'fiber type' cannabis, the content of THCA dose not exceed 0.2 % while in the 'drug type' the content on THCA is more than 2. 1 % and could exceed 4.1%. Therefore, molecular docking calculations could be interesting to explore enzyme affinity to CBGA/FAD and could be extended to search the binding conformation of the substrate in the active site pocket. These calculations could give an explanation of the behaviour of the others residues in the binding site in both 'drug type'and'fiber type'. To elucidate the THCA synthase interaction energy to CBGA and their structural basis for improving the affinity, a computational approach was used. In our study, we used molecular modeling techniques and Docking calculations of the CBGA using DockingServer **[5]**. The objectives were: **(1)** Modeling of THCA synthase proteins from the sequences of 'drug type' and 'fiber type' THCA synthase proteins **(2)** Structure refinement and evaluation of the modeled proteins in order to identify their quality **(3)** Docking of CBGA with the modeled proteins to examine the binding energy in the tow types (drug/fiber) **(3)** Propose an experiment design to explore the binding energy in the tow types drug-fiber to correlate between the enzyme affinity and THCA level production (**Figure 2**).
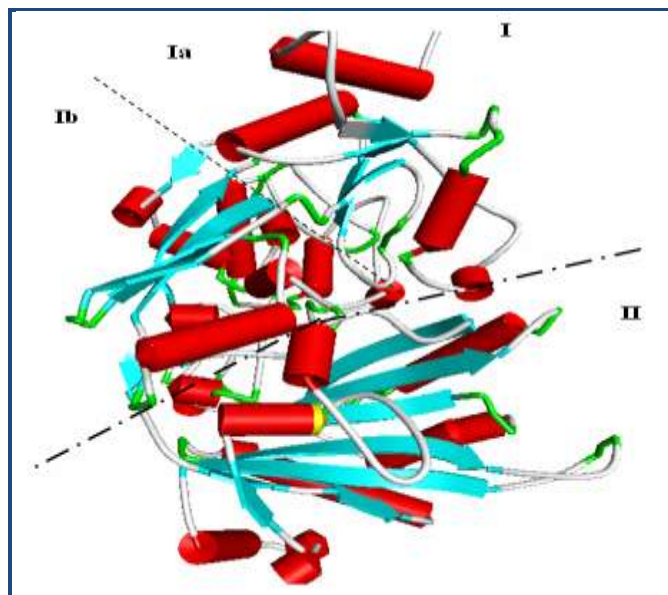


**Figure 1:** A schematic model of the THCA synthase (GenBank: JQ437482); α-helices are colored in cyan, β -strands are in red, and loops are in green.The two domains (I and II) and the two subdomains (Ia and Ib) are indicated.
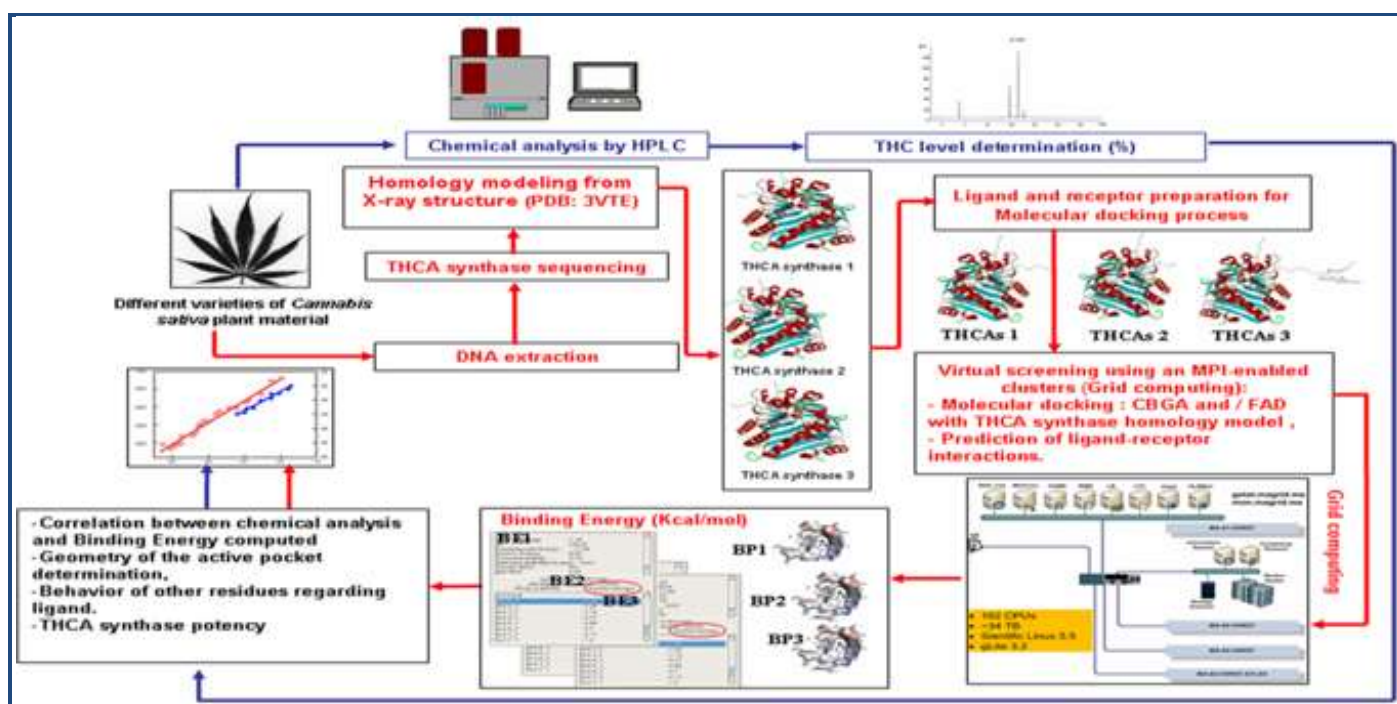


**Figure 2:** Proposed experimental design for the correlation between THCA synthase binding energy (CBGA and/FAD) and THCA level determination. BE (Binding energy); BP (Binding pocket).
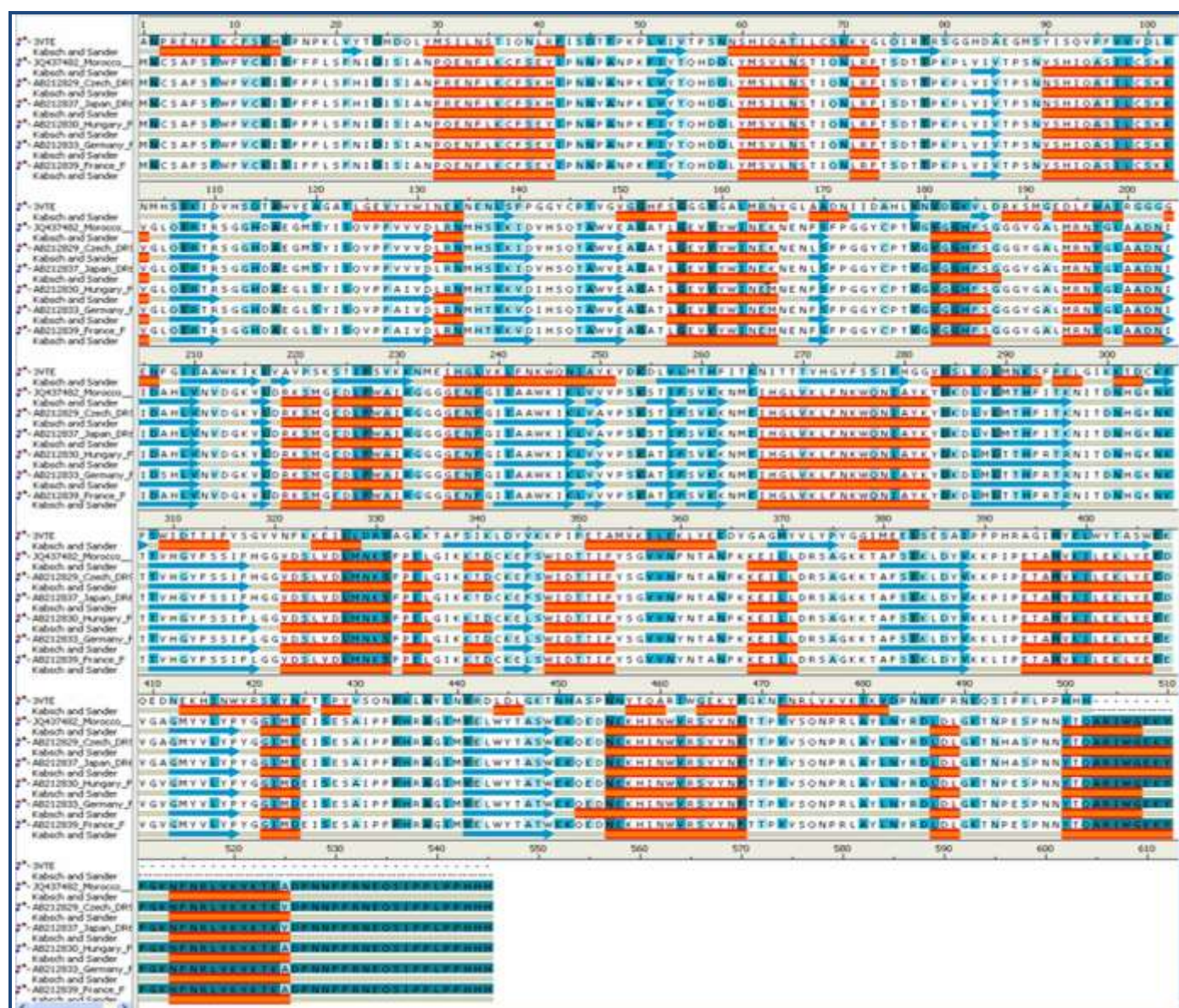
**Figure 3:** Alignment and secondary structure of the six modeld targets with the template. The deep blue color shows conserved residues in all sequences (α-helix are indicated in red and β-sheets in blue arrows).

**Methodology:**

*3D structure prediction*

The three dimensional homology models of the targets (**Figure 1**) were constructed using the X-ray structures of the template (PDB code 3VTE) assuming alignment between targets and template, all steps of homology modeling and refinement were accomplished by the program MODELLER9 **[6]**. MODELLER program generated several preliminary models which were ranked based on their DOPE scores. Some models with the lowest DOPE score were selected and the stereochemical properties of each model were assessed by PROCHECK **[7]**, and Errat **[8]**. The secondary structure of the modeled targets were visualized using discovery studio V3.1 **[9]** (**Figure 3**).

*Structure refinement and models evaluation*

The structural refinement of the THCA synthase models was checked for stability and energy minimization to avoid overlapping atoms using Gromacs software **[10]**. The refined structure was further evaluated to predict the overall quality of

the minimized geometries by the software program PROCHECK **[7]** and web server ProSA **[11]**. The graphical displays were generated with the Discovery Studio Visualizer V3.1 **[9]**.

*Structure superimposition*

Multiple sequence alignment was carried out to superimpose the obtained models and to identify the conserved regions by aligning the target with the templates structures. we superimposed and aligned homologous structures using discovery studio V3.1 **[9]**.

*CBGA-THCA synthase docking*

Crystallographic structures of the protein THCA synthase were retrieved from the RCSB database with PDB ID: 3VTE.Computational analysis was done to compute ligand protein binding affinity of the compound. Docking calculations were carried out using DockingServer **[5]**. The MMFF94 force field **[12]** was used for energy minimization of ligand molecule

(CBGA) using DockingServer. Gasteiger partial charges were added to the ligand atoms. Non- polar hydrogen atoms were merged, and rotatable bonds were defined.

Docking calculations were carried out on each modeled protein **Table 1 (see supplementary material)**. Essential hydrogen atoms, Kollman united atom type charges, and solvation parameters were added with the aid of AutoDock tools. Affinity (grid) maps of 20×20×20 Å grid points and 0.375 Å spacing were generated using the Autogrid program **[13]**. AutoDock parameter set and distance- dependent dielectric functions were used in the calculation of the van der Waals and the electrostatic terms, respectively. Docking simulations were performed using the Lamarckian genetic algorithm (LGA) and the Solis & Wets local search method **[14]**. Initial position, orientation, and torsions of the ligand molecules were set randomly. Each docking experiment was derived from 10 different runs that were set to terminate after a maximum of 250000 energy evaluations. The population size was set to 150. During the search, a translational step of 0.2 Å, and quaternion and torsion steps of 5 were applied **[15].**
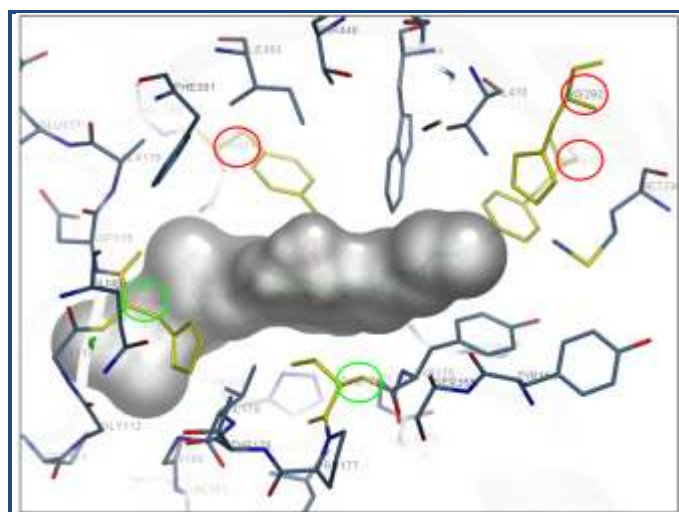


**Figure 4:** Predicted binding pocket of (CBGA) cannabigerolic acid in the target (GenBank code: JQ437482) Residues involved in the binding of the CBGA in the THCA synthase protein are indicated with a red circle. Residues involved in the binding of FAD (His114; Cys176) are indicated with a green circle.

**Results:**
*Assessment of resultants structures*
Results of the evaluation of different generated models attested for their high quality and the overall model quality of the targets proteins is shown in (**Table 1**). **Figure 1** is showing one of the generated models. Indeed, the overall calculated quality factor of ERRAT in all modeled proteins were ranging from 73 .624 to 80.998, thus indicating the errors values are negligible for individual residues (**Table 1**). The z-score for the modeled THCA synthase proteins indicates overall model quality score **Table 2 (see supplementary material)**. The z-score of the studied targets is ranging from-9. 72 to -9. 47, while the z-score for template was -9.26 which fill in the range of scores typically found for experimentally determined (X-ray, NMR) structure for native protein. Therefore, the 3D structure of modeled THCA synthase proteins is reliable. Structural identity of the obtained models based on different RMSDs for C-alpha atoms,

main-chain atoms, side-chain atoms and all atoms between the modeled structures and the corresponding template are indicated in **Table 2 (see supplementary material).** The RMSD values were ranged between 0.290A° to 1.252 A° which reflects the accuracy of the structures generated by homology modeling. Ramachandran plot of all obtained models indicated that 90.8%(GenBank : AB212839 ) to 92.5% (GenBank : JQ437482 and AB212830) of the residues have psi/phi angles falling in the most favoured regions (**Table 1**) and 6.5%  (GenBank : AB212830 ) to 8.8% (GenBank : AB212829 and AB212839) of the residues were in the additional allowed region. On the other hand 0.0%  (GenBank: AB212833; AB212839) to 1.0% (GenBank: AB212830) of the residues were in generously allowed region and 0.0% (GenBank: JQ437482; AB212829; AB212837; AB212830) to 0.4% of the residues were in the disallowed region (GenBank: AB212839) (**Table 1**). **Figure 4** shows the 2D sequences alignment of the superimposed proteins models with template.

*Molecular Docking*
The molecular docking analysis indicated that for 'drug types' proteins the energy of binding fluctuate between -7.16 kcal/mol to -6.76 kcal/mol.  On the other hand, the 'fiber types' protein had a energy of binding ranging from -4.29 kcal/mol to -5.95 kcal/mol **Table 3 (see supplementary material)** demonstrating that the affinity to CBGA for the 'fiber types' is lower than the 'drug types'. The differences observed in the binding energy computed maybe corroborated with the existing data concerning the level of the THCA production in 'fiber types' and 'drug types' cannabis. In 'fiber type' cannabis, the content of THCA does not exceed 0.2 % while in the 'drug type' the content on THCA is more than 2. 1 % and could exceed 4.1%. Probably if the affinity of the protein to the CBGA is high, the level of the THCA will increase and vice versa, hence the necessity to confirm this finding using the proposed experiment (**Figure 2**). The predicted binding pocket of (CBGA) cannabigerolic acid in all target proteins were successfully recognized using computational approach. **Figure 4** shows one of the target proteins (GenBank code: JQ437482). Residues involved in the binding of the CBGA in the THCA synthase protein are indicated and the residues involved in the binding of FAD (His114; Cys176) were successfully determined. The obtained results allowed us to propose an experiment design to explore the binding energies of the THCA synthase regarding the precursor and/or the cofactor and for the correlation between the THCA synthase-FAD/CBGA affinity and the level of the THCA production (**Figure 2**).

**Discussion:**
In this paper, we have explored for the first time the affinity of six modeled THCA synthase proteins (three 'drug type' and three 'fiber type') to CBGA, the precursor of the major cannabinoid in *Cannabis Sativa .L* plant which is THCA. The binding site of CBGA was successfully recognized and the calculated binding parameters confirm that the level of the THC produced in 'fiber type' and 'drug type' maybe explained by the affinity of the protein regarding the CBGA. The only data available to date, is the in vitro measurement of the enzymatic activity in wild-type and mutant THCA Synthase **[2]**. Separating inherited variation in cannabinoid content from environmental variation **[16, 17]**. The 3D structures  generated using (PDB code **3VTE** as template) from sequence data of

different *Cannabis Sativa* varieties representing a variable THCA level will help to understanding better the functional conformations of the THCA synthase binding pocket regarding their ligand and maybe will improve:**(i)** The behaviour of the other residues regarding the ligand **[18] (ii)** Differences in the active pockets geometry that will affect the binding energy of the ligand and consequently the level of the THCA produced, **(iii)** the correlation between THCA level and the binding energy computed, **(iv)** understanding the differences between THCA synthase potency in *Cannabis Sativa* drug type varieties.

The alignment of the studied THCA synthase amino acids sequences as well as other sequences retrieved from GenBank (data not shown) shows a conservation of the key residues implicated in the catalytic process, the substrate binding and the enzyme selectivity. This result fails to explain the difference in level of THCA between the *Cannabis Sativa* species. The results from homology modeling and docking study of the THCA synthase proteins with CBGA demonstrate that the active site pocket was successfully recognized using computational approach. The key residues involved in the catalytic process: (Tyr175, Tyr417), as well as the residue (His114) involved in the binding of the precursor FAD were successfully recognized (**Figure 4**). In order to correlate the THCA synthase activity and the level of the THCA synthase produced in different *Cannabis Sativa* varieties, we propose an experimental design starting from *Cannabis Sativa* plant material for the investigation of the enzyme affinity using a combined study which involve the chemical analysis [i.e., the level of THCA], the genetic analysis [the THCA synthase sequencing], the homology modeling and the dynamic study [interaction energies of the protein-ligand complexes] (**Figure 4**). The data generated from the computational study of the protein-ligand interaction and the chemical analysis such as High-performance liquid chromatography (HPLC) will be of great interest to understand better the functional conformations of the substrate-binding pocket and to correlate affinity and activity of the enzyme with THCA level produced. The ultimate goal of the THCA synthase prediction is to demonstrate that the binding pocket of THCA synthase can reach an accuracy comparable to the results achieved experimentally **[2]**. It also allowed performing mutational studies on the generated models and dynamic simulation using computational approaches.

**Conclusion:**
In this study THCA synthase of 'fiber type' and 'drug type' *Cannabis Sativa* were modeled based on the existing structure of THCA synthase X ray diffraction in order to perform a molecular docking of the CBGA on the tow types and to explore the binding energy between and within cannabis type. We demonstrate that the computational approach of protein ligand interaction using molecular docking calculations can reach almost the same results obtained experimentally. In this research we have proposed an experimental methodology for investigating the correlation between the protein affinity of THCA synthase with FAD/CBGA using accurate docking techniques and the THCA level production in the absence of any information regarding the binding energy of the THCA synthase to their substrate. These parameters can be a key to know more about the crucial interactions affinity in the binding pocket of the THCA synthase enzyme and THCA content in *Cannabis Sativa* plant material.

**Conflict of Interest:**
The authors declare that there is no conflict of interest.

**References:**
**[1]** http:/ /www.nap.edu /openbook. php?record _id =9586 Accessed 9August 2013.
**[2]** Shoyama Y *et al. J Mol Biol.* 2012 **423**: 96 [PMID: 22766313]
**[3]** Taura F *et al. Biochem Biophys Res Commun.* 2007 **361**: 675 [PMID: 17669365]
**[4]** Sirikantaramas S *et al. J Biol Chem.* 2004 **279**: 39767 [PMID: 15190053]
**[5]** Bikadi Z & Hazai E, *J Chemin form.* 2009 **1**: 15 [PMID: 20150996]
**[6]** Šali A & Blundell TL, *J Mol Biol.* 1993 **234**: 779 [PMID: 8254673]
**[7]** Laskowski RA *et al. J of Applied Cryst.* 1993 **26**: 283
**[8]** Colovos C & Yeates TO, *Protein Sci.* 1993 **2:** 1511 [PMID: 8401235]
**[9]** http://accelrys.com/products/discovery-studio /visualization-download.php Accessed 12 July 2012
**[10]** Pronk S *et al. Bioinformatics* 2013 **29**: 845 [PMID: 23407358]
**[11]** Wiederstein M & Sippl MJ, *Nucleic Acids Res.* 2007 **35** :W407 [PMID: 17517781]
**[12]** Halgren TA, *Journal of Computational Chemistry* 1996 **17**: 490
**[13]** Morris GM *et al. Journal of Computational Chemistry* 1998 **19**: 1639
**[14]** Abagyan R & Argos P, *J Mol Biol.* 1992 **225:** 519 [PMID: 1593634]
**[15]** Kuldeep Sahu SsB, *Int J Pharm Bio Sci.* 2013 **4**: 370
**[16]** De Meijer EP *et al. Genetics.* 2003 **163**: 335 [PMID: 12586720]
**[17]** Meijer EPM *et al. Euphytica* 1992 **62**: 187
**[18]** Punkvang A *et al. Molecules* 2010 **15**: 2791 [PMID: 20428080]

**Edited by P Kangueane**
**Citation**: **Alaoui *et al.*** Bioinformation 10 (1): 033-038 (2014)

# BIOINFORMATION

## Supplementary material:

**Table 1:** Ramachandran plot statistics based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions

|  | ID | Residues in most favoured regions [A,B,L] | Residues in additional Allowed regions [a,b,l,p] | Residues in generously allowed regions [~a,~b,~l,~p] | Residues in disallowed regions |
|---|---|---|---|---|---|
| Drug Type | GenBank : JQ437482 | 92.5% | 7.1% | 0.4% | 0.0% |
|  | GenBank : AB212829 | 90.8% | 8.8% | 0.4% | 0.0% |
|  | GenBank : AB212837 | 92.3% | 6.9% | 0.8% | 0.0% |
|  | GenBank : AB212830 | 92.5% | 6.5% | 1.0% | 0.0% |
| Fiber Type | GenBank : AB212833 | 91.6% | 8.1% | 0.0% | 0.2% |
|  | GenBank : AB212839 | 90.8% | 8.8% | 0.0% | 0.4% |

**Table 2:** Overall model quality of the targets proteins and RMSD calculation of aligned proteins against template 3VTE

|  | ID | Prosa | PROCHECK | | RMSD | | | |
|---|---|---|---|---|---|---|---|---|
|  |  | Z score | Verify % | Errat | C-Alpha | Main-chain | Side-chain | All Protein |
| Drug Type | Pdb : 3VTE (temple) | -9.26 | 95.23 | 92.813 | - | - | - | - |
|  | GenBank : JQ437482 | -9,47 | 92,12 | 77,799 | 0.319 | 0.389 | 1.189 | 0.884 |
|  | GenBank : AB212829 | -9,53 | 91,03 | 75,568 | 0.313 | 0.411 | 1.252 | 0.928 |
|  | GenBank : AB212837 | -9,51 | 92,31 | 80,998 | 0.330 | 0.395 | 1.200 | 0.894 |
| Fiber Type | GenBank : AB212830 | -9,72 | 90,11 | 74,472 | 0.298 | 0.363 | 1.199 | 0.881 |
|  | GenBank : AB212833 | -9,67 | 92,31 | 73,624 | 0.290 | 0.361 | 1.242 | 0.910 |
|  | GenBank : AB212839 | -9,49 | 93,41 | 78,476 | 0.295 | 0.389 | 1.221 | 0.902 |

**Table 3:** Energy values of docked CBGA with drug type and fiber type protein

|  | ID | Rank | Est. Free Energy of Binding | vdW + Hbond +dissolve Energy | Total Intermolec. Energy | Interact Surface |
|---|---|---|---|---|---|---|
| Drug Type | Pdb : 3VTE | 1. | -7.16 kcal/mol | -10.27 kcal/mol | -10.35 kcal/mol | 1071.369 |
|  | GenBank : JQ437482 | 1. | -7.02 kcal/mol | -9.85 kcal/mol | -10.00 kcal/mol | 1094.774 |
|  | GenBank : AB212829 | 1. | -6.75 kcal/mol | -10.05 kcal/mol | -10.05 kcal/mol | 1022.683 |
|  | GenBank : AB212837 | 1. | -6.76 kcal/mol | -9.80 kcal/mol | -9.89 kcal/mol | 1036.823 |
| Fiber type | GenBank : AB212830 | 1. | -4.29 kcal/mol | -7.18 kcal/mol | -7.24 kcal/mol | 988.945 |
|  | GenBank : AB212833 | 1. | -4.17 kcal/mol | -6.59 kcal/mol | -6.59 kcal/mol | 1241.8 |
|  | GenBank : AB212839 | 1. | -5.95 kcal/mol | -9.06 kcal/mol | -8.97 kcal/mol | 827.276 |