

Binding of 2CA1P (nocturnal inhibitor) to the active site of RubisCO using Genetic Algorithm (GA)

Ounissi Mourad*, Kameli Abdelkrim, Boudjeniba Messaoud, Kherraz Khaled, Redouane Nouredine

Biology department, Ecole Normale Supérieure, ENS-Kouba, PB 92, Algiers, Algeria; Mourad Ounissi – Email : ounissi_bio@yahoo.fr; Phone: +213778227733; Fax +21321282067; *Corresponding author

Received April 30, 2009; Revised August 01, 2009; Accepted September 11, 2009; Published November 17, 2009

Abstract:

Ribulose-1, 5- Bisphosphate carboxylase/ oxygenase (RubisCO) catalyzes the first step in net photosynthetic assimilation and photorespiratory carbon oxidation. The activity of this enzyme is modulated in response to changes in light intensity as suggested in a number of early reports. Several studies found that the natural inhibitor 2CA1P is involved in the inhibition of the enzyme under reduced light intensity in rice (*Oryza sativa*). Due to the lack of studies and information on the interaction between this inhibitor and the active site of the enzyme, we attempted to predict the interaction between the amino acids in the active site and the inhibitor using both Hyperchem7.5 and GOLD software. After the docking; three possibilities having the highest fitness score were found (65.71, 64.72, 62.04), in these possibilities the inhibitor was bound to the enzyme, the phosphate and carboxylate groups in the same positions with a clear difference in the position of OH. In order to confirm the accuracy of the genetic algorithm, the artificial inhibitor 2CABP was docked back in the active site of the enzyme using the same parameters used in the case of the 2CA1P and the algorithm's predictions were compared with the experimentally observed binding mode. The results showed that the difference in the active sites before and after the docking was in the range of 0.93 Å which indicated that the results were very accurate. Depending on this result it was concluded that the results obtained in the case of the 2CA1P were close to the experimental results.

Keywords: RubisCO, 2CA1P, CABP, GOLD, Hyperchem 7.5, Genetic algorithm.

Background:

Rubisco (Ribulose 1, 5-bisphosphate carboxylase/ oxygenase CE: 4.1.1.39), the most abundant enzyme on the earth, is the key enzyme in photosynthesis, which incorporates CO₂ and O₂ into substrate Ribulose -1, 5-bisphosphate to initiate photosynthesis and photorespiration, respectively. To be catalytically competent, the active site in Rubisco need to be open, carbamylated by addition of CO₂ to the lysine (Lysine 201 in spinach rubisco) and stabilized with Mg⁺². [4] This enzyme is composed of two types of subunits in most photosynthetic organisms (higher plants, algae and cyanobacteria), four small subunits cap the top and the bottom of the core of eight large subunits (encoded by *rbcS* gene located in nuclear genome) and look as though they hold together the four large subunit dimers (encoded by *rbcL* gene located in chloroplast genome), each of which forms two active sites. In higher plants, Rubisco has a hexadecameric structure being composed of eight large and eight small subunits. This is also known as Form I Rubisco. Each large subunit has two major structural domains, an N-terminal domain and a larger C-terminal domain which is an alpha/beta barrel. Most of the active site residues are contributed by loops at the mouth of the alpha/beta barrel with the remaining residues being supplied by two loop regions in the N-terminal domain of the second large subunit within a dimer. The availability of high resolution 3-D structure has provided detailed insight into the catalytic mechanism of the enzyme. [5] Activity of rubisco is modulated in vivo either by reaction with CO₂ and Mg⁺² to carbamylate a lysine residue in the catalytic site, or by the binding of inhibitors within the catalytic site, binding of inhibitors blocks either activity or the carbamylation of the lysine residue that is essential for activity. [5]

It has been observed that the reaction of catalytically competent Rubisco changed with light intensity. The principal cause of this change is the binding of inhibitor called 2-Carboxy-D-arabinitol 1-phosphate (2CA1P) which coexists in some plants such as *Phaseolus vulgaris*. [3]. This inhibitor is tightly bound to the carbamylated active site [5]. This tight binding property results from the resemblance to the transition state intermediate of the carboxylase reaction (gemdiolate). The 2CA1P inhibitor has an efficient role in the regulation of Rubisco, through the binding with the active site in activated state (Carbamylated + Mg⁺²) and keep it until the exposure to light. Another study showed that this inhibitor can protect Rubisco against proteolytic breakdown. [3] Due to the important role of this inhibitor in regulation and protection of Rubisco enzyme and the absence of published work on the binding mechanism of this inhibitor in the active site of Rubisco, and the amino acids attached to it. This study tries to predict the binding position of this inhibitor in the active site using bioinformatics tools consisting of GOLD (Genetic optimisation of Ligand Docking) and Hyperchem7.5 software. To achieve this study, first, a model of the inhibitor 2CA1P is built using Hyperchem7.5. Second, a search, selection and preparing of Rubisco enzyme structure with a higher resolution from PDB (Protein data bank). And third, docking the inhibitor model in the active site of Rubisco using genetic algorithm (GA). The molecular docking aims to predict the structure of a molecular complex from the isolated molecules, which is considerably easier to implement, cheaper and faster compared to experimental methods.

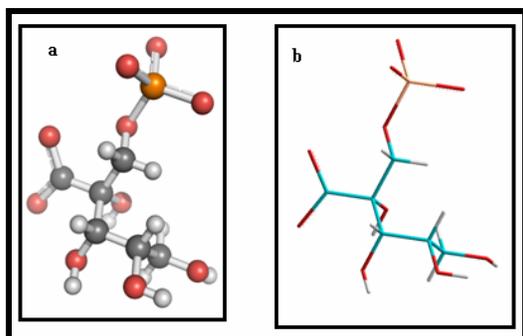


Figure 1: The three dimensional structure of 2CA1P inhibitor (a) shows in stick conformation (b) show in ball and stick (colored by CPK).

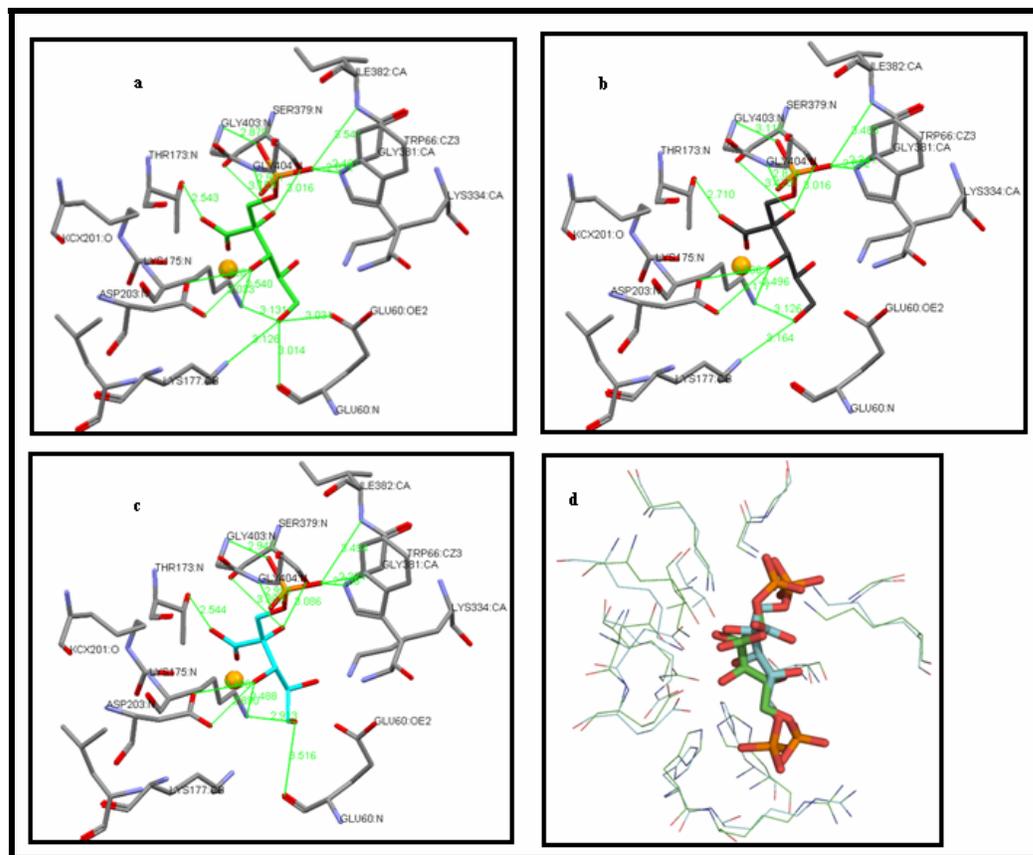


Figure 2: The binding modes of the 2CA1P in the active site of Rubisco; the three highest fitness score represented in **a**, **b**, **c** (65.71, 64.72, and 62.04 respectively) and **d** Superposition between the active site of rubisco before (colored green) and after (colored bleu light) the docking of the inhibitor 2CABP (Stick).

Methodology:

Building inhibitor model:

The inhibitor is known as 2-carboxy-D-arabinitol 1-phosphate. [6] It consists of carbon backbone including five atoms linked to five OH groups at the level of toms C2, C3, C4 and C5. The phosphate and craboxylate group linked to the carbon atoms C1 and C2 respectively.

Hyperchem software:

The inhibitor model was built using Hyperchem 7.5 software (www.hyper.com/) based on Lewis structure. Atoms have been chosen from dialog box (default elements) in build menu, the 3D structure of the inhibitor 2CA1P is shown in **Figure 1**.

Assigning Atomic Charges to the functional groups:

In this step we added the formal charge to the carboxylate and phosphate groups using the approximation of the charge method.

Energetic calculations:

Single point calculation:

A single point calculation was performed in order to compute the energy and gradient of the inhibitor model, this method allowed the calculation of the total energy and gradient before the geometry optimisation.

Optimizing the Structure of inhibitor model:

In this step, the inhibitor model structure was minimized by performing a molecular mechanics optimization using MM+ force field and Polack Ribier algorithm to obtain the most stable structure geometry.

Preparation of enzyme structure:

The crystal structure coordinates of the Rubisco enzyme (PDB_ID:1WDD with a resolution of 1.35 Å) was obtained from

the protein data bank (<http://www.rcsb.org>). In this case we obtained two large and two small subunits.

Enzyme structure optimization:

Two large subunits were selected for docking studies which included one active site at least. Water molecules and hetero atoms were removed. [2] All hydrogen atoms were added to the protein including those necessary to define the correct ionization and tautomeric states of amino acid residues such as Asp, Ser, Glu, Arg and His using Hyperchem software. A two step procedure was set up for the energy minimization of protein using the same software and CHARMM22 algorithm. In the first step, all hydrogen atoms in the protein were allowed to optimize. The hydrogen locations are not specified by the X-ray structure but these are necessary to improve the hydrogen bond geometries. In the second stage, all protein atoms were allowed to relax. Minimization in both stages was performed using 100 steps of steepest descent and 2000 steps of conjugate gradient algorithm. [1]

The docking procedure [9]:

Docking of the inhibitor in the active site of Rubisco enzyme was carried out using GOLD software (Genetic optimisation of ligand docking). The procedure consisted of tree main parts: (1) A scoring function to rank different binding modes; the Goldscore function is a molecular mechanics-like function with four terms were used. (2) A mechanism for placing the ligand in the binding site; GOLD use a unique method to do this, which is based on fitting point. (3) A search algorithm to explore possible binding modes; GOLD uses a genetic algorithm (GA). This method allows a partial flexibility of protein and full flexibility of ligand. [8] For each of the 10 independent GA runs, a maximum number of 100000 GA operations were performed on a set of five groups with a population size of 100 individuals. Operator weights for crossover, mutation,

and migration were set to 95, 95, and 10, respectively. Default cut-off values of 2.5 Å for hydrogen bonds and 4.0 Å for van der Waals distance were employed. When the top three solutions attained RMSD values within 1.5 Å, GA docking was terminated. [1]

Discussion:

The GA described in materials and methods required as input the approximate size and location of the active site, together with the coordinates of the protein and a ligand conformation. In any docking experiment it is required that the co-ordinate of the active site be known to reasonable accuracy, and the protein has high resolution. [2] Therefore, we used the highest resolution of RubisCO structure in the PDB (1.35Å), and defined the residues included in the active site by creating text file containing list of residues, and add the following commands to gold.conf file.

```
Floodfill_center = list_of_residues Cavity_file = path to tewt file
```

Gold will then read in the specified text file and extract the residues listed. After 10 GA runs, three highest fitness score were obtained (65.71, 64.72, 62.04) **Figure 2 a, b, c**. In these three solutions the phosphate and carboxylate groups of 2CA1P were bound to the active site of the enzyme in the same positions, however the positions of OH groups were different in the three solutions. When examining the binding place of the 2CA1P in the three solutions, we found that approximately the same binding place of the 4CABP with regards to the phosphate group P1. [7] **Table 1 (see supplementary material)**

In the results obtained after docking, the highest fitness score is 65.71 **Figure 2a** which makes it the closest result, compared to the other results (64.72, 62.04). When the fitness score increase, the molecular docking become better [2]. For each docking study, the solution with the highest fitness score was compared with the crystallographically observed binding mode. Depending on how close to the predicted binding mode, the result was assigned to one of four subjective categories. The first, good, was for those prediction where the binding mode, all hydrogen-bonding and metal coordination interactions and other close contacts between the protein and the ligand were reproduced correctly. If an acceptable result was generated with the ligand binding mode reproduced, but with some displacement of ligand groups from the experimental result, the GA prediction would be assigned to the close category. A third category, errors, was used for those predictions that were partially correct but contained significant errors. Finally, the fourth category, wrong, was used for completely incorrect prediction. [2] Due to the absence of any experimental results concerning the binding of the inhibitor 2CAIP in the active site and in order to classify the prediction results obtained and check the accuracy of the GA, the artificial inhibitor 2CABP was docked back in the active site of the enzyme using the same parameters and algorithm's used for the docking of natural inhibitor 2CA1P. The results were compared with the experimentally observed binding mode as illustrated in **Figure 2d**. We observed that the docked back and crystallographically binding inhibitor 2CABP occupy the same place

with some deviation in the functional groups **Figure 2d, Table 1(see supplementary material)**

The superposition of Ca atoms of the active site before and after docking shows that the RMSD between the two binding modes was in the range of 0.93 Å. It is a reasonably good result if compared with the result obtained by Jones and co-workers (1997) when they docked back the NADPH in Dihydrofolate reductase (DHFR). These authors found that the deviation between the two modes of binding (the crystallographically binding mode and the feed-back binding mode) are in the range of 1.2 Å, which is effective prediction. [2] Through this result we can say that the prediction obtained in the case of 2CA1P inhibitor is close to the real binding mode. This study may have important implications for the elucidation of photosynthesis regulation through the regulation of Rubisco activity by tight binding inhibitors. And we can consider it as a start point to study the binding mechanism of 2CA1P inhibitor, and also the residues that are involving in the liberation of this inhibitor from the active site of rubisco by Rubisco activase enzyme. Such regulation is important in biomass production and yield. Moreover, the modulation of Rubisco activity offers the possibility to control the stability of Rubisco under stress.

Conclusion:

In this work, we attempted to predict and define the position of 2CA1P inhibitor in the active site of Rubisco using the GOLD and Hyperchem softwares, in order to classify the prediction results obtained and check the accuracy of the GA, we docked back the artificial inhibitor 2CABP in the active site of the enzyme using the same algorithm and parameters used for the docking of 2CA1P inhibitor. The results were compared with the experimentally observed binding mode, we found that the docked back and crystallographically 2CABP inhibitor occupy the same place with some deviation in the functional groups. This result helped us to verify the accuracy of the algorithm in the case of the docking of the 2CA1P.

References:

- [1] MK Annamala *et al.*, *Bioinformatics* **9**: 339 (2008) [PMID: 17597917]
- [2] G Jones *et al.*, *J Mol Biol.* **267**: 727 (1997) [PMID: 9126849]
- [3] S Khan *et al.*, *Eur J Biocheur.* **266**: 840 (1999) [PMID: 10583377]
- [4] K Kim, AR Jr. Portis, *Photosynthesis Research* **87**: 195 (2006) [PMID: 16432666]
- [5] MAJ Parry *et al.*, *Experimental Botany* **59**: 1569 (2008) [PMID: 18436543]
- [6] JC Servaites *Plant Physiol.* **92**: 867 (1990) [PMID: 16667398]
- [7] TC Taylor *et al.*, *Biological Chemistry* **22**: 32894 (1996) [PMID: 8955130]
- [8] RD Taylor *et al.*, *Journal of computer-Aided Molecular Design* **16**: 151(2002) [PMID: 12363215]
- [9] ML Verdonk *et al.*, *PROTEINS: Structure, Function, and Genetics* **52**: 609 (2003) [PMID: 12910460]

Edited by P. Kanguane

Citation: Ounissi *et al.*, *Bioinformatics* 4(5): 206-209 (2009)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material :

Table 1 : Comparison between hydrogen bonding interactions of X-ray structure 2CABP and docked 2CABP at the active site of RubisCO

2CABP	after docking	Amino acids	Length of hydrogen bond	Crystallographically 2CABP	Amino acids	Length of hydrogen bond
	O3P1	Gly381 (N)	2.47	O2P1	Gly381 (N)	2.84
		Trp66 (NE)	2.81		Lys334 (NZ)	2.84
					Thr65 (OG1)	3.4
	O1P1	Gly404 (N)	2.66	O3P1	Gly404 (N)	2.73
	O2P1	Gly403 (N)	2.66	O1P1	Gly403 (N)	2.84
	O7	-	-	O1	Lys175 (NZ)	3.12
	O12 (CO2)	Lys175 (NZ)	3.04	O7 (CO2)	Glu60 (OE1)	3.33
		Lys334 (NZ)	3.09		Lys334 (NZ)	2.86
	O13 (CO2)	-	-	O6 (CO2)	Lys177 (NZ)	2.76
					Asn123 (ND2)	2.96
	O8	Lys175 (NZ)	2.85	O2	Lys175 (NZ)	3.01
		Asp203 (OD2)	2.81		Thr173 (OG1)	2.83
	O9	Thr173 (OG1)	3.3	O3	C-Lys201 (OQ1)	2.5
		Ser379 (OG)	2.6		His294 (NE2)	2.9
		Asp203 (OD2)	3.3			
	O5P2	His327 (ND1)	2.56	O4P2	His327 (ND1)	2.74
	O4P2	Arg295 (NE)	2.91	O6P2	Arg295 (NE)	2.35
	O6P2	-	-	O5P2	Arg295 (NH2)	2.89
O10	-	-	-	O4	Ser379 (OG)	2.86
					Ser379 (O)	2.87
O11	-	-	-	O5	-	-