Theoretical Simulation and Structural Analysis of a Protein

Ana Ros Camacho, Universitat de Barcelona
DESY Summer Student Program

September 16th, 2008

Supervised by Hans Bartunik, Grzegorz Domanski and Galina Kachalova
Max Planck Unit for Structural Molecular Biology, Hamburg, Germany

For Antonio José Palomares Díaz,
Requiescat in (Microbio)pace

Contents:

1.Introduction.....4
1.1. What is a protein.....4
1.2. Protein structure.....5
1.3. Function of a protein.....5
1.4. Our project.....8

2. Methods.....10
2.1. Software.....10
2.2. Algorithm.....11

3. Results.....17
3.1. EPSP Synthase.....17
3.2. S3P+PEP+EPSP Synthase.....18
3.3. EPSP+PO_4 20
3.4. Structure of the reaction.....21

4. Bibliography.....22

1-Introduction

1.1. What is a protein?

Proteins are large organic compounds made up of carbon, hydrogen, oxygen, nitrogen, phosphorus and sulfur. They are macromolecules with high molecular weight; structurally they are made of 20 amino acids arranged in a linear chain and joined together by peptide bonds:

- An amino acid is a nitrogenated compound that contains amine and carboxylic functional groups, separated by a carbon atom (the alpha carbon) united with an hydrogen atom and the R group. His general formula is $H_2NCHRCOOH$, (Figure 1) where R is an organic substituent: this R group is a side chain which makes the difference between all the amino acids (from size to properties). The only exception of this structure is proline, that lacks the NH_2 group because of the cyclization of the side chain. Amine and carboxyl groups can be attached in different ways. Alpha amino acids can form short polymer chains (less than 20 amino acids) called peptides (also oligopeptides) or proteins, also known as polypeptides.

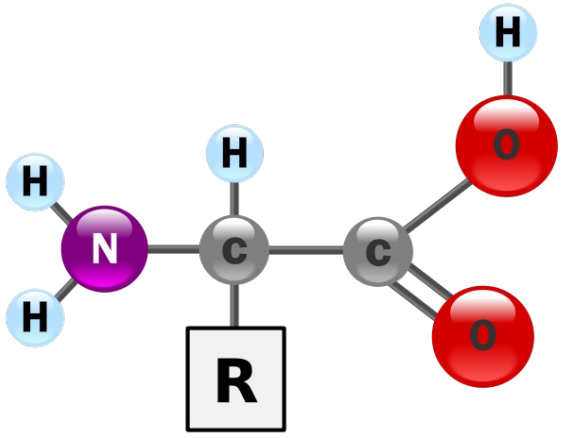


Figure 1: amino acid structure

- A peptide bond is a chemical bond formed between two molecules when the carboxyl group of one molecule reacts (Figure 2) with the amino group of the other molecule, releasing a molecule of water.

Figure 2: chemical reaction for peptide bond

Amino acids are joined end-to-end during protein synthesis by the formation of these peptide bonds, as much as the chain needs to elongate. One consequence is that the amino group of the first amino acid of a polypeptide chain and the carboxyl group of the last amino acid remain intact, and the chain is said to extend from its amino terminus (the so-called “N-terminal”) to its carboxy terminus (“C-terminal”). The formation of a succession of peptide bonds generates a “main chain”, or “backbone”, from which project the various side chains.

1.2. Protein structure

We can distinct four levels in a protein' structure (Figure 3):

- *Primary structure:* the amino acid sequence of the peptide chains.
- *Secondary structure:* highly regular sub-structures (alpha helix and strands of beta sheet) which are locally defined (there can be many different secondary motifs present in one single protein molecule).
- *Tertiary structure:* 3D structure of a single protein molecule; a spatial arrangement of the secondary structures. It also describes the completely folded and compacted polypeptide chain.
- *Quaternary structure:* complex of several protein molecules or polypeptide chains, which function as part of the larger protein complex.

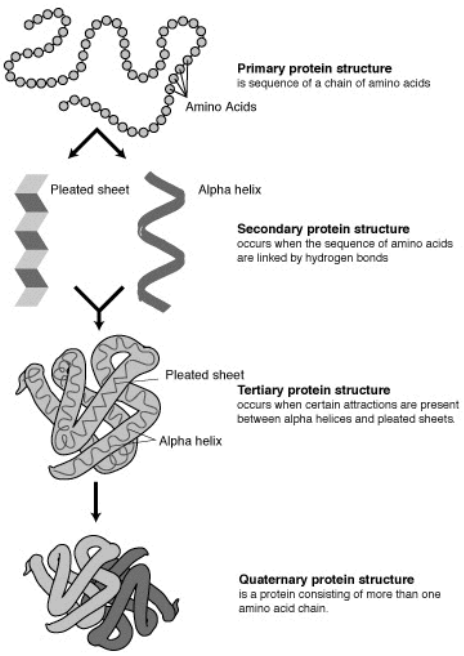


Figure 3: protein structures

1.3. Function of a protein

Proteins are essential parts of organisms and participate in every process within cells. Many of them are enzymes that catalyze biochemical reactions that are vital to metabolism. Some proteins are important in cell signaling, immune responses, cell adhesion, the cell cycle and in animals' diets (since animals cannot synthesize all the amino acids they need and must obtain essential amino acids from food). The chief characteristic of proteins that allows their diverse set of functions is their ability to bind other molecules specifically. The region of the protein responsible for binding a small molecule (“substrate” or “ligand”) is known as the binding site and is often a depression or "pocket" on the molecular surface. This binding ability is mediated by the tertiary structure of the protein, which defines the binding site pocket, and by the chemical properties of the surrounding amino acids' side chains. Protein binding can be extraordinarily tight and specific! Extremely minor chemical changes such as the addition of a single methyl group to a substrate partner can sometimes suffice to nearly eliminate binding.

The best-known role of proteins in the cell is their duty as enzymes. Enzymes are usually highly specific catalysts that accelerate a chemical reactions. They carry out most of the reactions involved in metabolism and catabolism, as well as DNA replication, DNA repair, and RNA synthesis. Some enzymes act on other proteins to add or remove chemical groups in a process known as post-translational modification. About 4,000 reactions are known to be catalyzed by enzymes. The rate acceleration conferred by enzymatic catalysis is often enormous - as much as 10^{17}-fold increase in rate over the uncatalyzed reaction in the case of orotate decarboxylase (78 million years without the

enzyme, 18 milliseconds with the enzyme).

Although enzymes can consist of hundreds of amino acids, it is usually only a small fraction of the residues that come in contact with the substrate, and an even smaller fraction - 3-4 residues on average - that are directly involved in catalysis. The region of the enzyme that binds the substrate and contains the catalytic residues is known as the active site, and both the enzyme and the substrate must be geometrically compatible for them to bind and perform a certain task (this is also called the “Lock and Key Theory”). A modern development of the Lock and Key Theory is the “Induced Fit Model” and it instead assumes that an active site is more flexible and that the presence of certain residues in the active site will encourage the enzyme to locate the correct substrate, after which conformational changes may occur as the substrate is bound. Substrates bind to the active site of the enzyme or a specificity pocket through hydrogen bonds, hydrophobic interactions, Van der Waals bonds or a combination of all of these to form the enzyme-substrate complex. Residues of the active site will act as donors or acceptors of protons or other groups on the substrate to facilitate the reaction.

Protein conformation is of paramount importance in understanding biomolecular interactions. Two molecules bind optimally with each other only after conformational changes have been done at their interface. Conformational changes may also take place away from the binding interface. This is often the prerequisite for functional activity. For protein like hemoglobin that shows allosteric behavior, the binding of small molecules at a region of the protein affects its binding affinity with other molecules at a distant region. In membrane receptors binding of ligand at the extracellular region causes changes at the cytoplasmic region, so that an extracellular signal is allowed to alter intracellular activity.

Conformational changes in proteins are made possible by their intrinsic flexibility and they may occur with only relatively small expenditure of energy. At the molecular structural level, conformational changes in single polypeptides are the result of changes in main chain torsional angles and side chain orientations. On the one hand, the overall effect of such changes may be localized with reorientations of a few residues and small torsional changes in the regional main chain. On the other hand, torsional changes localized at very few critically placed residues may lead to large changes in tertiary structure. The later type of conformational changes is described as domain motions, and they have two basic components.

Hinge motions may occur in the context of secondary structure interactions, within strands, beta-sheets and alpha-helices not constrained by tertiary packing forces. To qualify as fulcrum for hinge-motion, residue must bear very little tertiary structure packing constraints on its main chain. The hinge lies outside the interface between the two domains inter-connected by the hinge. On hinge-opening the motion is perpendicular to the plane of the interface, which is lost after opening. The closed conformation is usually stabilized by a bound ligand. This is necessarily so, for if the closed conformation is strongly held together without a ligand, then the hinge opening will have to cross a high energy barrier. Hinge motion at extended strand involves a few large changes in main chain torsion angles at the hinge connecting two domains, constrained only

by the Ramachandran allowance of torsional angles. As the range of (phi, psi) angles is relatively large for extended strand, the hinge angle can change by up to 60^0 with only torsional changes in two residues. In beta-sheets two adjacent strands can move like hinges of a door, with extra constraint of hydrogen bonds that hold the sheet together. To obtain the same hinge angular change, torsional changes at three or more residues are required. Alpha helices are further constrained with their more restrictive hydrogen bonding, thereby in need of more small-amplitude torsional angular changes to bend themselves significantly. Proline-kinked helix may allow larger torsional angular changes. Torsional angular changes may stretch an alpha helix by about 3 Angstroms into a 3_{10} helix. There are also cases where a long helix may split into two smaller helices inter-connected by a short extended strand that was previously in helical conformation.

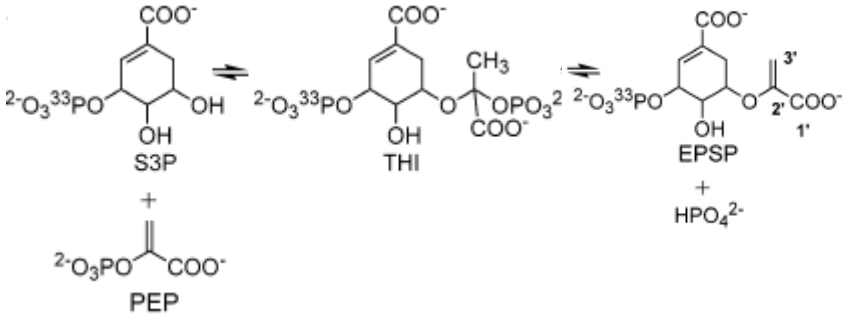
Shear motions occur parallel to the interface between closely packed segments of polypeptides, in the context of tertiary structure interactions. Proteins that shear often have layered architecture, with shearing that may occur across helix-helix, helix-sheet, helix-loop and sheet-loop interfaces. This type of motion is more severely constrained with additional packing contacts due to interdigitating side chains. A large enough sheared domain motion is due to the combination of a number of shear movements. Large shear movement that make the interdigitating lock from one state to another is not observed in domain motions, as in subunit interface of allosteric proteins. On the other hand small shear movements that do not require interdigitational repacking are common in domain motions. These shear movements are accommodated by small changes in side chain torsional angles with no significant deformation in main chain torsional configuration of the interface segments.

Multimeric proteins have an extra dimensionality to conformational transitions due to their quarternary structure. Haemoglobin is the classic prototype of allosteric proteins with cooperative behaviour. In the case of lamprey haemoglobulin, cooperativity is mediated by reversible dissociation and association of subunits. Packing at subunit interfaces are broken off all together. As for human haemoglobin this is achieved by equilibrium between two alternative quarternary structures of the tetramer. The overall structure changes are due to breaking and formation of electrostatic interactions at the tertiary and quaternary levels, as a result of binding to oxygen or other allosteric effectors. At the interface there is large shear motion that involves repacking in transition between tense and relax states. Cooperativity can also be realised singularly by saving expenditure of entropic energy, as in the binding of dimeric *trp* repressor and immunoglobulins to *operator* and antigen, respectively. After binding of one monomer to a binding site on the target molecule with energy expenditure to pay for the decrease in entropy, the binding of the other monomer to adjacent binding site on the same ligand molecule requires less energy for entropy reduction because the first binding has juxtapose the two interacting molecules, such that the second monomer is already placed in the favorable position to interact with the second binding site without having to increase the order of the bimolecular complexity much more.

1.4. Our project

The enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS; EC 2.5.1.19, gene AroA) is the sixth enzyme of the shikimate pathway essential for synthesis of aromatic amino acids. Uncommon features of this enzyme have received considerable attention by extensive studies over the last three decades. From biological point EPSPS is attractive potential target to design antimicrobial drugs as enzyme belongs to shikimate pathway which is absent in mammals but essential for a number of pathogenic microorganisms.

The enzyme catalyzed unique reaction (Figure 4) in which the transfer of enolpyruvyl moiety from phosphoenolpyruvate (PEP) proceeds by it addition to the 5-hydroxyl of shikimate-3-phosphate (S3P) through a stable tetrahedral intermediate (THI) to yield the enolpyruvyl-shikimate-3-phosphate (EPSP) via cleavage of C-O rather then more customary P-O bond by elimination of inorganic phosphate (P_i).



Step 1 Step 2 Step 3

Figure 4: EPSP Synthase reaction in three steps

Another uncommon peculiarities of this reaction are: (i) the formation of unusually stable reaction intermediate (THI) which has been isolated and characterized (ii) the distrubition of products in favour of EPSP and P_i whereas in solution PEP and P_i are the predominant products of THI breakdown. So far this novel catalytic mechanism was established for only one another enzyme MurA that transfer PEP to the 3-hydroxyl of UDP-*N*-acetylglucosamine (MurA - the first enzyme of bacterial cell wall biosynthesis is also attractive antimicrobial target).

Structural studies of EPSPS enzyme have depicted specific features of its folding pattern. This molecule consist of N and C-terminal domains, each of which comprises 3 subdomains with α/β topology. Deduced large-scaled domain motion to accompany transition from “open” - unliganded form (Figure 5) to “close” - substrate bounded form (Figure 6) is indicated that enolpyruvyl transfer reaction follows an induced-fit mechanism, which is formally described as relative rotation of N- and C-terminal domains around an axis located at the their interface. Up to now the only two structures of “open” form and a number of crystal structures of “close” form of the enzyme is determined. The most crystal structures of “close” form are complexes of EPSPS with one substrate, or with substrate and inhibitor, or with synthesized THI or it’s analogues that provide the valuable information about catalytically important residues in active site and suggestions about their role in addition and elimination enzymatic steps that however is still subject of debate.

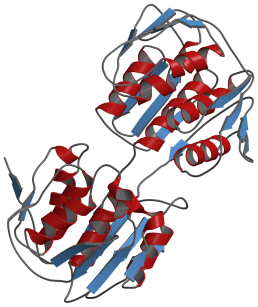


Figure 5: EPSPS open form

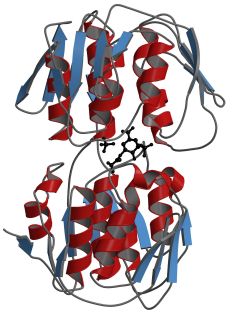


Figure 6: EPSP closed form

Since so far the structures of EPSPS complex with both substrates, or with natural intermediate, or products of reaction aren't determined, several important questions remain out of scope of structural studies, including the following : (i) how the catalytic course of the reaction is coupled with induced-fit mechanism (ii) how structure controls transition from precedent to subsequent the reaction step.

In our project we pretended to answer the first question. Our target was studying the catalyzing role of EPSP Synthase (such as the induced-fit mechanism of the process) in a reaction between S3P and PEP, obtaining after EPSP and phosphate groups: we look for the whole path of the reaction. Our start line was the two pdb files which contained the structural data from the first step ("Step 1")and the last step ("Step 3") of the reaction.

2- Methods

2.1. Software

To reach our objective, we need some computational tools for the treatment of the protein. It is necessary some kind of special software for these kind of super-collectivities, and we find it in Molecular Dynamics. MD is a form of computer simulation, wherein atoms and molecules are allowed to interact for a period of time under known laws of Physics giving a view of the motion of the particles; it can be understood as a “virtual experiment”, like a representation of an interface between laboratory experiments and theory. One of MD's key contributions is creating awareness that molecules, like proteins, are machines in motion. MD probes the relationship between molecular structure, movement and function.

AMBER allowed us to carry out MD simulations. Concretely, the version 9 of the AMBER software suite, released in March 2006, was used. AMBER is not a single program, but a collective name for a suite of programs, about 50. The major programs (such as the ones we used most) are as follows:

Preparatory programs:

- *LEaP*: LEaP is an X-windows-based program that provides for basic model building and Amber coordinate and parameter/topology input file creation. It includes a molecular editor which allows for building residues and manipulating molecules.
- *Antechamber*: This program suite automates the process of developing force field descriptors for most organic molecules. It starts with structures (usually in PDB format), and generates files that can be read into LEaP for use in molecular modeling. The force field description that is generated is designed to be compatible with the usual Amber force fields for proteins and nucleic acids.

Simulation programs:

- *Sander*: it is the "main" program used for molecular dynamics simulations, and is used for replica-exchange, thermodynamic integration, and potential of mean force (PMF) calculations.

Analysis programs:

- *Ptraj*: it is used to analyze MD trajectories, computing a variety of things, like RMS deviation from a reference structure, hydrogen bonding analysis, time-correlation functions, diffusional behavior, and so on.

More information about AMBER can be found in its website: http://amber.scripps.edu/. Although the current version is AMBER 10, it is possible for new users to get a number of tutorials of AMBER 9 in http://amber.scripps.edu/tutorials/ , and also the manuals for AMBER 8, 9 and 10. It is also possible to learn how to use AMBER with Ross Walker tutorials, available in http://www.rosswalker.co.uk/tutorials/amber_workshop/ .

Input files for simulation were prepared with Antechamber and LeaP along with GAFF force fields. Serial and parallel versions of Sander were employed to run the simulations and the data post-processing, like RMSd calculations or trajectory analysis, were done with Ptraj. For structure and trajectory visualization, the program PyMOL version 0.90 was used. The official website, www.pymol.org, provides all the information one can need, like tutorials or the

manual, such as the current version, PyMOL v.1.1.

The protocol we followed for the analysis of this data was:

1. Neutralization of the molecule adding sodium ions
2. Solvatation of the molecule in a box of water
3. Minimization of the molecule
4. Equilibration of the molecule
5. Production of the MD data
6. Analysis of energy, RMS, structure and others

Concerning minimization, we run a dual stage one: first, it is only allowed the minimization only for the hydrogen ions; after, for all the atoms. Minimization is a “fixing up” process of the positions of the atoms in order to remove any bad contacts that may lead to unstable molecular dynamics. Equilibration is another step where we remove any problem related to bad Van der Waals (non bond) or electrostatic interactions. We solvated the molecules in a box of water, but this water did not feel the influence of the solute or charges and moreover there may be gaps between the solvent and solute and solvent and box edges. If we are not careful such holes can lead to "vacuum" bubbles forming and subsequently an instability in our molecular dynamics simulation. Thus we need to run careful minimization before slowly heating our system to 300 K (we suppose that it is, in the beginning, at 0°K). It is also a good idea to allow the water box to relax during a MD equilibration stage prior to running production MD. In this phase it is another idea, since we use periodic boundaries, to keep the pressure constant and so allow the volume of the box to change. This approach allows the water to equilibrate around the solute and come to an equilibrium density. It is essential that we monitor this equilibrium phase in order to be certain our solvated system has reached equilibrium before we start obtaining results (production data) from our MD simulation.

Three different MD simulations were run at the same time in our project: one involving EPSP Synthase enzyme alone, “aroa_nos3p_nopep”; another one, “aroa_s3p_pep”, with EPSP Synthase, Shikimate-3-Phosphate, and phosphoenolpyruvate; and the last one, “aroa_po4”, involving EPSP with phosphates and sulfates as residues.

2.2. Algorithm

We describe the whole process of simulation as follows:

a) Preparing starting structures:

To start a simulation, the initial structure has to be generated, such as the topology/parameter and coordinate files for performing minimization or dynamics with sander. It can be done by Leap but before we need the files that it can read (pdb files).

Because our proteins contain some non-standard residues (S3P, PEP, sulfate, phosphate) which are not predefined in the Amber database, it is necessary to provide structural information and force field parameters for all the non-standard residues present in the simulation before can be created the Sander input files. So, we must copy the pdb coordinates from the original pdb file to different files containing each of the non-standard residues, and then we will be ready to load every residue as a unit in Leap.

b) Generating the parameter/topology and coordinate files:

First, a decision must be done about which force field shall we use. We chose GAFF ("general Amber force field"), as Antechamber is designed to be used with it. This force field has been specifically designed to cover most pharmaceutical molecules and is compatible with the traditional AMBER force fields in such a way that the two can be mixed during a simulation.

Like the traditional AMBER force fields, GAFF uses a simple harmonic function form for bonds and angles but unlike the traditional protein and DNA orientated AMBER force fields the atom types used in GAFF are much more general such that they cover most of the organic chemical space. The current implementation of the GAFF force field consists of 33 basic atom types and 22 special atom types. The charge methods used can be HF/6-31G* RESP or AM1-BCC.

By design, GAFF, is a complete force field (so that missing parameters rarely occur), it covers almost all the organic chemical space that is made up of C, N, O, S, P, H, F, Cl, Br and I. Moreover, wince GAFF is totally compatible with the AMBER macromolecular force fields it should prove to be an useful molecular mechanical tool for rational drug design. Specially in binding free energy calculations and molecular docking studies.

Second, we should generate the prepin and frcmod files for each molecule, and this was made with Antechamber: for example, for S3P, we can generate prepin file with

$AMBERHOME/exe/antechamber -i s3p.pdb -fi pdb -o s3p.prepin -fo prepi -c bcc -s 2

where the -i s3p.pdb specifies the name of the 3D structure file and the -fi pdb tells Antechamber that this is a pdb format file. The -o s3p.prepin specifies the name of our output file and the -fo prepi states that we want the output file to be of amber PREP format (this is an internal format supported by Leap). The -c bcc option tells antechamber to use the BCC charge model in order to calculate the atomic point charges while the -s 2 option defines the verbosity of the status information provided by antechamber. In this case we have selected verbose output (2).

(Similarly for the other residues, and once finished the analysis of all of them, for the whole protein).

For generating the frcmod file, we can use prmchk (another program from Antechamber):

$AMBERHOME/exe/parmchk -i s3p.prepin -f prepi -o s3p.frcmod

After generating prepin and frcmod files, we load them in Leap: it is possible to use the graphical version (XLeaP) or the terminal interface (tLeaP). It is possible to execute it typing:

$AMBERHOME/exe/xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff99

for XleaP, or:

$AMBERHOME/exe/tleap -f leaprc.gaff -f leaprc.ff99

for tLeaP. Before we do anything, we must ensure that Xleap knows about the GAFF force field:

source leaprc.ff99

And then, if we load the pdb file,

S3P=loadpdb s3p.pdb

we neutralise the molecule with sodium ions, solvate it in a box of water,

addions S3P Na+ 0
solvatebox S3P TIP3PBOX 12

and check if the unit is fine,

check S3P

we can obtain parameter/topology files and coordinates files typing:

saveamberparm S3P s3p.prmtop s3p.inpcrd

It is also possible, in Xleap, to see the molecules:

edit S3P

After analyzing all the residues, we can do the same for the whole pdbs, always with the same algorithm, but before loading the pdb file, we must load the residues frcmod and inpcrd files: for example, for aroa_s3p_pep, we have to load S3P and PEP files

loadamberprep s3p.prepin
loadamberparams s3p.frcmod
loadamberprep pep.prepin
loadamberparams pep.frcmod

c) Running a MD simulations:

A typical MD simulation consists of certain steps, which normally are realized in three stages: minimization, equilibration and long production run. Depending on the accuracy wanted and on the characteristics of the investigated system it can be needed to do more than one minimization or equilibration; in our case we chose two minimizations. One should also choose the adequate parameters in every stage in order to run the correct simulation.

In order to save time, we created run.sh files for every case that contained all the instructions for running MD: for example, for aroa_s3p_pep,

```
#!/bin/csh
#Script to run simple minimization/MD simulation using SANDER
setenv MOL 'aroa_s3p_pep'
setenv DO_PARALLEL 'mpirun -np 8'

source /opt/amber9/amber.setup

echo "Starting 1st minimization..."

${DO_PARALLEL} ${AMBERHOME}/exe/sander.MPI -O -i min1.in -p
${MOL}.prmtop -c ${MOL}.inpcrd -o ${MOL}_min1.out -r ${MOL}_min1.rst

#minimization
echo "Starting 2nd minimization..."

${DO_PARALLEL} ${AMBERHOME}/exe/sander.MPI -O -i min2.in -p
${MOL}.prmtop -c ${MOL}_min1.rst -o ${MOL}_min2.out -r ${MOL}_min2.rst

#equilibration
echo "Starting equilibration..."
```

${DO_PARALLEL} ${AMBERHOME}/exe/sander.MPI -O -i md1.in -p ${MOL}.prmtop -c ${MOL}_min.rst -o ${MOL}_md1.out -r ${MOL}_md1.rst -x ${MOL}_md1.mdcrd

#production run
echo "Starting production MD..."

${DO_PARALLEL} ${AMBERHOME}/exe/sander.MPI -O -i md2.in -p ${MOL}.prmtop -c ${MOL}_md1.rst -o ${MOL}_md2.out -r ${MOL}_md2.rst -x ${MOL}_md2.mdcrd

And analogously for aroa_po4 and aroa_nos3p_nopep. We can see that 8 processors were used. Approximately, every simulation took around 8 days to finish. Input files for minimization were:

- Min1.in

First minimization of our complex
&cntrl
imin=1, maxcyc=250, ncyc=150,
cut=16, ntb=0, igb=0,
/

- Min2.in

Second minimization of our complex
&cntrl
imin=1, maxcyc=1000, ncyc=250,
cut=16, ntb=0, igb=1,
/
–

Input file for equilibration was:

- Md1.in

Initial MD equilibration
&cntrl
imin=0, irest=0,
nstlim=20000, dt=0.001, ntc=1,
ntpr=20, ntwx=20,
cut=16, ntb=0, igb=1,
ntt=3, gamma_ln=1.0,
tempi=0.0, temp0=300.0,
/

And for the production run:

- Md2.in

Production MD
&cntrl
imin=0, irest=1, ntx=5,
nstlim=1000000, dt=0.001, ntc=1,
ntpr=20, ntwx=20,
cut=16, ntb=0, igb=1,
ntt=3, gamma_ln=1.0,
tempi=300.0, temp0=300.0,
/

Comment: for aroa_nos3p_nopep, for some reason we preferred to run a shorter simulation, so the input files were:

- Min1.in

```
First minimization
 &cntrl
  imin   = 1,
  maxcyc = 1000,
  ncyc   = 500,
  ntb    = 1,
  ntr    = 1,
  cut    = 10
 /
```

- Min2.in

```
Second minimization
 &cntrl
  imin   = 1,
  maxcyc = 2500,
  ncyc   = 1000,
  ntb    = 1,
  ntr    = 0,
  cut    = 10,
 /
```

- Md1.in

```
Equilibration
 &cntrl
  imin   = 0,
  irest  = 0,
  ntx    = 1,
  ntb    = 1,
  cut    = 10,
  ntr    = 1,
  ntc    = 2,
  ntf    = 2,
  tempi  = 0.0,
  temp0  = 300.0,
  ntt    = 3,
  gamma_ln = 1.0,
  nstlim = 10000, dt = 0.002,
  ntpr = 100, ntwx = 100, ntwr = 1000
 /
```

- Md2.in

```
Production MD
 &cntrl
  imin = 0, irest = 1, ntx = 7,
  ntb = 2, pres0 = 1.0, ntp = 1,
  taup = 2.0,
  cut = 10, ntr = 0,
```

ntc = 2, ntf = 2,
tempi = 300.0, temp0 = 300.0,
ntt = 3, gamma_ln = 1.0,
nstlim = 50000, dt = 0.002,
ntpr = 100, ntwx = 100, ntwr = 1000
/

d) Results analysis:

When simulations were finished, results can be analyzed. The information contained in the output files can be processed with a Perl script generating files with the values obtained during the simulation about energy, pressure, density, energy... vs time: for example, for aroa_s3p_pep,

./process_mdout.perl aroa_s3p_pep_md1.out aroa_s3p_pep_md2.out

To calculate RMS for all the protein and its backbone, we used ptraj:

ptraj aroa_s3p_pep.prmtop ptraj.in ptraj.out

where the ptraj.in file was, in the case of aroa_s3p_pep,

- ptraj.in for the whole protein

#ptraj file for first protein

trajin aroa_s3p_pep_md1.mdcrd 1 1000 50
trajin aroa_s3p_pep_md2.mdcrd 1 50000 50

*rms first out allprotein.rms ":1-424 & !@H**"

trajout allprotein.pdb PDB append

- ptraj.in for the backbone

#ptraj file for the backbone of the protein

trajin aroa_s3p_pep_md1.mdcrd 1 1000 50
trajin aroa_s3p_pep_md2.mdcrd 1 50000 50

rms first out backbone.rms ":1-424@Ca,N,C & !@H*"

trajout backbone.pdb PDB append

In every case it should be specified the correct number of atoms.

We finally plotted the results using the classic Gnuplot and also PyMOL.

3. Results

We obtained a quite good simulation of the structures of Step 1 and 3 from the reaction. Concerning energies and RMS, the following values for every simulation were shown:

3.1. EPSP Synthase:

In this case, we considered only Step 1 protein's structure alone, without any substrate (no S3P and no PEP); in other words, only EPSP Synthase structure. In figure 7, we can see its evolution along time of the energies (kinematics, potential and total):

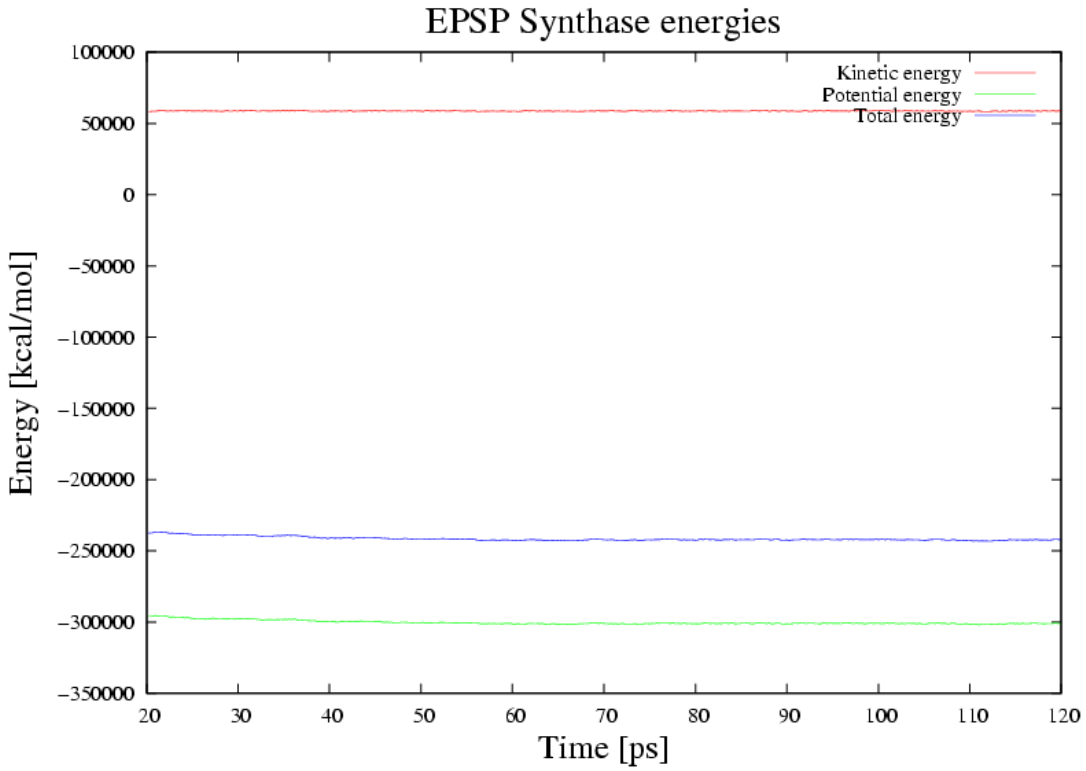


Figure 7: Energies for the structure at Step 1 without S3P nor PEP

Also, in figure 8 we can have a look at the RMS of the structure obtained after all the simulation process, compared with the data from original pdb file, obtained by X-Ray.

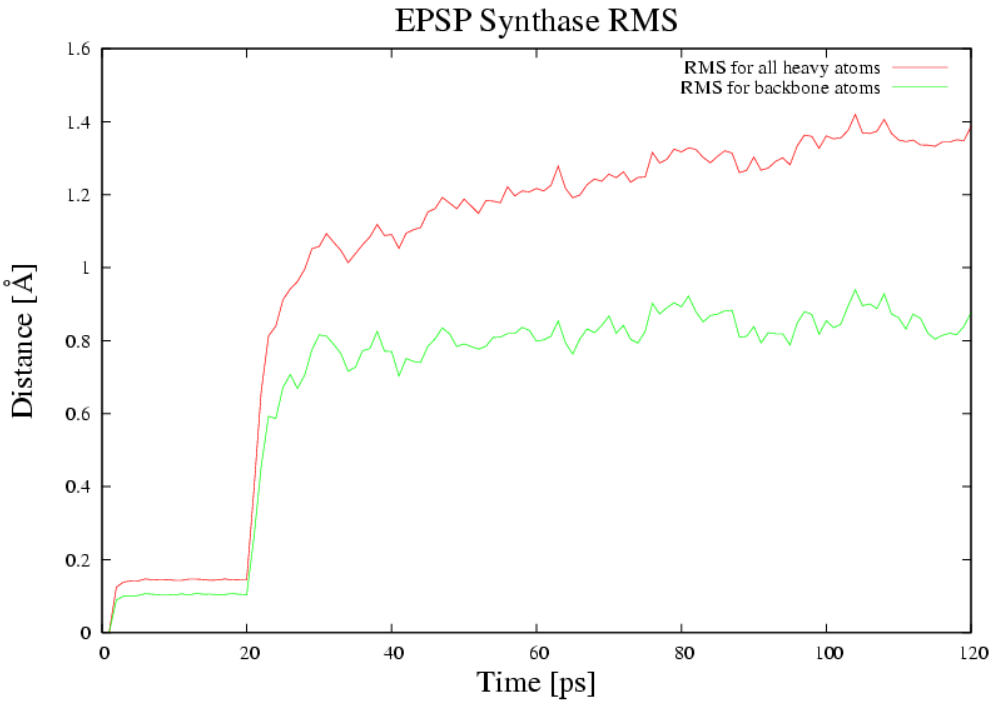


Figure 8: RMS for the structure at Step 1 without S3P nor PEP

3.2. S3P+PEP+EPSP Synthase:

Here we take into analysis the Step 1 of the reaction structure, S3P+PEP+EPSP Synthase, first again the energies (kinetics, potential and total, see Figure 9), and after the values of RMS of the structure obtained after all the simulation process, compared with the original pdb data file obtain by X-Ray (Figure10).

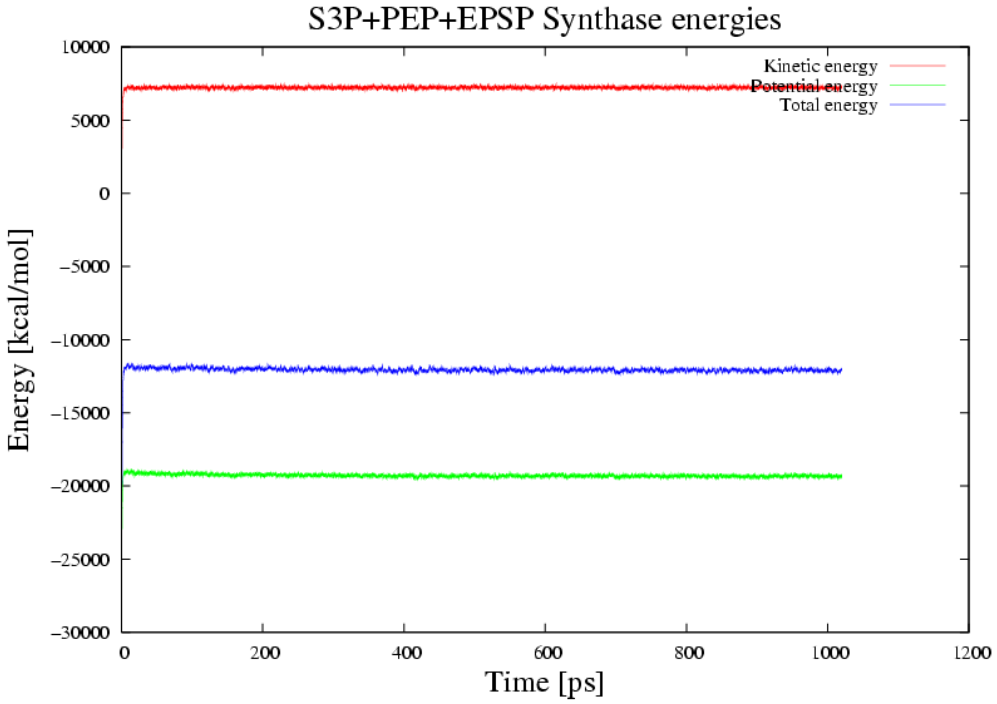


Figure 9: energies for the structure at Step 1 of the reaction

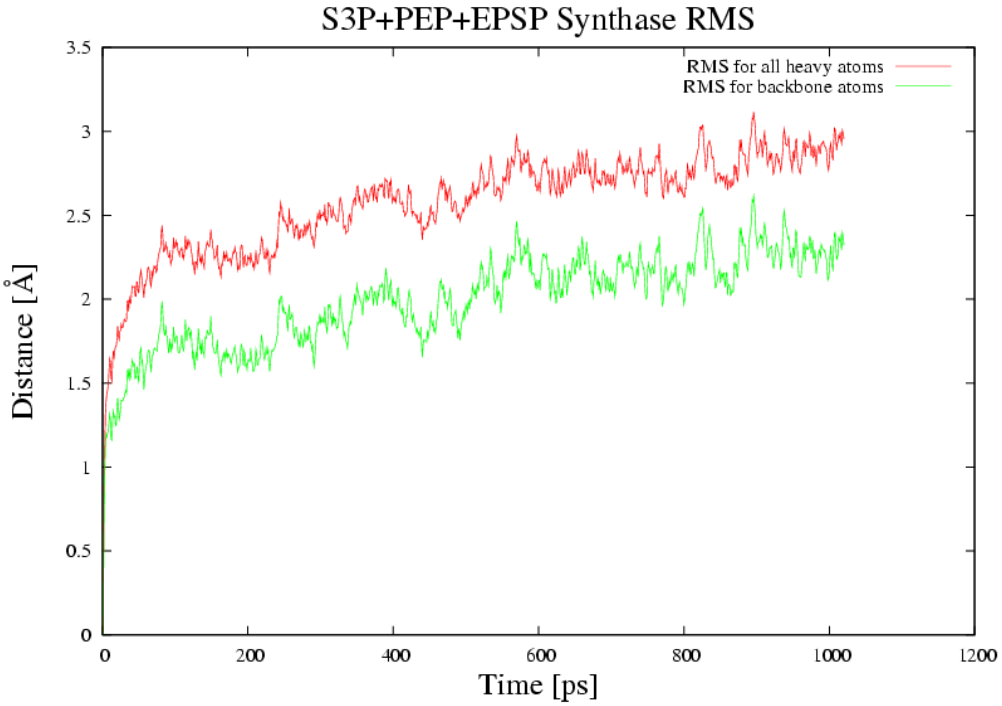


Figure 10: RMS for the structure at Step 1 of the reaction

3.3. EPSP+PO4:

In this last case we analysed the structure of the Step 3 of the reaction, EPSP+PO_4. In the beginning, we used a solvated phosphate molecule, but then during the simulation, we could see that it was produced a molecule distortion: oxygen atoms began to join together, something at least curious because the atoms should be repelled because of their charges. We had to change some parameters of the topology input file: charges for the atoms of the phosphate group, in order to avoid this joining between the oxygen atoms; and also bond force constant.

Then we could see our simulation was correct, meaning that no molecule distortion was observed in the phosphate nor the whole structure.

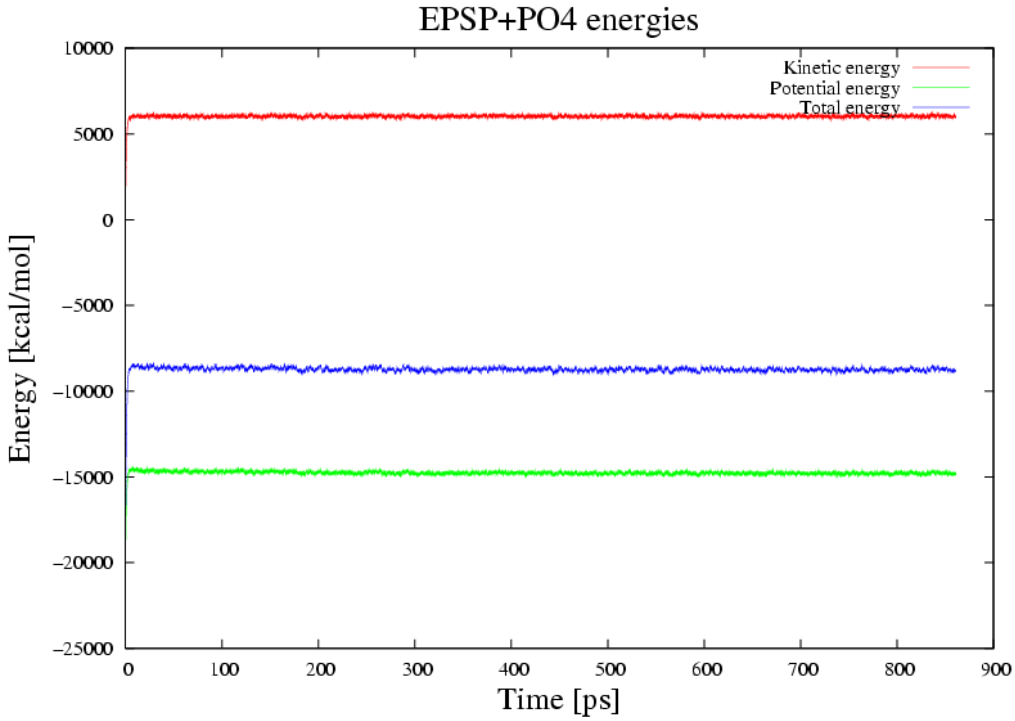


Figure 11: energies for the structure at Step 3 of the reaction

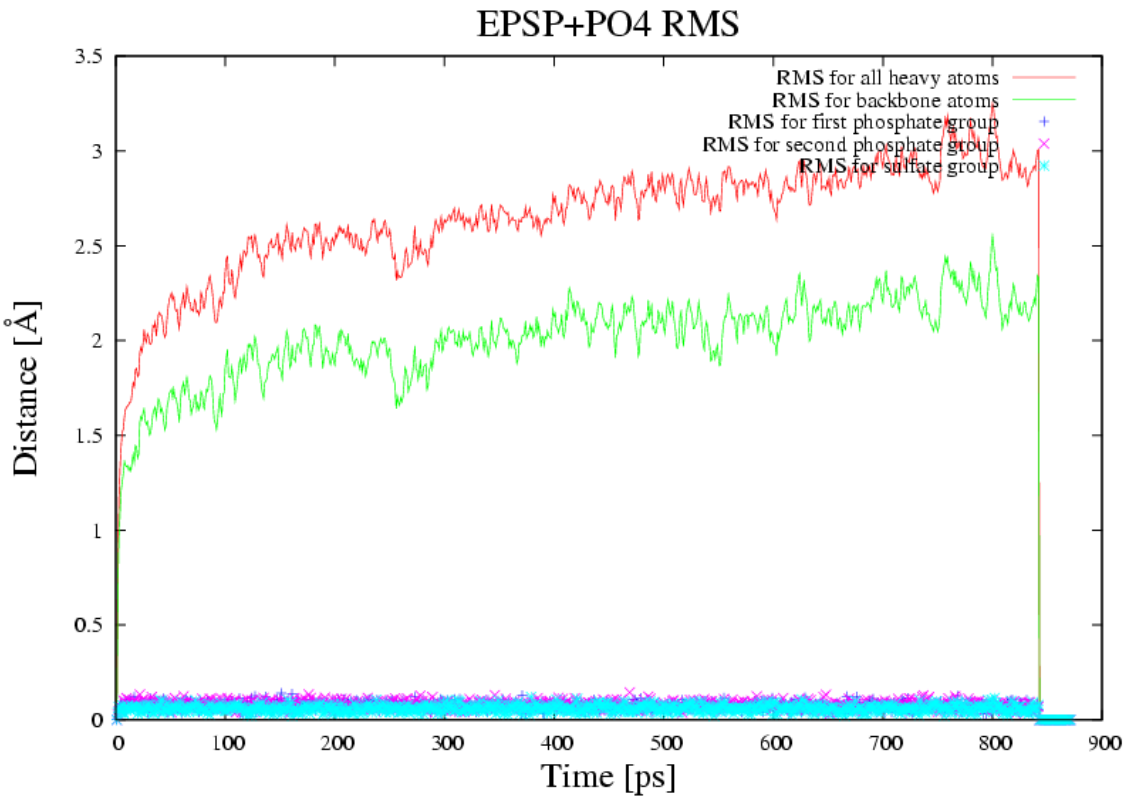


Figure 12: RMS for the structure at Step 3 of the reaction

3.4. Structure at Step 1 and 3 of the reaction:

Also, we can have a look at the structures of the two steps:

a) Step1: S3P+PEP+EPSP Synthase structure

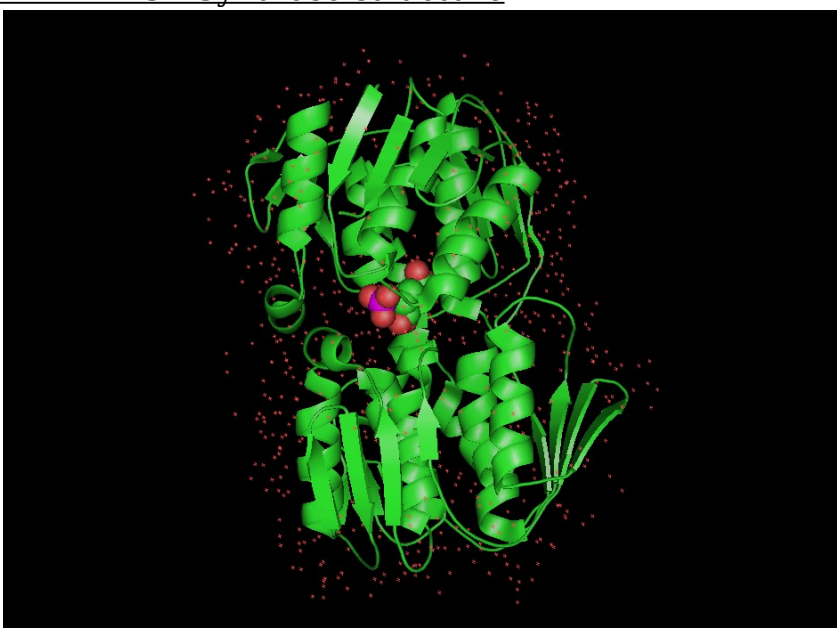


Figure 13: S3P+PEP+EPSP Synthase

b) Step3: EPSP+PO_4 structure

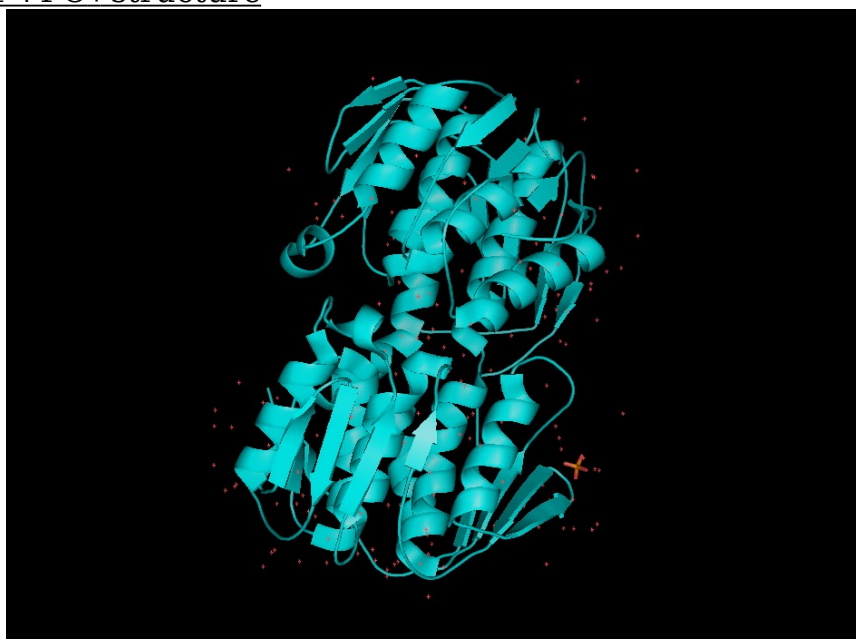


Figure 14: EPSP+PO_4

If we make a superposition between Step 1 and Step 3 structure,

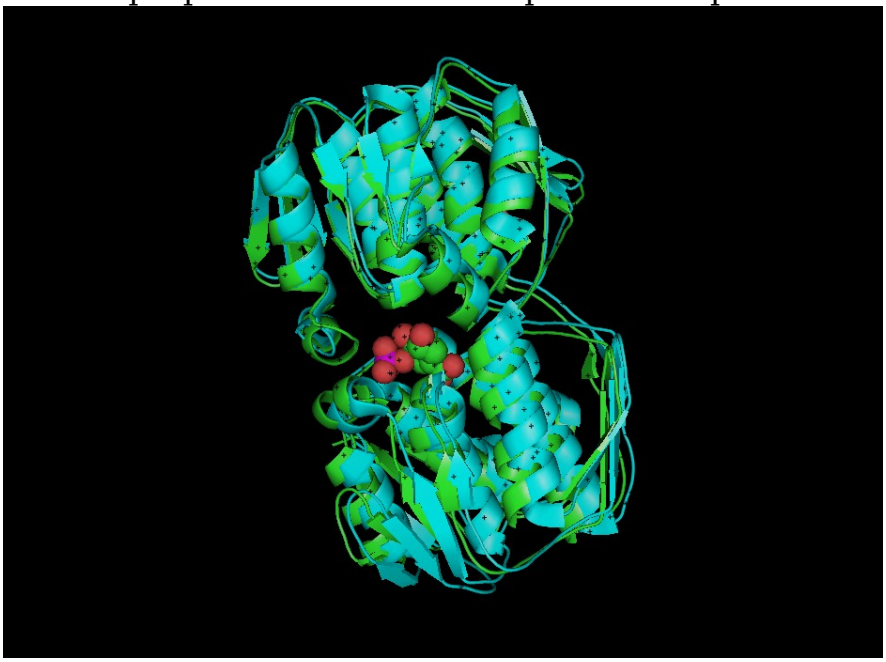


Figure 15: superposition Step1-Step3

we can see that the structure of the enzyme has changed, a hinge motion was produced between the two steps in an opening way.

4. Bibliography

Websites:

http://en.wikipedia.org/wiki/Proteins
http://en.wikipedia.org/wiki/Protein_structure
http://en.wikipedia.org/wiki/Amino_acid
http://en.wikipedia.org/wiki/Peptide
http://en.wikipedia.org/wiki/Peptide_bond
http://en.wikipedia.org/wiki/Protein_structure_prediction
http://en.wikipedia.org/wiki/X-ray_crystallography
http://www.rosswalker.co.uk/tutorials/amber_workshop/
http://en.wikipedia.org/wiki/Active_site
http://www.cryst.bbk.ac.uk/PPS2/projects/vun/PPS96proj_Vun.htm

Papers and books:

- “Enolpyruvyl Activation by Enolpyruvylshikimate-3-phosphate Synthase”, Clark and Berti, Biochemistry, Vol.46, No.7, 2007.
- “Molecular Dynamics Simulation of EPSP Synthase”, Maria Cano Colino, DESY Summer Student 2007 Report.
- Amber 9 Users'Manual
- “Introduction to Protein Structure”, Second Edition; Branden, Carl and Tooze, John, Garland Publishing, 1998.