Understanding Phosphoenolpyruvate carboxylase catalysis as a step to the engineering of a CO$_2$ capture device

Rita Maria Bolota Velho Pires da Fonseca
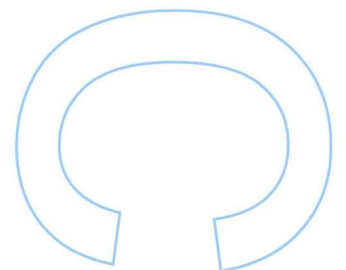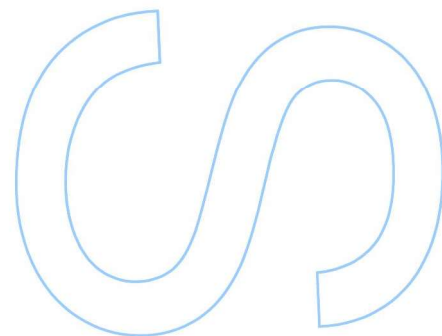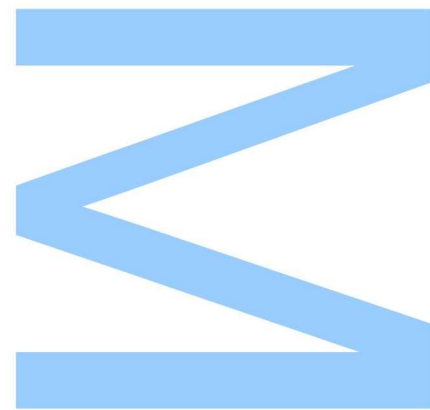
Bioquímica
Departamento de Química e Bioquímica
2020

Orientador
Pedro Manuel Azevedo Alexandrino Fernandes, Professor Associado, FCUP

Coorientador
Maria João Ribeiro Nunes Ramos, Professora Catedrática, FCUP

U.PORTO

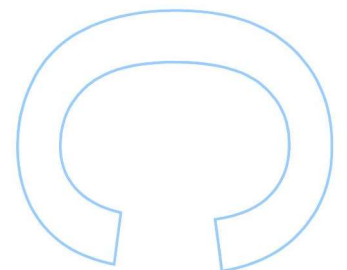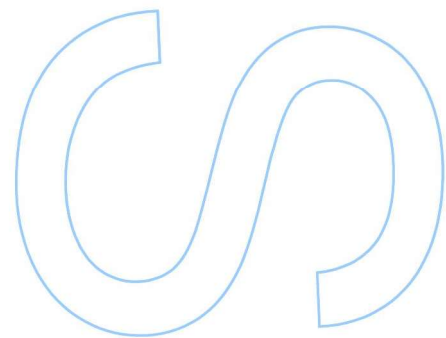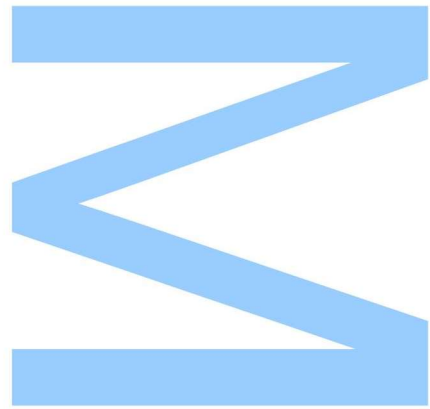**FC** FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |1

Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the engineering of a CO$_2$ capture device.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |2

Agradecimentos

Agradeço a Deus ter chegado aqui. Agradeço à minha mãe e ao meu pai toda a ajuda preciosa que me dão, e aos meus irmãos, que me enchem de alegria. Um agradecimento especial à minha irmã Sara, pela motivação que me dá, a cada dia; não sei onde estaria se não fosse ela.

Agradeço ao Professor Pedro a orientação ao longo da tese e a sua maneira clara de explicar. E o facto de tratar as pessoas de uma forma delicada e bem-disposta. Tem uma personalidade incrível. Agradeço também à Professora Maria João, por ser uma pessoa tão empática e por acompanhar o meu trabalho.

Obrigada ainda ao Pedro Paiva, por todo o esforço e dedicação a ensinar-me, discutir ideias e tentar (com sucesso) resolver os múltiplos problemas que surgiram ao longo do trabalho. Por fim, agradeço à equipa em geral, pois são pessoas simpáticas, em especial o Rui.

Apesar de não estarem tão presentes neste trabalho como previsto, agradeço ao Professor Luís, que me deu duas fontes excelentes para a introdução, e ao Ilde, simplesmente pela pessoa que é. Trocar ideias com ele anima-me e faz-me trabalhar melhor.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |3

Summary

CO_2 levels are being linked to global warming, which in turn is hypothesized to have damaging effects on people and living beings. There are different opinions, but if CO_2 levels can be harmful and be solved, efforts to reduce them are justified.

It is important to lower the emissions of CO_2, but due to industrialization, it is a task that may benefit from the help of Carbon Dioxide Removal (CDR) devices. For the design of new carbon capturing devices, natural processes of carbon removal may be considered, such as the ones promoted by plants. At a fundamental level, the enzymes are the key players in carbon fixation by plants. In this work, Phosphoenolpyruvate Carboxylase (PEPC), an enzyme that plays such a role in some photosynthetic species, is analysed. This works focuses on the reaction mechanism by PEPC.

A 3D model of the enzyme: substrates complex with an Mg^{2+} ion cofactor is proposed. This structure underwent simulations to study its stability in solution and to study the reaction mechanism. Most importantly, quantum mechanics/molecular mechanics (QM/MM) calculations were run to test the mechanism of reaction previously proposed in the literature.

The current mechanistic proposal was not verified computationally, since the energy barrier obtained for the first step of the reaction was too high to be feasible. However, the proposal cannot yet be dismissed. Other alternative mechanisms have also been evaluated, and some discussion about them is provided. Most of the attempts seem to have failed, yet some of them provided clues about how to approach the system and find the mechanism, which is expected to be figured out in the future.

Resumo

Os níveis de CO_2 atuais têm sido correlacionados com o aquecimento global, o que, por sua vez, tem sido apontado como uma causa possível de efeitos negativos em pessoas e outros seres vivos. As opiniões sobre este tema divergem, mas se os níveis de CO_2 podem, simultaneamente, causar problemas, mas ser diminuídos, uma procura de soluções para os reduzir é justificável.

Para atingir níveis mais seguros de CO_2, tentar reduzir as suas emissões é importante, mas devido ao grau de industrialização que existe, tal acarreta dificuldades. Um bom complemento desses esforços são os dispositivos que captam CO_2.

Com o intuito de obter novos dispositivos de captação de CO_2, os processos naturais de captura podem ser uma fonte de inspiração, nomeadamente os que ocorrem em plantas. Fundamentalmente, as enzimas são o componente fulcral da fixação de

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |4

carbono nas plantas. Neste trabalho, a enzima fosfoenolpiruvato carboxilase (PEPC), que participa na fixação primária de carbono em algumas espécies fotossintéticas, é analisada. O trabalho é essencialmente dedicado ao estudo do mecanismo da PEPC.

É descrito um modelo 3D do complexo enzima: substratos com um cofator metálico (Mg^{2+}). Nele, várias simulações foram feitas, para estudar a sua estabilidade em solução e o mecanismo de reação. Nomeadamente, foram feitos cálculos de mecânica quântica e molecular (QM/MM) para testar um mecanismo de reação proposto na literatura.

Tal mecanismo não foi verificado, pois a barreira energética obtida para o primeiro passo da reação foi demasiado alta para ser viável; ainda assim, a hipótese do mecanismo da literatura se verificar não está descartada. Foram experimentadas outras alternativas, sobre as quais se apresenta uma discussão. Não funcionaram, embora tenham providenciado alguma informação que poderá ser relevante.

Keywords

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |5

Index

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |6

List of Tables

List of Figures

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |8

List of abbreviations

PEPC: Phosphoenolpyruvate Carboxylase

CDR: Carbon Dioxide Removal

IPCC: Intergovernmental Panel on Climate Change

IEA: International Energy Agency

BECCS: Bioenergy with Carbon Capture and Storage

DAC: Direct Air Capture

PEP: phosphoenolpyruvate

OAA: oxaloacetate

DCO: 3,3–dichloro-2-phosphonometyl-acrylic acid

HIP: positively charged Histidine residue

LG: leaving group

QM/MM: quantum mechanics/molecular mechanics

MD: molecular dynamics

PES: potential energy surface

TS: transition state

ONIOM: n-layered integrated molecular orbital and molecular mechanics

AMBER: assisted model building with energy refinement

Pi – inorganic phosphate

RMSD: root mean square deviation

PDB: protein data bank

VMD: visual molecular dynamics

GAFF: general amber force field

RESP: restrained electrostatic potential

DFT: density functional theory

IRC: intrinsic reaction coordinate

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |9

A. Introduction

➢ Shall we capture CO_2?

There is some controversy about whether the current levels of CO_2 in the atmosphere are a matter of concern. The positions range from those who believe these levels do not carry major problems, to those who predict they have short and long term damaging consequences.

The sources of data are plenty and it may be confusing to organize them to extract a conclusion. In 1974, the United Nations created an organization – the Intergovernmental Panel on Climate Change (IPCC), which is currently devoted to the analysis of science related to climate change. They analyse articles covering its causes, effects, and plausible solutions, and they summarize their conclusions in reports called *Assessment Reports.* They try to collect information from different perspectives, and they also evaluate the rate of agreement on certain topics.

Currently, IPCC is writing its 6th Assessment Report – AR6, which will be released by 2022. The previous one – AR5 was written and published in 2013/14, and on it one may find some of the conclusions mentioned hereafter:

1. The concentration of greenhouse effect gases has risen, at least since 1970. In 2010, CO_2 emitted was larger than 50 Gigatons (Gt) and in 2011 they were the highest of the preceding 800 000 years. This comparison was made based on measurements of the levels of gases trapped in ice cores, which allowed the estimation of the levels of the greenhouse effect gases in the past thousand years. Yet in 2011, from the total CO_2 emitted, 9.5 Gt of CO_2 were from the burning of fossil fuels plus cement production. According to the report, more than 20% may remain in the atmosphere for the next thousand years.

2. Some problems may be correlated with high CO_2 levels: the frequency and intensity of droughts increased in Mediterranean West Africa. Yet, AR5 states that there is low confidence in assuming that droughts are caused by human activity.

3. The frequency of cyclones and the mean wave height in a large part of the Atlantic Ocean increased since 1970. It is also predicted that, in the Southern part of the ocean, the wind speed will rise and as so the wave heights. In the Arctic Ocean,

the smaller ice extent may lead to an increase in wave heights and longer wave seasons.

4. CO_2 in the atmosphere is in equilibrium with CO_2 in oceans. The more there is in the atmosphere, the more it will be absorbed by the ocean. By its reaction with water, CO_2 forms carbonic acid, which acidifies seawaters. Since 1750 until the release of the report AR5, the pH of seawater decreased by 0.1. In the Arctic Ocean, the acidification plus ice melting are being correlated with lower concentrations of calcium carbonate. Aragonite is a form of calcium carbonate found in plankton and some invertebrates, and the fact that carbonate fell to undersaturation levels suggests possible impacts on Arctic ecosystems.

   Still about the Arctic, since 1977, its sea ice decreased more than 9%, the summer retreat was the biggest ever seen, as well as the sea surface temperature. Sea ice also thickened from 1980 to 2008; it is predicted that in the upcoming years (before 2050) the Arctic Ocean may be ice-free seasonally (instead of being covered by ice all the year, as has happened so far).

5. Since the previous assessment report, in 2007, till AR5, most glaciers shrank. Ice was lost in the Canadian Arctic, Greenland, Asian Mountains, and Alaska. With nowadays temperatures, even with no further rise, they would keep on melting. The temperatures of permafrosts of many regions also rose: part of Alaska warmed up to 3 ˚C (since 1980) and, in Northern Russia, there were rises up to 2 ˚C.

6. Ice defrost also affects sea levels. In the last half of the 19th century, the sea level rise was larger than in the previous 2000 years. The upper part of the ocean – till 700 m depth – also warmed from 1971 to 2010 (with human activities being a probable cause).

For those opinions that temperature on Earth is cyclical and we will eventually enter a glacial age, there are also some considerations. Since Earth entered the interglacial age, around 120 000 years ago, temperatures were never 2 degrees higher than those registered at pre-industrial times. Now, the prediction is that such a threshold may be reached by 2050. As the next glacial age shall not happen in the next 1 000 years, even arguing that temperature on Earth cyclically oscillates between extreme warm and cold periods, it may get too warm far before the next glacial age.

IPCC also makes other predictions. It is forceful to mention that they, like any prediction, have some degree of uncertainty. The same is valid for evaluations of global scale properties based on measures on discrete points/locations. However, both

predictions and evaluations of impacts at a global scale are still useful, and particularly relevant if accompanied by its degree of confidence.

In line with this, it is worth to mention a Special Report covering Extreme Events and other topics of Climate Change, called SREX. Some of its conclusions follow:

SREX rates as *very likely* – 90 to 100% confidence – that, globally, the number of warm days rose, based on observations since 1950. It is *likely* – 66 to 100% that coastal high water occurrences rose due to higher sea levels. It is also *likely* - 66 to 100% confidence – that extreme daily maximum temperatures rose due to anthropogenic causes.

It is *virtually certain* – 99 to 100% confidence that the frequency and intensity of warm daily temperatures will increase during the 21st century (globally). It is *very likely* that warm waves will become longer and warmer.

It is also important to highlight that, if there is an increase of natural disasters (like droughts or flooding of coastal areas), the consequences may strongly vary according to the state of development of the affected country. Taking as reference data from 1970-2008, 95% of deaths caused by natural disasters took place in developing countries. So, what may be well tolerated in some areas, might be a problem in others. Lack of post-disaster recovery means may also reinforce the negative effects of natural disasters, which shall be taken into account when assessing the possible real impact of an increase in natural disasters due to climate change.

CO_2 levels may have negative effects which are highly probable and/or highly damaging. As so, part of it should be captured.

➢ Brief remarks about current levels of CO_2

As stated, IPCC makes regular reports about climate change. It is currently doing the Assessment Report 6, due for release in 2022. The previously published one – AR5, which was used as the source of most of the information provided above - was written in 2014. In the context of an analysis of decades, as the one done in AR5, a 6 years delay since the report till now is not much. Nevertheless, it is possible to fill this gap of some years with information from other sources.

International Energy Agency (IEA) is an organization that belongs to OECD (Organization for Economic Co-operation and Development) that provides some data related to CO_2 emissions. It attempts to promote energy security (as providing everyone with basic access to energy), economic development (as by means of efficient

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |12

technology), and environmental protection. It also attempts to encourage the use of energy sources other than fossil fuels (particularly renewable sources). IEA publishes annual reports about energy and CO_2, and the previous year's report - Global Energy & CO_2 Status Report 2019 - offers a more updated view of the situation.

According to it, in 2018, CO_2 emissions related to energy production were 33,1 Gigatons. The rise from the previous year was 70% larger than the average increase per year since 2010. The causes of this large amount of CO_2 were assigned to a *healthy* global economic state and to the weather conditions (which required energy for heating and cooling).

Interestingly, IEA mentions that between 2014-16, emissions did not grow, while economy did. They point out that it was due to an increase in both energy efficiency and the development of low carbon sources. In the following years, however, the economic growth surpassed them, and emissions restarted to rise.

The CO_2 levels in the atmosphere, in 2018, were 407.4 ppm (more 2.4 ppm than in 2017). As a comparison, in pre-industrial times, CO_2 levels were 180-280 ppm.

Unexpectedly, emissions may now face a break in the ascending trend. In a very recently released document – Global Energy Review 2020 - it is suggested that CO_2 emission will record a huge decrease due to the COVID-19 impact.

Closing this subsection, a citation is presented, regarding the impact that CO_2 levels may have. It is from a document of IPCC released in September 2019 – "Choices made now are critical for the future of our oceans and cryosphere" - and it states the following:

Global warming has already reached 1˚C above the pre-industrial level, due to past and current greenhouse gas emissions. (..) If global warming is stabilized at 1.5 ˚C above pre-industrial levels, the Arctic ocean would only be ice-free in September – the month with the least ice - once in every hundred years. For global warming of 2˚C, this would occur up to one year in three. (…) While sea level has risen globally by around 15 cm during the 20th century, it is currently rising more than twice as fast (..) It would reach 30-60 cm by 2100 even if global warming is limited to well below 2 ˚C, but around 60-110 cm if greenhouse gas emissions continue to increase strongly.

Beyond highlighting the impact that one degree may have on defrost, the document points out an additional issue: *Arctic and boreal permafrost hold large amounts of organic carbon, almost twice the carbon in the atmosphere, and have the potential to significantly increase the concentration of greenhouse gases in the atmosphere if they thaw.*

> ➢ What options do we have to cope with CO_2 levels in the atmosphere and their risks?

The majority of the CO_2 emissions come from energy production and industry, and many other human activities sum up to them. The industrialized countries also emit more CO_2 than the developing countries.

By the time AR5 was written, IPCC pointed out that the CO_2 levels required Carbon Dioxide Removal (CDR) strategies. However, it also stated that the already developed CDR devices were not practical or had problems to solve.

There are, according to AR5, alternative technology and procedures that would allow societies to evolve with less CO_2 emissions. Renewable and nuclear sources (which carry their own risks) are lowering the ratio [emission: energy demand], and more efficient technology lowers the energy demand. However, if they are not implemented to the required extent, large scale CO_2 capturing devices will be required to achieve lower levels of CO_2.

Nowadays, there is a variety of means for carbon removal. Growing and protecting forests allow plants to naturally remove CO_2. According to an IEA publication – *Going Carbon Negative: what are the technology options,* by Sara Budinis, from January 2020 – agriculture and forestry may remove between 1 billion to 11 billion tonnes of CO_2/year, by 2050 (considering that, in 2018, CO_2 emissions due to energy production were around 33 Gt, this may account for 3 - 33% of those emissions). This approach for CO_2 removal only requires land area.

Another natural option is biochar. Biochar is a result of the combustion of biomass in an environment with controlled levels of oxygen (in essence, it is like charcoal). Biomass is a reservoir of Carbon, and both combustion and decomposition make such carbon return to the atmosphere as CO_2, while charcoal retains it in its composition for longer times. Biochar use also has the advantage of generating fuel - syngas and bio-oil, alternatives to fossil fuels - in its production process. It may even be used as a soil quality enhancer, which promotes better growth of the crops and a consequent bigger capture of CO_2 by plants. This way, biochar has a very positive potential impact on lowering CO_2 emissions.

Biochar also has a few weaknesses: fuels from biochar emit CO_2; the transport of biomass to biochar production facilities, as well as from it to the disposal sites and the spread over the soils may require energy and, indirectly, CO_2 emissions related to such energy production. Burning of biochar may be regarded as a better alternative to fossil

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |14

fuels since the released CO_2 was the one previously captured (zero net emission). However, to correct high levels or to counterbalance other emitting sources, it would be better to use a negative emission strategy. Taking into account the amount of biomass that can be converted to biochar, without harming ecosystems or food security, and using modern, low emission technology for pyrolysis of biomass, biochar is predicted to be able to offset up to 12% of anthropogenic CO_2 [1].

Bioenergy with Carbon Capture and Storage (BECCS) is another promising solution for CO_2 removal. BECCS use biomass in processes that generate energy and release CO_2. The released CO_2 is captured and is then injected into geological formations.

Illinois Industrial Carbon Capture and Storage Project captures 1 million tons of CO_2 /year. The CO_2 is produced in a bioethanol facility and then stored it in a geological cavity. In Drax, UK, a BECCS pilot project planned to work in a power station (the biggest in the UK using renewable sources) may turn it into a negative emission power station (the first in the world). It is predicted that, by 2050, BECCS may remove up to 8 billion tons CO_2/year [2], which corresponds to around 24% of the CO_2 emitted in energy production (one of the largest emitting sectors). BECCs are quite appealing since CO_2 is captured in conditions where its concentration is high, which eases its trapping. Also, it is coupled with the generation of energy and products, which counterweights the costs. Disadvantages of BECCS are related to the availability of biomass, the cost of the transport from the capture site to the geological storage place, and also the availability of storage sites.

Finally, Direct Air Capture (DAC) devices are also being exploited. In Europe, the United States, and Canada, at least 15 DAC plants capture more than 9 000 tons of CO_2 /year [3] – this means 0.0003% of the CO_2 emitted for energy production. There are two strategies often employed for DAC. One is to force air to pass through a chemical solution that reacts with CO_2, such as alkaline or sorbent solutions. Some of these solutions are corrosive, which stands as a major drawback of this approach. Another strategy is by using sorbent materials (solids) that bind CO_2. After binding CO_2, the material may be heated (in a low pressure system) in order to release the bound CO_2 and to free the binding sites for a new cycle of capture. Solid sorbents are usually more energy demanding.

The released CO_2 may then be stored or used to produce other chemicals. When captured CO_2 is used for the production of synthetic fuels (which are then burned), part of it returns to the atmosphere. In this case, even if DAC ultimately does not reduce CO_2 levels, it may yet prevent their rise when compared to the use of fossil fuels for the same

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |15

purposes. The applications of captured CO_2 are indeed a key factor for the effectiveness of DAC in improving the state of CO_2 levels. An interesting option comes from Iceland: in a project lead by CarbFix, the captured CO_2 is injected in basalt rock. In a few years, CO_2 is incorporated in rocks by mineralization.

Since the concentration of CO_2 in the air is low, DAC devices require more energy than other means of CO_2 capture, as BECCS (where local $[CO_2]$ are higher). For a better negative net effect on CO_2 levels, the sources of the energy that supply DAC shall be clean, avoiding indirect emissions.

Yet, DAC devices offer some benefits. They are important to counterbalance the CO_2 from some sectors, in which is hard to avoid or reduce emissions, like the aviation sector. When compared to other CDR options, they require less area than forests and, if the trapped CO_2 is intended to be stored underground, DACs may be placed close to the storage site. This may be an advantage relative to BECCS, which may lie close to the facilities that provide their raw materials (otherwise the costs of transport of raw materials increase). Regarding the storage of CO_2 captured from DACs, as it happens with BECCS, both have the cost of compression and injection of CO_2 (in geological sites).

The cost of CO_2 removal by DAC is high [3]. Predictions are a bit uncertain, but some estimations for the future cost range from 100 to 1 000 USD/ton CO_2 captured. To date - 2020 - the current DAC devices are also of a small scale. The first large scale DAC is predicted to be operating in the US by 2023 and may capture up to 1 Megatonne CO_2/year. For comparison purposes, this is roughly 0.003% of the emissions to produce energy in 2018. It is being developed by a partnership between Carbon Engineering and Occidental Petroleum, and the captured CO_2 will be employed in enhanced oil recovery [3].

In conclusion, it is possible to assign advantages and drawbacks to the available CDR options. Their overall CO_2 removal efficiency also depends on details of their processes (as the source of energy chosen for DACs, or the use that is pretended for the captured CO_2). Current options need to be further developed. If not, other means of removing CO_2 shall be engineered to complement the already existing ones since, so far, the CO_2 is not at satisfactory levels.

➢ Phosphoenolpyruvate Carboxylase: a potential helper to reduce CO_2 levels

A great part of the plant organisms fixes CO_2 by an enzyme called ribulose -1, 5 – biphosphate carboxylase oxygenase (Rubisco). However, this enzyme can be regarded as a bit inefficient and error-prone, since it may fix O_2 instead of CO_2.

Some species of cactus of the family *Crassulacea*, as well as plants known as C_4-plants perform a special carbon fixation pathway, in which the enzyme Phosphoenolpyruvate Carboxylase - PEPC - primarily fixes HCO_3^- and stores it in organic compounds. These compounds are then transported to close to Rubisco and are decarboxylated there; the released CO_2 accumulates near the enzyme (Rubisco), that having a higher pressure of this gas, performs a more efficient catalysis.

This different pathway is quite often related to adaptation to dry climates, but it is a fact that PEPC is more efficient and selective than the more conventional Rubisco, and that plants having it, fix Carbon better. This way, PEPC can be regarded as a potential CO_2 scavenger.

In plants, PEPC adds HCO_3^- to an organic compound - phosphoenolpyruvate (PEP) - and produces oxaloacetate (OAA) and Pi; it also requires a metallic cofactor (like Mg^{2+}).

PEPC could be used to capture CO_2, as suggested in the literature [4]. The oxaloacetate that it produces could be used for the synthesis of other compounds or used itself. Notably, oxaloacetate is being studied for post-stroke depression treatment, and for other neuropsychiatric diseases [5].

Kinetic parameters also matter when considering practical applications. They vary for PEPC of different organisms, but for the enzyme in the leaves of *Crassula,* at pH=8 and in the presence of Mg^{2+}, v_{max} = 21.03 ± 0.01 umol·min^{-1}·mg^{-1} and K_M=0.35 ± 0.04 mM, and so v_{max}/K_M= 60.32 (± 0.05) min^{-1}·mg^{-1} [6]. The model described in this report used a structure of PEPC from *Zea mays* - for the enzyme of this organism, the k_{cat} is reported to be 9.0 x 10^3 min^{-1} [7].

Besides CO_2 capture, PEPC has other potential applications that would justify its study. PEPC is an enzyme that exists in plants, but also in bacteria, cyanobacteria, and some protozoa, like the one responsible for malaria. So, it can be engineered for producing more productive crops (including for bioenergy production), or used as a target for the design of herbicides (it does not affect animals, which do not own it) or antibiotics.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |17

Finally, from a fundamental view, PEPC is special in the way that, unlike other enzymes that perform carboxylation, only requires a divalent ion as a cofactor for the carboxylation (and not biotin).

➤ Phosphoenolpyruvate carboxylase: structure, mechanism, and some experimental findings

PEPC (of plants) is a giant enzyme composed of four chains of 110 kDa each. Two of the chains are more strongly attached, and in conditions of high dilution they detach from the other two ones; for this reason, PEPC may be regarded as a dimer-of-dimers.

Each monomer is composed mainly by α-helixes, but also displays a β-barrel composed by 8 β-strands, close to/on the which the active site lies. On it (active site), phosphoenolpyruvate (PEP) - a 3-C organic compound -, reacts with HCO_3^- to form oxaloacetate - a 4-C compound - and phosphate. A divalent cation, usually Mn^{2+} or Mg^{2+}, also binds to the active site [8]. The ion binding site structure, according to [8], is quite similar to the one of Pyruvate Kinase, despite the primary sequence being different for the two enzymes. A figure of the tetramer of the entry 5VYJ on PDB is shown in Fig. 1.
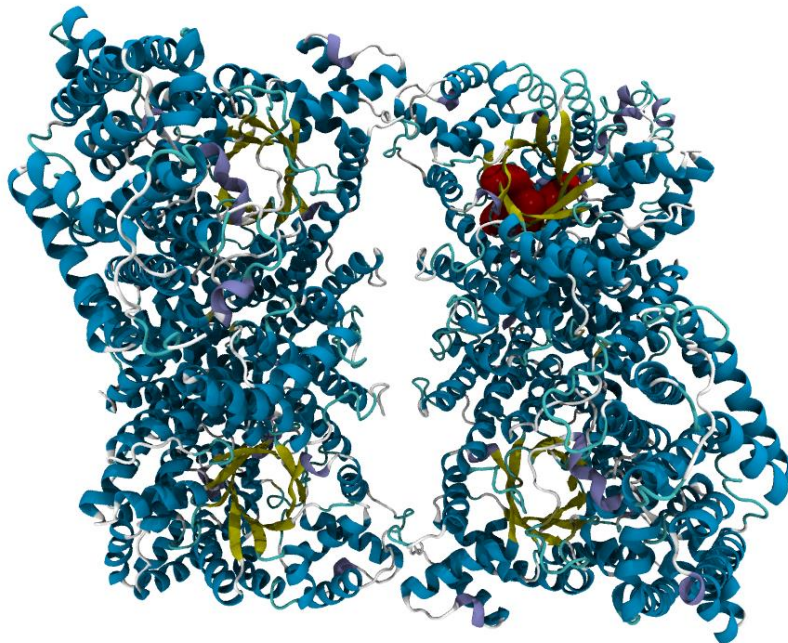
Fig. 1. **Structure of the PEPC tetramer**. β-barrels in each subunit are coloured yellow, and in one subunit, some residues of the active site are coloured red (surface representation of His177, Arg456, Arg647, Arg759, and Arg773).

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |18

The overall reaction is reported to be highly exergonic: $\Delta G^{o'}$=-30 kJ/mol (-7 kcal/mol) [9]. There is also a mechanism for it proposed in the literature [10] (see Fig. 2):
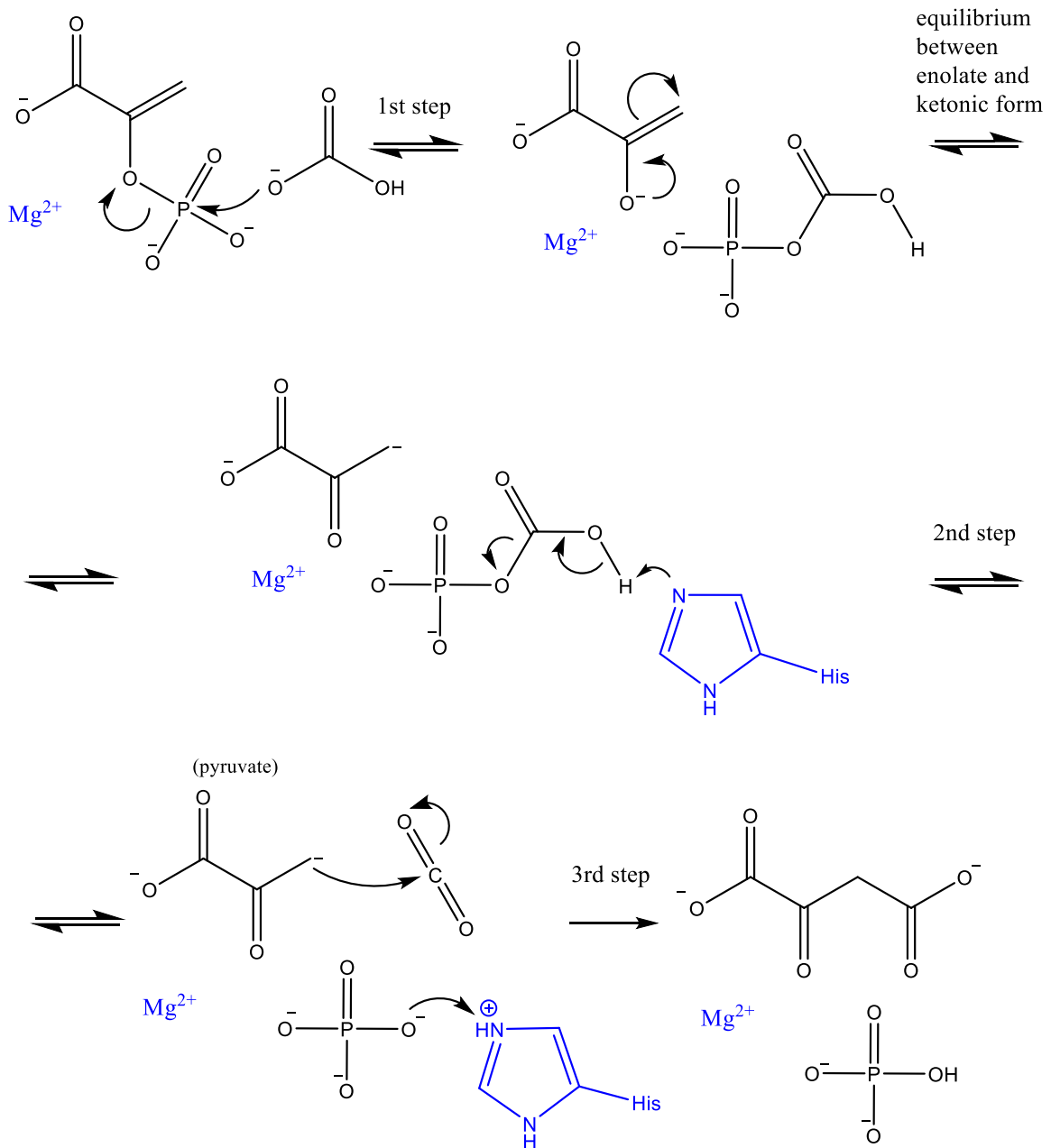
Fig. 2. **Mechanism proposed for the reaction catalysed by PEPC**. Adapted from reference *[10]*.

As depicted, the first and second steps are reversible, and the last one is quite irreversible.

Another structural characteristic of PEPC is that it displays allostery. The group of activating compounds that PEPC is sensitive to varies among organisms, but for *Zea mays,* for instance, Glycine, Serine, Alanine, and Glucose-6-phosphate are reported to activate the enzyme [11]. Ethylene glycol (EG) also seems to activate it (plants do not

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |19

usually have it, but they have other polyols that may be upregulated in stress response) [12]; truncation of the N-terminal may also have an activating effect [10]. PEPC is inhibited by Aspartate and malate.

There is a bit of controversy regarding where the allosteric binding site of the activators is. Most studies locates the binding site at the interface of the monomers within each dimer [10] [13]. This conclusion is based in the observation that: i) some molecules/activators were found at that site on the crystallographic structures; ii) mutagenesis studies on the which Arg residues in that site were mutated to Gln (R183Q, R184Q, R231Q, R372Q - numbering *Zea mays*) caused a marked decrease in the responsiveness to activators [13].

However, in an interesting article from Schlieper *et al* [12], this idea is questioned. New crystallographic structures show activators at active sites non-occupied by the substrates. Kinetic studies also suggested that activators act by competitive activation, that is, binding at the active site of one monomer causes the activation of the active site of the other monomer. Authors suggest that previously reported binding sites, out of the active site, might be promiscuous binding sites; moreover, the mutagenesis studies where Arg residues were mutated to Gln could owe its effect to the location of these residues at the monomer interface. They could simply affect the activation of the enzyme by the disturbance of structure they cause. Allostery, in the case of a competitive activation model, would require changes happening in one monomer to be transmitted to the other, and a profound change in the interface might impair it.

The conversion from the T state to the R state (*i.e.* the structural change from a *tense* conformation - with lower affinity for the substrate - into a *relaxed* conformation - with higher affinity to it), promoted by the activators, implies the movement of two loops [8]. One of them brings Arg647 closer to the substrates binding site; the other approaches His177 to close to it. The difference in the position of Arg647 in the T and R states is as large as 20 Å.

Besides the structural data, there is also some experimental information regarding the reaction mechanism of PEPC. Some experimental findings are listed below.

In 1965, in an experiment employing [^{18}O]-HCO$_3^-$, it was found that labelled oxygen was incorporated in the products oxaloacetate (OAA) and inorganic phosphate (Pi), and that OAA presented 2 times more labelled oxygen than Pi. This clarified that HCO$_3^-$, and not CO$_2$, was the substrate, and that one of its O was incorporated in phosphate, while the remaining two O atoms would be incorporated in OAA [14]. The mechanism proposal of *Fig. 2* fits such findings: HCO$_3^-$ attacks the phosphate group (see 1st step), and then the formed intermediate breaks, letting one of its oxygens in the released phosphate;

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |20

then, the CO$_2$ derived from HCO$_3^-$ would attack PEP and its oxygen atoms would be incorporated in the product OAA (see 3rd step).

There are two conserved histidine residues in PEPC. When one of them, His138 in *E. coli* (corresponds to His177 in *Z. mays),* is mutated to Asn, oxaloacetate does not form, and instead, pyruvate is formed. That is, the mutated enzyme catalyzes the dephosphorylation of PEP, but not its further carboxylation [15]. The current mechanism proposal (*Fig. 2*) includes a histidine residue (with an active role) in a step after pyruvate has formed – this possibility could explain the experimental findings.

When, instead of the normal substrate PEP, phosphoenol-alpha-ketobutyrate (P-alpha-ketobutyrate) is supplied to PEPC, in the presence of HCO$_3^-$ and Mg^{2+} (required), inorganic phosphate and alpha-ketobutyrate are formed (that is, dephosphorylation occurs, but no carboxylation). Experiments with PEPC with either PEP or P-alpha-ketobutyrate were also performed with [^{18}O]-HCO$_3^-$: When PEP was used, phosphate released had only one atom of ^{18}O, while for P-alpha-ketobutyrate, the released phosphate presented one or multiple atoms labelled [16]. These results provide conclusions similar to the ones reported in the previous two paragraphs. The study with ^{18}O suggests that HCO$_3^-$ (as hypothesized in the mechanism of *Fig. 2*), or any specie derived from it, attacks the P of the reactant, leading to a labelled phosphate. Similar to what happened with the mutated enzyme H177N, when P-alpha-ketobutyrate is used, only the dephosphorylation of the analog occurs, suggesting that its extra methyl group (compared with the true substrate PEP) impairs the attack by the remaining part of HCO$_3^-$ (step of carboxylation). It may further suggest that dephosphorylation and carboxylation are independent steps, as is proposed in the mechanism of *Fig. 2*.

Also from experiments performed with PEP analogs, it was reported that when phosphoenol-3-fluoropyruvate (the Z isomer) is used instead of PEP, a small amount of carboxylation occurred, but the majority of the substrate was simply dephosphorylated [17]. While the previously mentioned analog - P-alpha-ketobutyrate – presented an extra methyl group compared to PEP, this analog - phosphoenol-3-fluoropyruvate - presents a fluor. Considering the last reaction step of the mechanism proposal (in *Fig. 2*), where pyruvate is attacked by CO$_2$, it is possible that the analogs' substituents sterically hindered the access of the reacting atoms. As fluor is smaller than the methyl group, it could *hinder less*, allowing the small extent of carboxylation found (while for the methyl substitution no carboxylation was detected). Another reason that may contribute to impair the carboxylation in the analogs is the mesomeric effect of F and CH$_3$ – they increase the nucleophilicity of the C on the other side of the double bond (instead of the C where the attack should happen).

There is also one study [18] covering the use of a bicarbonate analog: formate. When used instead of the usual substrate, formyl phosphate and pyruvate were formed, despite that V_{max} was 1% of the rate found when employing HCO_3^-. If [^{18}O]-formate was used, Pi with one labelled O was formed. The type of molecules found is consistent with the mechanism proposal: the analog could perform the equivalent to the 1st step (the attack to PEP, leading to its dephosphorylation, see Fig. 2). However, if the reaction with the analog is indeed similar to the 1st step of the proposal mechanism (for the true substrate), the fact that its V_{max} is quite smaller is a bit intriguing.

Studies measuring Carbon isotope effects, and considering HCO_3^- the starting state showed that at pH=7.5, 25 ˚C, with Mg^{2+}, k_{12}/k_{13} was 1.0029 ± 0.0005 and it lowered with increased pH, achieving a value under 1 at pH=10. On the other hand, when considering CO_2 the starting state, the trend is the opposite. The authors of these experiments hypothesized that it could indicate that the mechanism of the reaction of PEPC is stepwise, with the rate-limiting step being the formation of carboxyphosphate (as intermediate) [19]. Despite this, the carboxyphosphate was not detected experimentally [15].

The same article [19] reports a pKa around 7 for the phosphate group of PEP. This way, for a pH above 7, this group should be mainly in the unprotonated form. According to the study, the activity of the enzyme was also dependent on the protonation of a group with a pKa of 10.

In a work later published by the same group [20], it was suggested that carboxylation occurred by an attack of CO_2 rather than by carboxyphosphate. They propose a mechanism in which carboxyphosphate is formed and then decarboxylated, and they also suggest that these steps are reversible. This assumption is based on experiments using fluoro-PEP and labelled bicarbonate, in which part of the remaining fluoro-PEP ended labelled. The mechanism proposal of *Fig. 2* satisfy this experimental evidence – on it, a Histidine residue induces a break of the carboxyphosphate into CO_2 (and Pi).

NMR studies with [^{16}O, ^{17}O, ^{18}O]-thiophosphoenolpyruvate suggested that phosphorus (at least the most part) undergoes inversion of configuration in the reaction [21], that is, the configuration of the phosphate group of PEP is the opposite of the one in the product phosphate. This supports an attack of P by an O in an SN_2 mechanism.

Also related to stereochemistry, it was proposed that the carboxylation of pyruvate (the intermediate after PEP dephosphorylation) is stereospecific, with the new carbon attacking at the *si* side of the double bond [22].

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |22

Concerning the activation energy, in a work reported in the literature [23], there is a plot of log (V_{max}) as a function of $1/T$, for both Maize PEPC and another plant. The authors attempted to extract the E_a from the slope of this Arrhenius plot, but report a change of slope on it, for the maize PEPC: lower than 25°C, the slope gives an E_a= 17 kcal/mol, and above it, E_a=2.6 kcal/mol. The latter value, however, is excessively low. The analysis of the results of the article suggests that the fit of the experimental points is a bit uncertain.

In a more recent paper [24], the E_a was evaluated as 68.1 kJ/mol, that is, 16.3 kcal/mol.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |23

B. Methods

➢ 1. Building the model

Structures of PEPC available in the Protein Data Bank (PDB) were compared. The structure in the entry 5VYJ [11], which comprises a PEPC belonging to the *Zea mays* plant, was selected. Besides the protein, this structure contains three glycine molecules (allosteric activators) and acetate ions co-crystallized with the protein. Its resolution is 3.3 Å. The PEPC structure comprises four chains, each of them having residues numbered 1-970, with some missing fragments: 1-34, 126-140, 933-935.

5VYJ displayed larger resolution values than other structures of PEPC in the PDB; yet, it was chosen, because it presented some residues absent in other structures, namely Lys762, Arg763, and Arg764, which according to [9], could be important for HCO$_3^-$ binding. It was also a protein belonging to a C$_4$ photosynthetic organism and, in opposition to some other structures available, it was expected to be in the R state. Moreover, there is some controversy about where the binding sites of the allosteric activators are, with two possible sites being pointed out in the literature [12] [13]. In the 5VYJ, the protein exhibits glycine molecules (activators) bound to both hypothesized sites, this way covering the two possibilities and increasing the confidence about it being in the R state. Finally, by the time of the choice of the starting structure, there was a recently released structure – PDB ID 6MGI - but the supporting article was not available. Nevertheless, an RMSD analysis of the active sites of 6MGI and 5VYJ showed they were similar (RMSD for Trp288, Arg456, Leu564, Glu566, Met598, Asp603, Arg759, Arg773, in the chain A, is 0.6 Å; the biggest differences - yet not large - are found for Arg759 and Arg773; the RMSD analysis included the sidechains of the residues).

For a summarized comparison of the structures available, see Table 1.

Table 1. **Available structures of PEPC in the PDB.** Information regarding their resolution, organism, and some comments about each PDB entry that influenced the choice of one of them for starting structure.

PDB entry	Resolution	Organism	Reason for <u>not</u> selecting the structure
1JQO	3.00 Å	*Zea mays*	This structure missed the residues K762, R763 e R764 that, according to [9], could be important for HCO$_3^-$ binding.
5VYJ	3.30 Å	*Zea mays*	Selected structure
1FIY	2.80 Å	*E. coli*	Non photosynthetic organism

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |24

1QB4	2.60 Å	*E.coli*	Non photosynthetic organism
1JQN	2.35 Å	*E.coli*	Non photosynthetic organism
3ZGB	2.71 Å	*Flaveria pringlei*	T state
3ZGE	2.49 Å	*Flaveria trinervia*	T state
4BXC	2.86 Å	*Flaveria trinervia*	Many residues missing
4BXH	2.24 Å	*Flaveria trinervia*	Many residues missing
6MGI	2.99 Å	*Zea mays*	Recently released structure; its article was not published when this work started
3ODM	2.95 Å	*C. perfringens*	Archaea-type protein

The structure of the entry 1JQN, a PEPC from *E. coli*, was also downloaded from the PDB. This structure contained an Mn^{2+} ion, as well as a substrate (PEP) analog – 3,3–dichloro-2-phosphonometyl-acrylic acid (DCO).

Some residues close to the active site in the entry 1JQN were used to align it with the chain A of 5VYJ - Trp288 [248], Arg456 [396], Leu564 [504], Glu566 [506], Met598 [538], Asp603 [543], Arg759 [699], Arg773 [713], in the numbering of the *Zea mays* PDB file [*E. coli* numbering is between brackets]. After aligning using *Visual Molecular Dynamics* (VMD) [25], coordinates of DCO, Mn^{2+}, and one water coordinated to it, were appended to the 5VYJ PDB file.

Only the chains A and B of the 5VYJ PDB file were kept, as well as the allosteric activators (glycine molecules) bound to them. In regions where there were missing residues, TER labels were inserted. Mn^{2+} was exchanged into Mg^{2+}, *i.e.,* coordinates were kept, but the element label was exchanged. The two atoms of Cl of the substrate analog DCO were exchanged by H, and one of its C atoms was exchanged into O, originating the substrate PEP, as represented in Fig. 3.
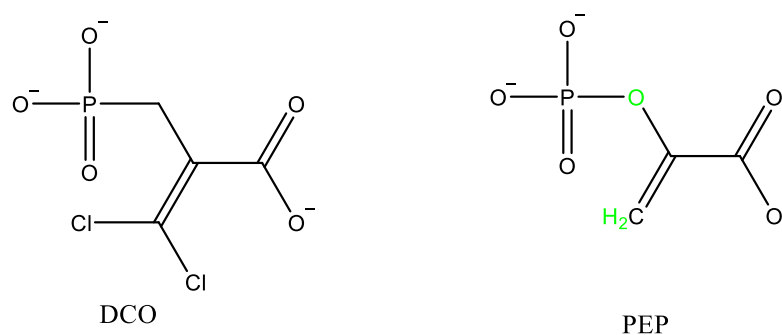

DCO   PEP

Fig. 3. **PEP analog (DCO) and PEP.** Green: atoms differing between both molecules.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |25

After that, parameters were calculated for the molecules PEP and HCO_3^-. Their structures were first drawn with the program *GaussView* 5.0.8. Using the *Antechamber* tool [26] and performing an optimization with Gaussian09 [27] (functional HF and basis set 6-31G*), the atomic point charges were calculated using a restrained electrostatic potential (RESP) scheme [28]. Force field parameters were assigned from General Amber Force Field – GAFF [29].

In the case of HCO_3^-, an additional change was made in the parameters of the force field file generated by *Antechamber*: the non-bonded parameters of the hydrogen atom were changed from 0 into the non-zero parameters of the GAFF2 library. This allowed to circumvent the problem found in the minimizations of the system, in which the proton of HCO_3^- tended to collide against one of the partially negative oxygens in the molecule.

Parameters for the Mg^{2+} ion were taken from the literature [30].

The propKa 3.1 software [31] [32] was then used to predict the protonation states of the residues for a pH = 7.3 (the condition employed in an experimental procedure reported in the article [33], in which kinetic parameters of PEPC were measured). When the predicted protonation states were different than usual, the environment surrounding the residue was inspected using VMD. Coming back to the system before it was protonated, the following residues were changed, accordingly to the propKa output:

- Glu675 of chain A and of chain B were changed into GLH (neutral form). They were close to an aromatic ring. Additionally, if the carboxylate group of the Glu sidechain was protonated, the H could establish a polar H - π interaction with the aromatic ring (one O of the Glu sidechain is at 3.89 Å of the closest C of the ring).

- Lys189 of chain A and of chain B were changed into LYN (neutral form). They were near Arg sidechains.

- Lys960 of chain B was changed into LYN. It was buried in the inner of the protein and there was no negative charge close for stabilizing it.

- Asp208 of chain A and chain B were changed into ASH (neutral form). They were close to a glutamate sidechain.

- Asp380 of chains A and B were changed into ASH. They lied close to tyrosine and tryptophan residues (aromatic rings). One O of the Asp sidechain was at 3.43 Å of the Tyr ring, and if it was protonated, an H – π interaction could be formed.

- His177 was changed into HID. According to the mechanism proposed in the previously mentioned article [10], the protonation state of this histidine at position δ was important.

The system and the parameters of the ligands and Mg^{2+} ion were loaded into the XLEaP module of the AMBER18 package. The AMBER force field ff14SB [34] was used to parametrize the protein. The program protonated the system, neutralized it and inserted it in a TIP3P [35] water box up to 12.0 Å of the protein.

In the reaction catalyzed by the enzyme, PEP reacts with HCO_3^-. This molecule was absent from the PDB structures, so it was inserted using *GaussView,* in the previously described model (scaffold of protein of *Zea mays* with the transferred metal ion and the pyruvate modelled from the inhibitor). Four different models were built, with HCO_3^- placed at different sites. In one of them, it was placed close to the Mg^{2+}. In other, it was inserted in one side of the His177. In another, it was placed on the other side of His177, close to the site of PEP where the carboxylation should occur (its double bond); a slight adjustment of this latter position to avoid a clash with the protein resulted in the forth site.

The systems were then subject to energy minimizations. First, the waters were minimized, while the remaining atoms were restrained (their position) by a harmonic potential with a force constant of 50 kcal.mol^{-1}. Å^{-2}. Following, hydrogen atoms were minimized, keeping the remaining atoms restrained by a potential of the same type as referred before. Then, sidechains of the residues were minimized, except that of His177 (an attempt that it interacted with the substrate molecule as proposed in the literature). Finally, the backbone of the protein was minimized (except for His177). An additional minimization of the entire system (totally free) lead to structure where HCO_3^- and PEP were a bit far from each other (in all the four systems with HCO_3^- at different positions). So, one of the systems – system with HCO_3^- close to the site of PEP where carboxylation should occur - minimized restraining the substrates and His177 position was used to start a Molecular Dynamics (MD) simulation. It was chosen since it presented an acceptable distance between the atom P (of PEP) and O (of HCO_3^-), which were supposed to bind in the mechanism proposed in the literature/article.

The first 2 minimizations referred (waters and hydrogens) took 10 000 cycles, and till the 5 000th one a steepest descent algorithm was employed; after it, the algorithm was switched to conjugate gradient. The latter two minimizations (sidechains and backbone) took 30 000 cycles and employed a conjugate gradient algorithm since the 7 500th step. All the calculations assumed a periodic system with constant volume and neglected nonbonding interactions more than 10 Å apart.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |27

➤ 2. Molecular Dynamics

The chosen system underwent molecular dynamics (MD) using Amber18. The MD protocol comprised 4 phases: Heating, Equilibration, Restricted Phase, and Free Production Phase.

In the Heating Phase, the system was heated up to 310.0 K (slight above the temperature of a plant). Its length was 50 ps and the simulation trajectory was saved at steps of 1 ps. The simulation of the heating phase was performed using a Langevin Thermostat [36] and assuming a collision of particles (between the system and a bath) of 2 collisions/ps. During this phase, His177, PEP, HCO$_3^-$, Mg^{2+}, and its remaining coordination shell (one water, an aspartate and a glutamate residues) were restrained by a harmonic potential with a force constant of 50.0 kcal·mol^{-1}·Å^{-2}. In the subsequent simulation stages, T was kept at 310 K.

During the equilibration stage, the system evolved along 500 ps, and at each 1 ps the coordinates were saved for further analysis. The restrained atoms were the same as during the heating phase, and the restraining constant was also the same. Periodic boundary conditions were always imposed and taken into consideration for calculations of non-bonded interactions. During the heating phase, volume was kept constant and, after it, the system was switched to a constant pressure scheme.

Then, an additional simulation of 5 ns in the same conditions (*i. e.* same atoms restricted) was run – restricted phase. Following, the system was set completely free and it was simulated for 5 ns - Production Phase. During these 10 ns, at every 4 ps, *frames* (geometry at the given time points) were saved.

Furthermore, an additional MD of 20 ns was performed (free system).

In all the simulations performed, the SHAKE algorithm was used to restrain the stretching of bonds involving hydrogens. The integration time was 0.002 ps. Non-bonded (electrostatic) interactions more than 10 Å apart were not accounted for.

➤ 3. Analysis of the MD simulation

The frames generated in MD were re-imaged using the *cpptraj* v18.01 tool [37]: the centre of a box was placed in the centre of mass of the atoms of protein + Mg^{2+} + PEP + bound activators (three molecules of glycine) and the molecules outside of this box were inserted back into it.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |28

Frames of the first 5 ns of the MD of the Production phase were then grouped in 5 clusters using *cpptraj*. The frames were grouped according to the lowest RMSD of [HCO$_3^-$, PEP, Mg^{2+}, His177 and the glutamate and aspartate residues on the coordination shell of the Mg^{2+}] between them. That is, initially all the frames were separated, and then frames having the smallest RMSD (considering the atoms referred above) were successively grouped (hierarchically agglomerated) until 5 groups have been formed.

Within the most populated cluster (*i.e.,* the group inclosing the biggest number of frames), the most representative frame (*i.e.,* lowest sum of RMSD to other frames in the group) was used for further steps.

➢ 4. QM/MM model: large system

Within the frame mentioned above, a sub-selection was performed: chain A of the protein (the one including the active site with the substrates), PEP, HCO$_3^-$, Mg^{2+}, and the residues of chain B up to 18 Å of the active site; waters up to 3 Å of these atoms were also included. The active centre considered (from which the 18 Å were measured) included His177, Arg456, Glu566, Ser602, Asp603, Arg647, Gln673, Arg759, and Arg773 (numbering from the original 5VYJ PDB file), as well as 3 waters close to the ion or substrate molecules, and PEP, HCO$_3^-$ and Mg^{2+}.

Since the chain B was truncated, TER labels were inserted manually in sites where the chain was interrupted. There were also some cases in which a single residue was within two TER labels, so such residues were exchanged to the zwitterionic state. To accomplish it, a library of zwitterionic amino acids – Zaa.off [38] – was downloaded and the residues names in the pdb file (the zwitterionic ones) were changed to match the name given in the library.

This smaller system was loaded into *XLEaP* (as well as the parameters files for the substrates, the ion, and the zwitterionic library) to generate parameter and topology files, that were then used to create a file suitable for *Gaussian* input. On it, residues more than 15 Å apart from the active site were frozen (*i.e.,* its coordinates were fixed). All the waters were also frozen, except 3 waters close to the ion and substrate molecules. *VMD* with the extension *MolUP* [39] was used in this step (of freezing).

The Quantum Mechanics (QM) layer was defined as: Mg^{2+}, PEP and HCO$_3^-$, and the residues directly stabilizing them - Asp603, Glu566 (coordinating ion), Arg456, Arg647, Arg759, Arg773, Gln673, His177 and Ser602, as well as three water molecules (one in

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |29

the Mg^{2+} coordination shell, and the other two ones interacting with PEP and HCO$_3^-$, respectively). In some cases, only part of the residues was included – those interacting with substrates or those atoms required for not placing the frontier of QM layer at resonance/peptide bonds. The final QM layer comprised 115 atoms - Fig. 4. The whole system had 21 879 atoms.
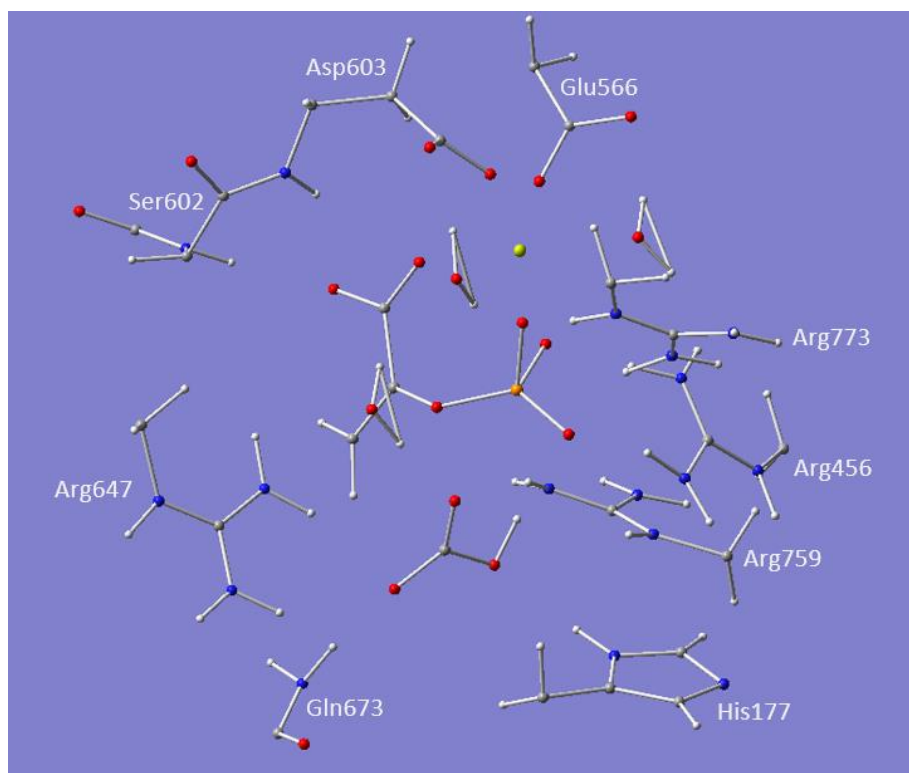
Fig. 4. **The QM layer of the QM/MM model.** Mg^{2+} is depicted as a green sphere and the remaining atoms are represented by ball-and-stick.

In the subsequent calculations, a subtractive method, ONIOM - n-layered integrated molecular orbital and molecular mechanics [40] - was employed, so there was the need of defining the whole system's charge and multiplicity and QM layer's charge and multiplicity. The whole system's charge was set as -17 and its multiplicity as 1; the QM layer's charge was set as 0 and its multiplicity as 1.

Some parameters for link atoms – H added to the QM layer to complete the valence at frontier atoms – were appended to the gaussian input file. Their source was the GAFF library.

The geometry of the system was optimized using Gaussian09.d [27] [41] [42]. The QM calculations used the theoretical model B3LYP [43] [44] [45] [46] (DFT hybrid functional) and the basis set 6-31g(d); the MM calculations employed the amber force field [47]. First, optimization was performed using a Mechanical Embedding scheme.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |30

Then, the resulting system underwent optimization with Electrostatic Embedding. In this calculation, a correction term was included to account for the dispersion of electrons – D3 version of Grimme's dispersion with Becke-Johnson damping (GD3BJ) [48].

Starting from the optimized system, a linear transit scan was performed: the proposed nucleophilic attack of one of the Oxygen of HCO$_3^-$ (the closest to P) to the Phosphorus of PEP.

Another attack was performed: the approach of the H$^+$ of HCO$_3^-$ to one of the oxygens of the phosphate group of PEP.

➢ 5. QM/MM model: small system

The size of the large QM/MM system – 21 879 atoms - was hampering the progression of simulations.

Hence, a smaller system was built by considering s sub-selection of the previous QM/MM model. It comprised the residues up to 22 Å of the active centre (took as the substrate molecules, the Mg^{2+} ion, and Arg456, Glu566, Ser602, Asp603, Arg647, Gln673, Arg759, Arg773, as well as three water molecules).

The previously referred truncation created several breaks in the chain of the protein. TER labels were inserted at each discontinuity.

Hydrogens of the amine groups after each TER label were removed. Where a single residue lied between TER labels, its name was changed to match the zwitterionic library.

The file was then loaded in *XLEaP*, as well as the required parameters. The program created the coordinates and parameters files for the system, which were then used to generate a *Gaussian* input file. The new QM/MM system comprised 11 779 atoms.

Residues more than 18 Å apart of the active site were frozen, as well as all the water molecules, except three of them close to the substrates. The atoms moved into QM layer were the same as the ones of the 1st QM/MM model described (see Fig. 4), and as so the additional parameters required for link atoms. The charge of the whole system was set to -2 and its multiplicity to 1; the QM charge was 0 and the multiplicity was 1.

Geometry optimizations were run as previously described. Linear transit scans similar to the ones ran in the larger (1st) QM/MM model were performed: 1) nucleophilic attack of the O of HCO$_3^-$ to the P of PEP; 2) the attack of H of HCO$_3^-$ to one O of the phosphate group of PEP. The results of the optimization and the scans were compared with those obtained with the larger system to validate the smaller system.

➢ 6. Addition and removal of water molecules

The linear transit scans of the attack of the O of HCO_3^- to the P of PEP presented a high energy barrier. It was hypothesized whether the addition of water molecules close to the oxygen of the leaving group could stabilize it and low the energy.

Two waters were modelled using *GaussView* 5.0.8. They were placed close to the intended atom, and also close to other molecules that could help to stabilize them - one modelled water was placed with its O close to a water of the system, and close to the H of HCO_3^-; the other water was modelled with its O atom near an Arg sidechain - Fig. 5.
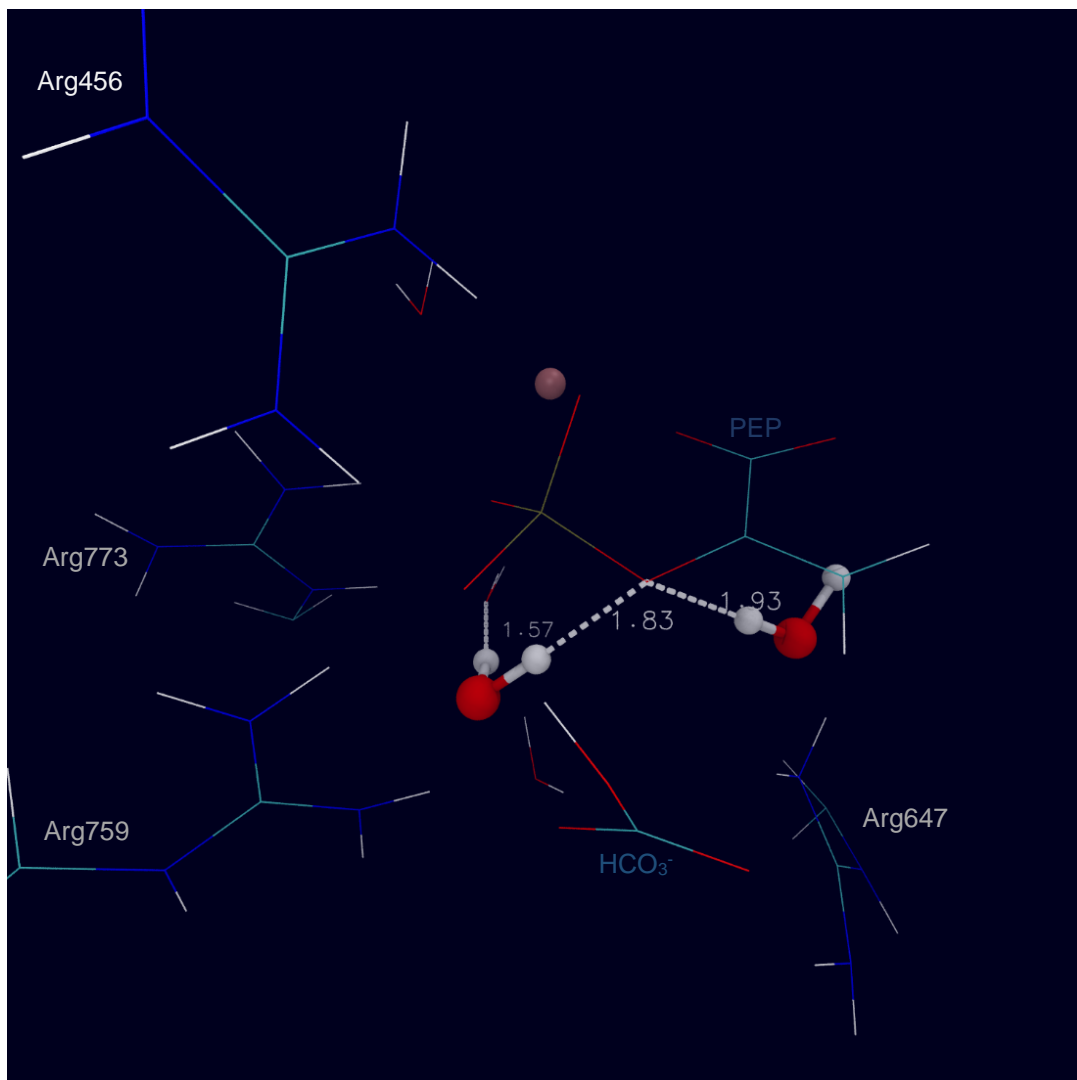
Fig. 5. **Two additional water molecules that were modelled in the active site.** The modelled water molecules are shown in ball-and-stick, and the remaining atoms as lines, including water molecules that were already in the system. Some distances are highlighted, namely the distance of the H of the modelled water molecules to the oxygen of PEP that, according to the proposed mechanism, should detach from P after its attack by HCO_3^-.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |32

The water molecules were modelled in the previous optimized system. After the addition of the water molecules, the structure was optimized with Mechanical Embedding. During the optimization, one of the modelled water molecules moved too far from the oxygen atom that it was intended to stabilize. Thus, it was deleted, and the system (with only one modelled water molecule) was optimized using Mechanical and then Electrostatic Embedding. The result was then submitted to a linear transit scan approaching the O of HCO_3^- to the P of PEP.

It was also noticed that one of the water molecules in the model (<u>not </u>considering the modelled water, but rather the water present since the MD simulations) was establishing two hydrogen bonds: one with Asp603 and another with PEP (Fig. 6). As this could be hampering PEP to adopt a conformation with lower energy, during the attack O (HCO_3^-) --- P (PEP), a new model was built, without this specific water. The model was optimized, first with a Mechanical Embedding scheme and then using Electrostatic Embedding. Then, a linear transit scan of the attack O (HCO_3^-) to P (PEP) was performed.
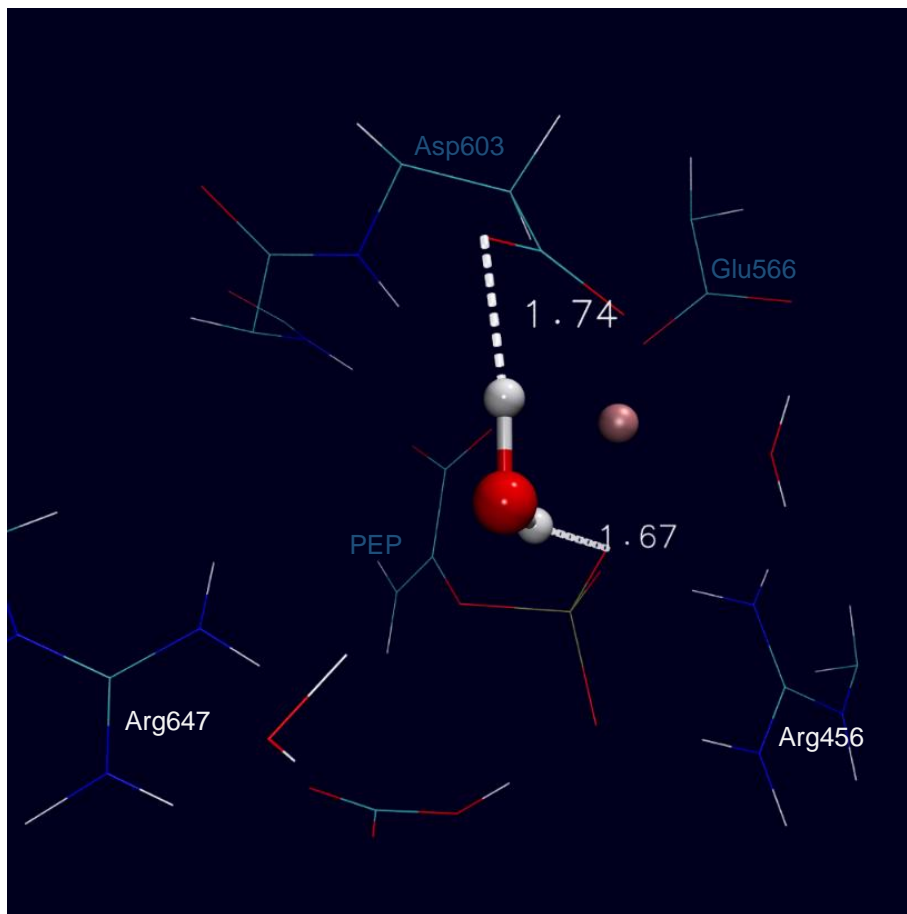

Fig. 6. **H-bond interactions between one water molecule in the QM layer and Asp603 and PEP.** The water molecule is represented by ball and stick and the remaining atoms by lines. All the atoms were present in the MD simulation (they were not modelled). Distances related to the H-bonds are also shown.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |33

> ➤ 7. New model with HCO_3^- placed at 180° of the Leaving Group

In an attempt to obtain an attack of HCO_3^- to the P of PEP with a lower energy transition state, a new model was built from the crystallographic structure. It was similar to the previously used (see Section 1. Building the Model), but the HCO_3^- was modelled with one of its O at 180° of the O of the leaving group of PEP (*i.e.*, angle O (HCO_3^-) – P – O (leaving group = 180°).

Bicarbonate was positioned among Arg456, Arg759, and Arg773 and with its hydroxyl group close to the backbone of an Asp residue, to allow the establishment of an H bond. The distance from P to the closest O of bicarbonate was 3.0 Å. After protonation, the system was as shown in Fig. 7.
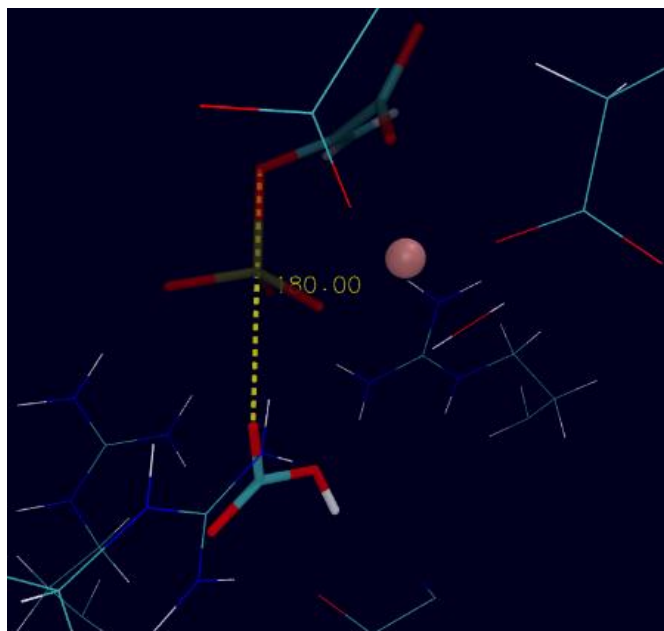
Fig. 7. **The system with HCO_3^- modelled at 180°of de Leaving Group.** The angle O (HCO_3^-) – P – O (LG) is coloured in yellow. Mg²° is depicted as a rose ball, protein residues are shown in lines and substrates in a thicker representation.

The system was neutralized and inserted in a water box up to 12.0 from the protein and underwent minimizations of H_2O, Hydrogens, and sidechains of the protein. The resulting structure was used as the starting point for MD simulations - 50 ps with heating till 298.15 K (constant volume), and 500 ps at such *T* for the equilibration of the system (constant pressure = 1 atm). During these simulation stages, the movement of the backbone of the protein, Mg²⁺ and its coordination shell, as well as HCO_3^- and PEP were restrained (by a harmonic potential with a constant force of 50.0 kcal.mol⁻¹.Å⁻²). The backbone was then free: 500 ps of simulation restraining only substrate molecules, the

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |34

ion, and its coordination shell. Finally, there was a 5 ns simulation with the system set free (Production Phase).

Most of the parameters were the same as described in section *2. Molecular Dynamics.*

During the Production Phase, at time points of 4 ps, the coordinates of the system were saved. One of the two frames presenting the shortest distance between the O of the HCO_3^- and the P of PEP was chosen: 128th frame (512 ps after the starting of Production Phase). The frame was re-imaged using *cpptraj*.

A sub-selection comprising residues up to 22 Å of Glu566, Asp603, Arg773, Arg456, Arg759, Arg647, Ser602, Gln673, PEP, HCO_3^-, and Mg^{2+} was saved. From it, a sub-selection discarding waters more than 3 Å apart from protein was saved. At sites where the protein chain was broken by the truncation, TER labels were inserted using the procedure described above (See section *5*). Where needed, residue names were exchanged to zwitterionic state (as described in Section *4*).

A *Gaussian* input file was generated as described previously; *VMD* with *MolUP* was used to freeze atoms more than 18 Å apart from the active site, considered as Glu566, Asp603, Arg773, Arg456, Arg759, Arg647, His177, Ser602, Gln673, PEP, HCO_3^-, and Mg^{2+}. All the waters were frozen, with the exception of three of them lying close to the ion or substrate molecules.

Some atoms of the residues mentioned above - Glu566, Asp603, Arg773, Arg456, Arg759, His 177, Ser602, Gln673, PEP, HCO_3^-, and Mg^{2+} - as well as of two additional residues close to the carboxylate group of PEP were assigned to the High Layer. Unlike the model described in section 4., this model did not include Arg647, since in this case it was not close to the substrates. The remaining atoms were kept in Low Layer. The High Layer comprised 104 atoms and the whole system was composed of 11 690 atoms. The distance from the O (HCO_3^-) to P (PEP) was larger than 4 Å, as depicted in Fig. 8.
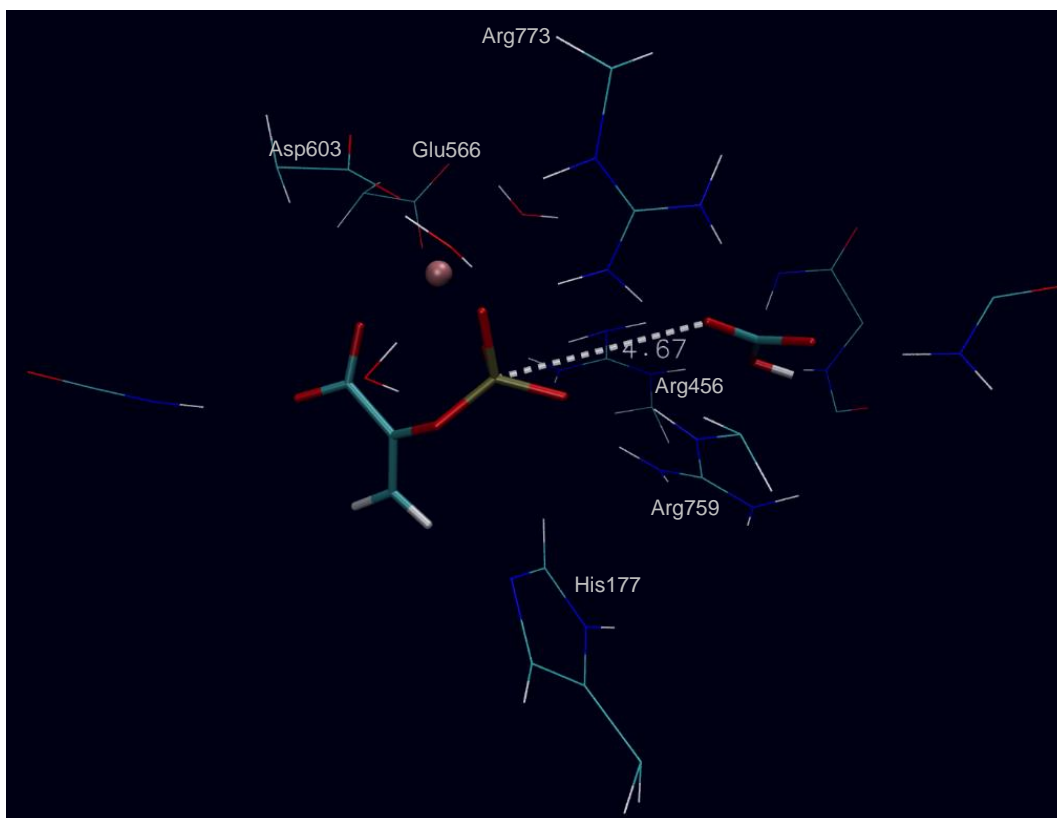
FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |35

Fig. 8. **High layer of the QM/MM system with HCO$_3^-$ in the opposite side of the leaving group.** Residues and water molecules as lines, substrates and the ion are thicker.

The valence of some atoms in High Layer was completed with H (link atoms), what in some cases gave rise to undefined parameters. Small molecules including the types of atoms whose angles or bonds were missing were drawn in *GaussView*, parametrized with a script using antechamber, and the parameters were copied/appended to the gaussian input file of the model.

The charge of the whole system was set as -6 and its multiplicity as 1. The charge of the high layer was -1 and its multiplicity was 1. The system was optimized with Mechanical Embedding, using the theoretical model B3LYP with the basis set 6-31g(d) for the QM subset of the calculations; the method used for Molecular Mechanics calculations was the amber force field. The system was then optimized with Electrostatic Embedding.

Starting from the optimized structure, a scan approaching the closest O of HCO$_3^-$ to the P of PEP was ran, as suggested by the proposed mechanism [10].
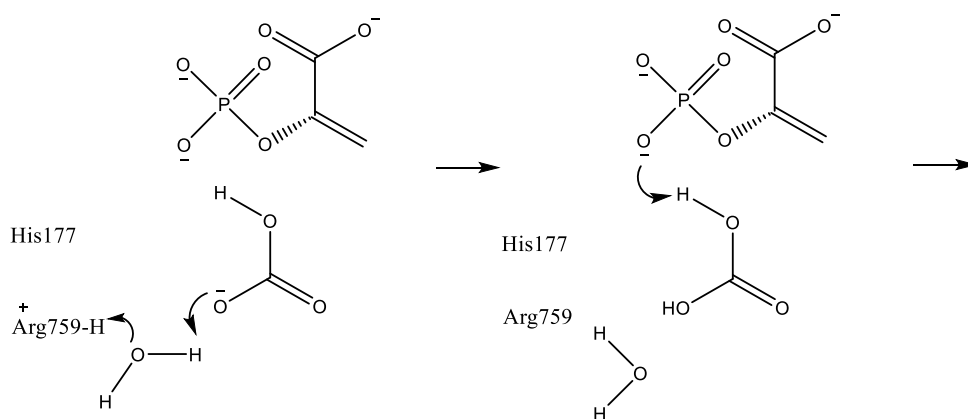
FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |36

➢ 8. Alternative mechanism

It was hypothesized whether water could transfer a proton to HCO_3^-. A scan approaching an H of the closest water to one of the O of HCO_3^- was performed. One of the final steps was optimized, ending up with a structure of H_2CO_3. The scan was repeated with smaller increments, an inflection was detected, and from it, a TS was optimized. The vibrational frequencies of this structure were calculated. VMD, with MolUP plug-in, was used to check the normal mode associated to the imaginary frequency found. Intrinsic Reaction Coordinates (IRC) jobs [49] [50] were run in both directions of the TS.

In another scan, the H of the hydroxyl group of H_2CO_3 was moved to the closest O of PEP. From the inflection detected, a structure was taken for optimization as TS. Its vibrational frequencies were computed and, after verifying it had solely one imaginary frequency whose normal mode was associated with the expected transfer, a bidirectional IRC was performed. The endpoints of the IRC calculations were further optimized.

With a new scan and an optimization for TS, as well as IRC from it, the next step of a new mechanism was characterized.

A Scheme of this alternative mechanism follows - Fig. 9. On it, HCO_3^- induces the breaking of the phosphate group, and only after the HCO_3^- attacks the P.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |37
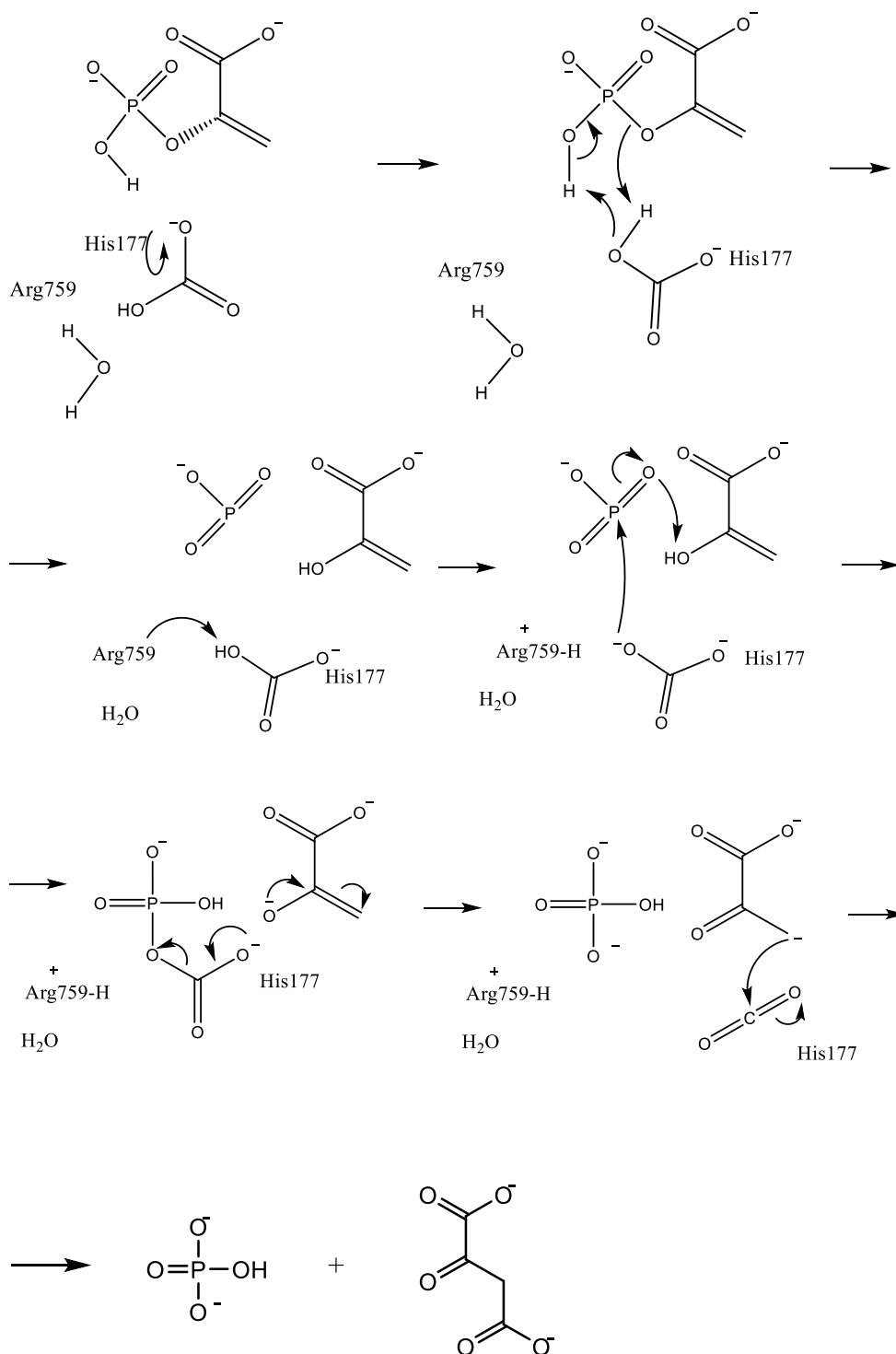
Fig. 9. **An alternative mechanism, in which phosphate of PEP breaks before the attack of HCO₃⁻ to P.** The presence of His177 is for the stabilization of the CO₃²⁻, after the removal of the proton by Arg759 at the 5th step. His177 can simply approach its proton, or share it, creating a more stable intermediate (HCO₃⁻); in the latter case, it could achieve a slight negative charge, that would be stabilized by Arg759, close to it and positive.

Only the first 3 steps of the mechanism of Fig. 9 were characterized and an overall energetic profile was drawn. For the highest energetic difference found, the two structures flanking it were evaluated with more accuracy regarding its energy: single-

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |38

point calculations were performed at the B3LYP/6-311 +G(2d,2p) – GD3BJ:AMBER level of theory to assess the electronic energy; the correction for the Gibbs free energy (which equals Gibbs free energy minus electronic energy) was gathered from a frequency calculation performed for each stationary point; both terms were summed to determine the Gibbs free energy of the points.

➢ 9. A model with a positively charged His177 - HIP

The initial choice of the protonation state of His177 as neutral charged and protonated at Hδ was based on the mechanism proposed in [10]. However, no experimental data was found that supported this protonation state.

Also, despite propKa 3.1 have not suggested to double protonate this histidine, it lies close to the ligands, where it is more difficult for the program to accurately predict the protonation states. Additionally, by the time it was run, it was being considered a pH close to 7 (condition employed in some experimental studies). However, some plants in which PEPC plays a role, are reported to be *acidic.* Indeed, PEPC is owned by two major groups of plants, C$_4$ and Crassulacean Acid Metabolism plants, a term coined by the acidic taste of their leaves in the morning (they fix Carbon by night).

To test if a double protonated His177 could help to figure out a plausible reaction mechanism, a new model was built.

As before, the chains A and B of the PEPC structure of the PDB entry 5VYJ were used in conjunction with the position for PEP, Mg^{2+}, and one water coordinated to it, taken from the entry 1JQN. All the procedure was the one described in Section *1.*, including the assignment of protonation states of the residues; an exception was made to the His177, which was renamed to be positive (HIP). The procedure differed also in the placing of HCO$_3^-$.

For the addition of HCO$_3^-$, the most representative frame of the 1st 5 ns of the free Production Phase, of the MD performed in the first model build (see Section *3.)* was aligned with the structure without HCO$_3^-$. The alignment was made based on the atoms of PEP whose type was P5, O6, O8, O1, or O3, and the Mg^{2+} ion. From this alignment, the new coordinates of HCO$_3^-$ were saved and appended to the structure with His177 as HIP*.*

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |39

The three bound glycine molecules - possible activators - were renamed to become zwitterionic glycines. Parameters for the force field used were those referred to in section *1*, plus those of a library for zwitterionic residues (see section 4.). The structure was protonated, neutralized, and inserted in a water box up to 12.0 Å. The system underwent minimization of H_2O, H, and sidechains and entered MD.

MD started with a Heating Stage of 50 ps, in which the system was heated till 298.15 °C, and then the system was allowed to equilibrate for 500 ps. During this time, backbone and substrate molecules, as well as the ion and its coordinating residues/water, were restrained. The system was then free, with the exception of PEP and HCO_3^- and it was equilibrated for further 500 ps. Finally, the system was completely free and equilibrated for 500 ps. A Production Phase of 5 ns followed.

The inputs provided, as restraining forces and integration times employed, were the same as those described in Section *2. Molecular Dynamics.*

Frames collected during the production phase were re-imaged and clustered into 5 groups based on the similarities of the position of HCO_3^-, PEP, Mg^{2+}, His177, Glu566, Asp603, Arg456, Arg647, Arg759, and Arg773. Within the most populated group, the most representative frame was selected and used in further steps. For details of the input, see Section *3.*

A sub-selection of residues and molecules up to 22 Å of HCO_3^-, PEP, Mg^{2+}, Glu566, Asp603, Arg456, Arg647, Arg759, Arg773, Ser602, and Gln673 was performed. TER labels were inserted where truncation caused discontinuities of the chain. In cases where one residue presented a TER label immediately before and after, its name was manually changed to the name of the corresponding amino acid in the zwitterionic library.

The structure and the required libraries or additional parameters were loaded on *xLeap*, which generated the coordinates and parameters files for the system. These were then used to create a Gaussian input file, with the aid of a script.

Residues more than 18 Å apart of HCO_3^-, PEP, Mg^{2+}, Glu566, Asp603, Arg456, Arg647, Arg759, Arg773, Ser602, and Gln673 were frozen, as well as all the water molecules with exception of three of them close to the ion or substrate molecules.

Using *Gaussview*, the ion, substrate molecules, and parts of the protein that stabilized them were moved into High Layer (ONIOM method) – 120 atoms. The remaining were kept in Low Layer. Atoms included in the High Layer are depicted in Fig. 10.

In atoms whose valence was incomplete (in High Layer), Link Atoms were employed. When it caused the need for new parameters, small molecules containing the

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |40

same type of bonds/angles where drawn on *Gaussview*, parametrized using
Antechamber and the parameters were appended to the gaussian input file.

As the subsequent calculations would be run with a subtractive scheme, the
charge and multiplicity of both the whole system and the High Layer were defined. The
charge of the whole system was set to -4 and its multiplicity to 1. Charge and multiplicity
of the High Layer were both 1.

The system was optimized: QM calculations used the model chemistry composed
by the B3LYP functional and the 6-31g(d) basis set; for the MM calculations, the amber
force field was used. First, optimization was carried considering a Mechanical
Embedding, and then an Electrostatic Embedding scheme. During the latter, the
correction for dispersion GD3BJ was used.

By mistake, unneeded waters had been included. Using the plugin *MolUP* for
VMD, a sub-selection that skipped waters more than 3 Å apart from the protein was
performed and saved as a Gaussian input file again. This allowed the system to become
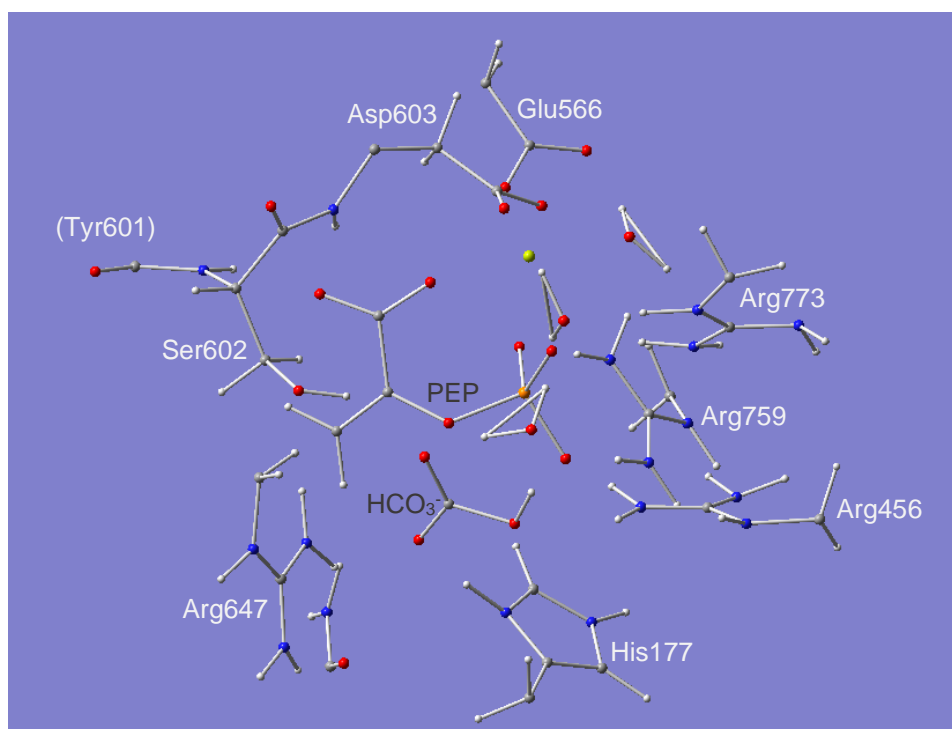smaller: 11 931 atoms.

Fig. 10. **The QM layer of the QM/MM model with a positive His177.** This structure was obtained after a geometry
optimization with Mechanical Embedding. Mg^{2+} is depicted as a green sphere and the remaining atoms are represented
by ball-and-stick. 3 water molecules are also present: one of them coordinates the ion, the other is at 1.82 Å of HCO$_3^-$ and
the remaining one is at 1.65 Å of the O of the phosphate group of PEP.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |41

From this structure, two scans were performed: in one, the angle that is shown in *Fig. 11* was changed, to bring the H of bicarbonate close to the O of PEP. This way, it was intended to achieve a conformation in which a mechanism similar to the reported above - the *Alternative Mechanism* in Fig. 9 - could take place. It includes a proton transfer from a specie derived from HCO_3^- to the phosphate group of PEP, so the proton of HCO_3^- should be close to the O of phosphate – <u>scan **A**</u>.

In another scan **,** the H of bicarbonate was moved to the closeness of an O of PEP coordinating the Mg^{2+}, by shortening the distance highlighted in *Fig. 11* – <u>scan B</u>.

After the scan A (and an optimization) the H of HCO_3^- interacted with an O of the phosphate of PEP. After the scan B (and an optimization), the H of HCO_3^- interacted with other O of the phosphate of PEP.
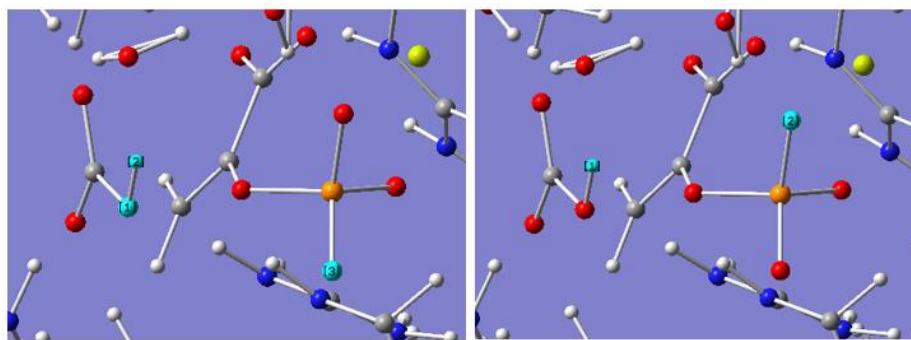

Fig. 11. **Paths to achieve different conformations, in the system with His177 doble protonated.** On the left side: angle varied till ~180˚, to achieve a conformation in which the H of HCO_3^- interacted with an O of the phosphate group of PEP. On the right side: distance shortened to achieve a conformation in which the H of HCO_3^- was interacting with other O of the phosphate group of PEP.

The final structures of the scans A and B were optimized. On the structure from the scan B, an extra water molecule was set free and included in QM (the closest water of Mg^{2+}) and the system was further optimized. This intended to allow the water to join the coordination shell of the ion, from the which O of the PEP had been partially separated during the scan (of the approach of it to HCO_3^-).

○ A path starting from A

A linear transit scan approaching H of the HIP to the O of the HCO_3^- was performed. From the inflection detected, a structure was taken and optimized for TS. One of the last structures of this scan was used for a new scan: the approaching of one H of the H_2CO_3 to the O of the LG of PEP.

○ A path starting from B

A linear transit scan approaching H of HIP to the O of HCO_3^- was performed. From its end, a scan approaching H of H_2CO_3 to the O of (the phosphate of) PEP was run.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |42

➢ 10. Other tests performed

1. According to the article [10] the atoms of the PEP which should coordinate to the metallic ion are shown in Fig. 12. That is, the ion coordinates 2 O of the phosphate group, one of which is the LG (the O bound to P and supposed to detach in the mechanism of Fig. 2), and 1 O of the carboxylate group. Some models were built in which PEP was modelled in this way of coordination with Mg^{2+} and were minimized (water, H, sidechains, and all the system).
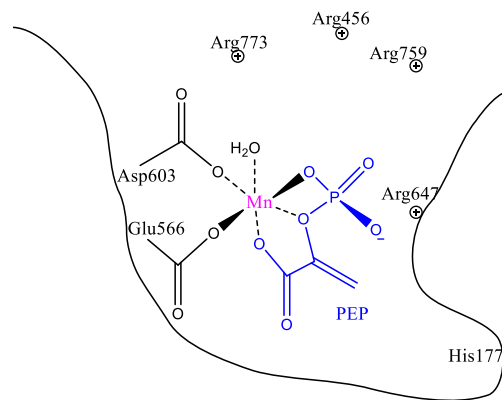

Fig. 12. **Coordination shell of Mg^{2+} according with article** *[10]*.

2. Linear transit scans were also performed to test alternative pathways for the reaction. Part of the hypothesized pathways are summarized in the scheme of *Fig. 13*.

3. A system with HIE177 was built, it underwent an MD simulation and, from a frame of it, an ONIOM model was built and optimized with Mechanical Embedding.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
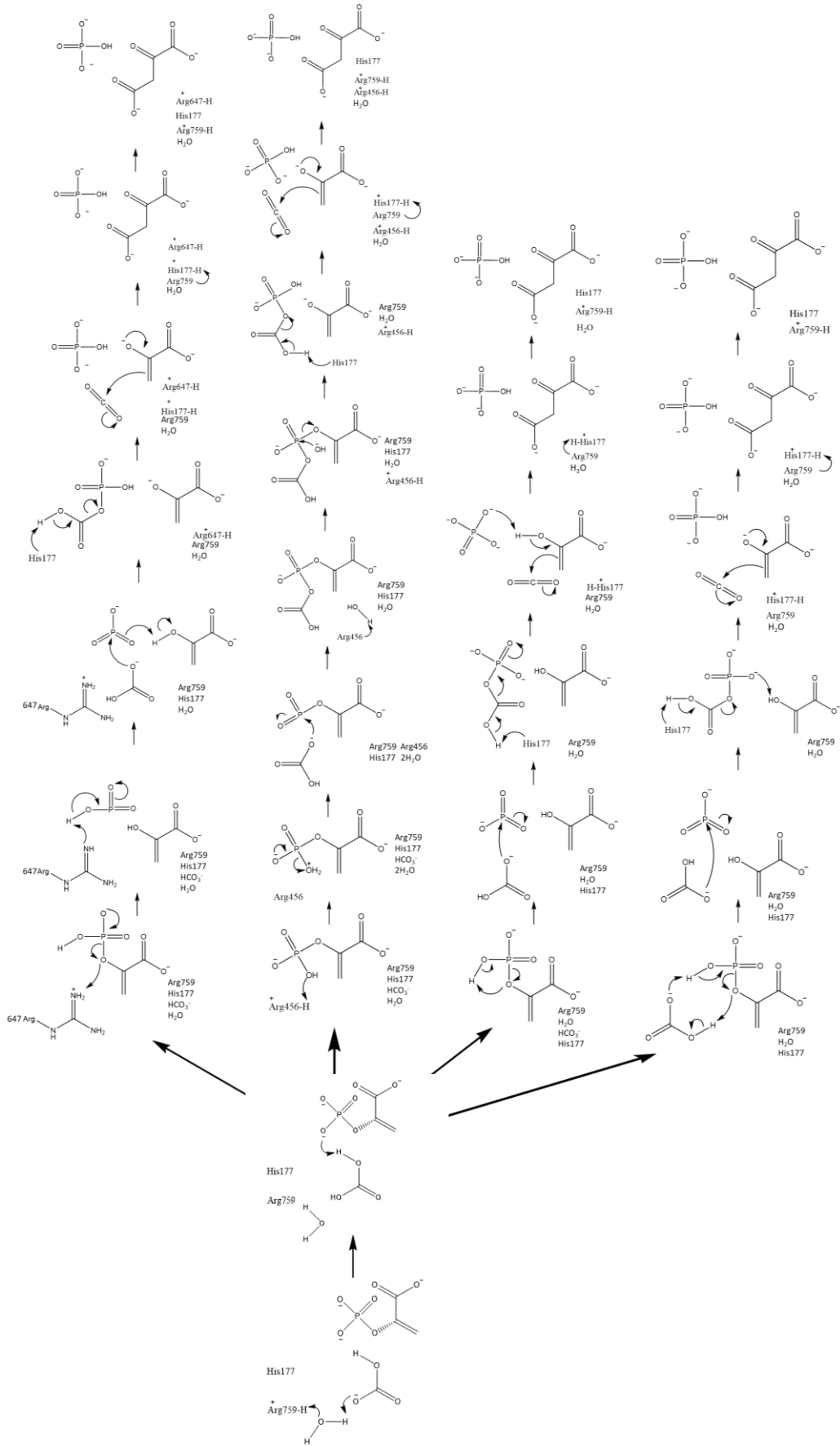engineering of a CO₂ capture device |43

Fig. 13. **Some reaction pathways hypothesized.**

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |44

C. Results and discussion

> ### 1. Building the model

PEPC structures 5VYJ (*Z. mays*) and 1JQN (*E. coli*) were aligned. The alignment of Trp288 [248], Arg456 [396], Leu564 [504], Glu566 [506], Met598 [538], Asp603 [543], Arg759 [699], Arg773 [713] - numbering of *Zea mays* PDB file [numbering of residues in *E. coli* PDB file] – is presented in Fig. 14. The RMSD of the residues (atoms of the sidechains included) after alignment was 0.96 Å.

It was from this alignment that the coordinates of Mn^{2+} , one water coordinating it, and a PEP analog were transferred to the system of the PEPC of *Zea mays.*
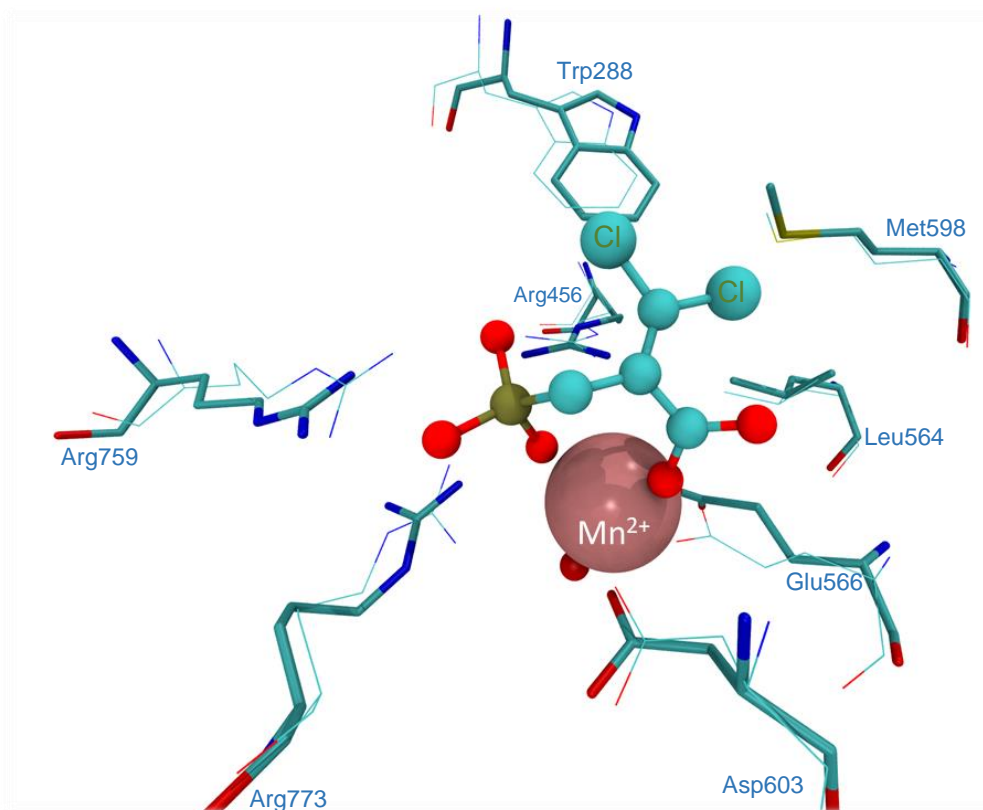

Fig. 14. **Superimposition of some residues of the active site of 5VYJ with those of 1JQN**. The residues of 5VYJ (*Z. mays*) are shown as sticks and the residues of 1JQN (*E.* coli) as lines. Also from 1JQN, PEP analog is represented in ball-and-stick, and the ion and the O of a water molecule as spheres. The numbering presented refers to *Z. mays* residues.

Using this scaffold, different systems were built, with HCO_3^- in different sites. In all of them, one of the oxygen atoms of HCO_3^- was close to the P of PEP. In one system, it was modelled close to the Mg^{2+} - site 1. In another model, it was placed on one side of the His177, a residue included in the mechanism proposed in the literature (see Fig. 2) - site 2. In another one, it was modelled on the other side of His177, and close to the site

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |45

of PEP where carboxylation should occur – site 3. Since in site 3 there was a clash with the H of a Gln sidechain, HCO$_3^-$ was slightly moved – site 4. The different sites are shown in Fig. 15. Despite four different sites were modelled, as sites 3 and 4 were similar, hereafter only sites 1, 2, and 4 will be considered.
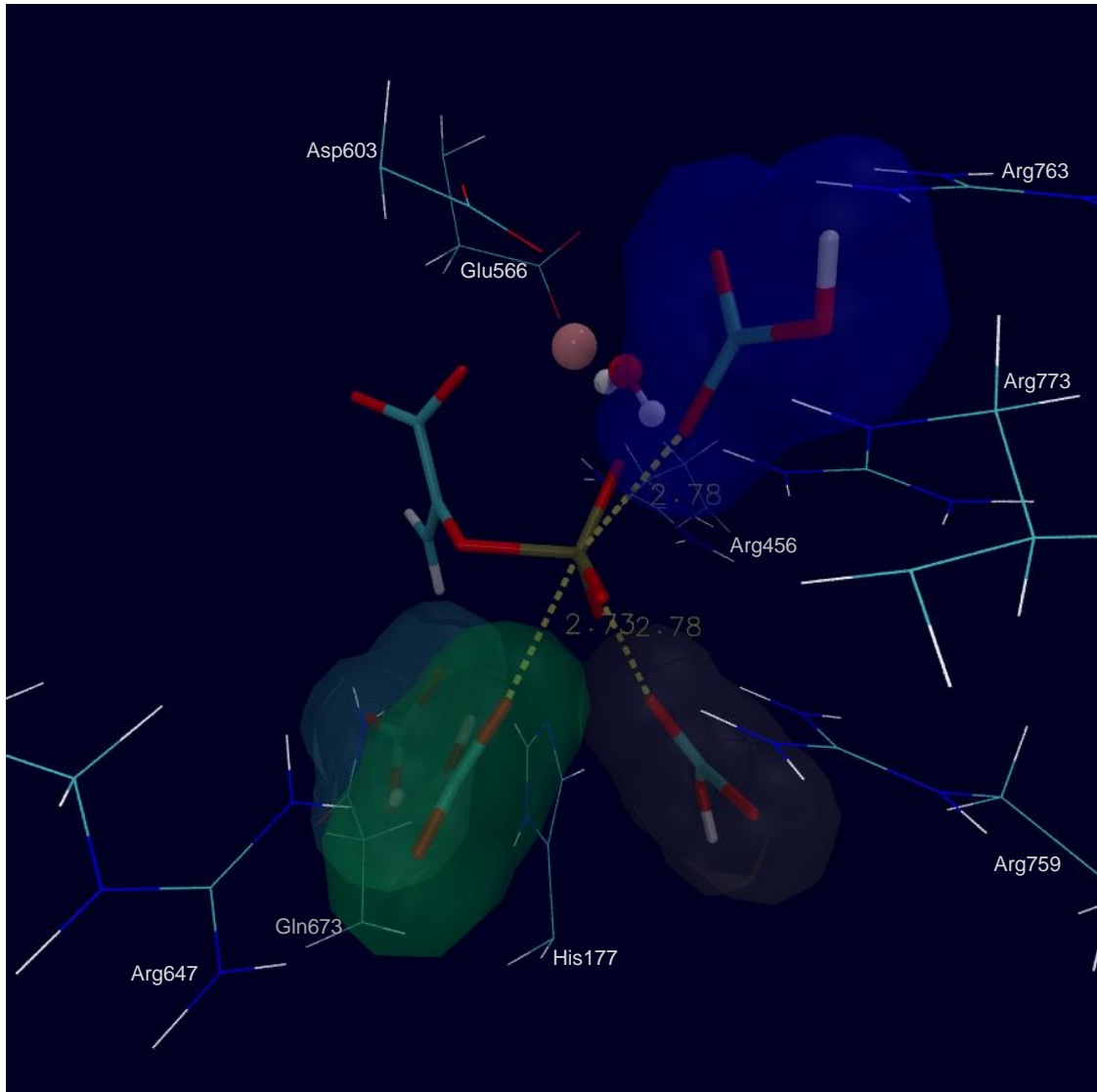

Fig. 15. **Four different sites where HCO$_3^-$ was modelled**. Blue-shaded (top right): site 1, with HCO$_3^-$ close to Mg^{2+}; dark grey (bottom right): site 2, with HCO$_3^-$ close to His177; light grey and green (bottom left): sites 3 and 4, respectively, with HCO$_3^-$ on the other side of His177. Site 3 and 4 are similar, but on 4 the molecule was adjusted to prevent the clash with the H of a Gln sidechain that occurred in site 3.

After minimizations of the entire system, for the different models, HCO$_3^-$ was found too far from the P of PEP – see Fig. 16. Thus, the result of the minimization of waters, H, and the backbone of the protein, in the model with HCO$_3^-$ in site 4 was used instead, to start an MD simulation. In the initial part of such MD simulation, the position

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |46

of the substrate molecules would be restrained, as an attempt that sidechains adjusted
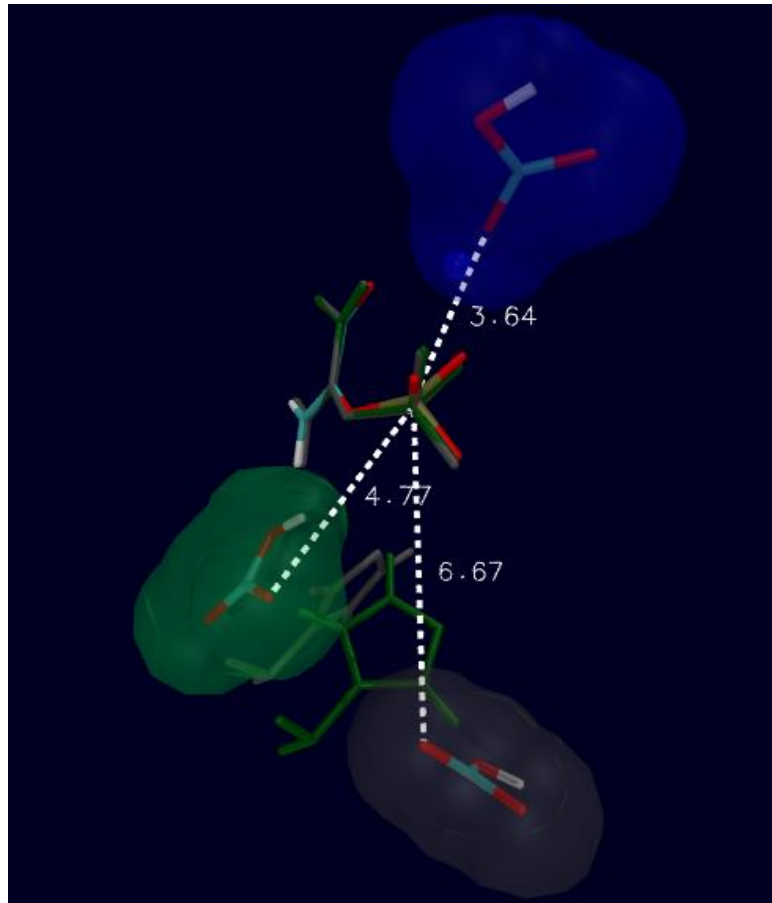to a conformation of the substrates in which they were close.

Fig. 16. **Different systems with HCO₃⁻ on different sites, after minimization.** Site 1: PEP is depicted in usual colours and HCO₃⁻ is shaded in blue, as well as Mg²⁺ ion. Site 2: PEP and His177 in grey, as well as the shading of HCO₃⁻. Site 4: PEP and His177 in green, as well as the shading of HCO₃⁻. The structures were aligned by PEP. His177 was not included for site 1 (it is far from the modelled HCO₃⁻ in such model).

➢ 2. Molecular Dynamics

After modelling HCO_3^-, changing of the analog to PEP, changing Mn^{2+} for Mg^{2+}, protonation, and partial minimization as reported in the Methods section, the system that entered MD simulation resembled like the one shown in Fig. 17.
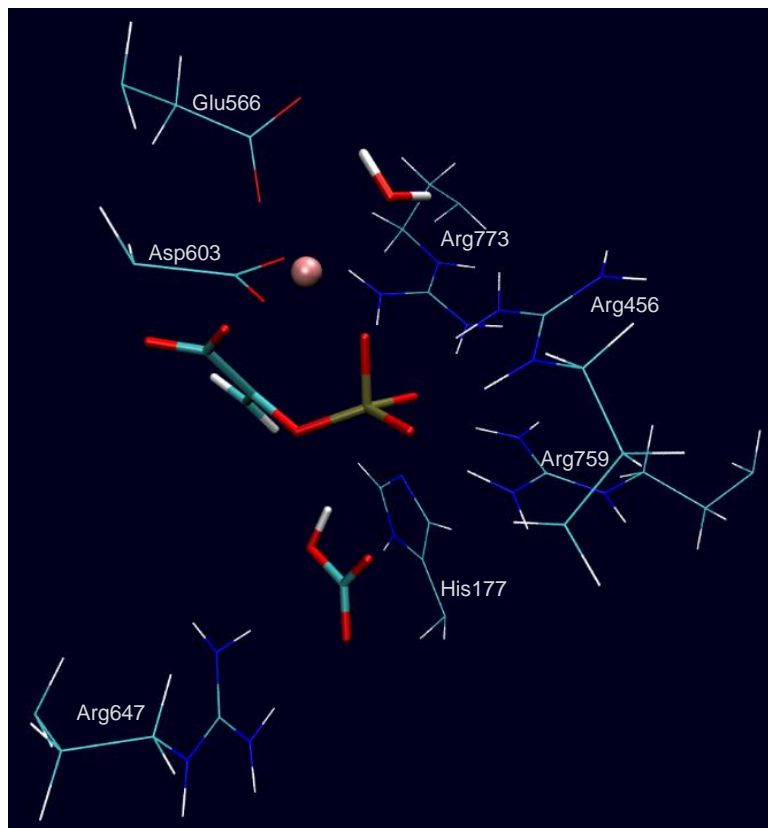
FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |47

Fig. 17. **Starting structure for the MD simulations.** Substrate molecules are shown in sticks, Mg^{2+} as a sphere, and a water molecule coordinated to it is also depicted in sticks. Some residues (sidechains) close to them are shown as lines. Distance of the P to the closest O of HCO$_3^-$= 3.18 Å.

During the MD simulation, the structure of the protein model seemed stable, as assessed by an RMSD analysis. The RMSD values were calculated after aligning the frames of the MD simulation with its first structure using VMD. The graph plotted in Fig. 18 presents the RMSD between frames of the MD and the initial structure, over the MD simulation time. In the top, it is shown the RMSD of the protein (atoms of the sidechains included in the RMSD calculation), over time. In the bottom of the figure, it is plotted the RMSD of Mg^{2+}, PEP, HCO$_3^-$, the residues Glu566 and Asp603, and one water molecule (coordinating the ion) over the MD. This MD included heating (50 ps), equilibration (500 ps), further equilibration with some atoms restrained (5 ns), and production (5 ns).

The distance between the two atoms supposed to bind in the proposed mechanism – the P (of PEP) and one of the O of HCO$_3^-$ was also measured over MD – it is plotted in Fig. 19. On it, it is observed that after the first 5 550 ps in which the substrates were restrained, the distance increases, but keeps within a reasonable range.

Finally, after clustering these MD frames in 5 groups, the most representative frame within the most populated group was selected to build a QM/MM model.
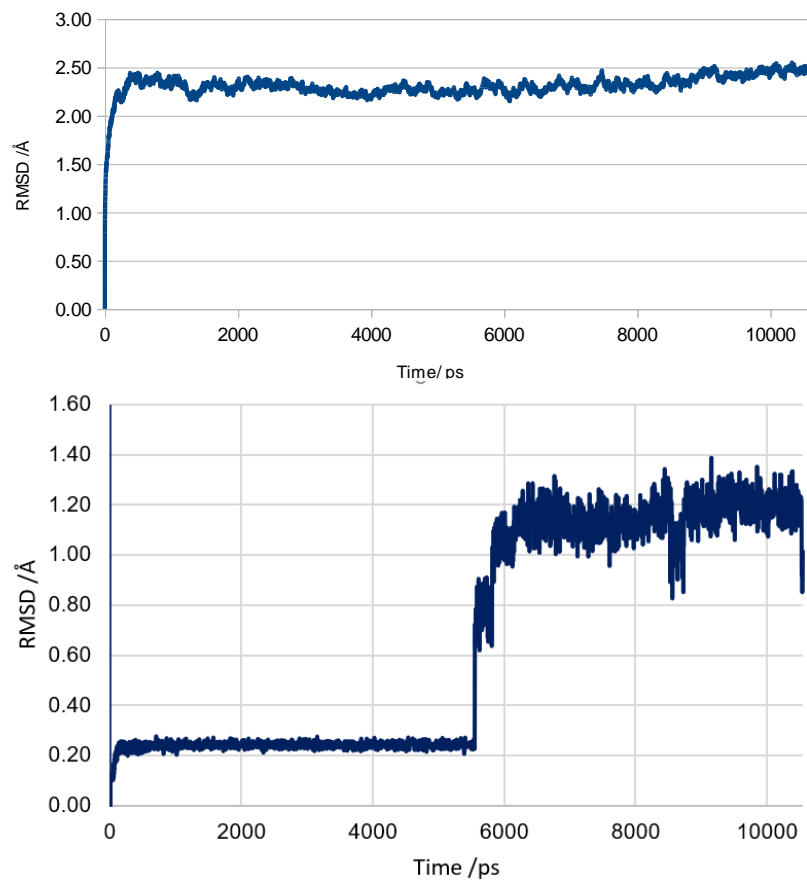
FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |48

Fig. 18. **RMSD analysis over the MD simulation.** Top: RMSD of the protein over MD. Sidechains included, first
structure of the MD used as reference. Bottom: RMSD of the atoms of PEP, HCO_3^-, Mg^{2+}, and 2 residues (Glu and Asp)
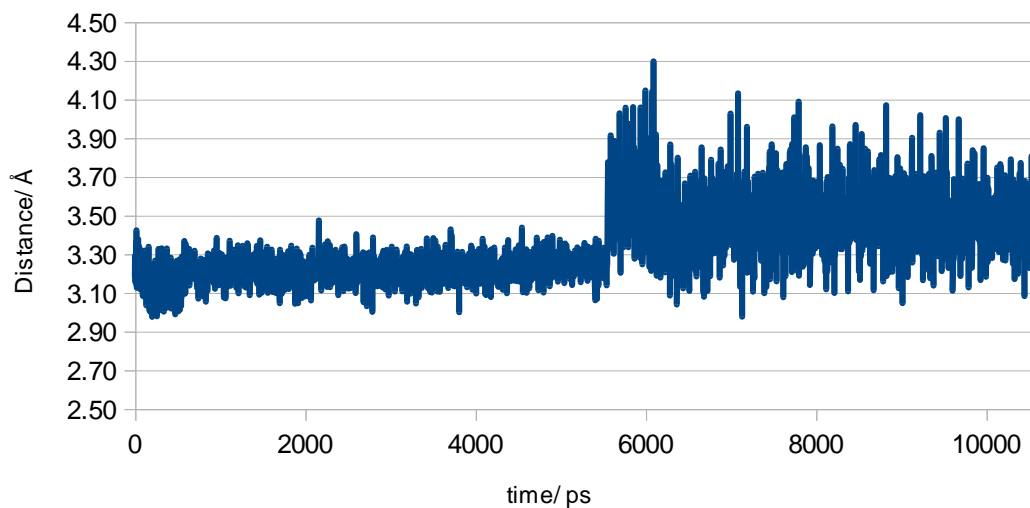and a water molecule coordinated to it, during MD. Note that their position was restrained during the first 5 550ps.

Fig. 19. **Distance from the P (PEP) to the O (HCO_3^-), over the MD simulation**. The O considered was the closest to
PEP, in the first frame of the MD simulation.

20 ns of simulation were additionally run: during them, the protein seemed stable (visually). However, one glycine molecule bound to PEPC (one of the allosteric activators) dissociated - Fig. *20*.
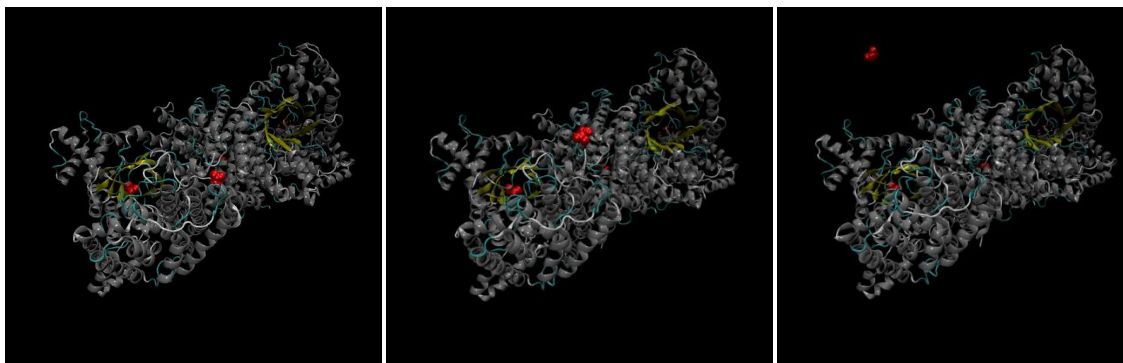

Fig. 20. **Detachment of one molecule of the allosteric activator.** Red: glycine molecules; remaining colours: PEPC; yellow: β-barrel, close to the which the active site is. Frames 1, 463, and 1875 of the 20 ns MD simulation. One of the three glycine molecules (red) detaches.

Interestingly, the molecule that detaches was bound to the site that Schlieper *et al* [12] report as an unspecific binding site, as opposed to the usually accepted idea that it is an activator binding site. It is a controversial topic in the literature. As reported in the Introduction, Schlieper assumes that glycine binds at unoccupied active sites (it is a competitive activator). The entry 5VYJ has one glycine molecule in such conditions. The 1st image above also has it: the red molecule (glycine) close to one of the β-barrels (the model used in the MD simulation is a dimer, in each monomer there is a β-barrel, and close to each of them there is an active site). Other authors claim that the effective binding location is out of the active site, what in the image above (1st) could refer to two glycine molecules (in red) away from the β-barrels. Like in the 1st frame represented above, in the PDB entry, glycine is found both in the active site and in other locations of the protein. However, Schlieper suggests that the binding out of the active site is unspecific. In this MD simulation, since one of the glycine molecules out of the active site detached, the binding could be weaker, as suggested by him.

It is also possible that this molecule binds at this site only in crystallization conditions (after evaporation), but not in bulk water. This would justify its presence in the crystallographic structure.

As reported above, from the cluster analysis of the frames of the first 5 ns of the free system in MD, a structure was picked for the crafting of a QM/MM model.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |50

> ## 3. QM/MM Model

The QM/MM model was built as described in Methods. After the optimization with electrostatic embedding, the structure of the QM layer was the one shown in Fig. 21.

This structure, when subject to an optimization and frequency calculation, displayed no imaginary frequencies, suggesting it was a minimum of the energy surface.
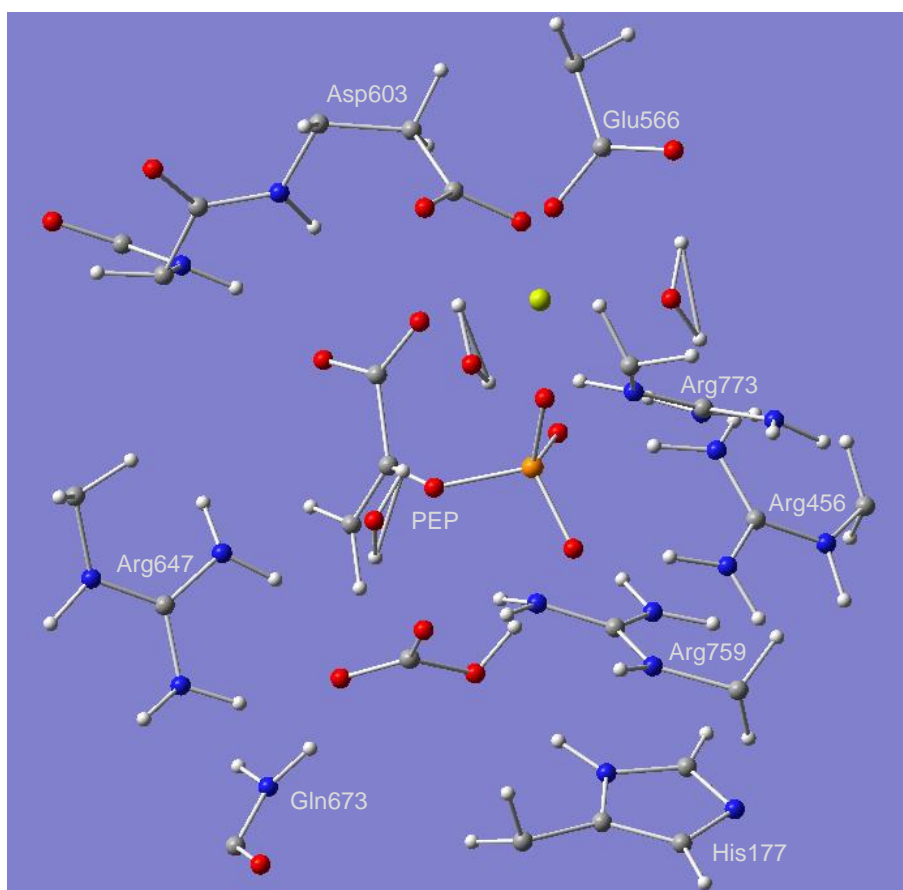
Fig. 21. **QM layer of the 1st ONIOM model built, after optimization with Electrostatic Embedding.**

> ## 4. Potential Energy Scans

In the 1st scan performed, the P of PEP and the O of HCO_3^- closest to it were approached. The initial distance was 3.80 Å and, at each step, the atoms were brought 0.1 Å closer. At the 12th step, the scan was restarted with a decrement of -0.08 Å per point of the scan. The potential energy profile is depicted in Fig. 22.
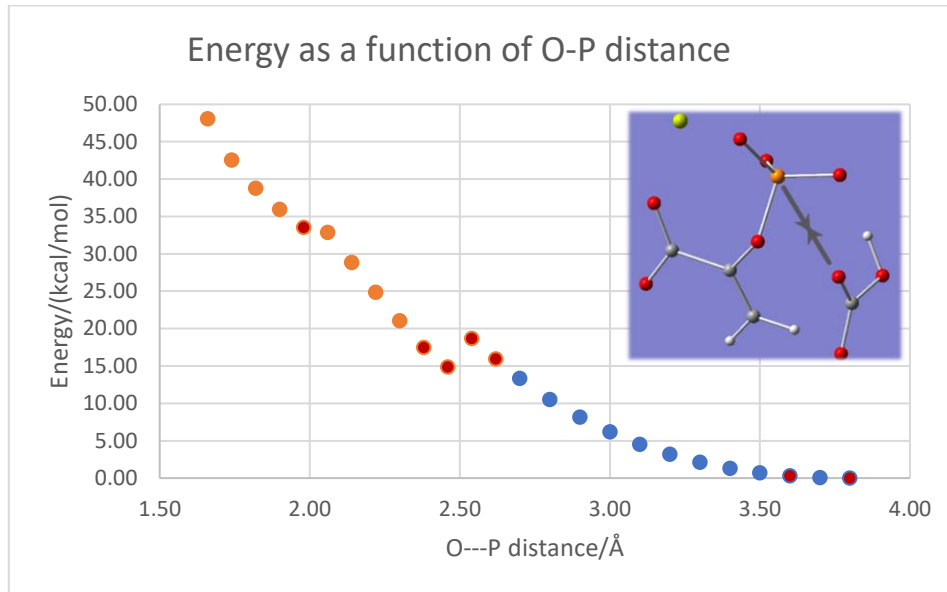
FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |51

Fig. 22. **Linear transit scan over O (HCO$_3^-$) - P (PEP) coordinate.** Red points are nonoptimized steps. Blue points are steps of the scan starting from the optimized model, and orange points are steps after the restart of the scan in blue.

The scan shows a high increase in total energy associated with the direct attack of O to P.

Falls in energy are due to structural rearrangements. The two events to be noticed in energy happened in the transition x=2.54 Å to x=2.46 Å and in the approach from x=2.06 Å to x=1.98 Å. On the first, there is a slight change in the coordination shell of Mg^{2+}, in the atoms of the PEP coordinated to it. On the second, the proton of HCO$_3^-$ is transferred to an O of PEP.

One of the structures after the mentioned fall at the 2.46 Å was optimized: the structure whose O-P distance was 2.30 Å (see *Fig. 22*). Then, a new scan O (HCO$_3^-$) --- P (PEP) was attempted: Fig. 23.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
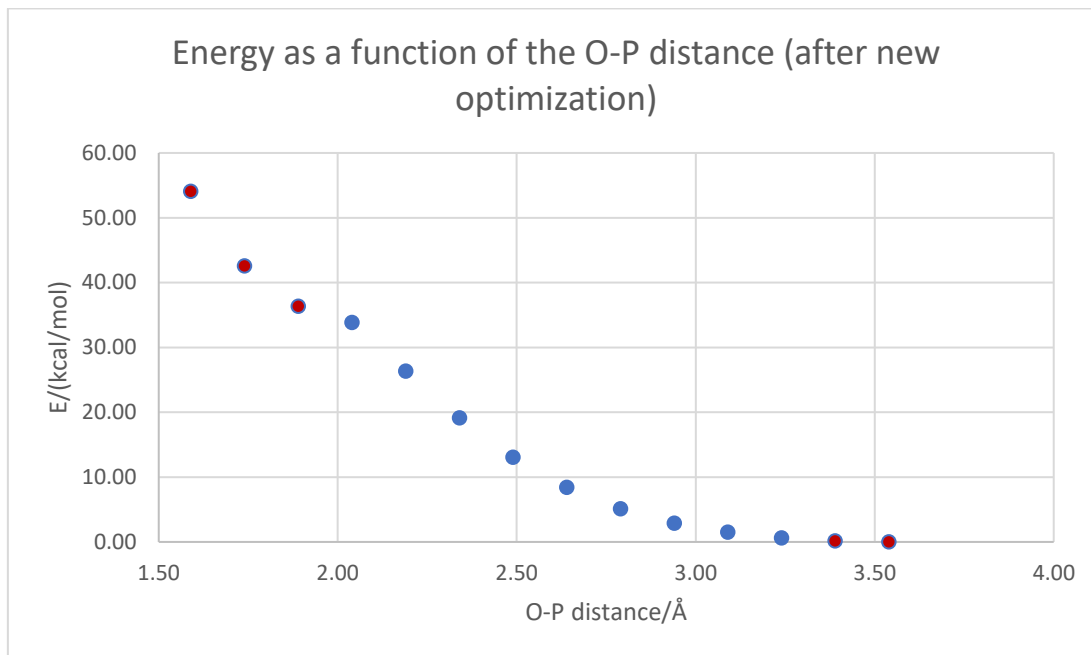engineering of a CO₂ capture device |52

Fig. 23. **Linear transit scan approaching the O (HCO3-) to the P (PEP), after new optimization**. A step of the scan of Fig. 22, after the conformational change that caused a discontinuity in the energy, was optimized, and used as the starting point for this scan.

The scan of Fig. 23 did not present the discontinuity (found in Fig. 22), caused by phosphate rotation, as intended. However, it showed that even for distances bigger than 2 Å, the energy rise was already above 30 kcal/mol. Thus, this approach was not continued.

A scan starting from the initial structure and decreasing the distance of the H of HCO_3^- to the closest O of PEP presented the energetic profile of Fig. 24. The atoms were initially 1.61 Å apart and they were approached 0.1 Å at each step.
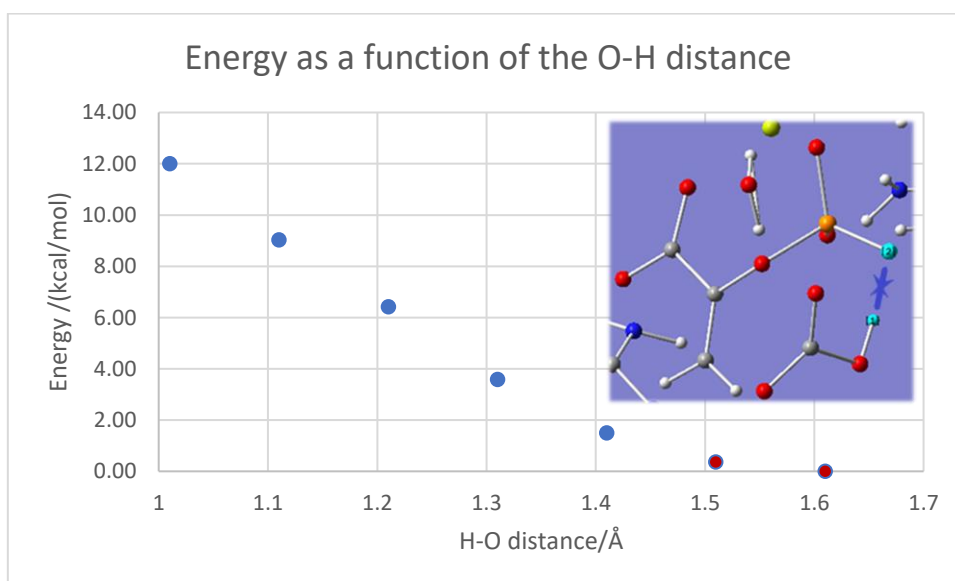

Fig. 24. **Linear transit scan approaching the H of HCO_3^- to the O of the phosphate group of PEP.**

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |53

On the scan above, the energy rising was not high. However, the scan presented no inflection, nor even a smoothing of the slope over the path followed. It also implied the formation of a doble charged product and this was not further explored.

➢ 5. QM/MM model: small system

The first ONIOM model built was comprised of 21 879 atoms. With the sub-selection of atoms up to 22 Å of the active site the size of the system decreased to 11 779 atoms.

To check if the two systems behaved similarly, after optimization with electrostatic embedding, the structures of the QM layer of both were aligned. They superimpose quite well, as shown in Fig. 25. The RMSD between them was 0.06 Å.

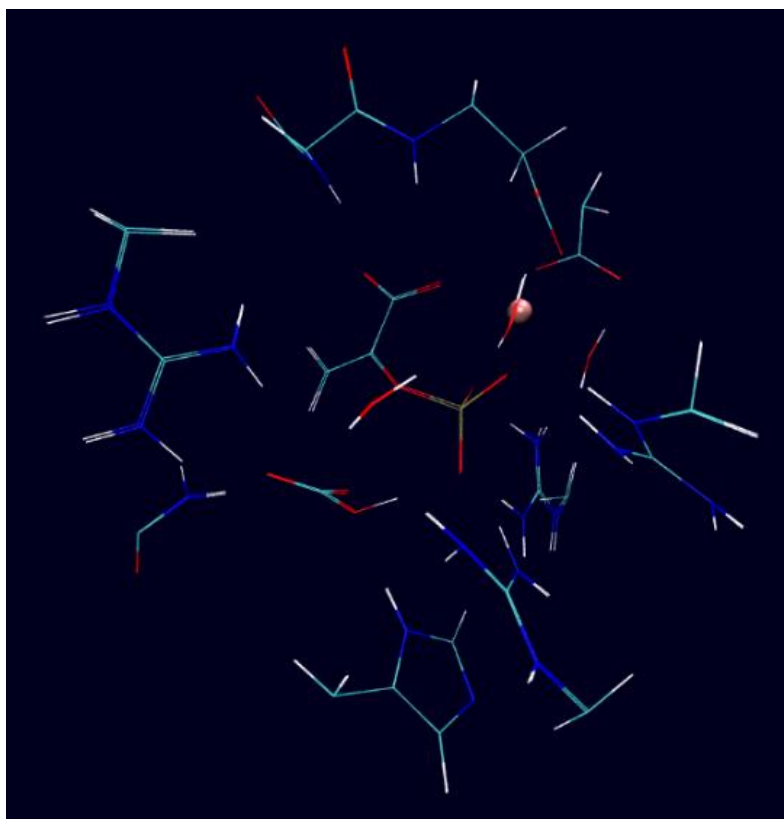Fig. 25. **Superimposition of the QM layers of the 1st ONIOM system built and of the smaller system.** Both had been optimized before this alignment.

Linear transit scans having the same reaction coordinate of the ones performed in the bigger system were run in the small system, to check if they showed similar profiles. In one of them, O (HCO$_3^-$) was approached to P, in decrements of -0.2 Å, and

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |54

in other, the H of HCO_3^- was approached to the closest O of the phosphate group of PEP, in decrements of -0.1 Å. as depicted in Fig. 26.
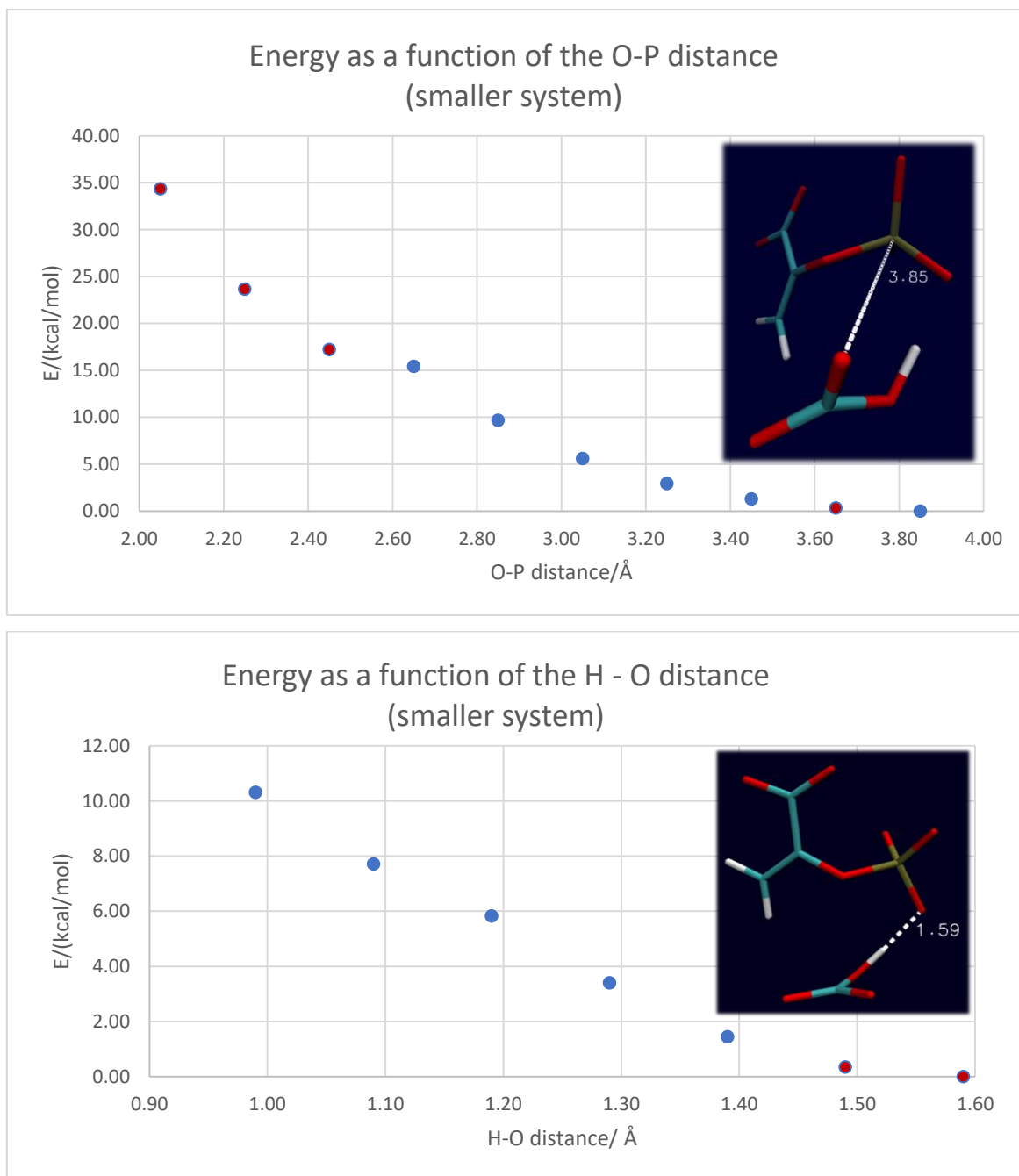
Fig. 26. **Linear transit scans performed in the small system**. Top: Linear transit scan approaching the O (HCO_3-) to the P (PEP). It is similar to the one in Fig. 22, which was performed in the larger system, but here the decrements by step are -0.2 Å, while in the scan of Fig. 22 they were -0.1 Å/step in the first part, and -0.08 Å in the restart of the scan. Bottom: Linear transit scan of the H+ transfer from HCO_3- to the O (of the phosphate group) of PEP. This scan is similar to the one in Fig. 24, which was performed in the larger system. The points in red didn't converge.

Comparing the energetic energy profiles of *Fig. 22* – the bigger system – with the one of Fig. 26, top – smaller system - they seem similar. For instance, in Fig. 22, when the atoms being approached are at 1.98 Å, the increase in energy relative to the first

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |55

structure of the scan is 33.52 kcal/mol, while in the plot of Fig. 26 (top), when the atoms reach the 2.05 Å distance, the energy rising is 34.33 kcal/mol.

Comparing the plots of *Fig. 24* and Fig. 26 (bottom), the energy of the larger system at x=1.01 Å is 12.01 kcal/mol above the first point, while the energy in the smaller system at x=0.99 Å is 10.32 kcal/mol above its first step. Part of this difference can be because the initial steps of both scans, where the 0 was set, didn't converge. Nevertheless, the profiles of the graphs seemed similar.

Both due to the small RMSD between the optimized structures of the large and smaller systems (truncated at 22 Å from the active site) and to the similar profiles of the scans presented above, the smaller system was used in the upcoming calculations.

➢ 6. Addition and removal of water molecules

Since the direct approach of O (HCO$_3^-$) to P, starting from the model built, seemed to require a high energy barrier, one water molecules was modelled close to the O of the leaving group of PEP, as reported in Materials and Methods. A close look at the optimized structure is shown below in Fig. 27.
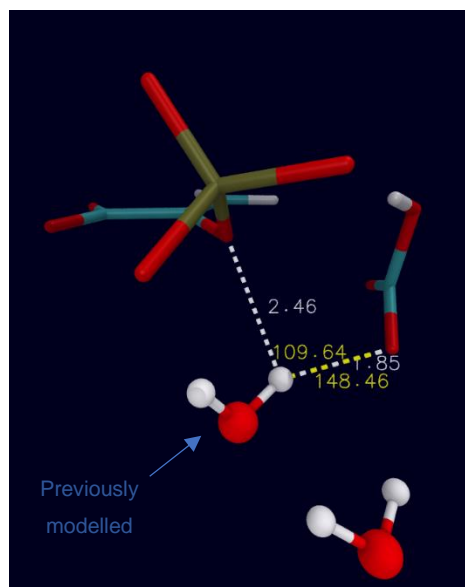
Fig. 27. **System in which one water had been modelled close to the O of PEP that was supposed to detach** (according to the proposed mechanism), after a geometry optimization. Values in white refer to distances in Å; values in yellow are related to angles between O (water) --- H (water) ---- O (X), being X the PEP (109.64˚) or HCO$_3^-$ (148.46˚).

After a geometry optimization, the modelled water seemed stable, establishing an H bond with another water of the system. It was also close to the O of the phosphate group

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |56

of PEP that it was intended to stabilize – the leaving group in the mechanism proposed in the literature (LG). However, the H of water that was close to the LG was also close to one O of the HCO$_3^-$ (1.85 Å). Moreover, the angle that the OH of water made with O of HCO$_3^-$ is compatible with an H bond interaction (120° < angle < 180°), while it was not the case for the LG (see Fig. 27). This way, the additional water molecule was not stabilizing the intended atom. Beyond it, the interaction with HCO$_3^-$ was undesirable, because the stabilized O of HCO$_3^-$ is the atom that should attack P: how more stabilized, how less nucleophilicity it would have.

Nevertheless, the water close to the leaving group could stabilize it after the breaking of the bond occur (in the attack of O of HCO$_3^-$ to P, see Fig. 2). A linear transit scan approaching the O of HCO$_3^-$ to the P was performed on this system and its profile is in Fig. 28. The initial distance between them was 3.86 Å and it was decremented by 0.15 Å/step.
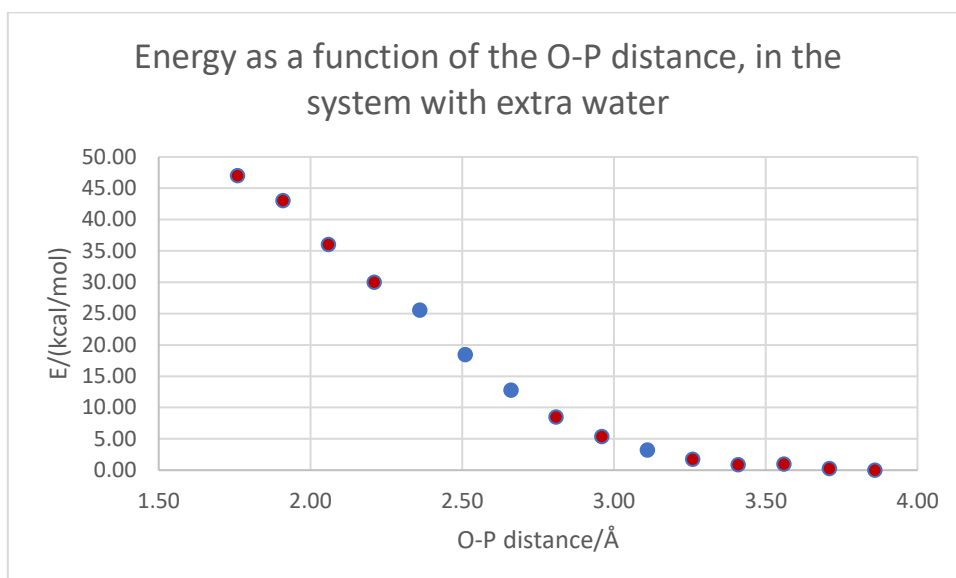
Fig. 28. **Linear transit scan starting from the optimized system with an extra water molecule** (the one modelled). Energies as a function of the reaction coordinate (the distance of O of HCO$_3^-$ to P of PEP).

The analysis of the graph above suggests that even with the additional water added to the system, the direct attack O-P requires a lot of energy – for instance, just for the approach until they are 2.06 Å apart, it takes 36.01 kcal/mol. This model was not further explored.

As reported in the Methods section, one water molecule (present since the MD; not modelled) was hypothesized to be preventing the substrate to adopt a conformation with lower energy over the reaction. A linear transit scan of the system without a specific

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |57

water molecule – see Fig. 6 - that was strongly bound to PEP follows (Fig. 29). In this scan, the O of HCO$_3^-$ was approached to P, in decrements of -0.2 Å.
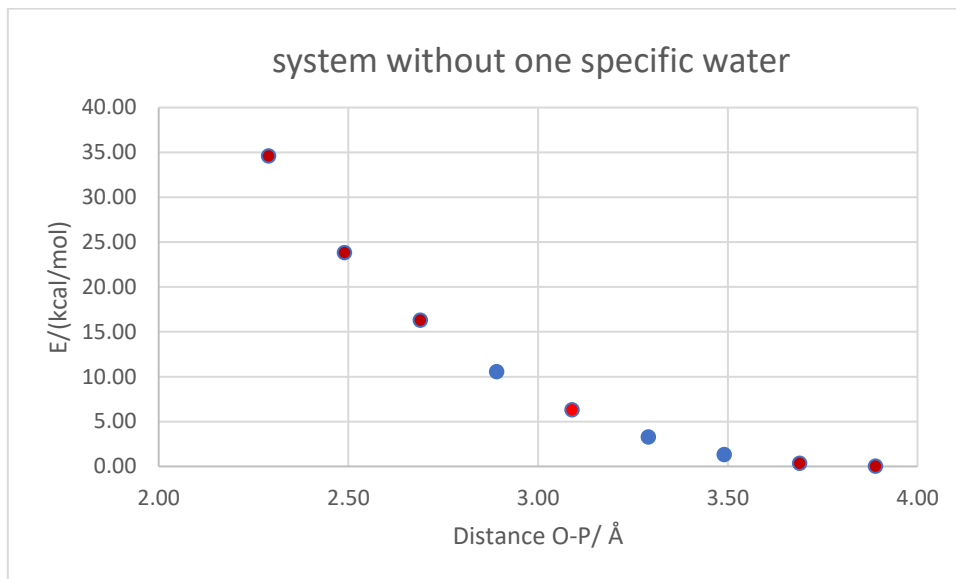
Fig. 29. **Linear transit scan starting from the structure where the water of Fig. 6 was deleted**. In this scan, O of HCO$_3^-$ was approached to P. Red points are steps that haven't converged.

As seen in *Fig. 29*, when one specific water was removed, there was a high energy associated with the approach of O to P: for instance, even for distances > 2 Å between the two atoms, the energy was ~35 kcal/mol above the energy of the initial step.

➤ 7. Model with HCO$_3^-$ placed at 180° of the Leaving Group (LG)

As reported in the Methods section, a model in which HCO$_3^-$ was modelled at 180° of the O atom that should detach of P after the attack by HCO$_3^-$ was built.

During the MD simulation reported in Methods, the distances from O of HCO$_3^-$ to P of PEP, during the production phase, are those shown in *Fig. 30.*
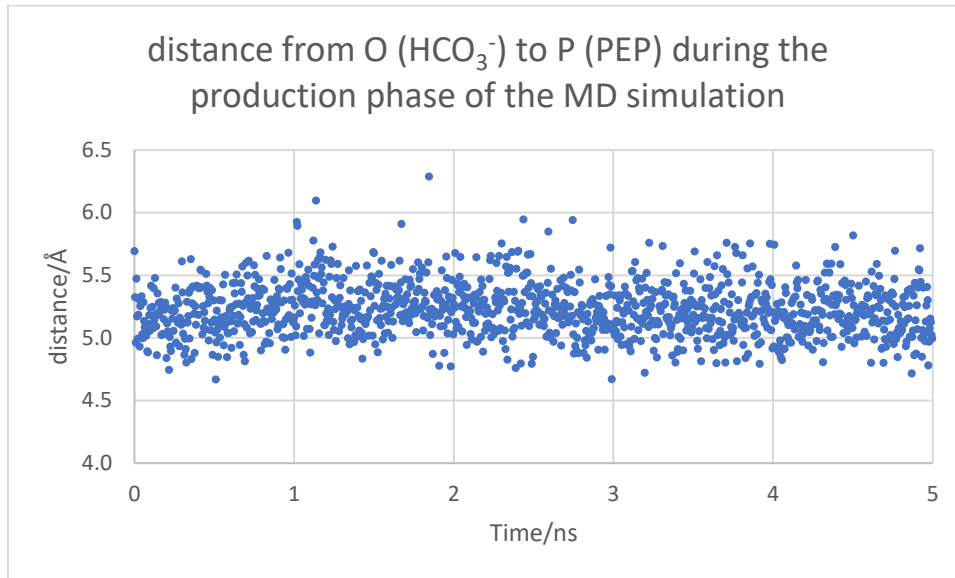
Fig. 30. **Distances of O (HCO$_3^-$) to P (PEP) over the MD simulation. System with HCO$_3^-$ at the side of P that was opposite to the LG.**

As shown above, the P of PEP and the O of HCO$_3^-$ were far from each other. The frame in which they were at the smallest distance (among all the frames acquired during the production phase) – 508 ps – also displayed some problems. As shown below (*Fig. 31*), sidechains of Arg759, Arg456, and Arg773 partially blocked the access of HCO$_3^-$ to PEP.

Fig. 31. **PEP and HCO$_3^-$, in the frame of MD where its O and P got closest**. Sidechains of Arg759, Arg456, and Arg773 are in VDW translucent representation. As can be seen, they partially block the contact between the substrates.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |59

Yet, there was the possibility of the sidechains of Arginine residues to adjust over the progression of HCO_3^- to PEP. As reported in the Methods, the scan of the approach of O of HCO_3^- (the closest O to PEP) to P was performed. It dis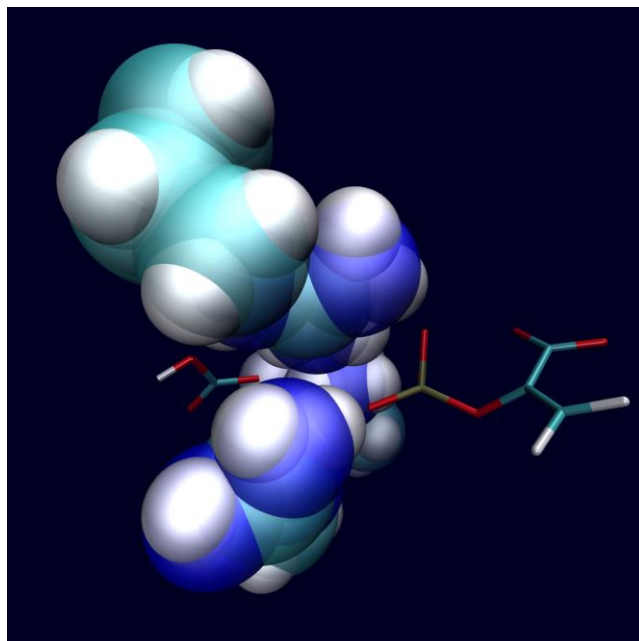played the energetic profile of Fig. 32. The scan stopped at its 9th step, so it was restarted (from its 8th point).
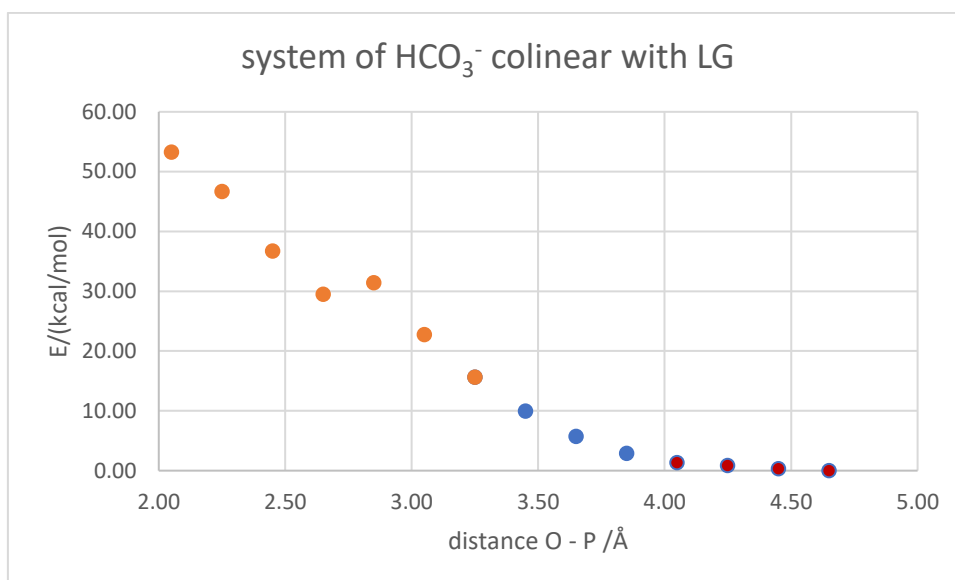

Fig. 32. **Linear transit scan in the system of HCO_3^- placed at 180º of the LG of PEP.** During this scan, the reaction coordinate explored was the distance O (HCO_3^-) to P (PEP). Red points mean non converged structures. Points in orange refer to the restart of the scan, and none of them converged.

Once more, the rise in energy seemed high to be feasible. Even if the last point acquired (2.05 Å, 53.27 kcal/mol) is discarded, since the calculation was stopped, the remaining points suggest that energy is high.

Similarly to what happened in other scans, the discontinuity between points of [distance O - P = 2.85 Å] and [distance O - P = 2.65 Å] is associated with a conformational change of the Phosphate group of PEP.

➢ 8. Alternative mechanism

The mechanism depicted in Fig. 9 is a reaction path that implies a cleavage of PO_3^- from PEP before HCO_3^- attacks the P. Since the previous attempts to attack P directly on PEP seemed to lead to too high energetic barriers, it was hypothesized if it could be due to the repulsion/steric hindrance (faced by the O of bicarbonate when approaching the Os of the phosphate group of PEP). This alternative mechanism could circumvent that; moreover, in experiments with PEPC in which His177 was mutated, it was detected

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |60

pyruvate, a result of dephosphorylation of PEP, so it was hypothesized if it could be one of the intermediates in the normal reaction.

The first 3 steps of the mechanism were exploited, and an energetic profile is depicted in Fig. 33. Attempts to find the TS for the next steps failed.

Fig. 33 presents, side by side with no energy adjustment between them:

- the TS of the 1st step, two IRC starting from it (forward and reverse), and a geometry optimization from the tail of IRC reverse (towards reagents);
- the TS of the 2nd step, one bidirectional IRC starting from it, and two geometry optimizations, one from each of the tails of the IRC;
- the TS of the 3rd step, two IRC starting from it, and a geometry optimization of the tail of the IRC reverse (towards 'products' in step 3 of Fig. 9).

Since the last step of IRC of 1st TS, forward, was similar in structure to the optimized structure of IRC tail (one of its sides) of 2nd TS, no further optimization of the result of the IRC of 1st TS was performed. RMSD between the proteins of the two structures was 0.002 Å, and the RMSD between the QM layer of them was 0.02 Å. Correlated with similarities in structure, the difference in energy between them was small: the non-optimized structure is 0.06 kcal/mol above.

The tail of the IRC forward (backward in reaction scheme of Fig. 9) of the 3rd TS was not optimized due to similar reasons. It was similar in structure to the optimized tail of the IRC of the 2nd TS (the side that connected with the same intermediate – B in the figure below), with an RMSD between the proteins of the two structures = 0.005 Å, and an RMSD between the QM layer of the two structures = 0.02 Å. The difference in the energies of the two structures was 0.1 kcal/mol.

The energies reported in this work, including the ones used for the construction of *Fig. 33*, were calculated using the functional b3lyp with basis set 6-31g(d). Other levels of theory were used only when explicitly mentioned.
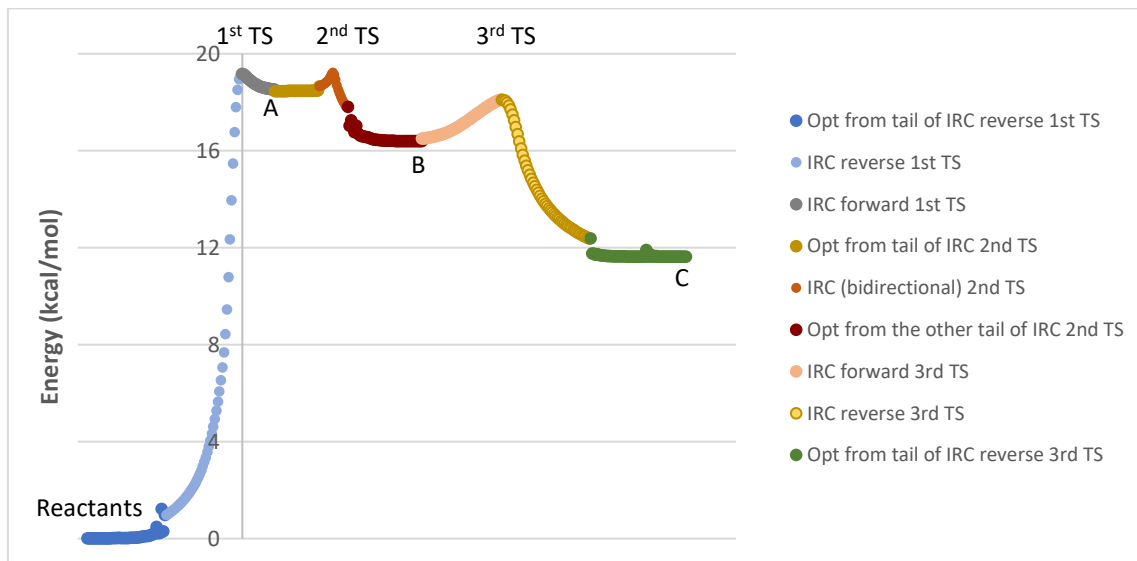
FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |61

Fig. 33. **Energetic profile of the first part of the alternative mechanism proposed** (Fig. 9). It includes the energy for the 1st, 2nd and 3rd TS found, as well as the IRC pathways from each of them. In some cases, from the tails of the IRC calculations, geometry optimizations were run. The X-axis is composed of the x values of Intrinsic reaction coordinates (for the IRC points) and optimization steps (in the case of the optimizations). Y-axis is composed of the original energy values of each calculation minus the underline constant value of the energy of the reactants (*i.e.*, all the values were translated in block, vertically). A, B, and C refer to intermediate species.

Figures of the transition states found follow: Fig. 34. They presented, each, one imaginary frequency, which are referred to in table 2. Additionally, the normal modes of such imaginary frequencies matched the processes proposed for the mechanism (of Fig. 9).

Intermediate species marked in the graph of Fig. 33 are also depicted in Fig. 34.

Energies of the reactants, intermediate species, and TS of Fig. 33 are summarized in table 3.
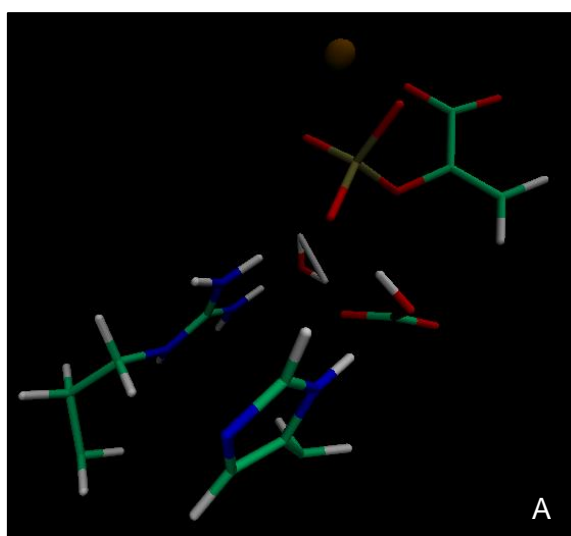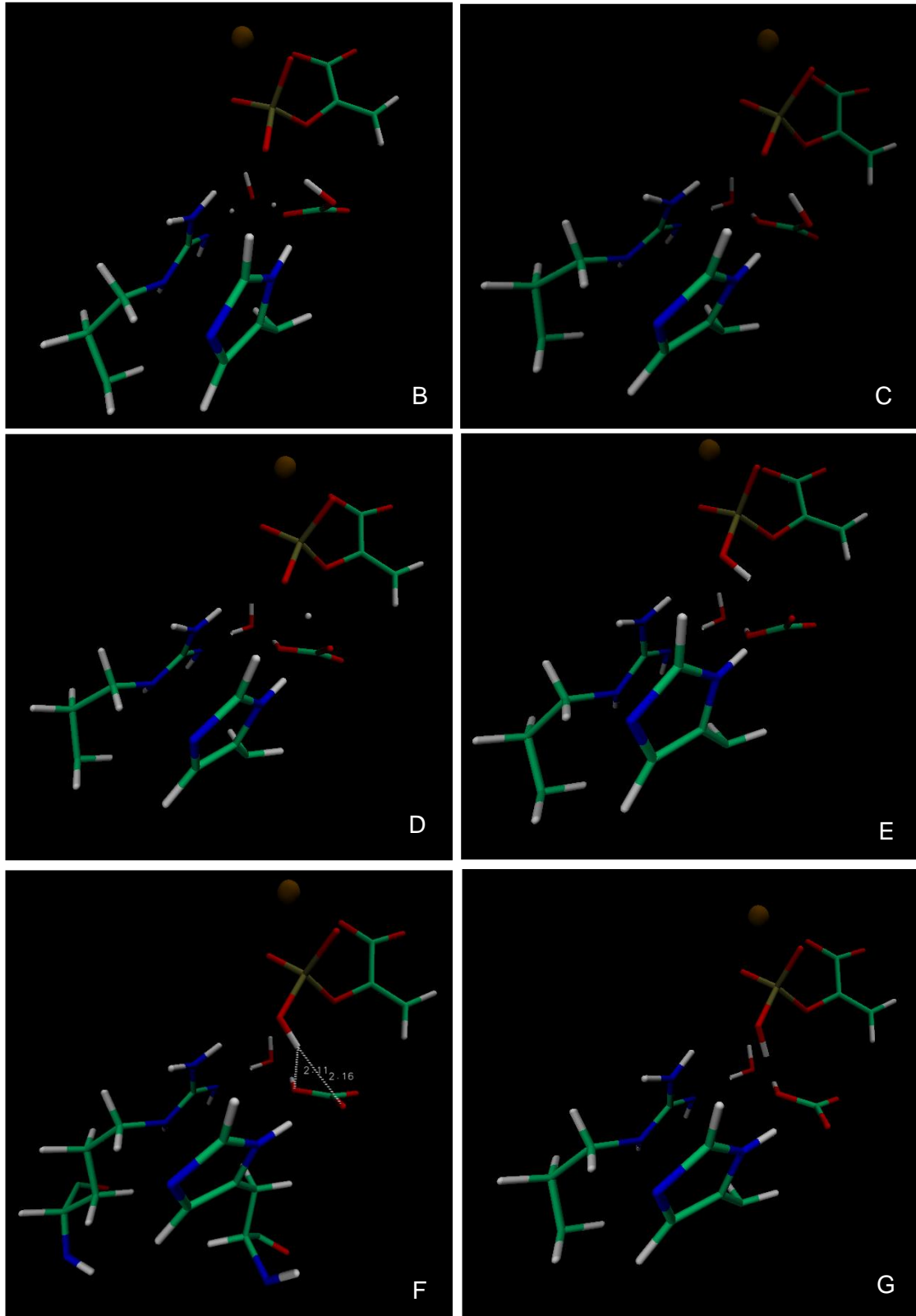
Fig. 34. **Structure of the Reactants, Transition States, and intermediate species.** A: reactants; B: change of a proton from an Arg to a water molecule, and of another proton from the same water molecule to HCO_3^-; C: H_2CO_3 and neutral Arg were formed; D: change of proton from H_2CO_3 to the phosphate group of PEP; E: PEP protonated at its phosphate group; F: change of interaction of the H of the phosphate group of PEP, from an interaction with the O where it came from (previous step), to a new interaction with the O of the hydroxyl group of HCO_3^-; G: H of the phosphate of PEP interacting with the hydroxyl group of HCO_3^-.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |63

Table 2. **Imaginary frequencies of reactants and TS of the reaction**. *Reactants refer to the structure optimized from the tail of the IRC of the 1st TS (one of its tails)

Structure	Imaginary frequencies (/cm^{-1})	Associated normal mode
Reactants*	-	-
1st TS	574.0 i	Transfer of one proton from Arg759 sidechain to water and transfer of another proton from the same water to HCO_3^-
2nd TS	1001.1 i	Transfer of one proton from H_2CO_3 to the phosphate group of PEP
3rd TS	258.4 i	The proton of PEP exchanges the interaction with the O of HCO_3^- that previously donated it, to a new interaction with the O of the hydroxyl group of HCO_3^-

Table 3. **Energies for the reactants, TS, and intermediate species.** *Reactants refer to the structure optimized from the tail of the IRC of the 1st TS (one of its tails). Energy, using oniom (b3lyp/6-31G(d):AMBER).

	Δ in Energy relative to the previous line/(kcal/mol)
Reactants*	-
1st TS	+19.16
Intermediate A	-0.71
2nd TS	+0.75
Intermediate B	-2.82
3rd TS	+1.71
Intermediate C	-6.47

The structure of the TS and the minima between the which there was the highest energetic gap (i. e., in the barrier for the first step, see table 3) were subject to single point energy calculations using a larger basis set: 6-311+g(2d,2p). The values of the electronic energies, from the single point calculations, were combined with a correction for Gibbs energy, from frequency calculations (which includes the zero-point energy, and the thermal energy associated to the translation and rotation at 298.15 K, as well as entropy contributions). The results are in table 4.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |64

Table 4. **Energies of the structures between which there is the largest energetic barrier** (considering the first 3 steps of the alternative mechanism that were characterized). *Reactants refer to the structure optimized from the tail of the IRC of the 1st TS (one of its tails).

Structure	Electronic energy calculated with ONIOM [B3LYP/6 -311+g(2d,2p):AMBER]/ Hartree	Correction to Gibbs energy (T=298.15K, p=1 atm) /Hartree	Total Gibbs Energy/ (kcal/mol)
Reactants*	-3937.75369	58.24643	-2 434 425.72
1st TS	-3937.72386	58.24080	-2 434 410.54
		$\Delta G°=$	15.18

➢ 9. A model with a positively charged His177

○ Pathway A

A QM/MM system with His177 double protonated was built and optimized, as reported in the Methods section . It underwent a scan to approach the H of HCO$_3^-$ to O of the phosphate group - an <u>O non coordinated with Mg^{2+}</u> - and it was optimized again.

A linear transit scan approaching the H of His177 (double protonated) to an O of HCO$_3^-$ showed an inflection and a low barrier (~5 kcal/mol), as shown in Fig. 35. The step with reaction coordinate 1.07 Å was then optimized successfully, leading to a structure containing H$_2$CO$_3$. From it, a new linear transit scan was performed: the approach of the H of H$_2$CO$_3$ (the H previously received from His) to the O of the LG of PEP - energetic profile in Fig. 36.

The potential energy profile (in Fig. 36) displayed a nice curve, so the step whose reaction coordinate was 1.63 Å was further optimized - the result is also shown in Fig. 36. On it, the H of H$_2$CO$_3$ was H-bonded to the LG of PEP.
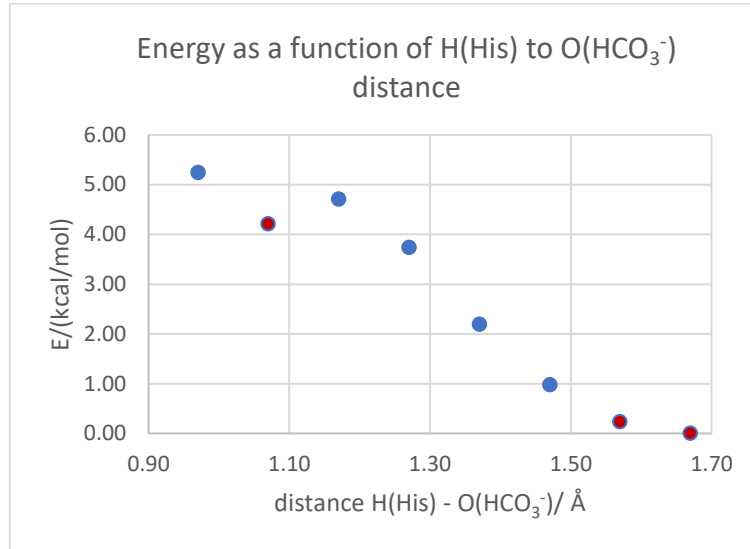
Fig. 35. **Linear transit scan, approaching the H of HIP177 to the O of HCO₃⁻.** Red points didn't optimize
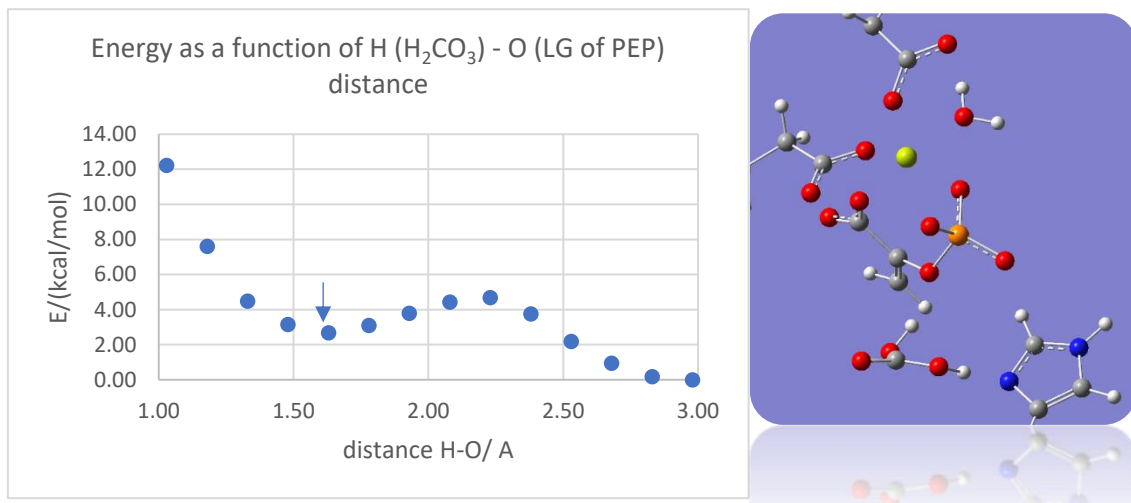
Fig. 36. **Linear transit scan, after the transfer of** *Fig. 35* **has occurred**. H of H_2CO_3 (the H previously received from His) was approached to the O of the LG of PEP. The structure of the marked point (arrow) was optimized, leading to the structure presented on the right side. On it, the H of the H_2CO_3 establishes an H-bond with the LG of PEP: angle O- H (H_2CO_3) – O (LG) = 158˚ and distance H (H_2CO_3) – O (LG) =1.61 Å.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |66

As this path seemed worthy, the structure of the inflection point in the scan of the (transfer of a proton from His to HCO$_3^-$; point with a reaction coordinate of 1.17 Å) was optimized for TS. Calculation converged to a structure and presented one negative frequency. However, when inspected using VMD and MolUP plug-in, the normal mode of such frequency was not the expected.

The linear transit scan of Fig. 36 derived from the scan of , so it was dependent on the transfer of H of His177 to HCO$_3^-$ to happen (whose TS was not found). This way, the hypothesis for this system was not further explored.

o   Pathway B

The initial QM/MM system with His177 double protonated, after optimization, underwent a linear transit scan, to approach the H of HCO$_3^-$ to an O of the phosphate group of PEP - O coordinated with Mg^{2+}. It displayed the potential energy profile of Fig. 37. One of the final steps was optimized, resulting in the structure shown in the same figure.
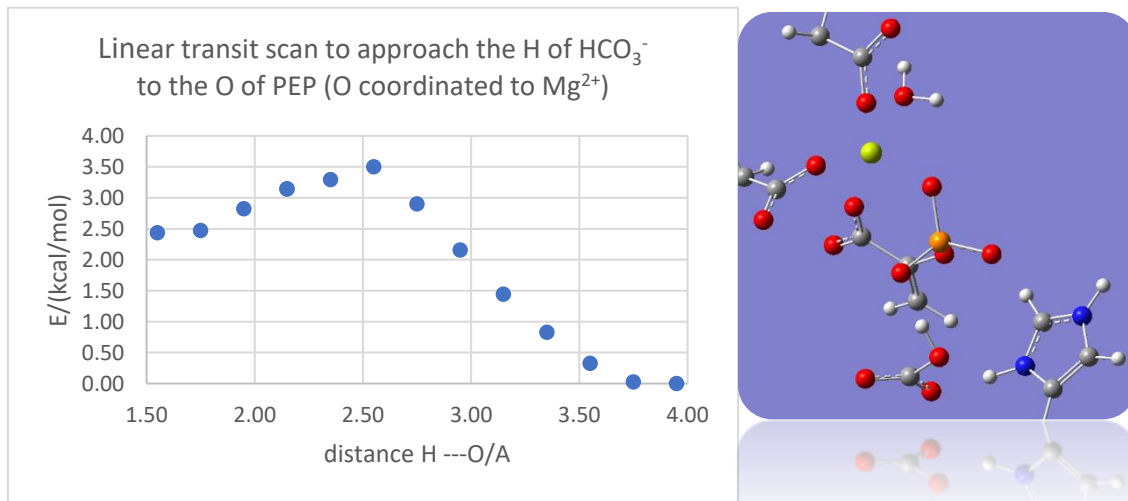
Fig. 37. **Linear transit scan ran to switch to a new conformation**, in which the H of HCO$_3^-$ interacts with the O of PEP that was previously coordinated to Mg^{2+}. After the scan, a geometry optimization was performed, resulting in the structure presented on the right side. On it, Mg^2 presents a coordination shell with 5 atoms.

As depicted above, Mg^{2+} switched to a penta-coordinated state. To allow it to recover the six-coordination state observed in many of the previous structures, the water of MM that was closest to the ion was unfrozen and included in QM. Despite it, after a new optimization, the Mg^{2+} retained its 5-coordination state.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |67

A linear transit scan approaching the H of His177 to the O of HCO$_3^-$ caused a rise
of ~9.5 kcal/mol and a scan starting from its last step and approaching the H of H$_2$CO$_3$
to the O of PEP (phosphate) showed a small rise in energy (2 kcal/mol) with inflection.
Their potential energy profiles are in *Fig. 38*, and the resulting structure is in Fig. 39.
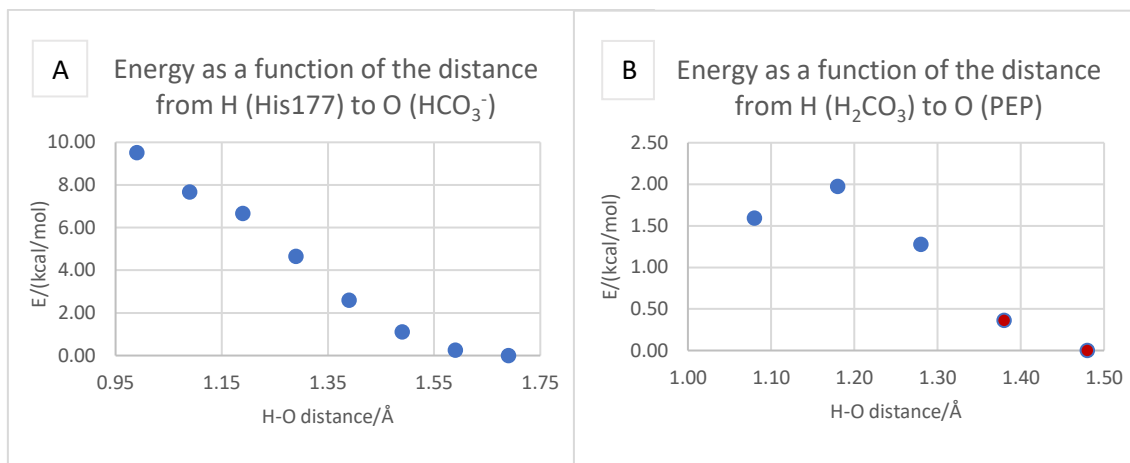
Fig. 38. **Linear transit scans performed in a system with His177 positive. A** – Approach of the H of His177 to the O
of HCO$_3^-$. **B** – Starting from the end of scan A, and approaching the H of H$_2$CO$_3$ to the O of the phosphate group of PEP.
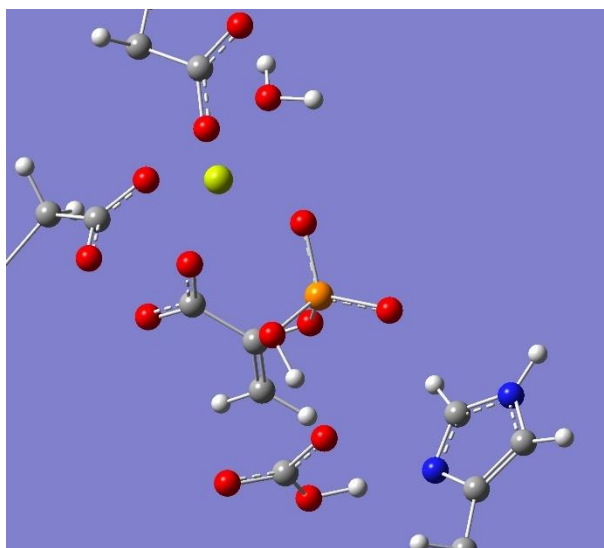
Fig. 39. **Structure after scans of Fig. 38**.

It is noteworthy the position of the P in the phosphate group relative to the O of
the HCO$_3^-$ – it seems favourable for an attack. Pathways using this structure could be
hypothesized.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |68

➢ 10. Other tests performed

1. As referred to in Methods, the coordination shell of the Mg^{2+} proposed in the article [10] was not the one used in the models of this work. In the article, the O of the supposed LG (the O that should detach from P) of PEP is coordinated to the ion (see Fig. 12). In opposition, the PEP analog crystallized in *E.coli* PEPC (that was used for building the models) had a C instead of the O of the leaving group of PEP (see *Fig. 3*). An attempt was made to test the coordination shell proposed in the article.

Using *GaussView,* after the exchanges of atom types required for turning the analog into the true substrate, PEP was modelled to coordinate Mg^{2+} in the way described in the article (*Fig. 40*, <u>left side</u>). However, after the minimization of the system, the O of the LG of PEP had left from the coordination shell, and another O of the phosphate replaced it (*Fig. 40*, <u>right side</u>).
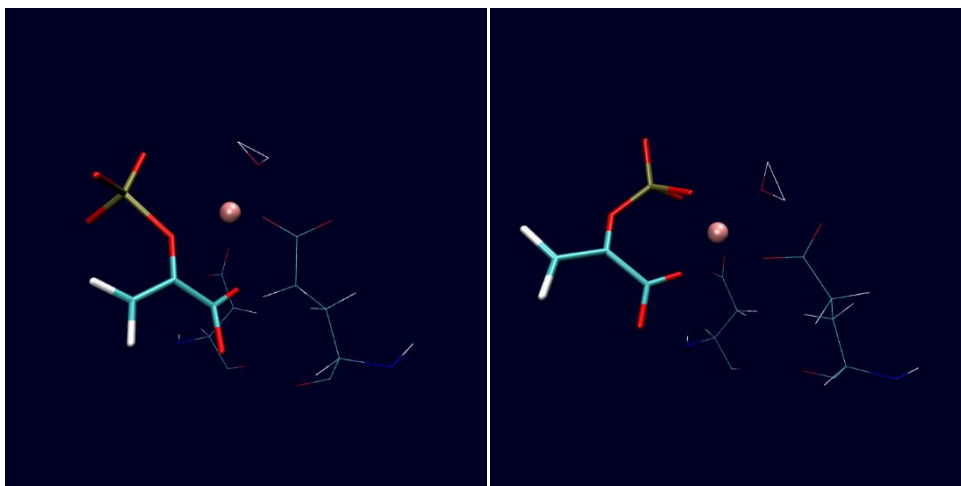
Fig. 40. **The coordination shell of Mg²⁺.** Left: PEP modelled to coordinate the ion by the atoms proposed in the article *[10]*. Right: system after minimizations.

2. The mechanisms of Fig. 13 didn't verify. Yet, starting from their initial steps whose TS were found, a different ending may be tested in the future.

3. A model was built with His177 as HIE, and MD simulations were ran using it. Bicarbonate seemed unstable. However, a QM/MM system build with a frame from it, after a geometry optimization with Mechanical Embedding, showed His177 interacting with the LG of PEP. In the future, models with this protonation state could be used.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |69

D. Conclusion

Using the model built, the mechanism proposed in the literature [10] was not verified - the direct attack from O of HCO$_3^-$ to P of PEP seems to carry a high energy barrier. Neither the addition of water to stabilize the leaving group, nor the removal of a water molecule to allow conformational adjustment were enough to bring the energy to feasible values. None of the other alternative reaction pathways exploited was able to explain the reaction of PEPC.

Despite what is stated above, the mechanism proposed in literature may yet be possible in conformations other than the ones employed in this work.

An alternative mechanism was exploited in this work and their first three steps were characterized. The highest of the energetic barriers on them was evaluated as 15.2 kcal/mol, which is lower than the experimental activation energy reported in one article [24]: 16.3 kcal/mol. This way, the reaction may have a mechanism that starts with the first 3 steps of the alternative mechanism described.

A summary scheme of the explored pathways is presented in the next page.

Further calculations for alternative reaction paths are required to establish the reaction path of PEPC; experimental data that could provide hints about the mechanism would also be desirable. Studies that evaluate the protonation state of His177, both at pH employed in the kinetic studies reported in the literature, and at physiological pH of plants using PEPC in carbon fixation, might also be valuable.

The reaction mechanism of PEPC is not established. Achieve it would fill a gap in the understanding of the enzyme; additionally, once it is established, further simulations may begin: simulations for biomimetics, simulations of conditions that allow or hamper reaction to take place, and assessment of the efficiency that an immobilized enzyme could achieve. For all of these, the mechanism of this enzyme deserves further research.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO₂ capture device |70

References

[1] D. Woolf et al., "Sustainable biochar to mitigate global climate change," *Nature Communications,* 2010.

[2] S. Budinis, "Going carbon negative: What are the technology options?," *IEA,* 2020.

[3] S. Budinis, "IEA Reports, Direct Air Capture," 2020.

[4] K. S. Chang e e. al., "Conversion of carbon dioxide to oxaloacetate using integrated carbonic anhydrase and phosphoenolpyruvate carboxylase," *Bioprocess and Biosystems Engineering,* 2013.

[5] B. F. Gruenbaum et al., "Blood glutamate scavenging as a novel glutamate-based therapeutic approach for post-stroke depression," *Therapeutic Advances in Psychopharmacology,* 2020.

[6] R. Wedding et al., "Kinetic studies of the form of substrate bound by phosphoenolpyruvate carboxylase," *Plant Physiology,* 1988.

[7] Y. Kai et al., "Three dimensional structure of phosphoenolpyruvate carboxylase: a proposed mechanism for allosteric inhibition," *Proceedings of the National Academy of Sciences of the United States of America,* 1999.

[8] K. Izui et al., "Phosphoenolpyruvate carboxylase: a new Era of structural biology," *Annual Review of Plant Biology,* 2004.

[9] Y. Kai et al., "Phosphoenolpyruvate carboxylase: three dimensional structure and molecular mechanisms," *Archives of Biochemistry and Biophysics,* 2003.

[10] H. Matsumura et al., "Crystal structures of C4 form maize and quaternary complex of E. coli phosphoenolpyruvate carboxylases," *Structure,* 2002.

[11] L. González-Segura et al., "Identification of the allosteric site for neutral amino acids in the maize C4 isozyme of phosphoenolpyruvate Carboxylase: the critical role of Ser100," *Journal of Biological Chemistry,* 2018.

[12] D. Schlieper, K. Förster, J. K. Paulus and G. Groth, "Resolving the activation site of positive regulators in Plant Phosphoenolpyruvate carboxylase," *Molecular Plant,* 2014.

[13] A. Takahashi-Terada et al., "Maize phosphoenolpyruvate carboxylase. Mutations at the putative binding site for glucose 6-phosphate caused desensitization and abolished responsiveness to regulatory phosphorylation," *2005.*

[14] H. Maruyama et al., "The Enzymatic Carboxylation of Phosphoenolpyruvate I. Purification and properties of phosphoenolpyruvate carboxylase," *The Journal of Biological Chemistry,* 1966.

[15] K. Terada and K. Izui, "Site-directed mutagenesis of the conserved histidine residue of phosphoenolpyruvate carboxylase His138 is essential for the second partial reaction," *European Journal of Biochemistry,* 1991.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |72

[16] N. Fujita et al., "Reaction mechanism of phosphoenolpyruvate carboxylase. Bicarbonate-dependent dephosphorylation of phosphoenol-alpha-ketobutyrate," *Biochemistry,* 1984.

[17] D. H. Gonzalez and C. Andreo, "Carboxylation and dephosphorylation of phosphoenol-3-fluoropyruvate by maize leaf phosphoenolpyruvate carboxylase," *Biochemical Journal,* 1988.

[18] J. Janc et al., "Mechanistic studies of phosphoenolpyruvate carboxylase from Zea mays utilizing formate as an alternate substrate for bicarbonate," *Biochemistry,* 1992.

[19] M. O'Leary et al., "Kinetic and isotope effect studies of maize phosphoenolpyruvate carboxylase," *Biochemistry,* 1981.

[20] J. Janc et al., "Mechanistic studies of phosphoenolpyruvate carboxylase from Zea mays with (Z)- and (E)-3-fluorophosphoenolpyruvate as substrates," *Biochemistry,* 1992.

[21] D. Hansen and J. R. Knowles, "The Stereochemical Course at Phosphorus of the Reaction Catalyzed by Phosphoenolpyruvate Carboxylase," *The Journal of Biological Chemistry,* 1982.

[22] I. Rose et al., "Stereochemistry of the enzymatic carboxylation of phosphoenolpyruvate," *The Journal of Biological Chemistry,* 1969.

[23] M.-X. Wu and R. Wedding, "Temperature effects on Phosphoenolpyruvate carboxylase from a CAM and a C4-plant. A comparative study," *Plant Physiology,* 1987.

[24] X. Yin et al., "Temperature response of bundle-sheath conductance in maize leaves," *Journal of Experimental Botany,* 2016.

[25] W. Humphrey, A. Dalke e K. Schulten, "VMD - Visual Molecular Dynamics," *J. Molec. Graphics,* 1996.

[26] J. Wang et al., "Automatic atom type and bond type perception in molecular mechanical calculations," *Journal of Molecular Graphics and Modelling,* 2006.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO$_2$ capture device |73

[27] M. J. Frisch et al., "Gaussian 09, Revision D.01," *Gaussian Inc., Wallingford CT,* 2016.

[28] C. I. Bayly et al., "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model," *J. Phys. Chem.,* 1993.

[29] J. Wang et al., "Development and testing of a General Amber Force Field," *Journal of Computational Chemistry,* 2004.

[30] O. Allnér, L. Nilsson e A. Villa, "Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations," *J. Chem. Theory Comput.,* 2012.

[31] M. Olsson et al., "PROPKA 3: consistent treatment of internal and surface residues in empirical pKa predictions," *Journal of Chemical Theory and Computation,* 2011.

[32] C. Sondergaard et al., "Improved treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa values," *Journal of Chemical Theory and Computation,* 2011.

[33] A. Tovar-Méndez et al., "Physiological implications of the kinetics of maize leaf Phosphoenolpyruvate carboxylase," *Plant Physiology,* 2000.

[34] J. Maier et al., "ff14SB: Improving the accuracy of Protein Side chain and Backbone Parameters from ff99SB," *Journal of Chemical Theory and Computation,* 2015.

[35] W. Jorgensen et al., "Comparison of simple potential functions for simulating liquid water," *Journal of Chemical Physics,* 1998.

[36] R. J. Loncharich, B. R. Brooks e R. W. Pastor, "Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide," *Biopolymers,* 1992.

[37] D. R. Roe e T. E. Cheatham, "PTRAJ and CPPTRAJ: Software for processing and Analysis of Molecular Dynamics trajectory data," *Journal of Chemical Theory and Computation,* 2013.

[38] A. Horn, "A consistent force field parameter set for zwitterionic amino acid residues," *Journal of Molecular Modeling,* 2014.

FCUP
Understanding Phosphoenolpyruvate Carboxylase catalysis as a step to the
engineering of a CO_2 capture device |74

[39] H. Fernandes, M. J. Ramos and N. Cerqueira, "molUP: A VMD plugin to handle QM and ONIOM calculations using the gaussian software," *Journal of Computational Chemistry,* 2018.

[40] M. Svensson et al., "ONIOM: a multilayered MO+MM Method for Geometry Optimizations and Single Point Energy Predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)3)2 + H2 Oxidative addition," *The Journal of Physical Chemistry,* 1996.

[41] C. Lee, W. Yang and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B,* 1988.

[42] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, "Ab Initio calculation of Vibrational Absorption and circular dichroism spectra using density functional force fields," *Journal of Physical Chemistry,* 1994.

[43] A. Becke, "A new mixing of Hartree-Fock and local density-functional theories," *Journal of Chemical Physics ,* 1993.

[44] C. Lee, W. Yang e R. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B,* 1988.

[45] S. H. Vosko, Wilk, L., Nusair e M., "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *Can. J. Phys.,* 1980.

[46] P. J. Stephens, F. J. Devlin, C. F. Chabalowski e M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields," *J. Phys. Chem.,* 1994.

[47] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. M. Jr, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, "A second generation force-field for the simulation of proteins, nucleic-acids, and organic molecules," *Journal of American Chemical Society,* 1995.

[48] S. Grimme, S. Ehrlich and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *Journal of computational chemistry,* 2011.

[49] K. Fukui, "The path of chemical reactions - the IRC approach," *Accounts of Chemical Research,* 1981.

[50] C. Dykstra, G. Frenking, K. Kim and G. Scuseria, Theory and Applications of Computational Chemistry: The First 40 years, Amsterdam: Elsevier, 2005.

[51] D. S. Williams et al., "Oxaloacetate supplementation increases lifespan in Caenorhabditis elegans through an AMPK/FOXO-dependent pathway," *Aging Cell,* 2009.

[52] A. Iglesias and C. S. Andreo, "On the Molecular Mechanism of Maize Phosphoenolpyruvate carboxylase Activation by Thiol Compounds," *Plant Physiology,* 1984.