Juliana Ferreira Rocha
Computational mechanistic studies in the synthesis of polyhydroxylated lactones obtained from a D-erythrosyl 1,3-dioxane template

UMinho|2017

Universidade do Minho
Escola de Ciências

Juliana Ferreira Rocha

Computational mechanistic studies in the synthesis of polyhydroxylated lactones obtained from a D-erythrosyl 1,3-dioxane template

julho de 2017

Universidade do Minho
Escola de Ciências

Juliana Ferreira Rocha

Computational mechanistic studies in the synthesis of polyhydroxylated lactones obtained from a D-erythrosyl 1,3-dioxane template

Dissertação de Mestrado
Mestrado em Química Medicinal

Trabalho realizado sob orientação do
Doutor Nuno Sousa Cerqueira
e da
Doutora Maria Jose Chão Alves

julho de 2017

DECLARAÇÃO

Nome: Juliana Ferreira Rocha

Endereço eletrónico: julianafdrocha@gmail.com

Número do cartão de cidadão: 14512201 8 ZY4

Título da dissertação: Computational mechanistic studies in the synthesis of polyhydroxylated lactones obtained from a D-erythrosyl 1,3-dioxane template

Orientadores: Doutor Nuno Sousa Cerqueira
Doutora Maria José Chão Alves

Ano conclusão: 2017

Designação do Mestrado: Mestrado em Química Medicinal

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE

Universidade do Minho, __ de __________ de 2017

Assinatura: __________________________

Para a minha avó Ana Joaquina

Acknowledgments

Esta é, talvez, a parte mais difícil desta tese, a parte menos exata. É um momento de reflexão, em que agradecer numas linhas a quem contribuiu para o meu sucesso, quer direta ou indiretamente, é pouco e apenas uma simbólica retribuição.

Em primeiro lugar, um grande agradecimento ao Professor Nuno Cerqueira e à Professora Maria José Alves pela oportunidade de trabalhar com eles que muito me enriqueceu, quer a nível profissional como pessoal, por toda a disponibilidade, apoio e simpatia demonstrada no decorrer do projeto. Em tudo na vida, os imprevistos acontecem e esta tese não foi exceção. Mas, devo-lhes um agradecimento especial por terem sempre arranjado uma alternativa viável que não permitisse que eu saísse prejudicada.

Aos meus colegas do Departamento de Química da Universidade do Minho por toda a simpatia vivida durante os curtos meses que aí estive, em que a boa disposição e o bom ambiente ao lanche e almoço era garantido. Em especial, aos meninos dos Laboratórios 6 e 15 que eram a companhia de todas as horas!

Aos meus colegas da Química Teórica do Departamento de Química e Bioquímica da Universidade do Porto, que tão bem me receberam e integraram quer à hora de almoço como nos profissionais jogos de basquetebol. Em especial, à Carla, ao Henrique, ao Marco e à Elizabeth pela boa disposição (e bolo!!) constante na sala 3.26.

À minha família pelo apoio incondicional e força que em muito contribuíram para o sucesso na minha vida académica. E, ainda, um especial obrigado aos meus pais que lutam diariamente para me dar a oportunidade de perseguir os meus sonhos.

Aos meus amigos, por todas as jantaradas e momentos de convívio que ajudam a relaxar aquando do stress que uma tese envolve.

Por último, mas não menos importante, não posso deixar de agradecer ao meu Pedro que surgiste no momento mais inesperado, mas como sabes, "és o meu euromilhões!". Aturaste muitas palermices e frustrações minhas sempre com um sorriso e pronto para me ajudar. Não sou pessoa de pedir muito, mas que eu saiba ser para ti, pelo menos metade daquilo que és para mim.

Abstract

Over time, the interest in studying molecular systems using computational methods has increased. These methods allowed to interpret and organize experimental results in order to solve or suggest new procedures to respond to a certain problem.

Synthesis of polyhydroxylated sugar-like compounds is difficult for organic chemists due to the necessary *stereo*- and *regio*- selectivity in the reactions. Monosaccharides are commonly employed as starting materials, but a multistep protection/deprotection manipulations are big disadvantage issues. The D-erythrose has been known for a long time but the poor chiral induction offered by this template prevent its large use in these types of reaction. Recently it was possible to obtain, a derivative of D-erythrose, the D-erythrosyl δ-lactone, that showed total facial selectivity in 1,3-dipolar cycloadditions and nucleophilic additions.

Two types of sugar-like scaffolds (five- and six-membered lactones ring) are readily accessed from nucleophilic attack to D-erythrosyl δ-lactone template. Both templates are obtained through a *stereo*- and *regio*-selective synthetic route and from which five- and six-membered lactones are obtained. Theoretical and computational means revealed that the reactions leading to both lactones share a common pathway but diverge after the second nucleophilic attack that is also the rate-limiting step in one of the reactions. The synthesis of five-membered lactone requires the presence of hard-nucleophiles and the full pathway requires 4 steps. The six-membered lactones require the presence of soft-nucleophiles and full catalytic process requires only 1 step. The origin of the *stereo*-selectivity is justified by an intrinsic characteristic of D-erythrosyl δ-lactone that forces the nucleophilic attack to occur only at one face of the lactone. The *regio*-selectivity is dependent on the nature of the attacking molecule, i.e, if hard or soft nucleophiles are used in the reactions.

The inhibitory activity of D-erythrose derivatives (five- and six-membered lactones) were tested against α-glucosidase, β-glucosidase, α-galactosidase and β-galactosidase. The biological assay revealed that some of the tested compounds present high specificity for some of these enzymes, an important feature in the development of new drugs targeting glycosidases.

Molecular docking studies with human-enzymes were also carried out to study the origin of the specificity of some of these compounds to these enzymes. The results indicate that the factor that plays a major role in the inhibitory activity of the D-erythrose derivatives is the position of the sugar-like moiety inside the binding pocket. When an efficient interaction is obtained with this part of the molecule and the active site of the enzyme, the compounds present satisfactory inhibitory activity otherwise, they are inactive. In α-glucosidase and β-galactosidase, this sort of interaction is possible with some of the tested compounds and these are the only ones that present inhibitory activity. In β-glucosidase and α-galactosidase all the compounds are inactive because the active site is located at the bottom of the binding pocket and none of the compounds can interact with it with sugar-like moiety.

Keywords: α (alpha), β (beta), glucosidase, galactosidase, D-erythrose derivatives, quantum mechanics, mechanism, *stereo*-selectivity, *regio*-selectivity, biological assays, molecular docking.

Resumo

Ao longo do tempo, o interesse em estudar sistemas moleculares usando métodos computacionais aumentou. Estes métodos permitiram interpretar e organizar resultados experimentais para resolver ou sugerir novos procedimentos para responder a um determinado problema.

A síntese de compostos semelhantes a açúcares polihidroxilados é difícil para os químicos orgânicos devido à necessária *estereo-* e *regio*-seletividade das reações envolvidas. Os monossacarídeos são comumente empregados como materiais de partida, mas os múltiplos passos de proteção/desproteção necessários são uma grande desvantagem apontada a estes métodos. A D-eritrose é conhecida há muito tempo, mas a baixa indução quiral oferecida por este modelo impede o seu uso em maior escala. Recentemente, foi possível obter um derivado de D-eritrose, a D-eritrosil δ-lactona, que mostrou total seletividade facial em ciclo-adições 1,3-dipolares e adições nucleofílicas.

Dois tipos de esqueletos semelhantes a açúcar (anel de lactona de cinco ou seis membros) são facilmente obtidos através do ataque nucleofílico ao modelo de D-eritrosil δ-lactona. Ambos os modelos são obtidos através de um método sintético *estereo-* e *regio*-seletivo, mas o mecanismo envolvido não era conhecido. Os resultados obtidos por métodos teóricos e computacionais revelaram que ambas as lactonas partilham a mesma via mecanística, mas divergem no segundo ataque nucleofílico que é o passo limitante numa das reações. A síntese do anel de lactona de cinco membros requer a presença de nucleófilos *hard* e exige 4 passos. As lactonas de seis membros requerem a presença de nucleófilos *soft* e a reação fica completa em apenas 1 passo. A origem da estereo-seletividade é justificada por uma característica intrínseca da D-eritrosil δ-lactona que força o ataque nucleofílico a ocorrer apenas por uma das faces. A regio-seletividade depende da natureza do nucleófilo, isto é, se são utilizados nucleófilos *hard* ou *soft* nas reações.

A actividade inibidora dos derivados de D-eritrose (lactonas de cinco e seis membros) foi testada contra α-glucosidase, β-glucosidase, α-galactosidase e β-galactosidase. Os ensaios biológicos revelaram que alguns dos compostos testados apresentam alta especificidade para algumas dessas enzimas, uma característica importante no desenvolvimento de novos fármacos com glicosidases como alvo.

Estudos de Docking Molecular com enzimas humanas foram realizados para avaliar e perceber a especificidade que alguns desses compostos apresentaram para algumas dessas enzimas. Os resultados indicam que o fator que desempenha um papel importante na atividade inibitória dos derivados de D-eritrose é a posição do anel que mimetiza o açúcar dentro da bolsa de ligação. Quando uma interação eficiente é obtida com esta parte da molécula e o centro ativo da enzima, os compostos apresentam atividade inibitória satisfatória, caso contrário eles são inativos. Nas α-glucosidase e β-galactosidase, esse tipo de interação é possível com alguns dos compostos testados e estes são os únicos que apresentam atividade inibitória. Na β-glucosidase e α-galactosidase todos os compostos são inativos porque o centro ativo está localizado na parte inferior da bolsa de ligação e nenhum dos compostos consegue interagir de forma eficiente.

Palavras-chave: α (alfa), β (beta), glucosidase, galactosidase, derivados de D-eritrose, mecânica quântica, mecanismo, *estereo*-seletividade, *regio*-seletividade, ensaios biológicos, Docking Molecular.

Preamble

This thesis is composed by four major chapter:

CHAPTER I. Introduction

In this chapter, a small review of what is known about glycosidases is made. It is briefly described its importance for the human being, as well as the organs in which it acts and which diseases they are involved in. Because they are important therapeutic targets, some drugs that are commercial available are described and the reasons for choosing D-erythrose derivatives to inhibit these enzymes are explained.

CHAPTER II. Theoretical Background

A brief review about the theoretical methods used during this work are described. Particular attention is given to quantum mechanics and molecular docking.

CHAPTER III. Results and Discussion

This chapter is divided in three sections:

1. Catalytic mechanism
2. Biological assays and Molecular Docking
3. Synthesis driven by theoretical results

The first two are organized as scientific articles, while in the third is provided a small discussion about the importance of theoretical and computational results in organic synthesis.

CHAPTER IV. Conclusions

In this chapter, the major conclusions of this work are presented.

CHAPTER V. Bibliography

List of all bibliographic references that were used. They are organized in the order they appeared in the text.

Index

ACKNOWLEDGMENTS .... **V**

ABSTRACT .... **VII**

RESUMO .... **IX**

PREAMBLE .... **XI**

INDEX .... **XIII**

ABBREVIATION LIST .... **XVII**

CHAPTER I - INTRODUCTION **1**

1. GENERAL ASPECTS 3
2. GLYCOSIDASES 3
   - *2.1. Reactions catalyzed by glycosidases* *4*
   - *2.2. Classification of glycosidases* *6*
   - *2.3. Glycosidases as Drug Targets* *7*
     - A. α-glucosidase 7
     - B. β-glucosidase 7
     - C. α-galactosidase 8
     - D. β-galactosidase 8
   - *2.4. Glycosidases Inhibitors* *8*
3. D-ERYTHROSE DERIVATIVES 11

CHAPTER II – THEORETICAL BACKGROUND **13**

1. INTRODUCTION 15
2. MOLECULAR DOCKING 15
   - *2.1. Search algorithm* *17*
     - 2.1.1. Genetic Algorithm methods 18
     - 2.1.2. Monte Carlo methods 18
   - *2.2. Scoring Function Algorithm* *19*
3. QUANTUM MECHANICS 22
   - *3.1. Schrödinger equation* *22*
   - *3.2. Hamiltonians* *23*
   - *3.3. Variational principle* *23*
   - *3.4. Born-Oppenheimer approximation* *24*
     - 3.4.1. Electronic Hamiltonian 25

3.4.1.1. Density-Functional Theory (DFT) ... 25
3.4.1.2. Basis Set ... 28
3.4.2. Nuclear Hamiltonian ... 29
3.5. Continuum models ... 30
3.6. Potential energy surface (PES) ... 31
CHAPTER III – RESULTS AND DISCUSSION ... 35
1. CATALYTIC MECHANISM ... 39
1. Introduction ... 41
2. Methodology ... 45
3. Results and Discussion ... 46
3.1. D-erythrosyl δ-lactone formation ... 47
3.2. Formation of the five-membered lactones (hard nucleophiles) ... 47
A. Pathways NA@C4 ... 48
B. Pathways NA@C2 ... 54
C. Pathways NA@C4 versus Pathways NA@C2 ... 58
D. Ring closure ... 60
3.3. Formation of the six-membered lactones (soft nucleophiles) ... 63
A. Pathway NA@C4 ... 65
1. Sulfur nucleophile ... 65
2. Hydroxylamine ... 67
4. Conclusions ... 72
2. BIOLOGICAL ASSAYS AND MOLECULAR DOCKING ... 75
1. Introduction ... 77
2. Methodology ... 79
2.1. Model enzymes ... 79
2.2. Biological Assays ... 80
2.3. Molecular Docking ... 81
2.3.1. Tunnel shape and SASA calculations ... 84
3. Results and Discussion ... 84
3.1. Biological Assays ... 84
3.2. Molecular Docking ... 86
A. α-glucosidase ... 86
B. β-glucosidase ... 88
C. α-galactosidase ... 89
D. β-galactosidase ... 90
4. Conclusion ... 92
3. SYNTHESIS DRIVEN BY THEORETICAL RESULTS ... 95
1. Introduction ... 97

2. Molecular docking studies addressed to lactams 100

A. α-glucosidase 100

B. Remaining enzymes.. 102

C. Conclusion 102

CHAPTER IV – CONCLUSION 105

CHAPTER V – REFERENCES 111

Abbreviation List

B3	Becke 3 Functional
B3LYP	Becke 3 and Lee, Yang and Parr Functional
CAZy	Carbohydrate-Active Enzymes Database
CGTO	Contracted Gaussian Type Functions
DFT	Density Functional Theory
DNJ	1-deoxy-nojirimycin
Ea	Activation Free Energy
Er	Reaction Free Energy
GGA	Generalized Gradient Approximation
GTO	Gaussian Type Functions
HF	Hartree-Fock
IEFPCM	Polarizable Continuum Model with the Integral Equation Formalism variant
IRC	Intrinsic Reaction Coordinate
IUBMB	International Union of Biochemistry and Molecular Biology
LDA	Local Density Approximation

LSDA	Local Spin-Density Approximation
LYP	Lee, Yang and Parr Functional
NJ	Nojirimycin
P	Product
PCM	Polarizable Continuum Model
PDB	Protein Data Bank
PES	Potential Energy Surface
R	Reagent
STO	Slater Type Functions
TS	Transition State

I

CHAPTER I

Introduction

If we knew what it was we were doing, it would not be called research, would it?

Albert Einstein (1879-1955)

1. General Aspects

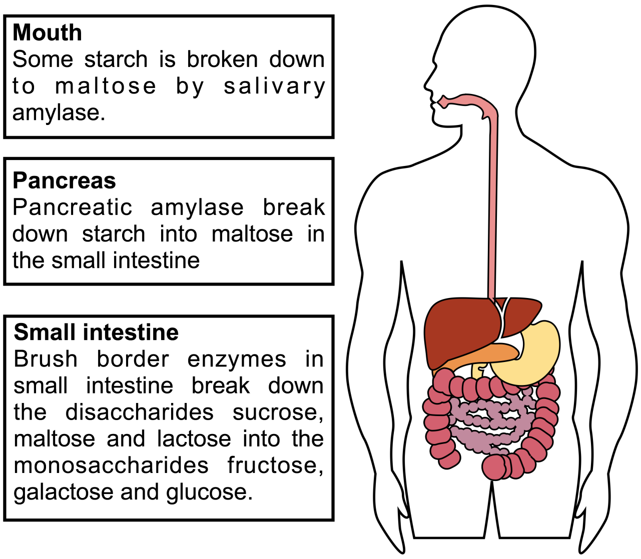
Carbohydrates are important biological molecules composed by carbon (C), oxygen (O) and hydrogen (H). They are the most abundant and structurally diverse class of biological compounds in nature and can be found from simple monomers to more complex oligomers, polymers or glycol-conjugates. This happens because, each sugar unit (monosaccharides) can accommodate multiple linkages and/or branches, and therefore several and different combinations among them can exist. In addition, and since each glycosidic linkage that connect two adjacent sugar units can have two anomeric configurations (α or β), this implies that even a small oligosaccharide can be quite complex [1].

Our current understanding about the biological importance of carbohydrates is still far behind from what is known regarding proteins and nucleic acids. In humans, carbohydrates are mainly used for energy storage molecules (e.g. starch and glycogen). They are digested into small pieces—either glucose or a sugar that is easily converted to glucose—that can be absorbed through the small intestine's walls. After a quick stop in the liver, glucose enters the circulatory system, causing blood glucose levels to rise.

2. Glycosidases

The digestion process of carbohydrates is not a straightforward process since the covalent bond that connects two sugar units in carbohydrates (glycosidic bonds) is one of the strongest bonds found in natural polymers [2]. These reactions can only be performed by very specialized enzymes that can break these bonds without requiring high energetic costs. These enzymes are called glycosidases and are among the most proficient enzymes found in nature because they are capable of breaking glycosidic bonds about 1017 times faster than the un-catalyzed reaction [1].

In humans, the digestion of carbohydrates involves several glycosidases. For example, the first step of the digestion of starch takes place in the mouth, where salivary amylase and mechanical motions come into play together. The next step takes place in the duodenum where the pancreatic amylase, that is released by the pancreas, decomposes the complex structure of starch into smaller units. The final step of hydrolysis of carbohydrates is performed on the surface of epithelial cells of the small intestine where other glycosidases, such as maltase, isomaltase, sucrase, lactase, among others hydrolyze these molecules into their corresponding monosaccharides. Only in this state, these molecules can be absorbed, carried out by the blood stream and enter the cells (Figure 1).



Mouth
Some starch is broken down to maltose by salivary amylase.

Pancreas
Pancreatic amylase break down starch into maltose in the small intestine

Small intestine
Brush border enzymes in small intestine break down the disaccharides sucrose, maltose and lactose into the monosaccharides fructose, galactose and glucose.

Figure 1 Schematic representation of the human digestive system and the location of some of the most important glycosidases.

2.1. Reactions catalyzed by glycosidases

The reactions that are catalyzed by glycosidases can be classified in two types: a retaining configuration mechanism, where the anomeric carbon has the same configuration in the reactant and product, and an inverting configuration mechanism where the configuration on the anomeric carbon is inverted (Figure 2A and Figure 2B).

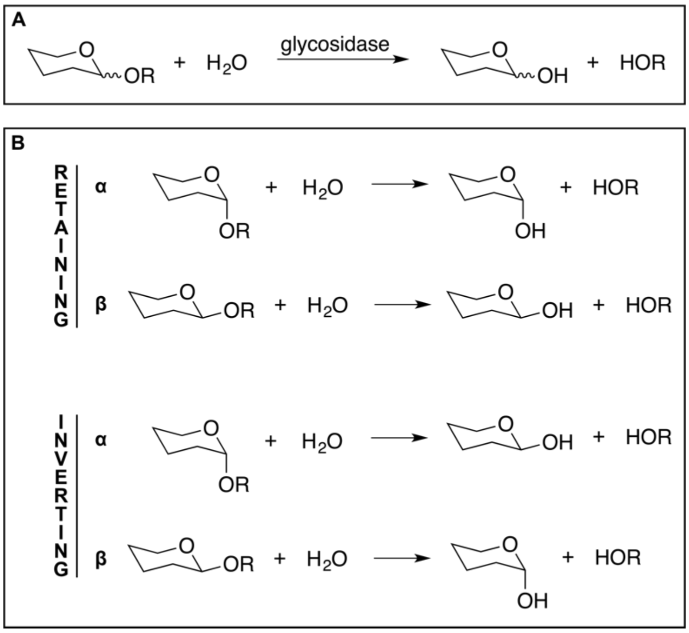


Figure 2 A. General reaction catalyzed by glycosidases. B. Schematic representation of retaining and inverting mechanism for both configurations (α or β).

Both mechanisms require the participation of two residues, Asp or Glu, that are in opposite sides of the active site of the enzyme. In the retaining mechanism, these residues are around 5.5 Å away from each other and the reaction happens in two steps. In the first step, one of the residues acts as a general acid and donates a proton to the sugar, while the second deprotonated residue functions as a nucleophile, attacking the anomeric carbon. This step, leads to the formation of a covalently linked glycosyl-enzyme intermediate that has an anomeric configuration opposite to that of the starting material. The second step of this reaction, involves the hydrolytic breakdown of the glycosyl-enzyme intermediate. The carboxylate that first behave as an acid catalyst, now acts as a base by abstracting a proton from the incoming nucleophile, usually through the participation of a water molecule. Simultaneously, the water molecule attacks the carbohydrate-enzyme linkage in a reverse mode of the first step. At the end of this reaction, the enzymatic turnover is obtained and a hemi-acetal is formed with the same anomeric configuration as the starting material [1 3].

In the inverting mechanism, the distance between the active site Asp or Glu is around 10.5 Å, which is larger than in the retaining mechanism. The differences on the distance between these two residues results from the necessity of accommodating a water molecule, between the residue that behaves as a base and the sugar, that is required for the catalytic process. In this case, the two residues act as general acid and base catalysts and the mechanism occurs in a single step. Just like in the retaining mechanism, one of the residues donates a proton to the sugar, but in the inverting mechanism, at the same time the other residue that is coordinated with the water molecule (that acts as a nucleophile), assists its deprotonation. In the end of this process, a hemi-acetal is formed owning the inverse anomeric configuration of the reactant [1 3].

In Figure 3, the retaining and inverting mechanism are schematically represented.

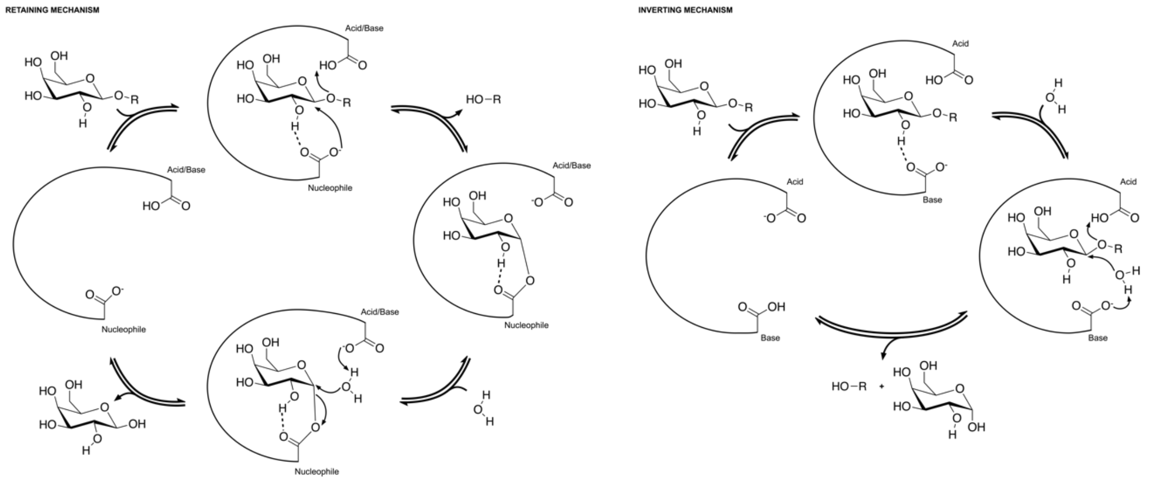


Figure 3 Schematic representation of retaining (left) and inverting (right) mechanisms of glycosidases. In this figure, it is illustrated a mechanism where the substrate has the anomeric carbon in the β configuration.

2.2. Classification of glycosidases

The Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) has classified glycosidases according to substrate specificity and/or molecular mechanism that is used to catalyze carbohydrates. For each enzyme, a specific code is attributed that is used to identify it. However, this classification is far from being perfect because, glycosidases can catalyze different reactions and therefore they can be classified in more than one EC group. To overcome this problem, a new classification for glycosidases was created based on amino acid sequence and folding similarities (Carbohydrate-Active Enzymes Database – CAZy http://www.cazy.org/) [4]. This new classification of the enzymes continues to have some drawback, like enzymes that catalyze distinct reactions belong to the same family/clan, but it makes the identification of the enzymes clearer and more precise.

2.3. Glycosidases as Drug Targets

The reactions that are catalyzed by glycosidases are indispensable for the normal functioning of most organisms. The malfunction of any of these enzymes can be the source of several diseases or its inhibition can be the cure of some of them. In this regard the α-glycosidase, β-glucosidase, α-galactosidase and β-galactosidase, have been receiving a lot of attention in the last decade.

A. α-glucosidase

α-glucosidase (EC 3.2.1.20) is responsible for the cleavage of 1,4-α-bonds of starch and disaccharide into α-D-glucose and can be found in the brush border of the small intestine and lysosomal deposits. Lysosomal α-glucosidase is essential for the degradation of lysosomal deposits of glycogen. Any dysfunction of this enzyme often leads to the Pompe disease [5]. α-glucosidase present in the brush border of small intestine participates in the degradation of glycogens and some types of carbohydrates. Inhibition of this enzyme delays carbohydrate digestion and therefore can be used in the treatment of type II diabetes [6].

B. β-glucosidase

β-glucosidases (EC 3.2.1.21) is a glycosidase that can be found in the liver, spleen and kidney. These enzymes are responsible for the hydrolysis of the glycosidic bonds to terminal non-reducing residues in β-D-glucosides and oligosaccharides, with the release of β-D-glucose. Malfunction of this enzyme is often related to glucosyl residues accumulation and it is currently being used as a successful biomarker for Gaucher's disease [7 8 9].

C. α-galactosidase

α-galactosidase (EC 3.2.1.22) is a lysosomal enzyme responsible for the hydrolysis of α-galactosyl terminals from glycolipids and glycoproteins, but mainly ceramide trihexoside. Insufficient levels of this enzyme results in an accumulation of this carbohydrate that often leads to the appearance of Fabry's disease [10 11]. This disease is characterized by an excessive deposition of neutral glycosphingolipids in several organs leading to their malfunction.

D. β-galactosidase

β-galactosidase (EC 3.2.1.23 is present in the small intestine and it is responsible for the breakdown of lactose into glucose and galactose. Lactose intolerance is a common disease associated with malfunction of these enzymes [12 13].

2.4. Glycosidases Inhibitors

Due to the therapeutic opportunity that is brought by the inhibition of glycosidases in the treatment of several diseases, different types of inhibitors have been developed targeting these enzymes. Most of the glycosidase inhibitors are carbohydrate molecules. Generally, one atom in the monosaccharide main scaffold is changed, which allows these compounds to disguise themselves and be accepted by the cells as normal substrates but, once they interact with glycosidases they specifically and efficiently inhibit their activity. Currently these compounds have been classified into five classes, accordingly to their structures: disaccharides, iminosugars, carbasugars, thiosugars and non-glycosidic [2].

Disaccharides occur naturally in nature. Two examples from this class of compounds are kojibiose and nigerose (Figure 4). These compounds are specific α-glucosidase inhibitors.

In iminosugars, carbasugars and thiosugars the oxygen of the pyranose ring is replaced by a different atom, a nitrogen (N), a carbon (C) and a sulfur (S), respectively. These compounds occur in nature but in different abundancies. Thiosugars are not very abundant in nature but two recognized examples are salacinol and kotalanol (Figure 4), that are two potent inhibitors of α-glucosidase. In contrast, carbasugars are highly found in nature and conduritols and cyclophellitols (Figure 4) are glycosidase inhibitors. Iminosugars are widespread in nature and the first one discovered was nojirimycin (NJ), that inhibits α- and β-glucosidase. Later, 1-deoxy-NJ (DNJ) was discovered and was found to be a potent inhibitor of α-glucosidase.

The glycosidase inhibitors described so far are based on a sugar scaffold however, there are also glycosidase inhibitors that are not based on that chemical structure. For example, tetrachlorophthalimide derivatives show a potent reversible inhibitory activity.

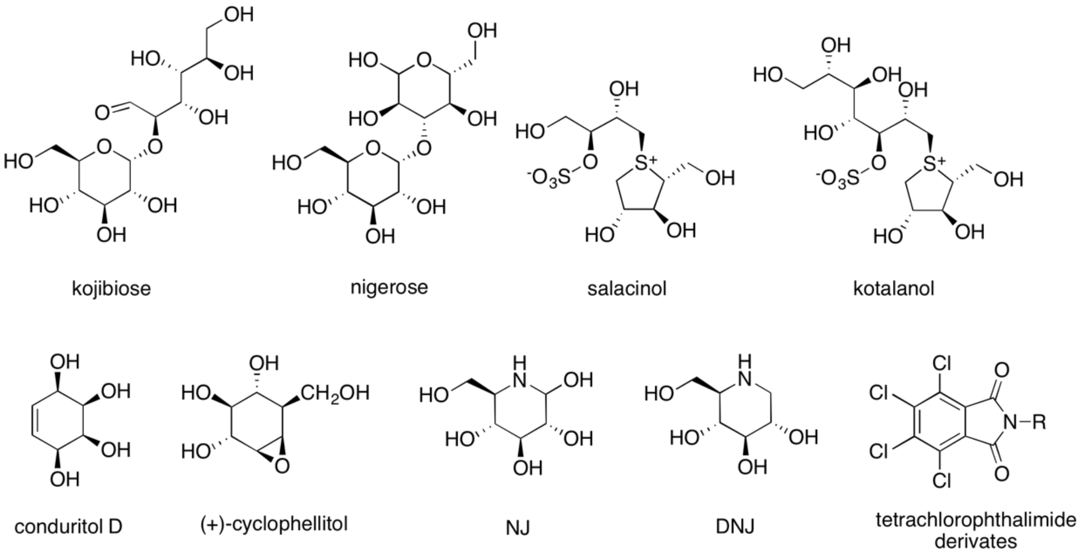


Figure 4 Natural compounds that exhibit potential inhibitory activity against glycosidases.

Among all the glycosidases inhibitors that were described before, the iminosugars derivatives are perhaps the most promising ones. Iminosugars, also called azasugars, are mono or bicyclic polyhydroxylated alkaloids and they are, essentially sugars in which the endocyclic oxygen is replaced by a nitrogen atom [14] [15]. Some of these compounds are already being used clinically in the treatment of some diseases. This is, for example, the case of Glyset and Zavesca (Figure 5) that are already commercialized for the treatment of diabetes and Gaucher's disease, respectively [16]. Amigal is also in clinical trials for the treatment of Fabry's disease.

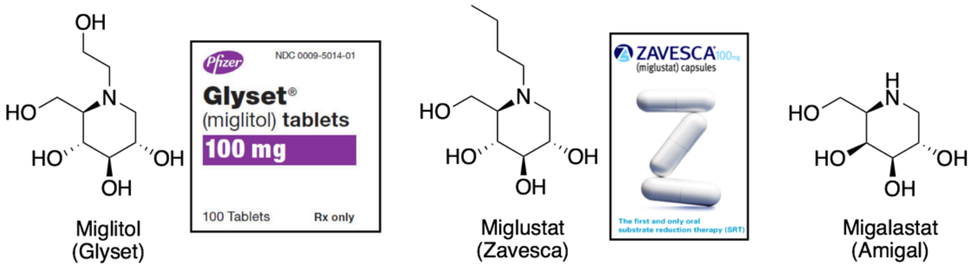

Miglitol
(Glyset)
Pfizer
NDC 0009-5014-01
Glyset®
(miglitol) tablets
100 mg
100 Tablets
Rx only
Miglustat
(Zavesca)
ZAVESCA 100mg
(miglustat) capsules
The first and only oral substrate reduction therapy (SRT)
Migalastat
(Amigal)

Figure 5 Glycosidase inhibitors used or in clinical trial for the treatment of some diseases. Glyset is used for diabetes, Zavesca for Gaucher disease and Amigal is on trial for Fabry disease.

Although these compounds have been successfully used in the treatment of a variety of diseases, many of them present a lack of specificity. This means that one inhibitor that has been, for example, developed to target α-glucosidase can also be a substrate of another glycosidase. This often leads to the appearance of several side effects, such as gastrointestinal tract distress [17] that prevent an effective treatment. Besides preventing an effective treatment, it can potentiate the development of symptoms from another disease, caused by malfunction of other glycosidase, as a side effect. The lack of specificity of these compounds is also one of the main causes of the early withdraw of these compounds from a drug development process.

3. D-erythrose derivatives

Preliminary studies in our laboratory have demonstrated that D-erythrose derivatives (Figure 6) are interesting and promising glycosidases inhibitors. These structures have in their core structure a six-membered lactone ring that is very similar to the structure of iminosugars but, have a carbonyl group attached to the anomeric carbon instead of a hydroxyl group. This carbonyl group has been described as an advantage since it mimics the transition state structure involved in the glycosidic bond cleavage on the wild type enzymes. This feature also allows the compounds to be substrates of either α- and β-glycosidases, overcoming the limitation of the substrate specificity imposed by the orientation of the hydroxyl group attached to the anomeric carbon.

Additionally, the five-membered lactone ring derivatives are a new scaffold that is worth to investigate. These compounds are a new opportunity to find a new scaffold that can be specific for one of the enzymes under study.

With the preliminary studies and the need to search for new and specific inhibitors of glycosidases, D-erythrose derivatives will be the main subject of this thesis.

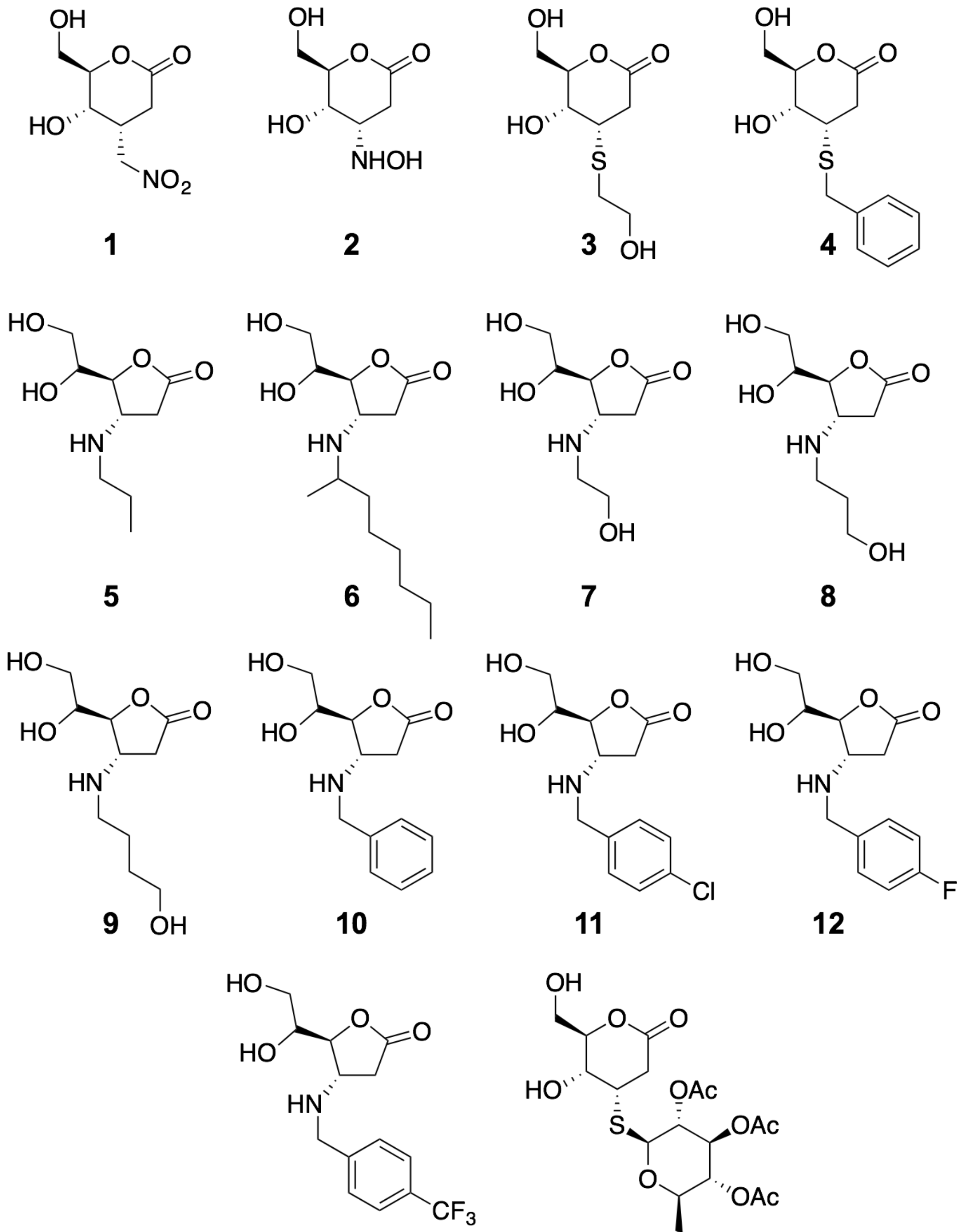

1
2
3
4
5
6
7
8
9
10
11
12
13
14

Figure 6 *D-erythrose derivatives in study.*

II

CHAPTER II

Theoretical Background

Both the man of science and the man of action live always at the edge of mystery, surrounded by it.

J. Robert Oppenheimer (1904-1967)

1. Introduction

Over time, the interest in studying molecular systems using computational methods has been increasing. This has been partially supported by the development of better algorithms, and optimized methodologies that allow to tackle with a high degree of accuracy, and with atomic detail, the cleavage and formation of chemical bonds, that is the genesis of chemistry. Currently, such knowledge is very difficult to obtain by experimental means (if not impossible) and therefore these methods offer new tools that can be used to re-interpret experimental results that are poorly understood or even to predict new data that has not been yet observed or measured. Such evolution of the computational methods has also been favored by the development of computers, whose performance during the last decades has made great progresses.

Currently, there is a plethora of computational methods that are already well established in the literature. These methods are often specific for the problem that is intended to solve. In the following section a brief review about these methods is done. Based on the work that was developed in this thesis, particular attention will be given to: molecular docking methods and quantum mechanics. The first are commonly used to predict the binding pose of small molecules in relation to a macromolecule. The quantum mechanics are more robust methods that allow to solve the Schrodinger equation and extract quantum information about the molecular system.

2. Molecular Docking

The field of molecular docking has emerged during the last decade driven by the needs of structural biology and structure based drug discovery [18]. The definition of molecular docking is a methodology that tries to predict the structure of intermolecular complexes formed between two or more constituent molecules, and that allows to distinguish, from an energetic point of view, complexes and/or forms of coordination that those molecules can adopt [19]. Generally, molecular docking methodologies can be divided in protein-protein docking and protein-ligand docking. In this work, the methodology that was used was the protein-ligand docking. This methodology tries to predict position and orientation of a small molecule, called ligand, when it is bound to a protein [20]. The docking protocol explore all the degrees of freedom of the ligand through a search algorithm and then, ranks and orders them by a score, that is obtained by a score function algorithm (Figure 7). Ideally, this quantitative value would correlate with the free energy of binding but due to the huge computational cost required, some simplifications are imposed to turn the methodologies faster, despite an accuracy cost. However, protein-ligand docking is quite reliable, and pharmaceutical research employs it to in everyday tasks to enhance the development of new drugs [19 21, 22 23].

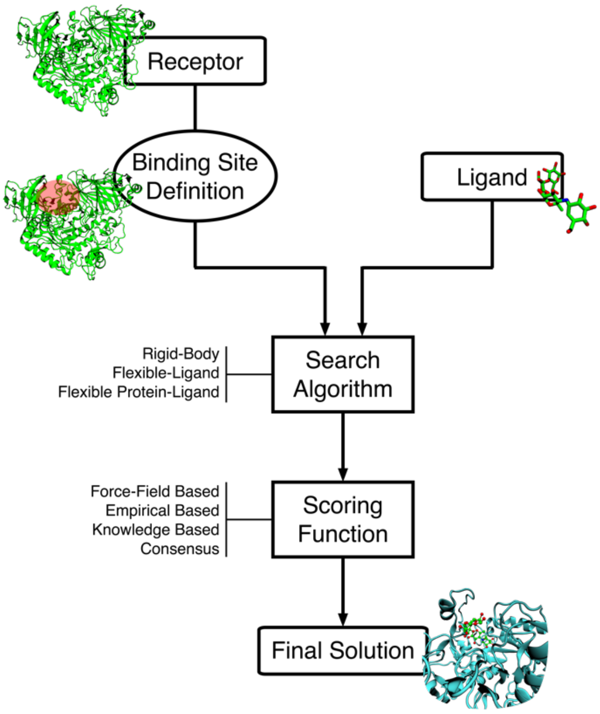

Receptor
Binding Site Definition
Ligand
Rigid-Body
Flexible-Ligand
Flexible Protein-Ligand
Search Algorithm
Force-Field Based
Empirical Based
Knowledge Based
Consensus
Scoring Function
Final Solution

Figure 7 Schematic representation of the molecular docking process.

The molecular docking algorithms are often characterized by two types of algorithms: a search algorithm and the scoring functions. Each of these methods are described briefly in the following sections.

2.1. Search algorithm

The search algorithm is a process where all the binding poses of the ligand and the receptor are generated. These algorithms can be divided into rigid, semi-flexible or full-flexible algorithms [24]. The rigid docking is, like suggested by the name, a search where both ligand and receptor are kept rigid during the molecular docking protocol. This means that they can only be rotated, translated against each other in order to find the best binding poses. This algorithm is based on the lock and key model. Although it is very fast, it is not very accurate. In the semi-flexible ligand model the receptor is maintained fixed during the molecular docking protocol while the ligand is flexible. The term flexibilized is employed to describe the rotation of all the single bonds that exist in the molecule. This fact improves the accuracy of the calculation but it is still inappropriate when large conformation rearrangements are involved in the protein. In the full-flexible search algorithms both the receptor and the ligand are flexibilized. This is the most accurate approximation however, it is computational demanding and therefore time expensive.

The most commonly employed search algorithms employed to study the protein ligand interactions are the semi-flexible search algorithms, since they provide the best compromise between accuracy and computational time. In these methods, the flexibilization of the ligand can be accomplished in several ways. In a systematic way, that tries to explore all the degrees of freedom in a molecule, and in a stochastic/random search, where the conformational space is sampled through random changes in the conformation of the ligand, which are accepted or rejected in agreement with a probability function that is previously defined. Inside the stochastic search algorithms, there are three types: Monte Carlo methods, genetic algorithm methods, and Tabu search methods. The genetic algorithm and the Monte Carlo methods have been employed with great success in a variety of problems and their used has become very popular in molecular docking.

2.1.1. Genetic Algorithm methods

This method flexibilizes the structure of the ligand based on the rules derived from genetics and biological evolution (Figure 8). They start from an initial population of different translations and orientations as well as conformations of the ligand in relation with the receptor. To this set of conformations are applied mutations, crossovers and migration (genetic operators) until a final population that optimizes a predefined fitness function is reached. The conformations that do not assemble a positive development are excluded and the process is repeated until a final population that contains the best conformations of the ligand [25].

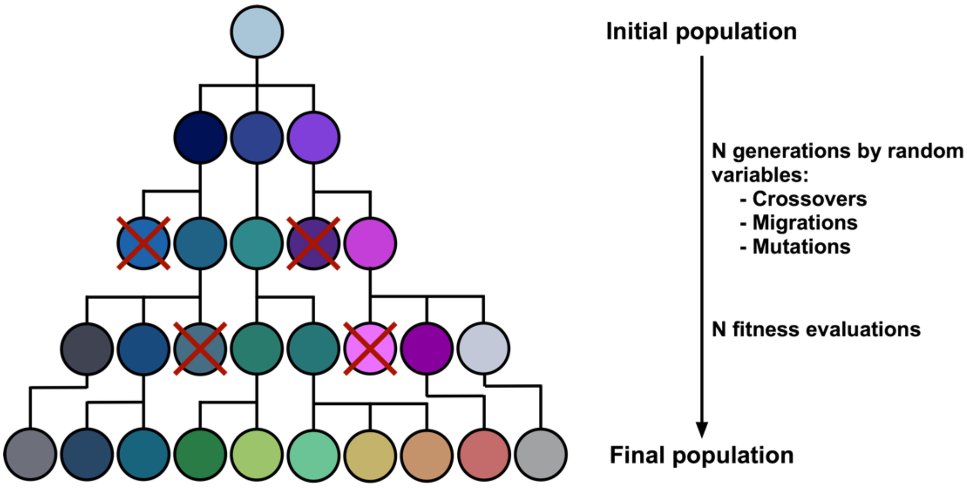


Figure 8 Scheme of search process applied by genetic algorithm method.

2.1.2. Monte Carlo methods

In this approach, we start with an initial conformation A and then a second conformation B is defined by taking random steps which are accepted or not based on a Boltzmann probability function and until a certain number of steps have been tried. In Figure 9 is illustrated a scheme of the Monte Carlo Method of search, were the different conformations (A to E) are in different energetic places [26].

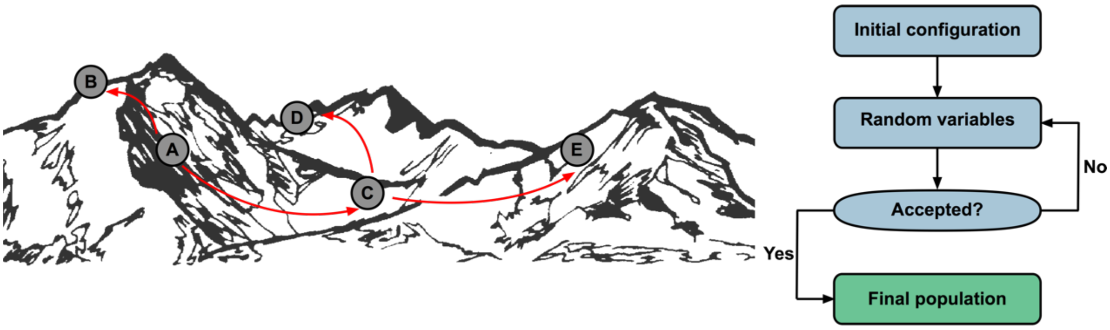


Figure 9 Scheme of search process applied by Monte Carlo method.

This method is very efficient to step energy barriers, allowing to perform a complete search of the conformation space of the ligand.

2.2. Scoring Function Algorithm

Once the search algorithm has generated several complexes between the ligand and the complex, it is time to evaluate which of the complexes are the most favorable from the energetic point of view. This difficult task, relies on the scoring functions, that tries to evaluate the interaction energy between the ligand and the receptor [27].

The scoring functions are the most challenging task in the molecular docking protocol, but also a central piece because it is based on that the ligand conformations will be scored with a good energy or a bad energy. However, a very rigorous scoring function that considers all aspects involved in protein ligand binding would be computationally too expensive, so several simplifications are used reducing the complexity and computational cost but with some loss of accuracy.

The scoring functions can be divided in three classes: force-field based, empirical, and knowledge based scoring functions [19].

Force-field based scoring functions are based on the application of force-fields to evaluate ligand-protein association and are quantified by the sum of two energies: the energy between the ligand and the protein; and the internal energy of ligand. These energies are calculated considering the van der Waals energy, that is described by a Lennard-Jones potential, and the electrostatic energy, that is given by a Coulombic function [27]. The absence of solvation and entropic terms, and the inaccurate treatment of the long-range effects involved in binding are some of the traditional limitations.

These scoring functions also include additional terms, namely a solvation term, a torsional term and hydrogen bond term. The solvation term describes the dessolvation and solvation effects that take place before and after the complex between the protein and a given ligand is formed. The torsional term tries to predict the energy required to rotate the flexible bonds of the ligand. The hydrogen bond term tries to describe in more detail the contribution of the hydrogen bonds in the binding process. In Equation 1, it is described the force field scoring function used by Autodock, version 3.0 [25].

$$\text{Score} = \Delta G_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbond} \right) + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{i_C + j} S_i V_j e^{\left(\frac{-r_{ij}^2}{2\sigma^2} \right)} \qquad \text{(Equation 1)}$$

Empirical based scoring functions, try to reproduce experimental data and are based on the idea that binding energies can be approximated by a sum of several individual uncorrelated terms. The coefficients for the various terms are normally determined by regression analysis considering experimentally determined binding energies [27 19].

AutoDock 4.2 uses a semi-empirical free energy force field to evaluate conformations during docking simulations. To the force-field equation, empirical parameters (weighting constant W) were added to give different weights to the several terms (Equation 2).

$$Score = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} \left(S_i V_j + S_j V_i \right) e^{\left(\frac{-r_{ij}^2}{2\sigma^2} \right)} \qquad \text{(Equation 2)}$$

Lastly, the knowledge based scoring functions that follow rules and general principles statistically derived to reproduce experimentally determined structures [19].

The typical formula of a knowledge based scoring function is exposed in Equation 3.

$$Score = \sum_{i,j}^{N} u_{ij}(r) \qquad \text{(Equation 3)}$$

The term u_{ij} is directly obtained from the occurrence frequency of atom pairs in a pre-defined database, using the inverse formulation of the Boltzmann law according to Equation 4.

$$u(r) = -k_B T \ ln\left(\frac{\rho(r)}{\rho^*(r)}\right) \qquad \text{(Equation 4)}$$

where k_B is the Boltzmann constant, T is the absolute temperature of the system, $\rho(r)$ is the number density of the protein-ligand atom pair at r distance and $\rho^*(r)$ is the atom pair density in a reference state where interatomic interactions are zero.

The various knowledge based scoring functions differ between each other in the sets of protein-ligand complexes used to obtain these potentials, the form of the energy function, the definition of protein and ligand atom types, distance and several additional parameters.

The molecular docking software AutoDock Vina, that is highly used, uses a knowledge based scoring function that is described in Equation 5.

$$Score = \Delta G_{gauss} + \Delta G_{repulsion} + \Delta G_{hbond} + \Delta G_{hydrophobic} + \Delta G_{tors} \qquad \text{(Equation 5)}$$

where, ΔG_{gauss} is an attractive term for dispersion, $\Delta G_{respulsion}$ is a function that is used to penalize close internal contacts, ΔG_{hbond} is a function that estimates favorable contribution from the hydrogen bonds between ligand and protein, $\Delta G_{hydrophobic}$is a Ramp function that treats hydrophobic contributions and ΔG_{tors} is a term proportional to the number of rotable bonds.

3. Quantum mechanics

When it is necessary to describe properties at the atomic level, theories of classical mechanics fail, and in response, quantum mechanics arose. The fundamental postulates of quantum mechanics assert that microscopic systems are described by wave functions that completely characterize all the physical properties of the system. In quantum mechanics, Schrödinger equation is analogous to Newton's second law in classical mechanics. It is a partial differential equation that describes how the quantum state of a physical system changes over time (Equation 6).

$$\widehat{H}\psi(r,t) = i\hbar\frac{\delta}{\delta t}\psi(r,t) \qquad \text{(Equation 6)}$$

Time dependent Schrödinger equation where $\widehat{H}$ is the Hamiltonian quantum operator; Ψ (r, t) is the wave function with respect to the position (r) and time (t); i is the imaginary unit; ħ is the Plank constant divided by 2π.

3.1. Schrödinger equation

In most computational methods, a simplification of the Schrödinger equation is used as the basis [28]. The time independent Schrödinger equation [29] is a function that depends only of the spatial coordinates of the nuclei and atoms (Equation 7).

$$\widehat{H}\psi(r) = E\psi(r) \qquad \text{(Equation 7)}$$

Time independent Schrödinger equation where $\widehat{H}$ is the Hamiltonian quantum operator; Ψ (r) is the wave function with respect to the position (r); E is the energy of the particle.

Both, time dependent and time independent Schrödinger equation do not have an analytical solution for chemical systems larger than a hydrogen, which makes their resolution very complex and, even impossible, from a computational point of view. The independence of time was the first simplification to try to solve this equation and, in the next sections, additional simplifications are described that are required to solve this equation for molecules.

3.2. Hamiltonians

Mathematically operators are used to get information about system properties from the wave function. The operator application to a eigenfunction results into a wave function multiplied by the correspondent eigenvalue. In this way, the Hamiltonian ($\hat{H}$) is the mathematical operator for the wave function that gives the total system energy [28].

Hamiltonian operator ($\hat{H}$) is the sum of the kinetic energy operator ($\hat{T}$) and the potential energy operator ($\hat{V}$) (Equation 8).

$$\hat{H} = \hat{T} + \hat{V} \qquad \text{(Equation 8)}$$

For the resolution of the Schrödinger equation, we can further decompose the Hamiltonian operator into the various energy components of the system composed of nuclei and electrons [29] [30]. Thus, we have the operator for the kinetic energy of the electrons ($\hat{T}_e$), the operator for the kinetic energy of the nuclei ($\hat{T}_n$), the operator for the attraction between nuclei and electrons ($\hat{V}_{ne}$), the operator for the repulsion between electrons ($\hat{V}_{ee}$) and the operator for the repulsion between nuclei ($\hat{V}_{nn}$) (Equation 9).

$$\hat{H} = \hat{T}_n + \hat{T}_n + \hat{V}_{ne} + \hat{V}_{nn} + \hat{V}_{ee} \qquad \text{(Equation 9)}$$

3.3. Variational principle

The variational principle is an approximation aiming a solution to the Schrödinger equation, and it uses an arbitrary wave function able to describe the electronic and nuclear coordinates appropriately. According to this principle, the energy value obtained by the wave function is always bigger than the energy of the real fundamental energy state. With this theory, several arbitrary wave functions are tested to reveal which ones have the lowest value. The lower is the value of the variational integral and closer will be this value to the real one. Consequently, the most accurate will be the wave function that will provide the better description of the closely "true" wave function. Taking this into account, we do not need to construct the wave function using a linear combination of orthonormal wave functions. Instead any random wave function could be built and tested [30].

3.4. Born-Oppenheimer approximation

One factor, which greatly contributes to the complexity of the Schrödinger equation, is the dependence of coordinates on nuclei and electrons. The Born-Oppenheimer approximation postulates the decoupling of electronic and nuclear motions. It means that the total wave function could be a result of the product of the nuclei wave function with the electrons wave function [30] [31] (Equation 10).

$$\psi_{total} = \psi_{nuclear} \times \psi_{electronic} \qquad \text{(Equation 10)}$$

Since the mass of the nucleus is much larger than the mass of the electrons, its velocity is correspondingly small. In this way, the nucleus experiments the electrons as if they were a charge cloud, whereas the electrons feel the nuclei as if they were static. So, the electrons adapt instantly to any position of the nuclei without affecting their movement.

Therefore, the Hamiltonian operator can be divided into electronic Hamiltonian ($\hat{H}_{elec}$) and nuclear Hamiltonian ($\hat{H}_{nuc}$) (Equation 11).

$$\hat{H} = \hat{H}_{nuc} + \hat{H}_{elec} \qquad \text{(Equation 11)}$$

3.4.1. Electronic Hamiltonian

The electronic Hamiltonian is calculated with nuclei positions fixed, considering kinetic energy of each electron under the nuclei force field ($\hat{T}_e$), the potential energy involved in the electrons repulsion ($\hat{V}_{ee}$) and in the attraction between nuclei and electrons ($\hat{V}_{ne}$) (Equation 12).

$$\hat{H}_{elec} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne} \qquad \text{(Equation 12)}$$

Since electronic energy is difficult to calculate, several methodologies are used to try to calculate it. There is the wave function based methodologies, such as the Hartree-Fock methods, semi-empirical methodologies and post-Hartree-Fock. Moreover, there is the Density-Functional Theory (DFT) based methodologies which are the ones that will be focused in this work.

3.4.1.1. Density-Functional Theory (DFT)

The premise of DFT is to use the electronic density, rather than the wave function, to obtain molecular properties. The main advantage of working with density, over the wave function, is that the electronic density is a function of three coordinates, independent of the number of electrons, while the wave function is much more complex, with 3N coordinates where N is the number of particles. While the complexity of a wave function increases exponentially with the number of electrons, the electron density has the same number of variables, independent of the system size [29]. But this is only possible due to the theorems developed by Hohenberg and Kohn [29 30]. The first theorem states that any property of a system in the fundamental state (non-degenerated) can be calculated from the electronic density of the ground state [32]. That is, any property can be written as a functional of the electronic state density ($\rho(r)$) of the fundamental state. Hohenberg and Kohn showed that it is possible to obtain the external potential (nuclei position) from the electronic density. Having knowledge of the external potential ($\nu(r)$) and the number of electrons, which can be obtained from the derivation of the electron density, it is possible to obtain the Hamiltonian ($\widehat{H}$) and consequently the wave function (Ψ) (Equation 13).

$$\rho(r) \rightarrow \nu(r) \rightarrow \widehat{H} \rightarrow \Psi \qquad \text{(Equation 13)}$$

The second theorem is, basically, a version of the variational principle applied to DFT. The electronic density that yields the lower energy is the closer one to the real ground state energy [33].

However, these theorems do not tell us how to calculate the energy of the system from $\rho(r)$, since some electronic variables are unknown functional of $\rho(r)$. To overcome this difficulty, Kohn and Sham proposed a method of simplification. By considering a fictitious reference system formed by non-interacting electrons, with each electron subject to an unknown effective potential, the Hamiltonian can be described as the sum of the mono electronic operators. The electronic density should be similar to the ground state electronic density of the system under study, with interacting electrons.

Based on DFT, Equation 14 describes the density with an orbital approximation:

$$E[\rho(r)] = \sum_{i}^{N}\left(\left\langle\chi_i\middle|-\frac{1}{2}\nabla_i^2\middle|\chi_i\right\rangle - \left\langle\chi_i\middle|\sum_k^M \frac{Z_k}{|r_i - r_k|}\middle|\chi_i\right\rangle\right) + \sum_{i}^{N}\left\langle\chi_i\middle|\frac{1}{2}\int\frac{\rho(r')}{|r_i - r'|}dr'\middle|\chi_i\right\rangle + E_{XC}[\rho(r)] \qquad \text{(Equation 14)}$$

where N is the number of electrons and M is the number of nuclei and electronic density is given by $\rho = \sum_{i=1}^{N}\langle\chi_i|\chi_i\rangle$.

The $E_{XC}[\rho(r)]$ is called exchange-correlation energy term and contributes for the energy calculations with electronic exchange and correlation. It should compensate the approximations made to other terms by doing a correction to the kinetic energy which consider the interacting electrons of the real system [29]. To have a useful electronic density energy, it is necessary to have a good approximation of the exchange-correlation term. There are different functional that tries to solve this, the simple one are the local density approximation (LDA), where electronic density is treated as a uniform distribution of electrons. These functional presented good results for systems containing metals, but it is inefficient for systems containing many different molecules because LDA requires a well-defined electronic density. LDA considers that each electron feels the same field regardless its position. Later on, the local spin-density approximation (LSDA) functional emerge, and is related with LDA but includes the spin. Other functional is the generalized gradient approximation (GGA), which are based on LSDA with the addition of a gradient dependent term.

However, the existing functional were not yet able to meet all the needs of studied molecular systems, and then came the hybrid methods that take advantage of the best from LDA and GGA with a Hartree-Fock exchange term.

B3LYP is the most used hybrid functional in computational chemistry and combines Becke's three parameters functional (B3) and Lee-Yang-Parr correlation functional (LYP) [34] with Hartree-Fock term contribution (about 20%). The exchange-correlation energy, through B3LYP functional, is calculated by Equation 15.

$$E_{XC}^{B3LYP} = (1-\alpha)E_X^{LSDA} + c_0E_X^{HF} + c_x\Delta E_X^{B88} + (1-c_c)E_C^{LSDA} + c_cE_C^{LYP} \qquad \text{(Equation 15)}$$

where c_0, c_x and c_c are empirical parameters determined by Becke.

These functional provides good and fast results comparing with other exchange-correlation functional, and the free energy calculation error is about 3 kcal.mol^{-1}.

3.4.1.2. Basis Set

Basis Set are a set of known functions that tries to describe and characterize the behavior of an unknown wave function. A better description can be obtained with a greater and more complete basis set [29]. However, the more complex the basis set used, the higher will be the associated computational cost. The Gaussian (GTO) and Slater (STO) type functions are the most relevant [29].

The STO functions varies exponentially with the electron-core distance and describes with accuracy the orbital of the hydrogen atom. However, its use is limited to small systems, because the calculation with many atoms is computationally difficult to execute. GTO functions are mathematically simpler to handle, therefore computationally less demanding. However, it does not correctly describe the behavior of electrons, specially near the nucleus. Then, the contracted Gaussian functions (CGTO) have emerged that take advantage of the best of STO and GTO. These functions try to simulate a good description of atomic orbitals in a computationally efficient way. Each CGTO results from the contraction of a set of primitive GTOs. A common basis set, and used in this work, is 6-31G, where the "6" means that were used six primitives to contract GTO of the core orbitals and the "31" means that the valence orbitals are divided into two, from which the most internal is described through three primitive GTO and the most external ones only use one primitive GTO. Letter "G" means that the functions used are from the Gaussian type (http://gaussian.com/basissets/).

However, CGTO functions fails in the description of molecular mechanism since they are not polarized and diffuse functions. To solve this issue, the CGTO functions were combined with polarization and diffusion functions to represent the intermolecular interactions.

Polarization functions are modeled through the inclusion of a higher angular momentum of the valence orbitals. When added a "d" on the basis set, this means the inclusion of polarization of orbital p through orbital d, and a "p" denotes the inclusion of p polarization to atoms with s orbital valence electrons.

The other way to improve the basis set is the inclusion of diffusion functions. With the addition of plus signs between the numbers that describe the division and the treatment assigned to the valence orbitals and the letter G. One plus sign means the inclusion of diffuse functions in heavy atoms, while a second plus sign denotes the inclusion of diffuse functions in hydrogen atoms.

3.4.2. Nuclear Hamiltonian

The nuclear Hamiltonian is added lately but it is important for the calculation of activation and reaction free energies. Even the minor effect of the nuclear movement into the global free energy can originate divergence between experimental and theoretical data. The nuclear Hamiltonian, due to the Born-Oppenheimer approximation, can be described by neglecting the kinetic energy of the electrons in relation to the kinetic energy of the nuclei (Equation 16).

$$\hat{H}_{nuc} = \hat{T}_n + \hat{V}_{nn} \qquad \text{(Equation 16)}$$

Considering the Schrödinger equation and that the nuclear Hamiltonian can be decomposed as $\hat{H}_{nuc} = \hat{H}_{trans} + \hat{H}_{rotational} + \hat{H}_{vibrational}$, it is possible to demonstrate that for empirical approximations, nuclear Hamiltonian can be expressed in energy and it is calculated by the sum of translational energy ($E_{translational}$), rotational energy ($E_{rotational}$) and vibrational energy ($E_{vibrational}$) at a certain temperature and pressure (Equation 17).

$$E_{nuclear} = E_{translational} + E_{rotational} + E_{vibrational} \qquad \text{(Equation 17)}$$

Furthermore, we can obtain all the frequencies in the system from the vibrational energy. With these frequencies, we are able to fully characterize the transition state (TS) and both associated minima (product (P) and reactant (R)) on a potential energy surface (PES). The TS is characterized by having one and only one imaginary frequency, while P and R have only positive frequencies. The imaginary frequency associated to the TS, should be correlated to the atoms that are involved in the reaction coordinate.

3.5. Continuum models

Chemical and Biochemical reactions, both in the organism or in the laboratory, occur mostly in aqueous phase. The presence of a solvent interferes with the reaction and influences the stability of the solutes, the reaction rate, the activation free energy, the reaction free energy, and may even alter the product of the reaction.

In computational chemistry and biochemistry, there are two ways that are used to include the contribution of the solvent in a chemical reaction: using an explicit model or an implicit model [35] (Figure 10).

The most rigorous one, but also more computationally expensive approach is the explicit one (Figure 10A), in which the molecules of the solvent are included in the calculation explicitly. In the implicit approach (Figure 10B), the solvent is simulated by a continuum model that is well defined by macroscopic properties, such as the dielectric constant. This approximation considerably decreases the computational cost however, if the solvent interferes directly in the reaction, this approach is not suitable to use.

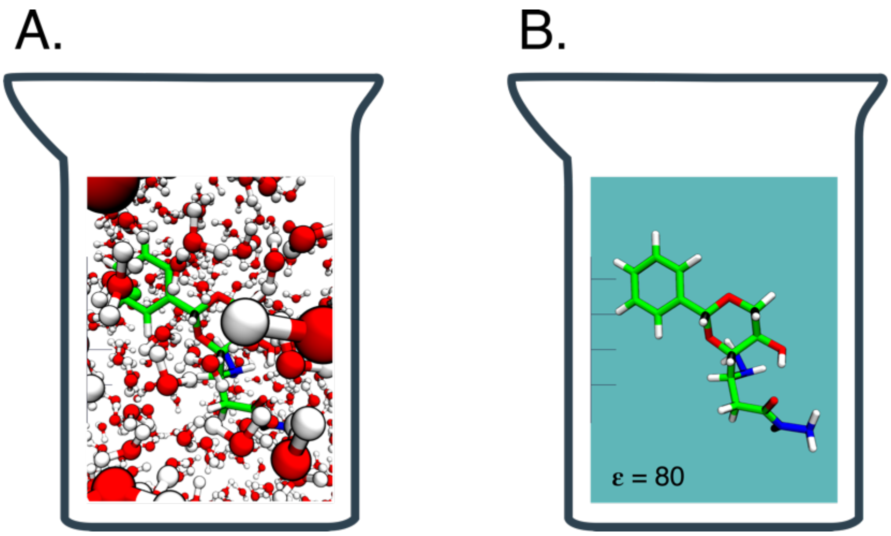


Figure 10 Representation of both methods to predict solvent influence. A. Explicit model. B. Implicit model.

The Polarizable Continuum Model (PCM) with the integral equation formalism variant (IEFPCM) is the most used method in implicit model calculations. This method creates the solute cavity through a set of overlapping spheres. In the Table 1 are described the dielectric constants of some of the most used solvent in organic chemistry (http://dev.gaussian.com/scrf/?tabid=7).

Table 1 List of dielectric constants of some of the most used solvents in organic chemistry.

Solvent	Dielectric constant (ε)
Water	78.355
Acetonitrile	35.688
Methanol	32.613
Ethanol	24.852
Diethyl Ether	4.240

3.6. Potential energy surface (PES)

The potential energy surface (PES) is the relation (mathematical and graphical) between the energy of a system and its geometry. For example, in Figure 11 it is graphically represented the energy of a system (for all the geometries) in function of two reactional coordinates (nuclei position and bond length for example).

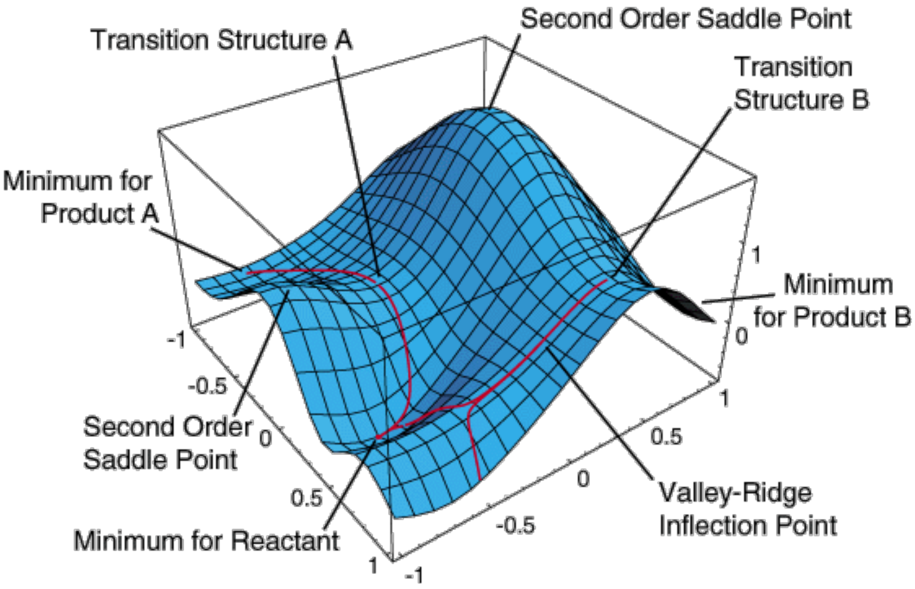

Transition Structure A
Second Order Saddle Point
Transition Structure B
Minimum for Product A
Minimum for Product B
Second Order Saddle Point
Minimum for Reactant
Valley-Ridge Inflection Point
-1
-0.5
0
0.5
1
-1
-0.5
0
0.5
1
0
1

Figure 11 Example of a 3D potential energy surface.

When looking for a mechanism, we try to find the so-called stationary points on PES. The reactant (R) and product (P) correspond to stationary points and should be both a minimum, with all forces (first derivatives) of the energy with respect to the spatial coordinates equal to zero. On the other hand, transition state (TS) should be a maximum (point of highest energy) that connects the R and P, but still a minimum on the PES. TS is a stationary point with all forces equal to zero and all force constants positive, except one and only one that must be negative.

The computer cannot predict the mechanism and do not know which are the most favorable pathways, the user must search for it. The researcher should explore all the possibilities and draw several models of the pathways that the reaction can follow. Once all the hypothesis has been drawn, it is necessary to model the reactions by computational means and understand which ones are possible or not. In this way, it is possible to locate and characterize the reactants, transition state and products for all the hypothetical pathways, and calculate the Gibbs activation and reaction free energies of each pathway. Analyzing the Gibbs free energies of the pathways and comparing them with experimental data, if available, we should conclude which are the most favorable and decide if new alternatives should be studied.

To draw conclusions, some key considerations have to be taken into account:

- all the studied pathways must be characterized by a reactant, transition state and product.
- both activation and reaction free energies should be used to decide which reaction is favored from the kinetically (lower activation free energy) and thermodynamically (more exergonic) points of view.

III

CHAPTER III

Results and Discussion

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

Marie Curie (1867-1934)

The present chapter is divided in two main sections, a molecular mechanism section and a biological one.

In the first section, we tried to understand the mechanism beyond the synthesis of the two new scaffolds using molecular mechanics for molecular mechanism study. The mechanism was not known at all, and there were several issues. Why the same synthesis method could give two kinds of compounds, five- and six-membered lactones? And what is beyond the *stereo*- and *regio*- selectivity in these reactions? These questions are answered in the quantum mechanics studies.

In the biological section, these several compounds were tested as specific inhibitors of α-glucosidase, β-glucosidase, α-galactosidase and β-galactosidase. To this end, the inhibition rate and IC_{50} values of these compounds targeting each of these enzymes were calculated. Still in this section, a Molecular Docking study was conducted to understand, based on the binding poses of the compounds, the reasons beyond the selectivity of some of the compounds for one of the enzymes.

In a third section, some exceeded difficulties during the work are discussed, showing the importance and robustness of the theoretical and computational calculations.

III

1. Catalytic Mechanism

The art and science of asking questions is the source of all knowledge.

Thomas Berger (1924-2014)

1. Introduction

The synthesis of polyhydroxylated sugar-like compounds is very rich since, due to the large variety of *regio*- and *stereo*-configurations, multiple compounds can be obtained from a single starting material.

The potential associated to these reactions to produce multiple products is both a curse and an opportunity. It is a curse because it has the potential to create multiple products and each of which must be separated from each other. But it is also an opportunity because if one can develop reactions that can yield one isomer over the other (and vice versa) it produces a very useful tool. This means that one can start with a simple starting material – like D-glucose – and transform it into several complex products through a series of selective reactions. And, that is extremely powerful.

The reality is that the synthesis of specific polyhydroxylated sugar-like compounds is very difficult to achieve in organic synthesis. Several strategies are used to limit the number of *regio*- and *stereo*-configurations that can be obtained in the final product. Most of these strategies involve specific or selective chemical routes. A selective chemical route implies that there are factors which favor one product over the other, while the term "specific" is usually a sign that there is something inherent to the mechanism of the reaction that leads only to one product. This last type of reactions is the most difficult to obtain and are the ones that are constantly pursued by organic chemists.

There are quite a few examples of organic reactions that can be used to ensure *regio*- and *stereo*–selective characteristics of the polyhydroxylated sugar-like compounds. Most of these strategies imply the use of specific monosaccharides as starting materials that already possess the correct *stereo*- and *regio*-selective configuration that is desired for the final product. However, this strategy implies a constant protection and deprotection multistep manipulations that limits their application in real conditions. One of these examples is the synthesis of D-erythrose derivatives that have been known for a long time as a potential precursor of polyhydroxylated sugar-like compounds. However, the poor chiral induction offered by the template often limits their application. Recently, however, it was published the synthesis of a derivative of D-erythrose, the D-erythrosyl δ-lactone **15**, that overcomes this issue and offers a template that provides a total facial selectivity towards the C=C unit during the 1,3-dipolar cycloadditions [36]. The same template was also used with success with a series of nucleophiles and from which several polyhydroxylated sugar-like compounds were obtained (Figure 12). This new chemical route continues to guarantee the *stereo*-selectivity of the reaction as it was observed in the 1,3-dipolar cycloadditions reactions, but it also provides *regio*-selectivity and the formation of two new chemical scaffolds: a six-membered **20** and a five-membered ring **18** lactones (Figure 13). The formation of these compounds can be manipulated experimentally, using different nucleophiles. Hard nucleophiles tend to generate five-membered lactones, whereas soft nucleophiles tend to generate six-membered lactones.

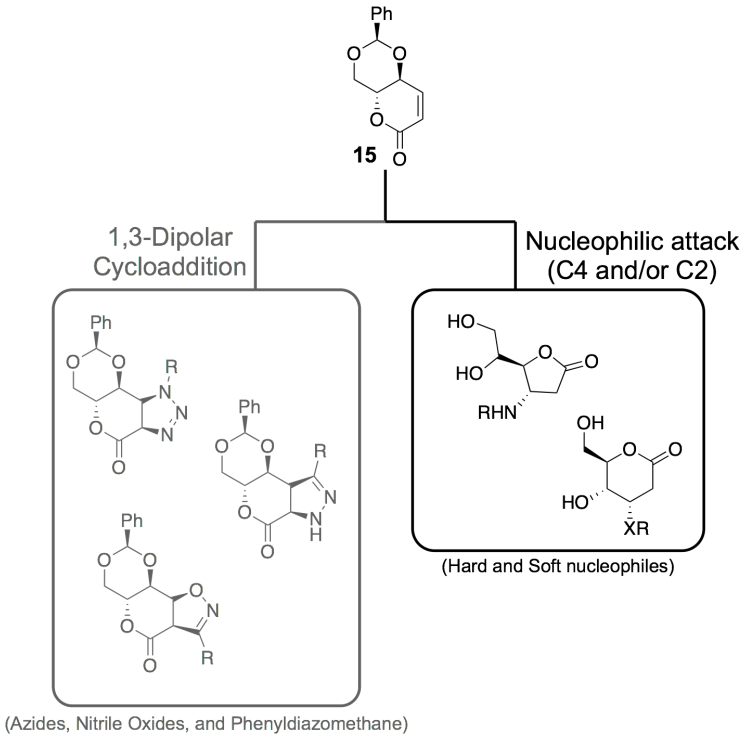

Ph
O
O
O
O
15
1,3-Dipolar Cycloaddition
Nucleophilic attack (C4 and/or C2)
Ph
R
N
N
N
Ph
R
N
N
H
Ph
O
N
R
HO
HO
RHN
O
O
OH
HO
XR
(Hard and Soft nucleophiles)
(Azides, Nitrile Oxides, and Phenyldiazomethane)

*Figure 12 Representation of some of the compounds that is possible to obtain from D-erythrose derivative, D-erythrosyl δ-lactone **15**, through 1,3-dipolar cycloaddition at gray, and through nucleophilic attack to C4 and/or C2 ate black.*

The hard and soft nomenclature is used in this thesis to characterize the nature of the nucleophiles. This nomenclature is also known as the Pearson acid base concept. Hard nucleophiles have high charge-density. They are usually charged or highly polarized. Their orbitals do not necessarily overlap that well with the accepting orbital of the electrophile, but charge attraction directs them and benefits the reaction. Soft nucleophiles favor an orbital interaction over charge. They are usually not particularly polarized species, but have large and available orbitals for nucleophilic interaction with the accepting orbital of electrophile.

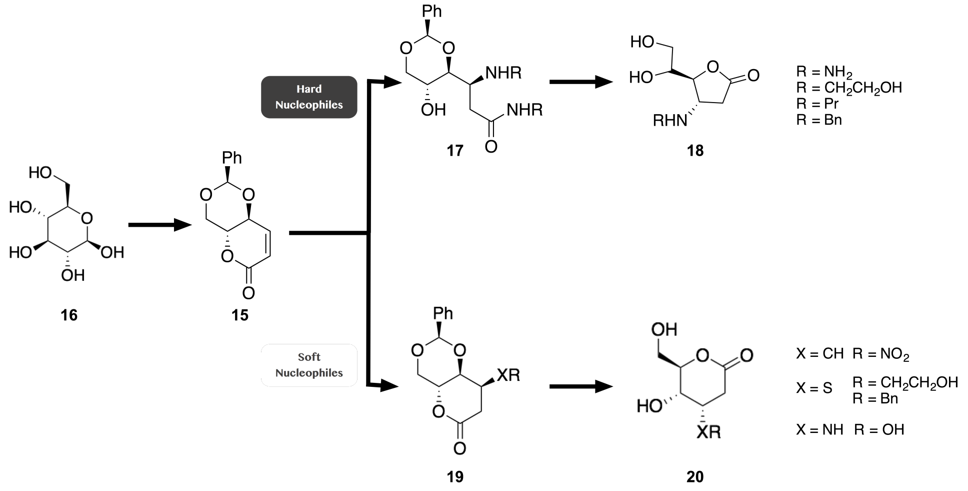


Figure 13 General schematization of the synthetic route for five- and six-membered lactone rings.

Despite of the success in applying these reactions in the synthesis of *regio*- and *stereo*-selective lactones, the mechanism is still poorly understood. This precludes an efficient use of this new chemical route to improve the yields of the chemical reactions.

In general terms, the reaction involves the nucleophilic attack of the nucleophile species to carbon C4 and/or C2 of the D-erythrosyl δ-lactone **15**. However, the nature of the nucleophile species was observed, experimentally, to change the mechanism of the reaction. The formation of five-membered ring lactones **18** is favored when D-erythrosyl δ-lactone **15** reacts with harder nucleophiles, such as hydrazine and primary amines. When D-erythrosyl δ-lactone **15** react with soft nucleophiles, such as, sulfides, base catalyzed nitromethane and hydroxylamine, six-membered ring lactones **20** are obtained instead.

These results indicate that both reaction should follow different mechanisms. The formation of five-membered ring lactones **18** requires the nucleophilic attack of the harder nucleophiles at carbon C4 and C2, whereas in the six-membered ring lactones **20** the nucleophilic attack of the soft nucleophiles occurs only at carbon C4. From these reactions also result different reaction intermediates that are detected experimentally. In the formation of five-membered ring lactones **18** it is detected the formation of a reaction intermediate **17** that adopts an open form, whereas in the formation of the six-membered ring lactones **20**, only the reaction intermediate **19** in a close form is observed. In addition, the formation of the six-membered ring **20** is complete after the cleavage of the acetal, whereas, in the case of the five-membered ring lactone **18** an additional step is required. It involves the cleavage of the acetal protection followed by an intramolecular cyclization.

All of these facts taken together show that there are many open questions about the mechanism of these reactions:

i) what factors define the *stereo* and *regio*-selectivity of these reactions?

ii) it is not known how the hard and soft nucleophile species control the mechanism that is followed and lead to the formation of different products. How they influence the mechanism?

iii) How is the nucleophilic attack at C4 and/or C2 of D-erythrosyl δ-lactone **15** controlled?

iv) Why it is detected, experimentally, the formation of a reaction intermediate in an open form in the formation of five-membered lactones whereas in the case of six-membered lactones only a close form is observed?

v) the reaction mechanism of the five- and six-membered lactones are competitive or different reactions are involved?

In order to answer all of the questions raised above, theoretical and computational means will be used to study the mechanisms of those reactions in particular the formation of the D-erythrosyl δ-lactone **15** and the five- and six-membered lactones (**18** and **20**).

Considering the biological importance of lactones, that offer new molecular scaffold that can be used to develop new and more selectively inhibitors for α- and β-glycosidases, the chemical routes that can speed up their synthesis in a *stereo*- and *regio*-selective way have become a major demand.

2. Methodology

All geometry optimizations were performed with Gaussian 09 [37], applying density functional theory [32]. Becke's three-parameter exchange functional together with the 6-31G(d) [38] basis set was used for all atoms.

In all geometry optimizations, we first searched for the transition state starting from a structure like the reactant model. This was generally obtained with unidimensional scans along the particular reaction coordinate in which we were interested. Once a putative transition structure was located, and thus was fully optimized, the reactants and the products associated with it were determined through intrinsic reaction coordinate (IRC) calculations. In all cases, the geometry optimizations and the stationary points were obtained with standard Gaussian convergence criteria. The transition state structures were all verified by vibrational frequency calculations, having exactly one imaginary frequency with the correct transition vector. The ZPE, thermal and entropic energy corrections were calculated using the same method and basis set (T = 310.15 K, P = 1 bar).

The final electronic energies were calculated using the all-electron 6-311++G(3df,2pd) basis set and the functional B3LYP [38] for all atoms. A conductor-like polarizable continuum model using the integral equation formalism variant (IEF-PCM) [39], as implemented in Gaussian 09 [37], was used to simulate the water solvent with different dielectric constant (Table 2).

Table 2 Values of the dielectric constants used with the IEF-PCM calculations.

Solvent	Dielectric Constant
Acetonitrile	ε = 35.688
Ethanol	ε = 24.852
Mixture of acetonitrile and ethanol	ε = 30.000
1,4-Dioxane	ε = 2.210

All the activation and reaction free energies provided in the text and figures refer to free energy differences calculated at the B3LYP/6-311++G(3df,2pd) level of theory, while the atomic charge distributions were calculated at the B3LYP level of theory employing a Mulliken population analysis, using the 6-31G(d) basis set.

3. Results and Discussion

The complete synthetic route that yield the five- and six-membered ring lactones involves two main stages. The first one compromises the synthesis of the D-erythrosyl δ-lactone **15** that is accomplished in four sequential steps having as starting material D-glucose **16**. The five- and six-membered ring lactones are obtained afterwards, through the reaction of the D-erythrosyl δ-lactone **15** with hard and soft nucleophiles.

From the literature, it is reported that the reaction of unsaturated carbonyl compounds with soft and hard nucleophiles deviated in different mechanism. Soft nucleophiles are known to operate by a thermodynamic control, connected with a nucleophilic attack at C4, whereas the harder nucleophiles act by a kinetic control, connected with attack at C2. Since, this critical mechanistic aspect might be the reason behind the deviations in the mechanism that allow the formation of two different compounds, the five- and six-membered rings, both mechanism were studied in detail by theoretical and computational means.

In order to make the discussion of the results clearer, we divided the discussion in three main sections. The first one involves the formation of D-erythrosyl δ-lactone **15**. The second and third sections involves the formation of the five- and six-membered lactones, respectively. Particular attention will be given to the last two sections, since they are the ones that are the source of the *regio*- and *stereo*- specificity of the chemical route and allows the specific formation of lactones with two different scaffolds.

3.1. D-erythrosyl δ-lactone formation

The formation of the D-erythrosyl δ-lactone **15** can be achieved in four sequential steps starting from D-glucose **16**. In the first step takes place the protection with the acetal group from the sugar moiety with dimethylacetal benzaldehyde **21** under *p*-toluene sulfonic acid catalysis to produce compound **22**. This step is very important to protect these groups from the following reactions and preserve the *stereo*-selective of the D-glucose in the final products. The second step involves an oxidative cleavage of the acetal derivative **22** with sodium periodate **23** from which results the aldehyde **24**. The third step involves the Wittig olefination of the aldehyde **24** with methyl (triphenylphosphoranylidene)acetate **25** and from which results an isomeric mixture of compounds **26** *cis* and **26** *trans*. The last step of the reaction involves the selective intramolecular cyclization of the Z isomer (**26** *cis* form) and from which D-erythrosyl δ-lactone **15** is obtained (Figure 14).

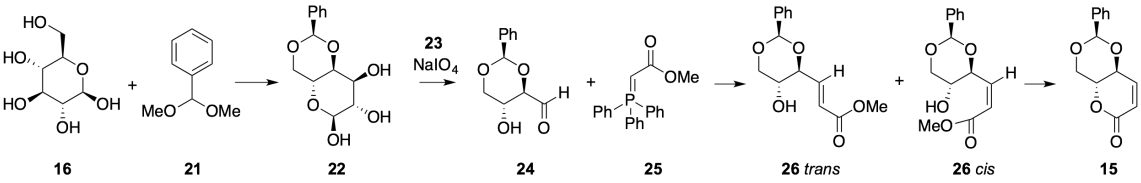


*Figure 14 Schematic representation of the D-erythrosyl δ-lactone **15** synthesis.*

3.2. Formation of the five-membered lactones (hard nucleophiles)

Reaction of D-erythrosyl δ-lactone **15** with hard nucleophiles lead to open amide structures which are then cyclized under strong acid conditions.

The theoretical and computational calculations show that the mechanism requires several sequential steps involving concerted reactions in each one. It was also found that the reaction can occur through two competitive mechanisms depending where the nucleophilic attack takes place. These pathways were called pathway NA@C2 when the first Nucleophilic Attack occurs at carbon C2 of the D-erythrosyl δ-lactone **15** and pathway NA@C4, when the first Nucleophilic Attack takes place first at carbon C4 of the lactone (Figure 15).

Independently of the pathway that is followed, both mechanisms lead to the formation of the same reaction intermediate **17** that adopts an open configuration. The formation of the five-membered ring **18** requires an additional step involving an intramolecular cyclization of compound **28** and from which the five-membered lactones **18** are obtained (Figure 15). Each of these pathways are described in detail in the following sections.

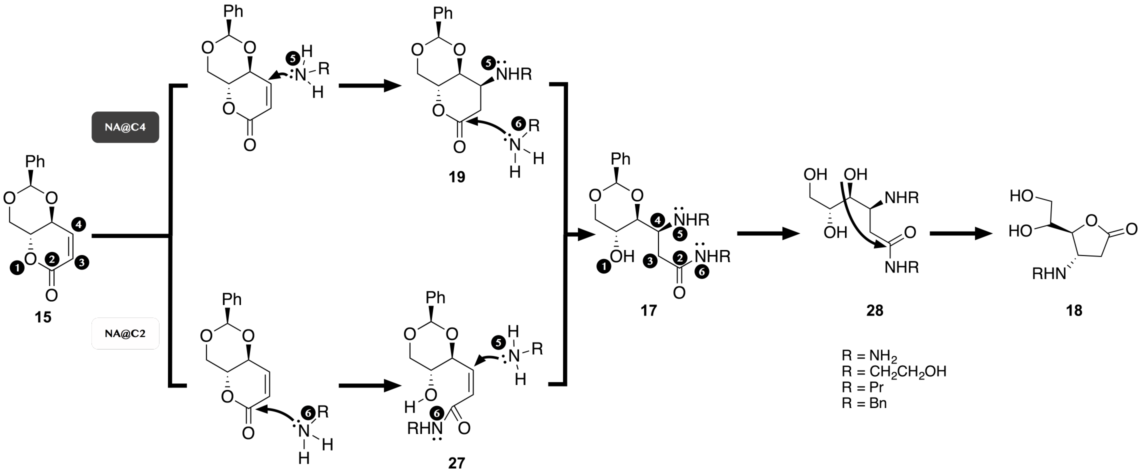


Figure 15 Pathway NA@C4 and Pathway NA@C2 of the five-membered ring lactone formation with hard nucleophiles.

A. Pathways NA@C4

The computational calculations devoted to this pathway revealed that the NA@C4 pathway requires three steps and for each molecule of D-erythrosyl δ-lactone **15** three nucleophile molecules (primary amines) are required (Figure 16).

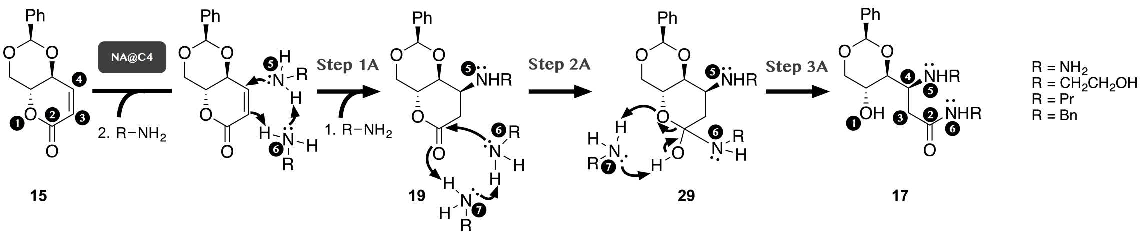


Figure 16 Mechanism of pathway NA@C4 of D-erythrosyl δ-lactone with hard nucleophiles.

The first step of this pathway requires the presence of two nucleophile molecules and one molecule of D-erythrosyl δ-lactone **15**. One of the nucleophile molecules is involved in the nucleophilic attach at carbon C4 of the lactone **15**. The second nucleophile molecule catalyzes the proton transfer between the lactone **15** and the other nucleophile molecule.

An important question about this reaction is related on how the nucleophilic attack takes place. Experimentally, only one *stereo*-isomer is isolated at the end of the reaction, but the true is that there are two possibilities for this reaction. The nucleophilic attack at C4 can occur in the bottom or in the top-face of the double bond that links carbons C3 and C4 of the lactone (Figure 17). When the attack occurs in the bottom face of the double bond, in the end of the reaction the nucleophile molecule becomes aligned in the same face where the hydrogen H4a is located at. If the attack occurs in the top-face, then it becomes aligned in the opposite face of the hydrogen H4a.

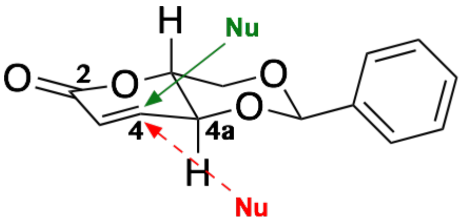


Figure 17 Stereo-specific nucleophilic attack at carbon C4 of the lactone. In red is represented the nucleophilic attack through the bottom face and in green the nucleophilic attack through the top-face.

The theoretical and computational results have shown that the nucleophilic attack can only occur in the opposite face where the hydrogen H4a is facing. This happens because when the attack occurs in the bottom face, this proton is on the way of the nucleophile molecule creating a steric hindrance (activation energies above 50 kcal/mol). The attack on the other face of lactone **15** is on the other hand more exposed and free from any steric hindrance. This means therefore that the nucleophile molecule becomes always covalently bonded to the lactone in the *S*-configuration, favoring by this way the *stereo*-selectivity of the reaction (activation energies below 23 kcal/mol).

The computational results also shown that the nucleophilic attack to lactone **15** is favored due to the formation of a network of hydrogen among the nucleophile molecules and lactone **15**. This effect stabilizes the two nucleophile molecules close to each other and one of them with carbon C3 of D-erythrosyl δ-lactone **15**. Such rearrangement also potentiates the proximity of carbon C4 from lactone **15** to nitrogen N5 of one of the nucleophile molecule (4.16 ± 0.58 Å) making it ready for the nucleophilic attack. The same type of rearrangement was found in all the hard nucleophile species that were studied. What differentiates each nucleophile molecule is the R group (Figure 15) but it does not influence the network of hydrogen bonds or the orientation of the molecules in relation to the lactone **15**.

The transitions states of all the studied reactions are very similar among each other and are characterized by one and only one imaginary frequency (Table 3). The results shown that in all the TS structures, nitrogen N5 is already very close to carbon C4 (1.56 ± 0.01 Å at TS versus 4.16 ± 0.58 Å in the reactants) (Figure 18). At the same time, one proton is equally shared between the two nitrogen atoms from the amines (1.39 ± 0.01 Å to N5 and 1.24 ± 0.01 Å to N6) and the proton that was bonded to nitrogen N6 of the amine is now very close to carbon C3 (1.92 ± 0.06 Å ate TS against 5.40 ± 1.43 Å in the reactants) of the lactone **15**.

In the product of this reaction **19**, the formation of the covalent bond between carbon C4 of the D-erythrosyl δ-lactone **15** and nitrogen N5 of the one of the nucleophile molecules is complete (1.47 ± 0.00 Å in the product against 4.16 ± 0.58 Å in reactants). It is also observed the concerted proton transfer between N5 to N6 from the two nucleophile molecules and the proton transfer from nitrogen N6 of the amine to carbon C3 to the lactone.

Table 3 Frequencies of the transition states of all the reaction studied in pathway NA@C4 with hard nucleophiles (data in cm^{-1}).

Hard Nucleophiles	**Pathway NA@C4**		
	Step 1	Step 2	Step 3
NH_2NH_2	896.42i	832.01i	492.68i
$NH_2(CH_2)_2OH$	1026.76i	816.53i	469.53i
NH_2Pr	1006.77i	794.25i	392.36i
NH_2Bn	1048.41i	857.85i	461.43i

The first step requires an average activation free energy of 22.41 kcal/mol and a reaction free energy of -3.33 kcal/mol (Table 4).

The second step of pathway NA@C4 involves the nucleophilic attack of one amine molecule to carbon C2 to the product of the first step (compound **19)**. Similarly, to what was observed in the first step, this reaction requires two nucleophile molecules for each of compound **19**. Similarly to what happened in the previous reaction, this step involves the formation of a network of hydrogen bonds between the amine molecules and the product **19**. In this case, the carbonyl group establishes a hydrogen bond with amine N7 (2.04 ± 0.02 Å) and the later one with the other amine molecule (2.11 ± 0.03 Å). Such rearrangement allows the close proximity of one of the amines (N6) to carbon C2 (3.34 ± 0.10 Å) and makes it ready for the second nucleophilic attack at carbon C2. Again, no significant differences on the mechanism are observed when different nucleophiles are used.

The transition state of all these reactions are characterized by one and only one imaginary frequency (Table 3). In all the analyzed structures, nitrogen N6 and carbon C2 are already very close to each other (1.57 ± 0.02 Å in TS against 3.34 ± 0.10 Å in reactants) (Figure 18). The proton from nitrogen N7 of one of the amines is very close to oxygen O2 of the product of the previous step (1.60 ± 0.04 Å in TS against 2.04 ± 0.02 Å in reactants), and the proton from nitrogen N6 is very close to nitrogen N7 from the other amine (1.25 ± 0.01 Å in TS against 2.11 ± 0.03 Å in reactants).

In the product of this step (compound **29)**, the formation of the covalent bond between carbon C2 and nitrogen N5 of the amine is complete (1.45 ± 0.01 Å in products against 3.34 ± 0.10 Å in reactants). It is also observed the concerted proton transfer between N7 and N6 from the two amine molecules and the proton transfer from nitrogen N7 from one of the amines to oxygen O2, transforming the carbonyl group into a hydroxyl group.

The second step of pathway NA@C4 requires an average activation free energy of 30.49 kcal/mol and a reaction free energy of 12.93 kcal/mol.

The third, and last, step of pathway NA@C4 involves the cleavage of the covalent bond between oxygen O1 and carbon C2 of the intermediate **29**. This requires the participation of only one amine molecule, that works as a proton shuttle.

In the reactant of this reaction, the amine molecule established two hydrogen bonds with the lactone **29**: one with the oxygen O1 (2.65 ± 0.73 Å) and another with the hydroxyl group attached to carbon C2 (1.66 ± 0.04 Å).

The transition states of these reactions are characterized by one imaginary frequency (Table 3). These structures reveal an increase in the bond length between carbon C2 and oxygen O1 (1.89 ± 0.04 Å in TS against 1.55 ± 0.04 Å in reactants) from the lactone intermediate **29**. It is also observed the proximity of the proton bonded to nitrogen N7 from the amine to oxygen O1 (1.75 ± 0.03 Å in TS against 2.65 ± 0.73 Å in reactants), as well as, the proximity of the proton from the hydroxyl group bonded to carbon C2 of the intermediate to nitrogen N7 of the amine (1.28 ± 0.04 Å in TS against 1.66 ± 0.04 Å in reactants) (Figure 18).

In the product of the reaction, the bond between carbon C2 and oxygen O1 from the lactone intermediate is broken (3.19 ± 0.06 Å in products against 1.55 ± 0.04 Å in reactants) and the proton that was bonded to oxygen O2 is now bonded to nitrogen N7 of the amine (1.03 ± 0.00 Å in products against 1.66 ± 0.04 Å in reactants). The proton that was bonded to the amine is now bonded to oxygen O1 of the lactone intermediate **29** (1.02 ± 0.00 Å in products against 2.65 ± 0.73 Å in reactants), from which an open reaction intermediate **17**, that is detected experimentally, is obtained.

The third step of pathway NA@C4 requires an average activation free energy of 10.44 kcal/mol and a reaction free energy of -19.98 kcal/mol.

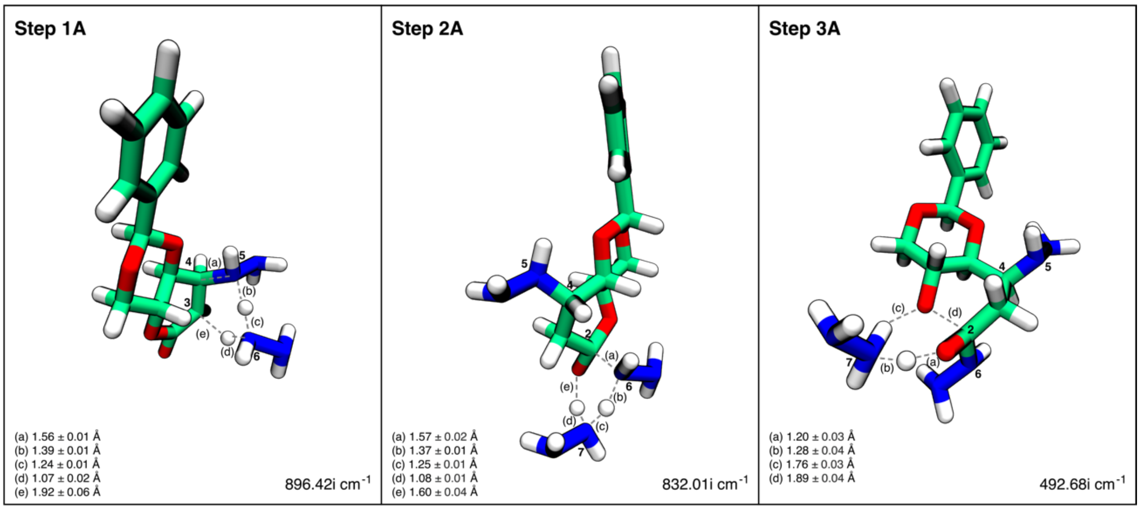

Step 1A
4
(a)
5
(b)
(c)
3
(e)
(d)
6
(a) 1.56 ± 0.01 Å
(b) 1.39 ± 0.01 Å
(c) 1.24 ± 0.01 Å
(d) 1.07 ± 0.02 Å
(e) 1.92 ± 0.06 Å
896.42i cm-1
Step 2A
5
4
2
(a)
6
(e)
(b)
(d)
(c)
7
(a) 1.57 ± 0.02 Å
(b) 1.37 ± 0.01 Å
(c) 1.25 ± 0.01 Å
(d) 1.08 ± 0.01 Å
(e) 1.60 ± 0.04 Å
832.01i cm-1
Step 3A
4
5
(c)
(d)
2
7
(b)
(a)
6
(a) 1.20 ± 0.03 Å
(b) 1.28 ± 0.04 Å
(c) 1.76 ± 0.03 Å
(d) 1.89 ± 0.04 Å
492.68i cm-1

Figure 18 Transition state of the three steps from pathways NA@C4. In this case, it is represented the reaction having as hard nucleophile the compound with R=NH_2.

As can be seen in Table 4, the activation free energy, for the studied substituents, of the first step are quite similar, being slightly smaller in the smaller groups, and the most polar group is the one where the reaction is more exergonic. In the second step, there are not significant differences. In the third step, the picture seems to be inverted, the smaller and the polar nature of the group increases, the less favored is the reaction.

Table 4 Free activation and reaction energies of all the reaction studied in pathway NA@C4 with hard nucleophiles (data in kcal/mol).

Hard Nucleophiles	**Pathway NA@C4**					
	Step 1		Step 2		Step 3	
	Ea	Er	Ea	Er	Ea	Er
NH_2NH_2	22.80	-6.15	28.32	11.48	14.42	-10.83
$NH_2(CH_2)_2OH$	24.21	-1.25	31.49	13.47	10.54	-21.89
NH_2Pr	20.10	-2.01	30.52	13.37	7.66	-24.77
NH_2Bn	22.53	-3.92	31.52	13.39	9.13	-22.41
Mean	22.41	-3.33	30.49	12.93	10.44	-19.98

As it was already discussed, no substantial geometrical differences are observed in the reactions when different amine nucleophiles are used.

B. Pathways NA@C2

The main difference between pathway NA@C4 and pathway NA@C2 is the order through which the reactions take place. While in NA@C4 the nucleophilic attack at carbon C4 occurs prior to nucleophilic attack at C2 followed by the lactone ring opening, in pathway NA@C2 the reaction occurs in the opposite way. First occurs the nucleophilic attack at C2, and from which results the lactone ring opening and only afterwards occurs the nucleophilic attack at carbon C4 of the lactone **15** (Figure 19).

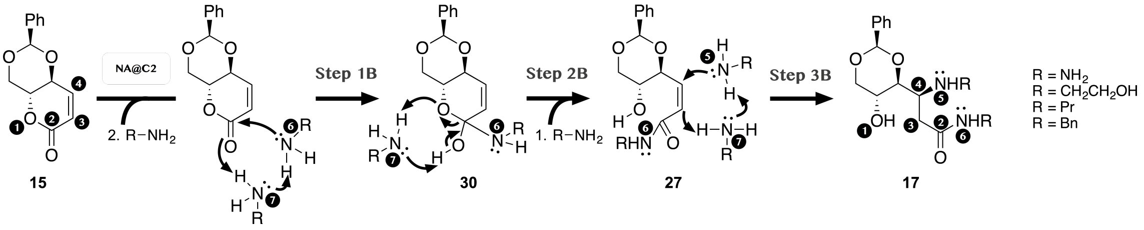


Figure 19 Mechanism of pathway NA@C2 of D-erythrosyl δ-lactone with hard nucleophiles.

Comparing both pathways, step 1B and step 2B from the NA@C4 pathway correspond to step 2A and 3A from NA@C2 pathway, respectively. In both cases for each molecule of lactone **15**, three molecules of amines are required to complete the reaction.

The first step of the pathway NA@C2 involves the nucleophilic attack of the amine molecule to carbon C2 (4.02 ± 0.56 Å) of the D-erythrosyl δ-lactone **15**. At the same time, one proton is transferred from the attacking amine (N6) to oxygen O2 of the lactone, through a second amine (N7) molecule that behaves as a proton shuttle. This step is favored by the formation of a network of hydrogen bonds that in the reactants stabilizes the two amines in relation of oxygen O2 and carbon C2 of the derivative **15**, making it ready for the reaction.

The transition states (Figure 20) of all the studied reactions are characterized by one and only one imaginary frequency (Table 5). In all of these structures nitrogen N6 and carbon C2 are already very close to each other (1.56 ± 0.00 Å in TS against 4.02 ± 0.56 Å in reactants). The proton from nitrogen N7 of one of the amines is very close to oxygen

O2 of the lactone (1.61 ± 0.03 Å in TS against 2.34 ± 0.37 Å in reactants), and the proton from nitrogen N6 is very close to nitrogen N7 from the other amine (1.25 ± 0.01 Å in TS against 2.07 ± 0.03 Å in reactants).

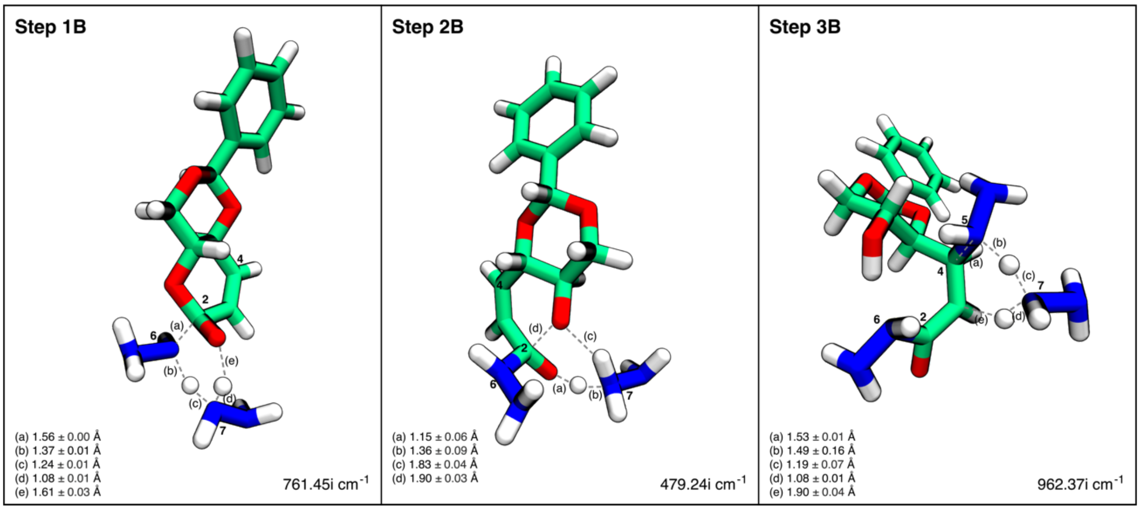

Step 1B
4
2
6
(a)
(b)
(c)
(d)
(e)
7
(a) 1.56 ± 0.00 Å
(b) 1.37 ± 0.01 Å
(c) 1.24 ± 0.01 Å
(d) 1.08 ± 0.01 Å
(e) 1.61 ± 0.03 Å
761.45i cm^{-1}
Step 2B
4
(d)
(c)
2
6
(a)
(b)
7
(a) 1.15 ± 0.06 Å
(b) 1.36 ± 0.09 Å
(c) 1.83 ± 0.04 Å
(d) 1.90 ± 0.03 Å
479.24i cm^{-1}
Step 3B
5
(b)
(a)
4
(c)
7
6
2
(e)
(d)
(a) 1.53 ± 0.01 Å
(b) 1.49 ± 0.16 Å
(c) 1.19 ± 0.07 Å
(d) 1.08 ± 0.01 Å
(e) 1.90 ± 0.04 Å
962.37i cm^{-1}

Figure 20 Transition state of the three steps from pathways NA@C2. In this case, it is represented the reaction having as hard nucleophile the compound with R=NH_2.

In the product **30** of the reaction, a quaternary intermediate is obtained with amine N6 already covalently bonded to C2 (1.44 ± 0.01 Å in products against 4.02 ± 0.56 Å in reactants). The concerted proton transfer between N7 and N6 from the two amine molecules, and between N7 from one of the amines and O2 is complete and the carbonyl group converted into a hydroxyl group.

Table 5 Frequencies of the transition states of all the reaction studied in pathway NA@C2 with hard nucleophiles (data in cm^{-1}).

Hard Nucleophiles	Pathway NA@C2		
	Step 1	Step 2	Step 3
NH_2NH_2	761.45i	479.24i	962.37i
$NH_2(CH_2)_2OH$	823.63i	173.35i	926.26i
NH_2Pr	801.41i	215.70i	751.86i
NH_2Bn	830.73i	318.54i	183.43i

The first step of pathway NA@C2 requires an average activation free energy of 27.20 kcal/mol and a reaction free energy of 10.10 kcal/mol (Table 6).

The next reaction involves the cleavage of the covalent bond between oxygen O1 and carbon C2 (1.54 ± 0.01 Å) of the quaternary intermediate **30**. This requires the participation of only one amine molecule (N7) that catalyzes the proton migration between oxygen O2 and oxygen O1 of the intermediate.

The transition state of this reaction is characterized by one imaginary frequency (Table 5). These structures reveal an increase in the bond length between carbon C2 and oxygen O1 (1.90 ± 0.03 Å in TS against 1.54 ± 0.01 Å in reactants) from the intermediate **30**. It is also observed the proximity of the proton bonded to nitrogen N7 from the amine to oxygen O1 of the lactone (1.83 ± 0.04 Å in TS against 3.36 ± 1.02 Å in reactants) and the proton from the hydroxyl group bonded to carbon C2 of the lactone to nitrogen N7 of the amine (1.36 ± 0.09 Å in TS against 1.66 ± 0.02 Å in reactants).

In the product of the reaction (compound **27**), the bond between carbon C2 and oxygen O1 from the intermediate **30** is broken (3.08 ± 0.11 Å in products against 1.55 ± 0.01 Å in reactants). The proton that was bonded to oxygen O2 is now bonded to nitrogen N7 of the amine (1.03 ± 0.00 Å in products against 1.66 ± 0.02 Å in reactants) and the proton that was bonded to the amine is now attached to oxygen O1 of the lactone intermediate (1.02 ± 0.00 Å in products against 3.36 ± 1.02 Å in reactants), resulting an opened form amide with no substitution at carbon C4.

The second step of pathway NA@C2 from all the studied reactions requires an average activation free energy of 10.76 kcal/mol and a reaction free energy of -21.61 kcal/mol.

The final step of this pathway requires the participation of two amine molecules for each reaction intermediate **27**. It involves the formation of a covalent bond between carbon C4 of the open intermediate **27** and nitrogen N5 of the amine. It is also observed the concerted proton transfer between N5 to N7 from the two amine molecules and the proton transfer from nitrogen N7 of the amine to carbon C3. Like in the first step of the NA@C4 pathway, these three molecules form a network of hydrogen bonds that stabilize the two amines close to each other and one of them with carbon C4 of derivative **27** making it ready for the reaction.

The transition states of the studied reactions (Figure 18) are characterized by one and only one imaginary frequency (Table 5). The results shown that in all the TS structures, nitrogen N5 is already very close to carbon C4 (1.53 ± 0.01 Å in TS against 1.64 ± 0.03 Å in reactants). At the same time, one proton is equally shared between the two nitrogen atoms (1.49 ± 0.16 Å from N5 and 1.19 ± 0.07 Å from N7) of the amines and the proton that was bonded to nitrogen N6 is now very close to carbon C3 (1.90 ± 0.04 Å in TS against 2.99 ± 0.55 Å in reactants).

In the product of this reaction, the formation of the covalent bond between carbon C4 and nitrogen N5 of the amine is complete (1.47 ± 0.01 Å in products against 1.64 ± 0.03 Å in reactants). It is also observed the concerted proton transfer between nitrogens N5 and N7 from the two amine molecules and the proton transfer from nitrogen N7 of the amine to carbon C3 (1.09 ± 0.00 Å in products against 2.99 ± 0.55 Å in reactants) obtaining the opened form amide **17.**

The final step of all the studied reactions requires an average activation free energy of 6.66 kcal/mol and a reaction free energy of -34.01 kcal/mol (Table 6).

In the case of pathway NA@C2 and as it was also observed in the other pathway there are no substantial differences between the different groups attached to the amines, either from the geometrically or energetically point of view (Table 6).

Table 6 Free activation and reaction energies of all the reaction studied in pathway NA@C4 with hard nucleophiles (data in kcal/mol).

Hard Nucleophiles	Pathway NA@C2					
	Step 1		Step 2		Step 3	
	Ea	Er	Ea	Er	Ea	Er
NH_2NH_2	22.95	7.47	12.07	-20.04	5.31	-37.82
$NH_2(CH_2)_2OH$	29.83	12.04	10.78	-22.27	7.84	-32.16
NH_2Pr	28.16	11.34	10.05	-22.88	6.41	-32.97
NH_2Bn	27.87	9.53	10.14	-21.23	7.06	-33.07
Mean	27.20	10.10	10.76	-21.61	6.66	-34.01

C. Pathways NA@C4 versus Pathways NA@C2

The major difference between both studied pathways are observed in step 3B and 1A, where the nucleophilic attack to the double bound (at C4) of the lactone happens and that we expected to have a similar behavior, however, the obtained energetic profiles revealed a different picture (Figure 21).

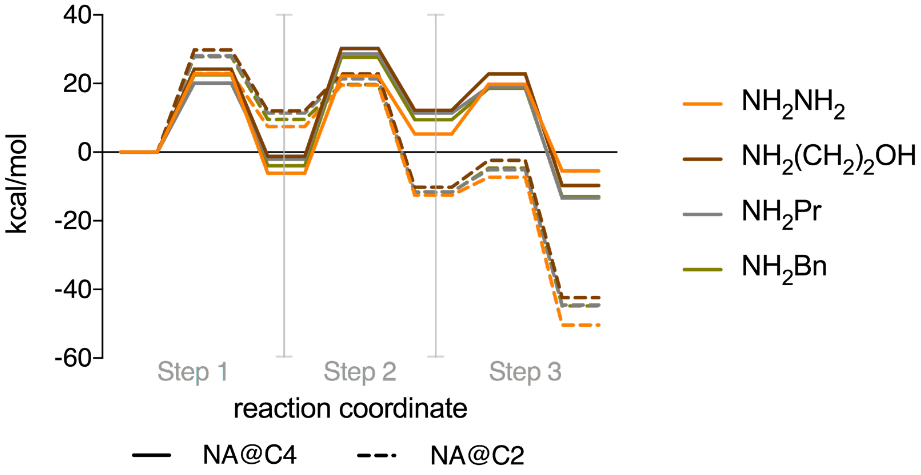


Figure 21 Energetic profile of pathway NA@C4 at full line and pathway NA@C2 at dashed line with hard nucleophiles.

In general terms both pathways present energetic profiles that makes them possible to occur in experimental conditions, which can make us think that both mechanism are competitive. However, and from Figure 21 it is clear that in the initial steps, pathway NA@C4 is favored from the kinetic and thermodynamic point of view, whereas pathway NA@C2 is kinetically and thermodynamic favored in the last steps of the mechanism. These results suggest therefore that pathway Na@C4 should be favored in relation to pathway NA@C2. This goes in line with the literature that indicates that reaction of hard nucleophiles has a kinetic control, being the first step the rate limiting step of these reactions.

It is worth mentioning that the product of both reactions is the same (Figure 22). However, and as the energetic profile (Figure 21) reveals in the end of each pathway, the same molecule is obtained but with different energies. The one obtained from pathway NA@C4 has higher energy than the one that is obtained from pathway NA@C2. These energy differences come from the different conformation that these molecules adopt in the end of each pathway. With a closer inspection of these structures, one can observe that there is a slight rotation of the dihedral angles involving atoms C1a, C4a, C4 and N5, that differentiates both compounds. In order to explain the energetic difference of approximately 24 kcal/mol between both compounds, we divided both products in portions and calculate the corresponding energies of each portion. In the Figure 22, it is described the contribution of each group for these energetic differences.

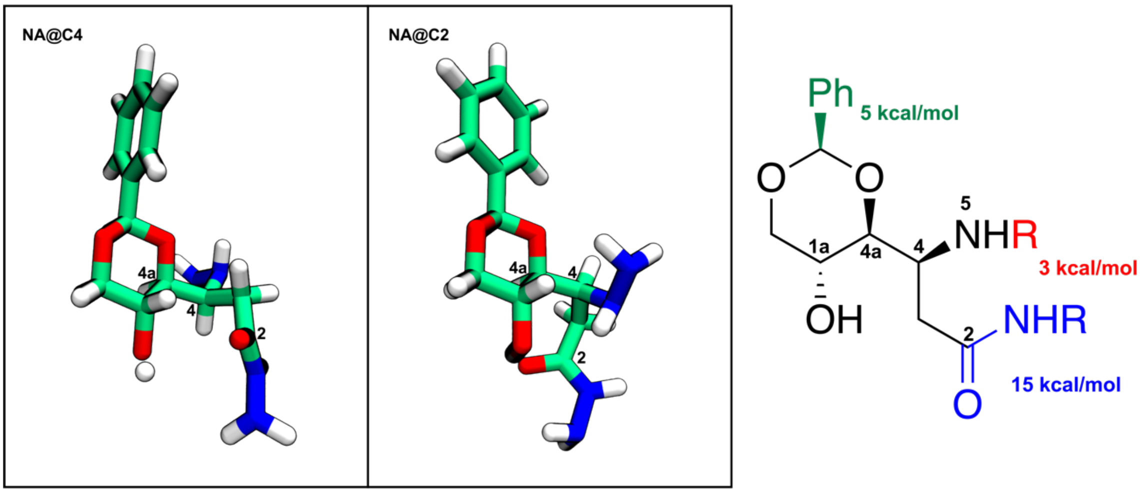
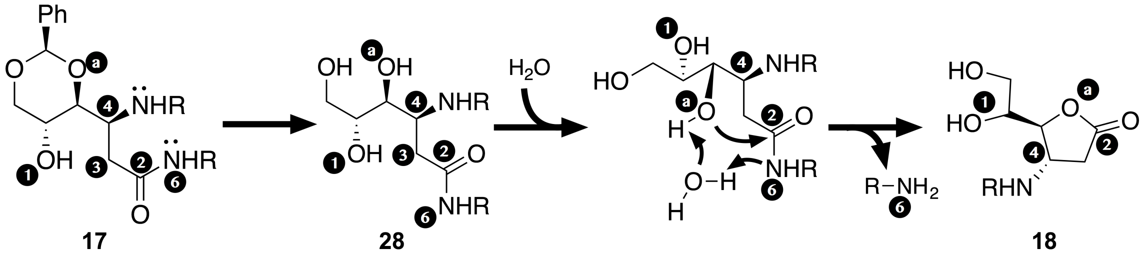


Figure 22 Representation of products of pathway NA@C4 (on the left) and pathway NA@C2 (at the middle). On the right, it schematically represented the energetic contribution of each group.

The results have shown that main energetic different between both compounds comes from the conformation adopted by the amine group. In the case of the product of pathways NA@C2, the conformational rearrangement of this group establishes a hydrogen bond with one of the hydroxyl groups from the lactone and this feature has an impact of 15 kcal/mol in the final energy of the compound.

D. Ring closure

The formation of the five-membered ring **18** involves the cleavage of the acetal group and subsequent intramolecular cyclization under acid catalysis (Figure 23).



17 **28** **18**

R = NH_2
R = CH_2CH_2OH
R = Pr
R = Bn

Figure 23 Mechanism of the 5-membered ring closure.

The cleavage of the acetal was not computationally studied, and we focused our attention instead in the intramolecular cyclization of the reaction intermediate **28**. The reaction involves the formation of a covalent bond between oxygen Oa and carbon C2 and the cleavage of the covalent bond between carbon C2 and nitrogen N6. In the reactants of this reaction, the water molecule forms a network of hydrogen bonds with hydrogen of Oa and N6 making it ready for the proton shuttle that is required for the cyclisation process.

The transition state of this reaction (Figure 24) is characterized by one imaginary frequency (Table 7). The results shown that in the TS structure, oxygen Oa is already very close to carbon C2 (2.17 ± 0.05 Å in TS against 3.19 ± 0.47 Å in reactants) and N6 is already quite distant from C2 (1.57 ± 0.01 Å in TS against 1.40 ± 0.03 Å in reactants). At the same time, the proton from oxygen Oa is equally shared with the oxygen from the water molecule (1.27 ± 0.01 Å from Oa and 1.16 ± 0.01 Å from water oxygen) and the proton from water is now very close to nitrogen N6 (1.12 ± 0.01 Å in TS against 1.91 ± 0.10 Å in reactants).

Table 7 Frequencies of the transition states of all the reaction studied for ring closure with hard nucleophiles (data in cm^{-1}).

Hard Nucleophiles	**Step 1**
NH_2NH_2	927.87i
$NH_2(CH_2)_2OH$	680.84i
NH_2Pr	677.76i
NH_2Bn	659.28i

In the product of the reaction, the formation of the covalent bond between carbon C2 and oxygen Oa is complete (1.35 ± 0.01 Å in products against 3.19 ± 0.47 Å in reactants) and the amine molecule N6 is released from carbon C2 (3.44 ± 0.49 Å in products against 1.40 ± 0.03 Å in reactants). It is also observed the concerted proton transfer from the water molecule to amine molecule that is then released.

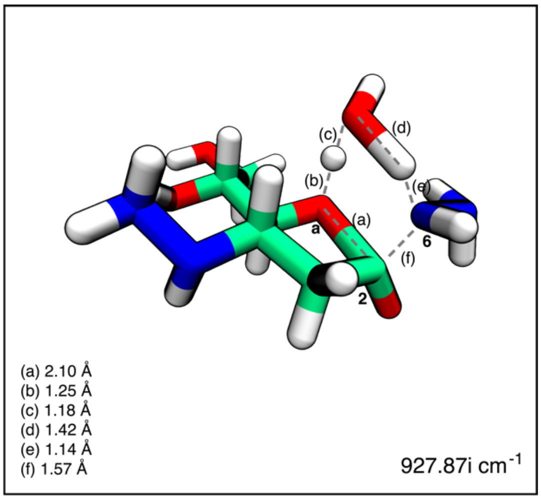


Figure 24 Representation of transition state of the closure of the 5-membered lactone ring.

This reaction requires an average activation free energy of 31.23 kcal/mol and a reaction free energy of -6.04 kcal/mol (Table 8).

Table 8 Free activation and reaction energies of all the reaction studied for ring closure with hard nucleophiles (data in kcal/mol).

Hard Nucleophiles	**Step 1**	
	Ea	Er
NH_2NH_2	29.55	-4.78
$NH_2(CH_2)_2OH$	30.20	-14.34
NH_2Pr	29.22	-6.01
NH_2Bn	35.95	0.98
Mean	31.23	-6.04

This energetic profile is higher than the formation of the intermediate **17**. This was expected since this reaction is more time consuming and, experimentally, requires several days for the reaction to be complete.

3.3. Formation of the six-membered lactones (soft nucleophiles)

The reaction of the lactone **15** with soft nucleophiles is also possible by the two pathways that were described in the hard nucleophiles. However, neither it is observer the formation of the reaction intermediate with an open form **17**, neither the final product of the reaction is similar. This suggests that the mechanism of the reaction of the lactone **15** with the soft nucleophiles is different from the one that is followed by the hard nucleophiles or perhaps the mechanism deviates at some point and this is the reason why the final products of both reaction are different. In order to investigate this aspect, we started by studying the same mechanism that was done with the hard nucleophiles with the soft nucleophiles (Figure 25).

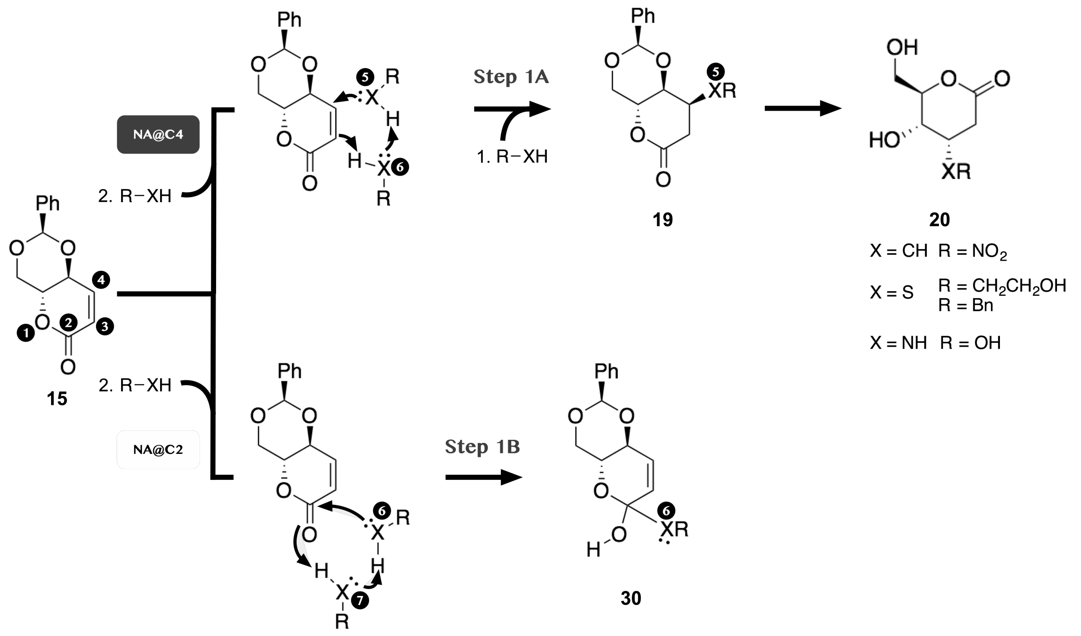


Figure 25 Mechanism beyond six-membered lactone formation.

The theoretical and computational results show that the first step of pathway NA@C2 and NA@C4 requires very high activation free energies, above 35 kcal/mol. The major difference between the two pathways is that the nucleophilic attack at carbon C2 is endergonic whereas the nucleophilic attack at carbon C4 is slightly exergonic. The second step reveals the opposite trend. The activation free energy of the pathway

NA@C4 requires higher activation free energies and the reaction is endergonic, whereas the NA@C2 has lower activation free energy and the reaction is exergonic.

The cumulative energetic profile between the two first steps of each pathway reveals that the pathway NA@C4 is much more favored from the energetic point of view than pathway NA@C2. The energetic profile of pathway NA@C2 turns this reaction almost impossible to take place due to the high cumulative activation free energies that it requires (above 40 kcal/mol).

As it is described in the literature, soft nucleophiles act by a thermodynamic control which is true in this reaction, since the product of the first step of NA@C4 is thermodynamically more stable than the product of the first step of pathway NA@C2 (Figure 26).

The energies that were calculated with the soft nucleophiles and, in particular, for pathway NA@C4 are higher than the ones that were calculated with the hard nucleophiles. The reason is due to the presence of the sulfur atom in the nucleophilic attack than when compared with the nitrogen atom of the harder nucleophiles, has more difficulties to perform the nucleophilic attack. This effect is even stronger in the NA@C2 because of the closer proximity of the sulfur atom with the oxygen of C2 from the lactone **15**, which creates a steric hindrance that turn the reaction very difficult to take place.

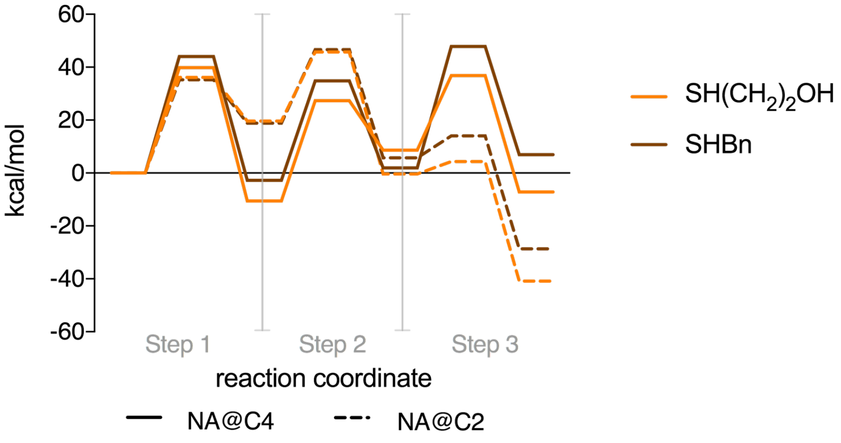


Figure 26 Energetic profile of hypothetical pathway A and B. Pathway A is represented in full line and it is the nucleophilic attack in C4 first. Pathway B is represented in dashed line and it is the nucleophilic attack in C2 first.

Another difference of this mechanism when it is compared with the one that was obtained with the harder nucleophiles, is that it is not observed the formation of a reaction intermediate in an open form (**17**). Indeed, after the second step of the NA@C4 pathway does, it not occurs due to high activation free energy that is required, and the reaction stays trapped in compound **19**. At last, the six-membered lactone **20** is obtained by the cleavage of protective acetal group under acid catalysis.

Since the steps were already discussed in the previous sections, in this section, only the most favorable ones from pathway NA@C4 are described.

A. Pathway NA@C4

Due to some particularities of the mechanism beyond the nucleophilic attack by hydrazine to the D-erythrosyl δ-lactone **15**, it will be, first, described the characterization of sulfur nucleophile reaction and, then the characterization of hydrazine reaction.

1. Sulfur nucleophile

In the reactant, D-erythrosyl δ-lactone **15** and the two sulfur molecules become aligned along a network of hydrogen bonds. These interactions allow both nucleophiles to become well aligned in relation to carbons C2 and C4 of the lactone ring (4.50 ± 1.41 Å and 5.83 ± 1.17 Å, respectively), making them ready for the reaction.

The calculated transition state structures from all the modelled reactions are characterized by one and only one imaginary frequency (Table 9) and present similar structural features. In all the studied cases, sulfur S5 from one of the nucleophile molecules is already very close to carbon C4 of the derivative **15** (2.30 ± 0.01 Å against 5.83 ± 1.17 Å in reactants) and the proton from the second nucleophile specie is also close to carbon C3 from compound **15** (1.37 ± 0.01 Å against 4.50 ± 1.41 Å in reactants). The proton involved in the hydrogen bond between both nucleophile molecules is also equally shared between both compounds (2.04 ± 0.01 Å and 1.69 ± 0.00 Å) and the bond length between carbon C3 and C4 of the D-erythrosyl δ-lactone **15** increases (1.44 ± 0.00 Å against 1.34 ± 0.00 Å in the reactants). In Figure 27 it is illustrated the TS in pathway

NA@C4 in contrast with the first step of pathway NA@C2. As it is possible to see, in Figure 27B the distance between C2 and S6 is very large, due to steric hindrance, which makes the reaction very difficult to occur.

Table 9 Frequencies of the transition states of all the reaction studied in pathway NA@C4 and NA@C2 with soft nucleophiles (data in cm^{-1}).

Soft Nucleophiles	Pathway NA@C4			Pathway NA@C2		
	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3
SHBn	1064.47i	626.20i	383.82i	770.06i	297.44i	1100.79i
$SH(CH_2)_2OH$	1015.08i	164.23i	397.50i	736.68i	333.84i	1196.31i

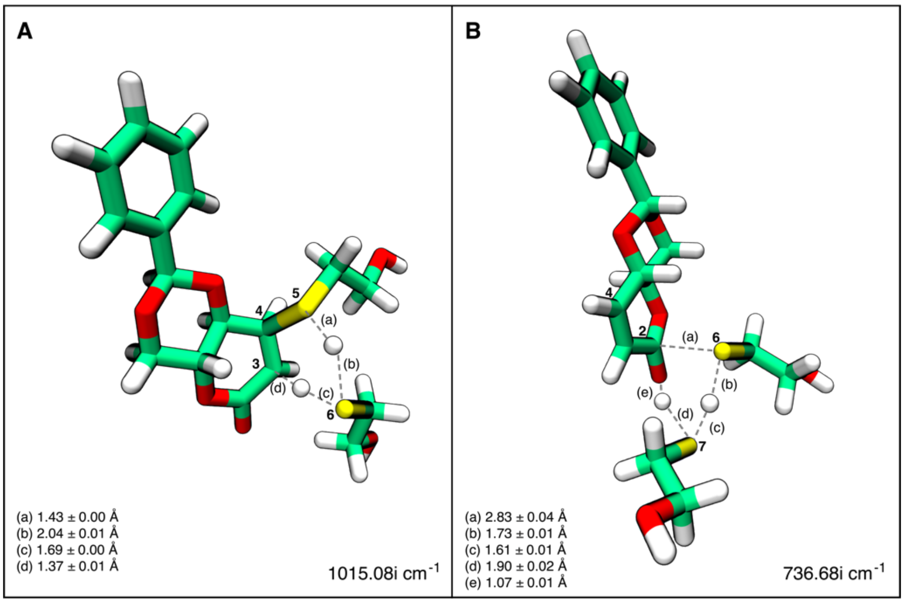

A
4
5
3
6
(a)
(b)
(c)
(d)
(a) 1.43 ± 0.00 Å
(b) 2.04 ± 0.01 Å
(c) 1.69 ± 0.00 Å
(d) 1.37 ± 0.01 Å
1015.08i cm-1
B
4
2
6
7
(a)
(b)
(c)
(d)
(e)
(a) 2.83 ± 0.04 Å
(b) 1.73 ± 0.01 Å
(c) 1.61 ± 0.01 Å
(d) 1.90 ± 0.02 Å
(e) 1.07 ± 0.01 Å
736.68i cm-1

*Figure 27 Representation of transition state of the first step. **A**: pathway NA@C4. **B**: pathway NA@C2.*

In the product of this step, one of the sulfur molecules becomes covalently attached to carbon C4 of the D-erythrosyl δ-lactone **15** (1.85 ± 0.00 Å against 5.83 ± 1.17 Å in the reactants), carbon C3 becomes protonated and acquires a sp3 configuration. Carbons C3 and C4 from the lactone **15** become bonded by a single covalent bond instead of the double bond found in the reactants (1.54 ± 0.00 Å against 1.34 ± 0.00 Å in reactants) and the closed form intermediate **19** is obtained.

This concerted reaction requires an average activation free energy of 41.96 kcal and the reaction is exergonic in -6.63 kcal. The energetic profile of both complete pathways is represented in Figure 26.

2. Hydroxylamine

This mechanism is very similar to the one that was described before. However, in this case and due to the smaller size of the molecule, the distances in the transition states are different. In Figure 28 it is represented the transition states of the first step of both studied pathways. Although the distances are quite similar in both pathways, the activation free energies are completely different. In the case of pathway NA@C4, it is exergonic in -24.66 kcal/mol and the activation free energy is 15.55 kcal/mol. On the other hand, pathway NA@C2 is endergonic in 16.28 kcal/mol and the activation free energy is 32.69 kcal/mol. It becomes clear that the pathway that is more favored is pathway NA@C4.

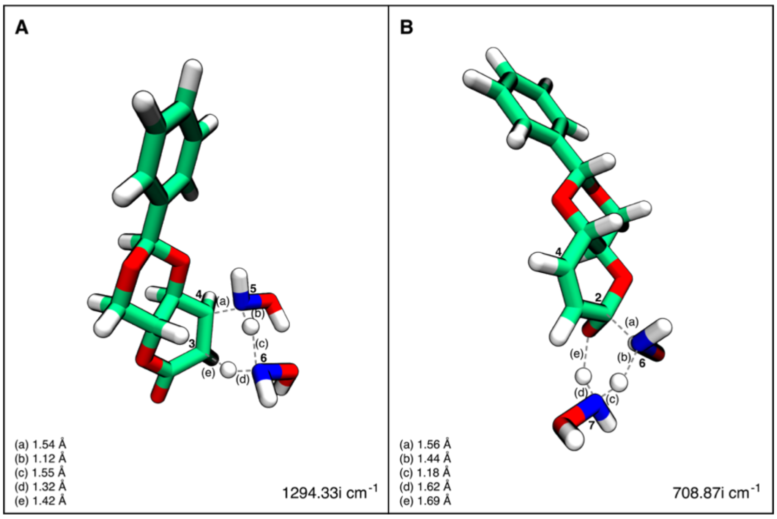


Figure 28 Representation of transition state of the first step. ***A****: pathway NA@C4.* ***B****: pathway NA@C2.*

Hydroxylamine has, also, the particularity to have two groups that can make the nucleophilic attack and two atoms that can behave as good hydrogen bond donors/acceptors (nitrogen and oxygen). Therefore, there was the need to explore which of these two atoms can act as better proton shutters and explore the two nucleophilic attacks, one performed by the nitrogen and the other by the oxygen (Figure 29).

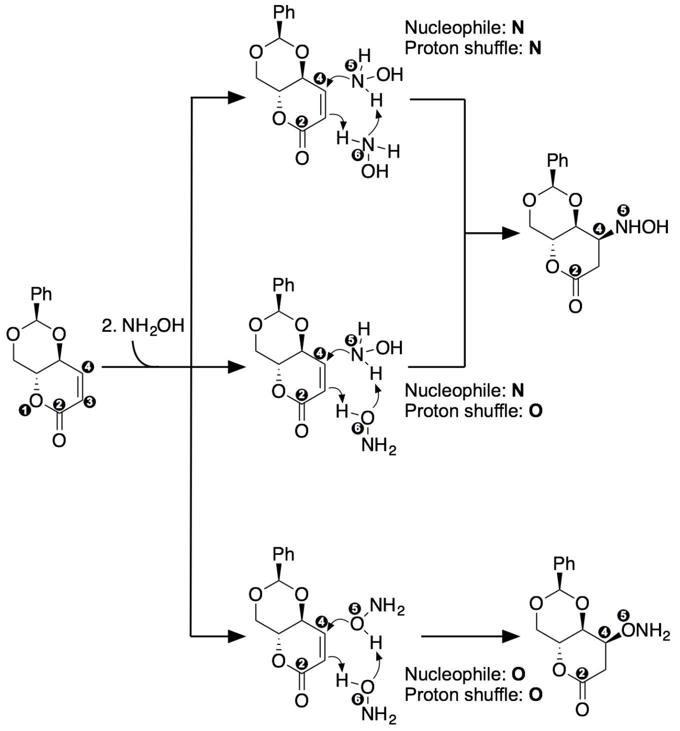


Figure 29 Schematic representation of the proton shuffle through the nitrogen and through the oxygen and the nucleophilic attack by the two atoms.

The main difference between the studied reactions is observed in the energetic profile. From the geometric point of view both reactions are very similar (Figure 28, Figure 31, Figure 32).

The theoretical results reveal that the reaction is favored when the oxygen donates and accepts the protons (Figure 30 and Figure 31). Considering this result, it was necessary to explore the case of $SH(CH_2)_2OH$ and $NH_2(CH_2)_2OH$ (Figure 30). Nevertheless, in these cases there was not any geometric or energetic advantage. In the case of $SH(CH_2)_2OH$ (Figure 30 E and F), the reaction is slightly compromised when oxygen participates as proton shuffle. It is suggested that this occurs due to the large size of sulfur orbitals, which deforms the six-membered ring formed by the hydrogen bonds. In the case of $NH_2(CH_2)_2OH$ (Figure 30 A and B), it seems to have neither a positive or negative influence. While in the case of NH_2OH (Figure 30 C and D), it has a great energetic advantage. It is suggested that this advantage is related to the proximity of the nitrogen and oxygen in the hydrazine molecule. Oxygen has a bigger electronic cloud, since by inductive effect it withdraws the charge from the nitrogen. This larger electron cloud will provide a greater facility in capturing the proton and thus benefiting the proton exchange, and therefore activation free energy of the step. When the nitrogen and oxygen are far away, this inductive effect is less pronounced and the proton exchange is not benefited.

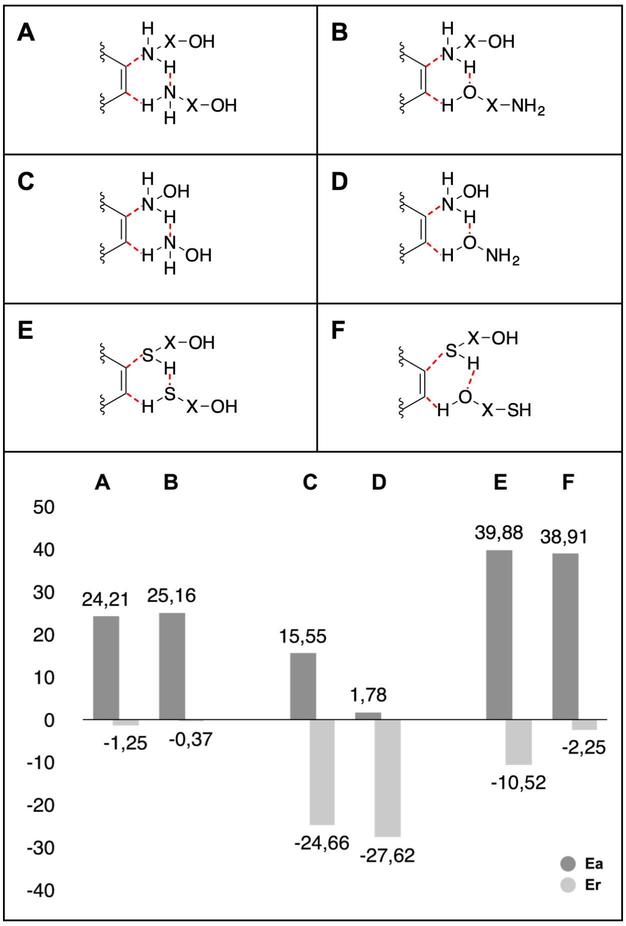


Figure 30 Schematic representation of pathway NA@C4 with $NH_2(CH_2)_2OH$, NH_2OH and $SH(CH_2)_2OH$, when N and S are the proton shuffle (A, C and E) and when O is the proton shuffle (B, D and F). $X=CH_2CH_2$.

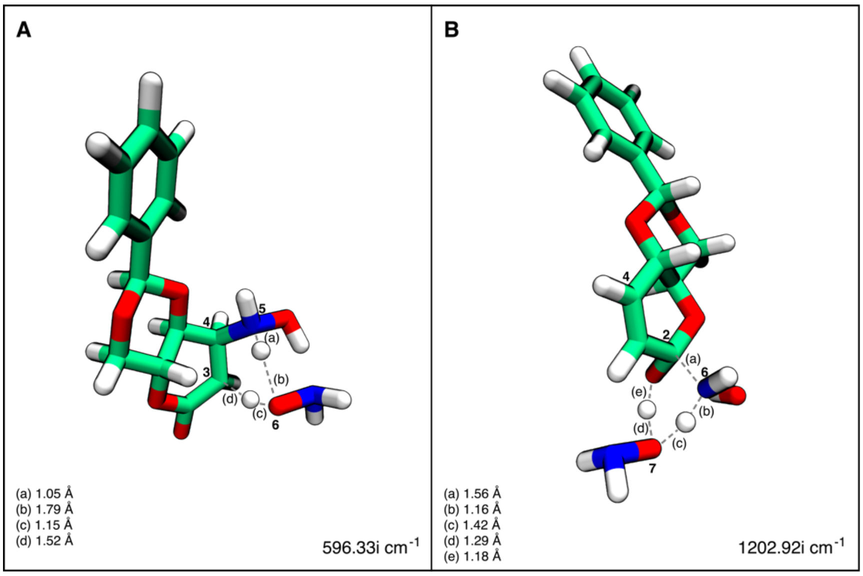

A
5
4
3
6
(a)
(b)
(c)
(d)
(a) 1.05 Å
(b) 1.79 Å
(c) 1.15 Å
(d) 1.52 Å
596.33i cm-1
B
4
2
6
7
(a)
(b)
(c)
(d)
(e)
(a) 1.56 Å
(b) 1.16 Å
(c) 1.42 Å
(d) 1.29 Å
(e) 1.18 Å
1202.92i cm-1

*Figure 31 Representation of transition state of the first step when oxygen is the proton donor. **A**: pathway NA@C4. **B**: pathway NA@C2.*

The next step was to find out if the oxygen could perform the nucleophilic attack in a more efficient way. As it is illustrated in Figure 32, the geometric organization is similar to the previous ones, but the free energy that is required for the reaction to happen is too high to compete with the nucleophilic attack by the nitrogen.

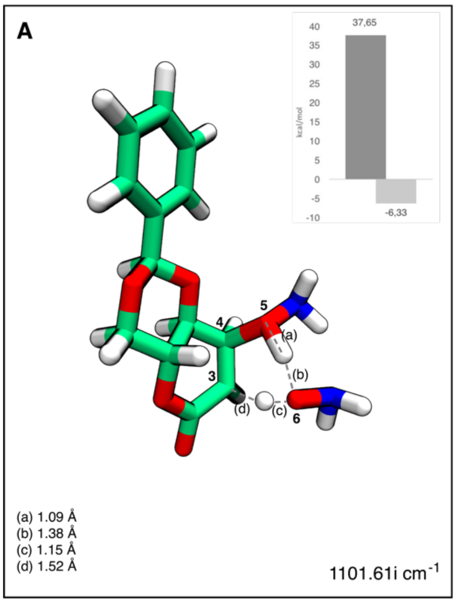

A
40
35
30
25
20
15
10
5
0
-5
-10
kcal/mol
37,65
-6,33
5
4
3
6
(a)
(b)
(c)
(d)
(a) 1.09 Å
(b) 1.38 Å
(c) 1.15 Å
(d) 1.52 Å
1101.61i cm-1

Figure 32 Representation of transition state of the first step of pathway A when nucleophile atom is the oxygen. Activation free energy is represented in dark gray and reaction free energy in soft gray.

From these results, it can be concluded therefore that hydrazine takes advantage of the proximity of the two atoms in the nucleophilic attack by the nitrogen to the D-erythrosyl δ-lactone **15**, making it more efficient from the energetic point of view. However, oxygen does not have the ability to perform the nucleophilic attack and there will be no competition under experimental conditions.

4. Conclusions

The main difference between the reactions that were studied is the nature of the nucleophiles. When the nucleophile atom is a sulfur (soft), the reaction has higher activation free energies and the compounds where the nitrogen (hard) is the nucleophile atom have lower activation free energies. This can be justified looking at both TS and comparing the distances. The sulfur nucleophiles, due to the larger size, promote a steric hindrance that difficult their approach to the carbon and, consequently, larger distances are involved in the TS structures when comparing with the nitrogen nucleophiles (Figure 33).

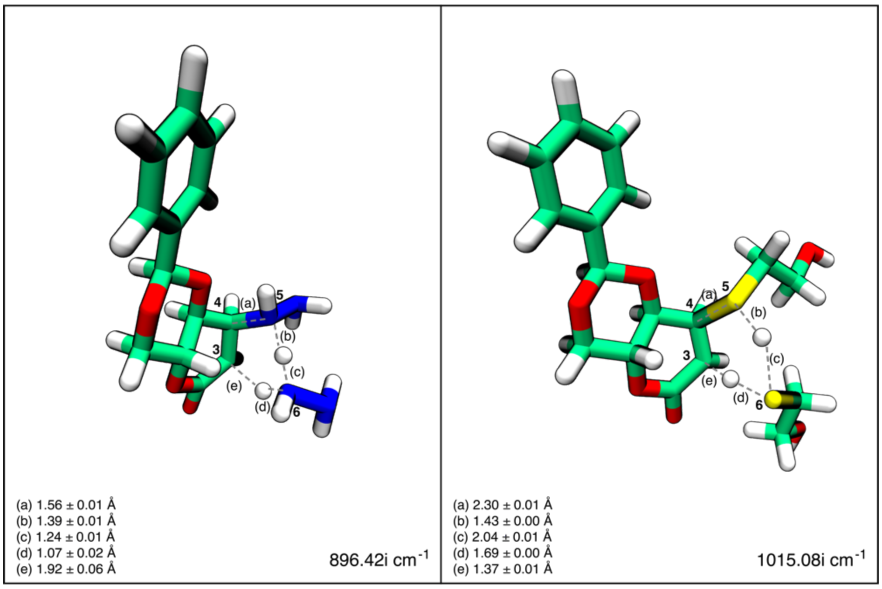


Figure 33 Comparison of the first step of NA@C4 for hard nucleophiles (on the left) and soft nucleophiles (on the right).

The studied reactions bring a great advantage to organic chemistry since they are *stereo*- and *regio*-specific, and allow us to obtain five- and six-membered lactone ring (Figure 34). D-erythrosyl δ-lactone **15** offers a new and rich template for synthesis of several compounds due to the particularities that allow the specificity of these reactions. *Stereo*-selectivity was observed in two kinds of reactions, 1,3-dipolar cycloadditions [36] and 1,4-nucleophilic additions to α,β-unsaturated esters, and can be justified by the proton that precludes the nucleophile molecule to approach from the bottom face. *Regio*-selectivity is justified based on the nature of the nucleophiles molecules. Both, hard and soft nucleophiles follow pathway NA@C4 but, interesting, by different controls. While soft nucleophiles follow this pathway due to thermodynamic control, hard nucleophiles are due to kinetic control.

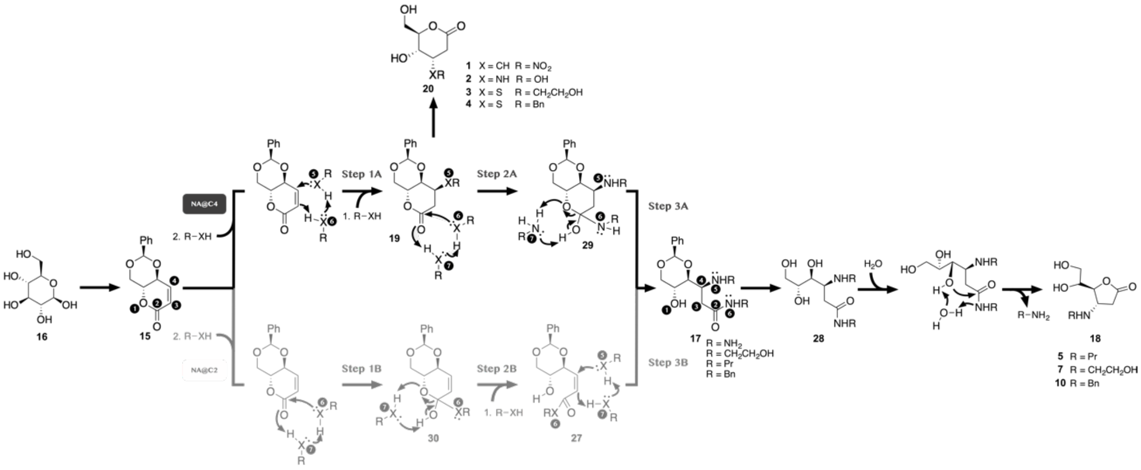

20
1 X = CH R = NO$_2$
2 X = NH R = OH
3 X = S R = CH$_2$CH$_2$OH
4 X = S R = Bn
NA@C4
NA@C2
2. R−XH
Step 1A
1. R−XH
Step 2A
Step 3A
Step 1B
Step 2B
Step 3B
16
15
19
29
30
27
17
R = NH$_2$
R = CH$_2$CH$_2$OH
R = Pr
R = Bn
28
H$_2$O
R−NH$_2$
18
5 R = Pr
7 R = CH$_2$CH$_2$OH
10 R = Bn

*Figure 34 Scheme of the general mechanism studied during this work. At gray, it is the hypothetical pathway NA@C2 that is not followed. Through pathway NA@C4 it is possible to obtain five (**25**) and six (**26**) –membered lactone rings.*

The two different scaffolds, five- and six-membered that are obtained in the end of the reactio, are also related to the nucleophiles nature. While hard nucleophiles require a second step which is possible under the experimental conditions, this step with soft nucleophiles are not capable. This difference translates in two different intermediates before the final product is obtained: an open amide form one (hard nucleophiles) and a closed one (soft nucleophiles). After the acetal cleavage, the final product of the reaction with soft nucleophile product is directly obtained, while the product from hard nucleophiles is obtained only after an extra intramolecular cyclization.

Considering the questions that were raised at the beginning, we were able to answer them and other ones that emerge during the study:

i) The *stereo*-selectivity of these reactions is controlled by the hydrogen of D-erythrosyl δ-lactone **15** that precludes the nucleophile molecule to approach from the bottom face. *Regio*-selectivity is controlled, thermodynamically or kinetically, by the nature of the nucleophile molecules.

ii) Hard and soft nucleophiles control the *regio*-selectivity of the studied reactions. Hard nucleophiles act by a kinetic control, while the soft nucleophiles act by a thermodynamic control.

iii) Once again, the nucleophilic attack at C4 and/or C2 of D-erythrosyl δ-lactone **15**, is controlled based on the nature of the nucleophiles, hard or soft.

iv) The open form is detected in the reaction with hard nucleophiles, since they attack both C4 and C2 and as it is a lactone, a cyclic ester, the opening is facilitated. In the case of soft nucleophiles, they only attack at C4 and do not interacting directly with the ester group of the lactone ring.

v) The reaction mechanism of the five- and six-membered lactones are not competitive since they follow the same pathways, but involve different reactions, that results from the different nucleophiles molecules that are used.

III

2. Biological Assays and Molecular Docking

Science is not only a disciple of reason but, also, one of romance and passion.

Stephen Hawking (1942-)

1. Introduction

In this study, it was our aim to study the inhibitory activity of the D-erythrose derivatives targeting human α- and β-glucosidases and α- and β-galactosidases. These enzymes are involved in the breakage of the glycosidic bonds of carbohydrates and employ similar retaining type of catalytic mechanisms [40 3 41]. These reactions involve the preservation of the anomeric stereochemistry of the carbohydrate and require the presence of two negatively charged amino acid residues in the active site of those enzymes. Interestingly, in the β-glucosidase and β-galactosidase, these amino acid residues are two glutamates whereas in the α-glucosidase and α-galactosidase they are two aspartates. This difference is justified by the difference at the anomeric carbon of the carbohydrates that in the case of the β isomers would place the hydroxyl group far apart from the negatively charged residues of the active site. Nature has resolved this issue placing a glutamate instead of an aspartate in β-glucosidase and β-galactosidase since these amino acid residues have a longer side chain [42 43].

All of these similarities suggest that these enzymes should share many similar among each other. However, the fact is that all of them are quite different. All of them have different three-dimensional structures and different active sites.

In Figure 35, it is shown the shape and nature of the binding pocket of each of these enzymes, together with their active sites. From all the analyzed enzymes, α-glucosidase is the one that has the deepest binding site (13Å away from the protein surface). The smallest binding site is the one from α-galactosidase (7Å away from the protein surface), whereas the ones from β-glucosidase and β-galactosidase have similar dimensions (10Å away from the protein surface). From these four enzymes, the binding site of α-glucosidase is the only one that is populated with several charged and polar amino acids. In all the other enzymes, the walls of the binding site are mainly hydrophobic in particularly the one from β-galactosidase.

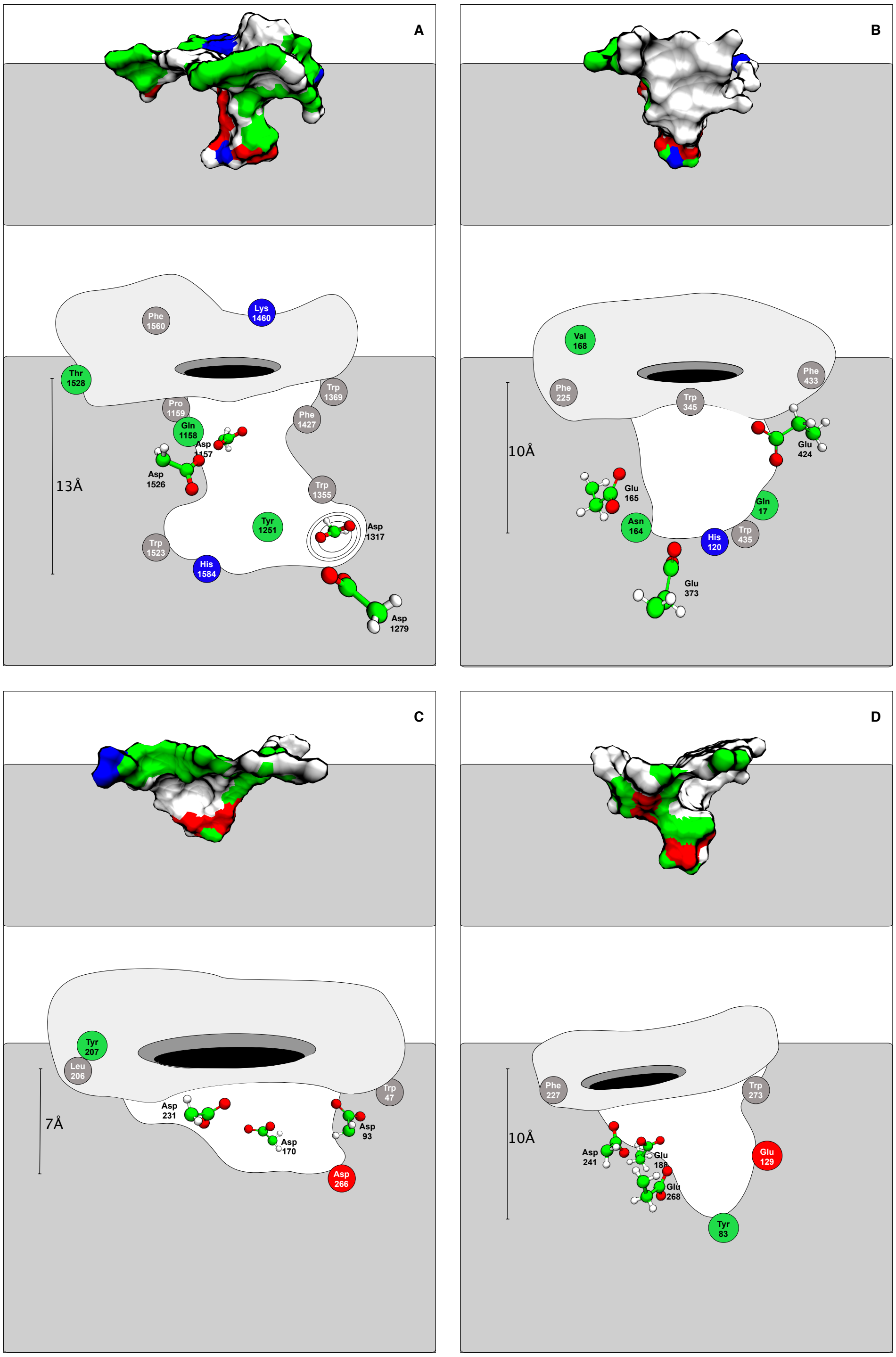

A
Phe 1560
Lys 1460
Thr 1528
Trp 1369
Pro 1159
Gln 1158
Phe 1427
Asp 1157
Asp 1526
13Å
Trp 1355
Tyr 1251
Asp 1317
Trp 1523
His 1584
Asp 1279
B
Val 168
Phe 433
Phe 225
Trp 345
10Å
Glu 424
Glu 165
Gln 17
Asn 164
His 120
Trp 435
Glu 373
C
Tyr 207
Leu 206
Trp 47
7Å
Asp 231
Asp 93
Asp 170
Asp 266
D
Phe 227
Trp 273
10Å
Asp 241
Glu 188
Glu 129
Glu 268
Tyr 83

Figure 35 Shape and nature of the binding pocket of A: α-glucosidase, B: β-glucosidase, C: α-galactosidase and D: β-galactosidase, together with their active sites.

These features have tuned the active sites of each one of these enzymes specific for certain types of substrates. For example, α-glucosidase accept carbohydrates containing α-glucose monomers, β-galactosidase accepts only substrates with β-galactose monomers and so one. Other substrates can also be accepted by these enzymes, as it is the case of several inhibitors that are currently used clinically to impair the function of these enzymes. However, in these cases these compounds should mimic the behavior of the natural substrate of the enzymes, otherwise, the enzyme will not recognize them and they would not reach the active site where they exercise their inhibitory activity.

The D-erythrose derivatives presented here, have many interesting characteristics. First, the substitution of the hydroxyl group from the anomeric carbon of the carbohydrate by a ketone group can allow these compounds to overcome the substrate specificity imposed by the active sites. This means that they can interact with the active site of all of these enzymes and therefore exploit their inhibitory activity there. Additionally, this substitution allows to the compounds mimic the positive charged transition state of natural substrate. Secondly, the long side chains that some of the compounds contain can be important to improve their inhibitory efficiency. Indeed, the binding site of α-glucosidase, β-glucosidase and β-galactosidase are quite deep and therefore this feature may help to improve the inhibitory efficacy of these compounds.

In the following section, it will be presented the biological assays performed with all the fourteen D-erythrose derivatives. This will allow to understand which compounds are active to each enzyme and it will be particularly important to known if the inhibition process is specific for any of the tested enzymes.

2. Methodology

2.1. Model enzymes

The enzymes involved in this study are not easily expressed and therefore, in the biological assays, the same enzymes from different organisms were used instead.

To perform the enzymatic assays, enzymes from commercial sources were used: α-glucosidase from Saccharomyces cerevisiae (G-5003 Sigma-Aldrich, USA), β-glucosidase from Almonds (G-0395 Sigma-Aldrich, USA), α-galactosidase from green coffee (G-8507 Sigma-Aldrich, USA) and β-galactosidase from bovine liver (G-1875 Sigma-Aldrich, USA).

In order to understand if these enzymes differ substantially from the human ones, a sequence alignment of these enzymes was done and the sequence identity and homology was evaluated. All the calculations were performed in an adequate software Blast from NCBI (National Center for Biotechnology Information) website.

In Table 10, are presented the enzymes that share a higher degree of sequence identity and homology to the human ones. The α-glucosidase from *Saccharomyces cerevisiae* is the one that shown higher similarity (56%) to the *Homo sapiens*. For β-glucosidase, *Almonds* and *Homo sapiens* (41%) were also the ones that showed higher similarity. For α-galactosidase, *Coffea arabica* (green coffee) and *Homo sapiens* also showed a good percentage of similarity (45%). For β-galactosidase, *Bovine liver* and *Homo sapiens* showed a strong sequence identity (80%).

Table 10 Identity similarity and homology (in brackets) performed by Protein Blast NCBI for α-glucosidase, β-glucosidase, α-galactosidase and β-galactosidase.

α-glucosidase		**β-glucosidase**		**α-galactosidase**		**β-galactosidase**	
H. sapiens	56%	***H. sapiens***	41%	***H. sapiens***	45%	***H. sapiens***	80%
S. cerevisiae	(77%)	***Almonds***	(58%)	***C. arabica***	(59%)	***B. liver***	(88%)

Although some of these enzymes do not present a high sequence identity, they share high sequence homology. In fact, a visual inspection to the active sites of the related enzymes reveals that they can be almost superimposed.

2.2. Biological Assays

The activity of each enzyme was measured using the respective substrate, which is degraded to *p*-nitrophenol (colorimetric compound – yellow). For α-glucosidase was used the substrate *p*-nitrophenyl-α-D-glucopyranoside (N1377, Sigma-Aldrich, USA); for β-glucosidase was used *p*-nitrophenyl-D-β-glucopyranoside (N7006, Sigma-Aldrich, USA); for α-galactosidase was used *p*-nitrophenyl-α-D-galactopyranoside (N0877, Sigma-Aldrich, USA) and for β-galactosidase was used *p*-nitrophenyl-β-D-galactopyranoside (86004-4, Sigma-Aldrich, USA).

Varioskan Flash 96-well microplate reader (Thermo Scientific, USA) and Synergy HT reader (BioTek Instruments Inc., USA) were used for absorbance measuring.

Enzymatic assays were conducted in a total volume of 100 µL reaction mixture, containing 70 µL of phosphate buffer (50 mM, pH= 6.8), 10 µL of test compound (0.5 mM) and 10 µL of enzyme (0.2 units/mL). As control, the test compound was replaced with phosphate buffer. The contents were mixed, pre-incubated for 10 min at 37 ºC and pre-read at 400 nm (T0). The reaction was initiated by the addition of 10 µL of the correspondent substrate (0.5 mM) and then incubated at 30 ºC. After 10 min (T1), absorbance was measured at 400 nm. All the experiments were carried out in three independent trials, each one in triplicate.

After subtracting T0 to T1, absorbance was obtained for the experimental group (Abs (E)) and control group (Abs (C)). The percentage of inhibition was calculated by the following Equation 18:

$$Inhibition\ (\%) = \frac{Abs\ (C)\text{-}Abs\ (E)}{Abs\ (C)} \qquad \text{(Equation 18)}$$

Compounds with more than 65% of inhibition, were submitted to IC_{50} tests. IC_{50} values were calculated varying the concentration of the test compound. All the calculations were realized on GraphPad Prism 5 and presented as mean value ± standard error of the mean (SEM).

2.3. Molecular Docking

The proteins used for the molecular docking were the ones from the human origin. For α-glucosidase and β-glucosidase the enzymes with the PDB code 3TOP {Guce:2011gl} and 2E9L {Ohto:2012kv} were used, respectively. For the α-galactosidase and β-galactosidase the enzymes with the PDB code 3S5Z {Guce:2011gl} and 3THC {Ohto:2012kv} was used, respectively. The enzymes were selected based on the resolution of the protein structure and whenever the enzymes exist co-crystallized with inhibitors or the natural substrate. This last parameter is important since these structures can then be used in the validation process of the molecular docking protocol.

All the compounds were studied using the molecular docking software AutoDock [44] and the vsLab plug-in [45]. The structure of the enzymes was protonated at pH 7 using the MolProbity software (molprobity.biochem.duke.edu). All the water molecules were deleted from the structure as well as any other organic molecule. The ligands were builted with GaussView, protonated at physiological pH and optimized with gaussian09 (HF/6-31G(d)). In the docking process, the Lamarckian genetic algorithm (LGA) was used. The number of generations, energy evaluations, population size, and docking runs was set to 27000, 2500000, 150, and 100, respectively. The types of atomic charges were taken as Kollman-all-atom for the protein and Gasteiger for the compounds. The final solutions were retrieved from the molecular docking process according to the criteria of interacting energy.

For each enzyme, different areas from the active site were evaluated in the docking protocol. For α-glucosidase, the area used to dock the ligands was center on: x=-30.78, y=34.35 and z=26.64, and with the following dimensions: width=16.83, height= 16.83 and depth=19.38; fitting points: 0.170. For β-glucosidase, the box used to dock the ligands was center on: x=4.83, y=39.87 and z=-6.12, and with the following dimensions: width=18.99, height=19.87 and depth=19.77; fitting points: 0.170. For α-galactosidase, the box used to dock the ligands was center on: x=30.14, y=40.30 and z=15.98, and with the following dimensions: width=18.21, height=18.27 and depth=18.62; fitting points: 0.170. For β-galactosidase, the box used to dock the ligands was center on: x=-5.63, y=-7.58 and z=7.36, and with the following dimensions: width=14.05, height=18.70 and depth=18.36; fitting points: 0.170

In order to evaluate if the molecular docking was giving results accordingly to the experimental ones, a validation process was carried out. This protocol involves the docking of molecules whose co-crystallized X-ray structure is already known (Figure 36). In the case of α-glucosidase we tested the acarbose present in the PDB file 3TOP. For β-glucosidase, we tested the β-D-glucose present in the PDB file 2E9L. For α-galactosidase we tested the α-D-galactose and for β-galactosidase we tested the β-D-galactose, present in the PDB file 3S5Z and PDB file 3THC, respectively.

In general, the results were quite satisfactory and an RMSD below 3.5Å (except for acarbose) was obtained between the ligand conformation inside the active site from the molecular docking and the ones observed in the co-crystalized X-ray structures.

In all cases, the binding poses of ring A and ring B (when more than one ring is available) of the compounds that were obtained from the molecular docking could be almost superimposed to what is observed in the co-crystallized X-ray structures. Significant deviations are observed in the compounds that have more than 2 rings in their structure (ring C and ring D, which account for an RMSD of 6.58 A). Such behavior was expected since this part of the molecule does not completely fulfill the binding pocket, and therefore a higher displacement should be observed. This is evident for example on ring D of acarbose.

Some displacements are also observed in the orientation of the hydroxyl groups of the compounds. This occurs because during the molecular docking protocol the protein remains fixed and only the ligand is flexibilized. However, the hydroxyl groups that are important in catalytic process, and interact with the negatively charged residues of the active site, are in most of the cases almost superimposed.

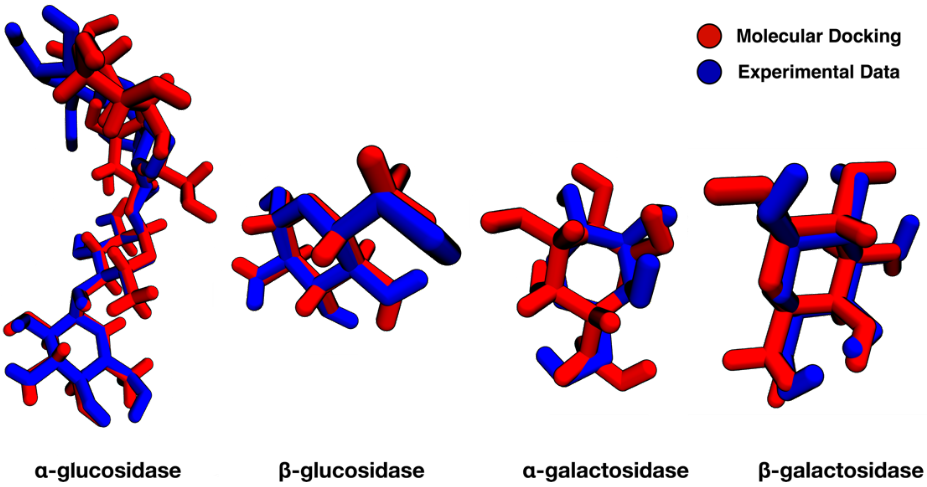


Figure 36 Overlap between molecular docking solution and the conformation of the ligand available on the co-crystalized X-ray structure. The co-crystallized structure of α-glucosidase contains acarbose (PDB code 3TOP, RMSD: 6.58Å), and in β-glucosidase the PDB structure containing β-D-glucose (PDB code 2E9L, RMSD: 3.32Å) was used. The co-crystallized structure of α-galactosidase contains α-D-galactose (PDB code 3S5Z, RMSD: 3.27Å), and in β-galactosidase the PDB structure containing β-D-galactose (PDB code 3THC, RMSD: 3.27Å) was used.

2.3.1. Tunnel shape and SASA calculations

The images of the tunnels were made measuring the free volume of the binding site of each isoform using the software VolArea [46]. The solvent accessible surface area (SASA) corresponds to the surface area of a molecule that is accessible to a solvent. In this article, we used it to measure the potential area of the binding site that can interact with a potential substrate. In all calculations, a probe of 1.4 Å was used.

3. Results and Discussion

3.1. Biological Assays

The biological assays addressed to α-glucosidase, β-glucosidase, α-galactosidase and β-galactosidase, have shown that from the fourteen D-erythrose derivatives that were tested, five could selectively inhibit α-glucosidase and β-galactosidase (Figure 37).

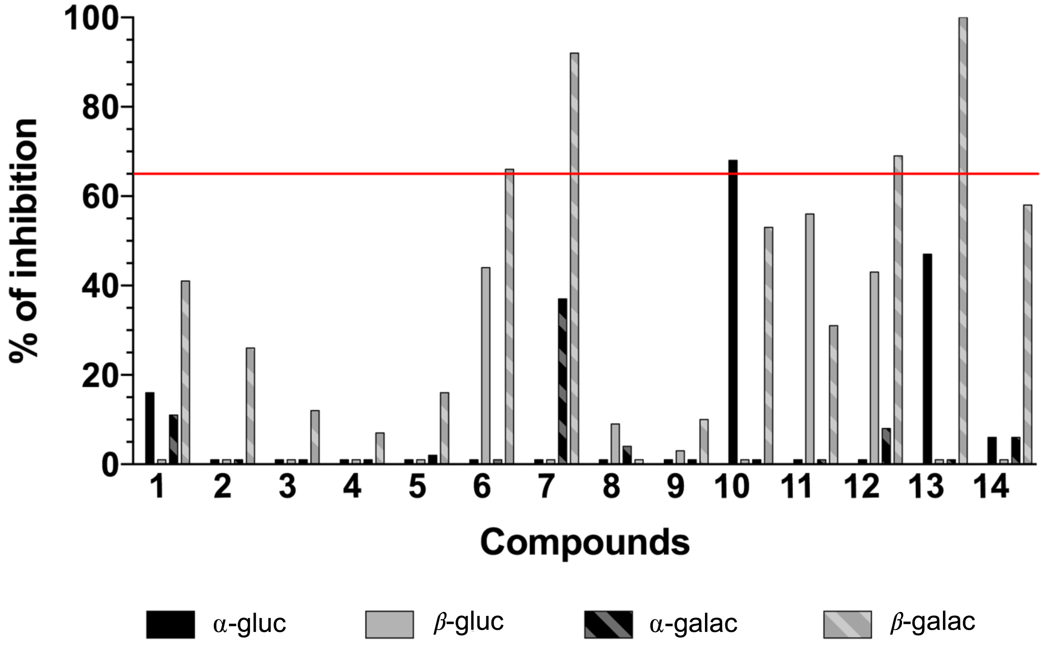


Figure 37 Inhibition rate of the compounds in study for the enzymes α-glucosidase (black), β-glucosidase (gray), α-galactosidase (black with gray lines) and β-galactosidase (gray with light gray lines).

All the other compounds did not present any inhibitory activity against any of the tested enzymes. From the active compounds, one of them was specific for α-glucosidase (compound **10**) and the remaining four, for β-galactosidase (compounds **4**, **6**, **11** and **12**). In Table 11 are depicted the IC50 values of the compounds that showed inhibitory activity targeting α-glucosidase, β-glucosidase, α-galactosidase and β-galactosidase.

The obtained results indicate that five D-erythrose derivatives are promising compounds to specifically inhibit α-glucosidase and β-galactosidase. This is a very important feature since the compounds that are currently used clinically to inhibit α-glucosidase and β-galactosidase lack this feature and often lead to the appearance of many side effects or, in a drug development stage, preclude their use in clinical trials.

These results also show that small modifications at the D-erythrose substituents can easily turn a good and specific inhibitor of one enzyme into an inactive compound or a good inhibitor of another enzyme. These specific molecular features that are behind such behavior are very important to understand since they can provide important clues for the development of new and more specific compounds as well as to optimize the pre-existing ones.

Table 11 IC50 values of the compounds that showed inhibitory activity targeting α-glucosidase, β-glucosidase, α-galactosidase and β-galactosidase. All the values are in μM and with ± SEM.

ID	α-glucosidase	β-galactosidase
4		247.7 ± 42.5
6		206.2 ± 53.2
10	180.6 ± 31.8	
11		213.0 ± 58.4
12		205.2 ± 54.9

In order to understand how the compounds, interact with the binding site of each of the studied enzymes and understand how small changes on the D-erythrose derivatives can promote different inhibitory activities, molecular docking studies were performed.

3.2. Molecular Docking

A. α-glucosidase

Human Maltase-Glucoamylase (MGAM) is present in human brush border small intestine and it is responsible for degradation of starch and disaccharide into α-D-glucose. This enzyme is composed by a double active site, nominated C-terminal (PDB code 3TOP) and N-terminal (PDB code 2QMJ). C-terminal MGAM (C-MGAM) has a higher affinity for longer saccharides and is responsible for its degradation, while N-terminal MGAM (N-MGAM) is responsible for degradation of smaller saccharides since it has higher affinity for these. Considering results from the Blast, size and number of rings of the compounds in study, we chose to perform molecular docking studies in C-MGAM. However, we proceed to the analysis of both active sites and they are very similar.

C-MGAM (PDB code 3TOP) is a dimer, each one containing one independent active site. Each active site is deeply buried 13Å away from the protein surface of each monomer. The binding site is a thin tunnel that diverges in two larger cavities. The active site of this enzyme is located at the middle of the binding pocket right above the region where the two cavities are formed (Figure 35).

The binding pocket is populated with several polar (Gln1158, Tyr1251, Lys1460, Thr1528 and His1584) and charged amino acid residues (Asp1157, Asp1279, Asp1317 and Asp1526), whose side chains are pointing towards the center of the binding pocket.

The molecular docking results have shown that the majority of the compounds bind inside the binding pocket, close to the active site region where the two aspartates, Asp1526 and Asp1157 are located. These interactions are important since these amino acid residues are directly involved in the cleavage of the glycosidic bonds and therefore those interactions mimic what happens with the natural substrate. The only exceptions are compounds **1** and **2** that due to their small volume, lack a specific binding pose inside the binding pocket. The other exception is compound **14** whose entrance inside the binding pocket is precluded due to the large volume of the compound.

However, the biological assays indicate that this feature is not enough since only one compound presents an acceptable inhibitory activity against α-glucosidase.

The molecular docking results have shown that the compounds containing polar groups attached to the lactones derivatives (compounds **3**, **7**, **8**, **9**), are the ones in which the hydrogen bonds with the two Asp residues are the less efficient. This happens because these compounds become folded inside the binding site. This last feature compromises the correct alignment of the compounds inside the active site, in particular with the two catalytic Asp residues. The binding pose of the compounds with flexible hydrophobic substituents (compound **4**, **5** and **6**) also present a similar behavior and this may explain their lack of inhibitory activity.

Compounds **10**, **11**, **12** and **13** present a phenyl group attached and the only difference among these compounds is the group that is attached to the *para* position of each phenyl group. Compound **10** has a hydrogen in this position, compound **11** a chloride, compound **12** a fluorine and compound **13** a CF_3 group. This bulky group is aligned in the entrance of the binding pocket due to π stacking with Phe residues, forcing the lactone ring and attached hydroxyl groups to interact with the charged residues at the bottom of the binding pocket. The amine group is aligned with the catalytic aspartates, which appears to be favorable for inhibitory activity. When the carbonyl from lactone ring is interacting with Asp1279, amine group is correctly aligned with catalytic Asp1526 and the activity is improved (Figure 38A), which happen with compound **10**. In the case of the compounds with a substituent in *para* position, the lactone ring accommodates in the binding pocket in a different way, losing crucial interactions (Figure 38B).

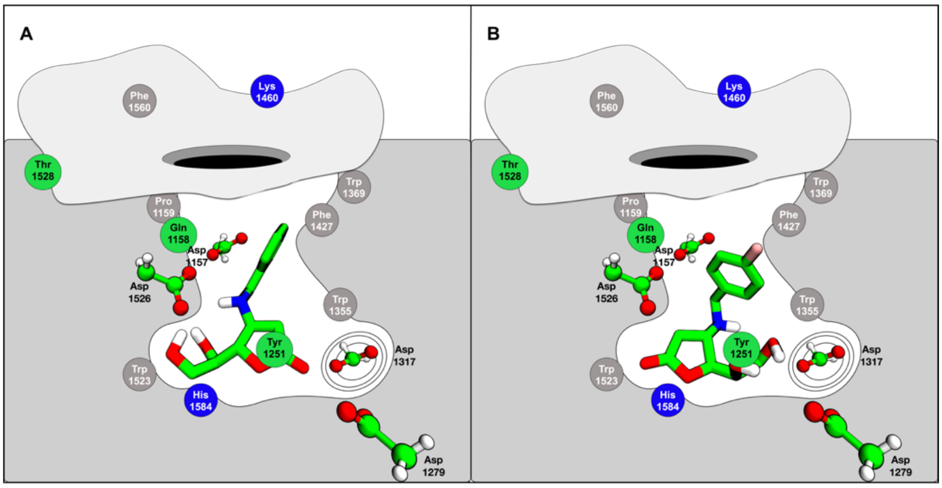

A
Phe 1560
Lys 1460
Thr 1528
Trp 1369
Pro 1159
Gln 1158
Phe 1427
Asp 1157
Asp 1526
Trp 1355
Tyr 1251
Asp 1317
Trp 1523
His 1584
Asp 1279
B
Phe 1560
Lys 1460
Thr 1528
Trp 1369
Pro 1159
Gln 1158
Phe 1427
Asp 1157
Asp 1526
Trp 1355
Tyr 1251
Asp 1317
Trp 1523
His 1584
Asp 1279

*Figure 38 Binding poses of **A**: compound **10** which has inhibitory activity against α-glucosidase; and **B**: compound **12** which does not have inhibitory activity against α-glucosidase.*

B. β-glucosidase

β-glucosidase from *Homo sapiens* (PDB code 2E9L) is a monomer. The active site of this enzyme is found 10Å away from the protein surface at the bottom of the binding pocket. The binding pocket is a thin and long tunnel that is populated by several polar (Gln17, His120, Asn164 and Glu424) and charged amino acid residues (Glu165 and Glu373) (Figure 35).

All the D-erythrose derivatives that were tested against this enzyme showed no inhibitory activity in the biological assays. These results go in line with the molecular docking results that shown that these compounds cannot interact efficiently with the active site of this enzyme through the sugar-like moiety. Instead, it becomes lodged in the majority of the cases above the active site region, which decreases considerably the binding affinity of them. This happens because the active site is in the deepest region of the binding pocket (as we can see in Figure 39 with the natural inhibitor) and the presence of bulkier or flexible substituents attached to the sugar-like moiety precludes any interaction of the latter one with it.

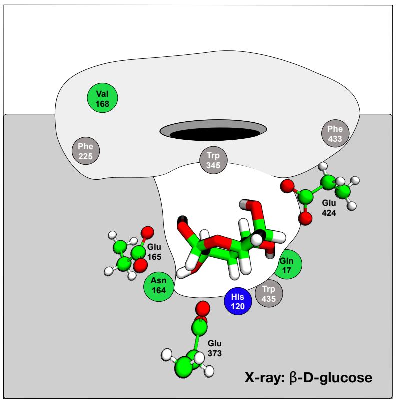


Figure 39 Natural inhibitor of β-glucosidase, β-D-glucose, showing the deep active site of the enzyme.

C. α-galactosidase

α-galactosidase from *Homo sapiens* (PDB code 3S5Z) is a dimer. Each monomer contains one binding site that contrarily to the other studied enzymes is very superficial and highly exposed to the solvent, as it is presented in Figure 35. The binding site has almost a hydrophobic nature, apart from the bottom region, where the active site is located and three aspartate residues are found, Asp93, Asp170 and Asp231. Two of these residues are involved in the cleavage of the glycosidic bonds (Asp170 and Asp231).

The biological assays revealed that none of tested D-erythrose derivatives have an efficient inhibitory activity against α-galactosidase, i.e., all the compounds are inactive. The molecular docking results indicate that there are two reasons that can explain such behavior. First, the binding site is very small and in the majority of the cases, the compounds either do not fit efficiently inside the binding pocket or become aligned along the protein surface. Secondly, since the active site is located in the bottom region of the binding pocket and the compounds have attached to their main scaffold bulkier groups, the sugar-like moiety of the compounds cannot interact efficiently with the aspartate residues that are catalytic important for the cleavage of the glycosidic bonds, and were shown to be important for the inhibitory activity of α-galactosidase. This is evident in the co-crystallized X-ray structure of α-galactosidase with the inhibitor α-D-galactose (Figure 40). In this structure, it is clear that the binding site can only accept compounds that are composed by only sugar-like structure that do not contain bulkier groups attached to the main structure.

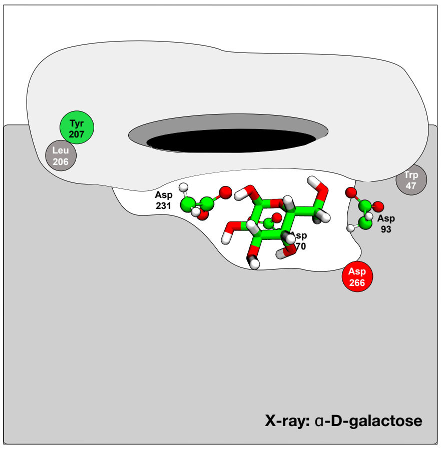


Figure 40 Natural inhibitor of α-galactosidase, α-D-galactose, showing the location of active site and its exposure to solvent.

D. β-galactosidase

β-galactosidase from *Homo sapiens* present in the PDB structure with the code 3THC is a tetramer. The active site presented in each monomer is buried 10Å away from the protein surface along connected to the solvent through a tight binding pocket similar to what is found in β-glucosidase. (Figure 35).

The binding pocket is mainly hydrophobic, apart from the region of the active site that is populated by one polar residue (Tyr83) and three charged amino acid residues (Glu129, Glu188 and Glu268). Glu188 and Glu268 are the ones that are involved in the cleavage of the glycosidic bonds.

One key feature that differentiate the binding pocket of β-glucosidase and β-galactosidase is the location of the active site. While in the β-glucosidase, the active site is located in the bottom of the binding pocket, in the β-galactosidase it is located at half of it, similar to what is found in the α-glucosidase.

The biological assays revealed that four compounds could efficiently inhibit this enzyme: compound **4**, **6**, **11** and **12**. The molecular docking results have shown that most of these compounds are anchored to binding pocket due to the formation of several hydrogen bonds between the hydroxyl groups of the sugar-like moiety and Glu188. The interactions with Glu188 are important since this amino acid residue is directly involved in the cleavage of the glycosidic bonds and therefore those interactions mimic what happens with the natural substrate.

Compounds **10-13** present similar binding poses. However, only compounds **11** and **12** showed a significant inhibitory activity. The only difference among these compounds is the group that is attached at the *para* position of each phenyl group. Compound **10** has a hydrogen atom in this position, compound **11** a chloride, compound **12** a fluorine and compound **13** a CF_3 group. The molecular docking results reveal that the presence of chloride and fluorine (compounds **11** and **12**) in this position stabilize, somehow, the interaction between these compounds and the Glu and Asp residues, and an extra interaction with Tyr83 and/or Glu129 is accomplished by the lactone ring (Figure 41A). In compounds **10** and **13**, these extra interactions are lost and the inhibitory activity is compromised. This lack of extra interaction dislocates the lactone ring and amine group which in turn compromises their binding affinity and therefore their inhibitory activity (Figure 41B).

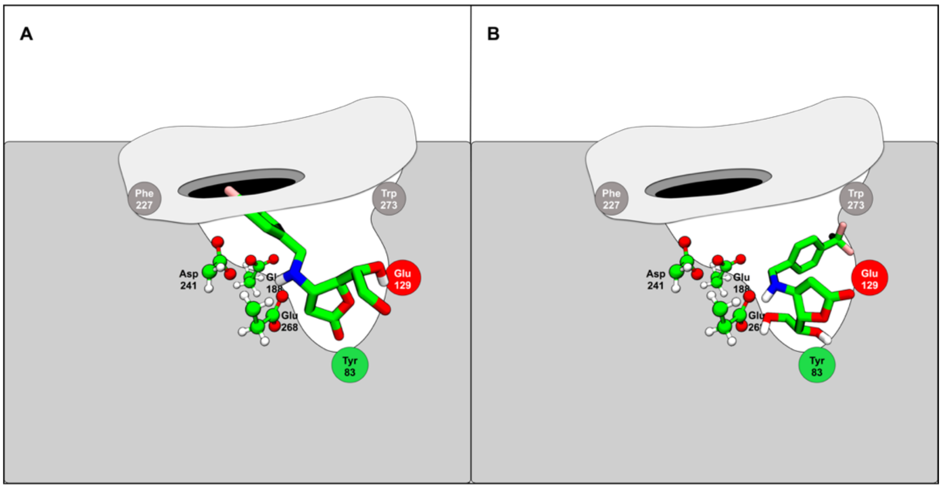


*Figure 41 Binding poses of **A**: compound **11** which has inhibitory activity against β-galactosidase; and **B**: compound **13** which does not have inhibitory activity against β-galactosidase.*

4. Conclusion

In this study, the inhibitory activity of 14 D-erythrose derivatives were tested against four glycosidases: α-glucosidase, β-glucosidase, α-galactosidase and β-galactosidase. These compounds have a carbonyl group attached to the anomeric carbon of the sugar-like moiety that mimic the transition state structure involved in the glycosidic bond cleavage on the wild type enzymes. This feature also allows the compounds to be substrates of either α- and β-glycosidases, overcoming limitation of the substrate specificity imposed by the orientation of the hydroxyl group attached to the anomeric carbon.

The biological assays performed in this study revealed that from the fourteen compounds that were tested, one compound has inhibitory activity against α-glucosidase and four against β-galactosidase at μM scale. These inhibitors are specific for each of these enzymes, which means that they can be used to selectively inhibit each of these enzymes, without affecting the activity of the other three enzymes. The biological assays also indicate that none of the tested compounds present inhibitory activity against β-glucosidase and α-galactosidase.

The molecular docking results revealed that the inhibitory activity of the tested compounds is mainly dependent on the specific interaction that the sugar-like moiety of the D-erythrose derivatives play with the negatively charged aspartate (in α-glucosidase or α-galactosidase) or glutamates (β-glucosidase or β-galactosidase) residues that are present in the active site of the enzymes. Such type of interactions is only observed in α-glucosidase and β-galactosidase because, the active site is located at the middle of the binding pockets, and there is enough room to accommodate the compounds in it. In the case of β-glucosidase and α-galactosidase, the binding pocket is located at the bottom of the binding pocket where such type of interactions is not possible (Figure 42). This happens because the sugar-like moiety of the tested compounds is, in most the cases, linked to a bulkier group and therefore impairs the proximity with the active site of these enzymes. In the case of α-galactosidase the binding site is also very small and highly exposed to the solvent, two conditions that also precludes an efficient binding of the tested compounds in the active site of this enzyme.

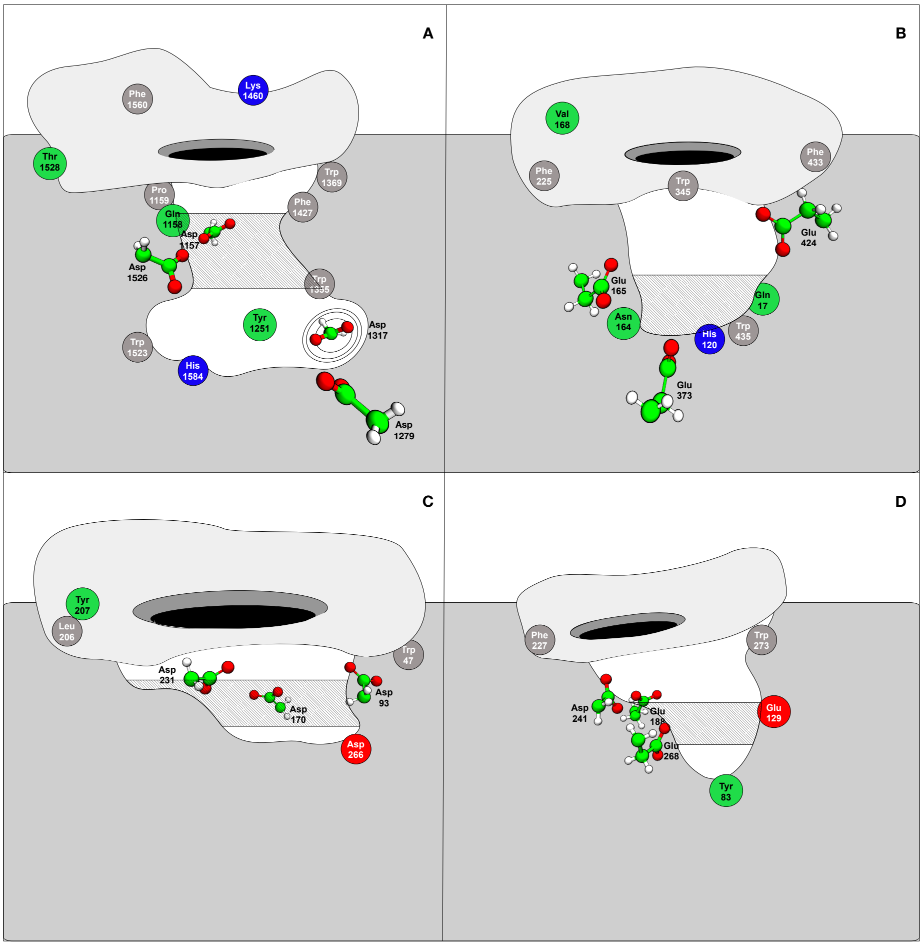

A
Phe 1560
Lys 1460
Thr 1528
Trp 1369
Pro 1159
Gln 1158
Phe 1427
Asp 1157
Asp 1526
Trp 1355
Tyr 1251
Asp 1317
Trp 1523
His 1584
Asp 1279
B
Val 168
Phe 433
Phe 225
Trp 345
Glu 424
Glu 165
Gln 17
Asn 164
His 120
Trp 435
Glu 373
C
Tyr 207
Leu 206
Trp 47
Asp 231
Asp 93
Asp 170
Asp 266
D
Phe 227
Trp 273
Asp 241
Glu 188
Glu 129
Glu 268
Tyr 83

Figure 42 Binding pocket and localization of the active site of A: α-glucosidase, B: β-glucosidase, C: α-galactosidase and D: β-galactosidase.

III

3. Synthesis driven by theoretical results

You shouldn't turn away from the evidence just because it points in a direction you don't want to go.

Stephen King, "End of watch" 2016

1. Introduction

The work developed in this thesis have shown that theoretical and computational chemist are powerful tools that can aid organic experimentalists to improve their knowledge about the mechanism of the reactions that they use in the lab. This knowledge can then be used to improve the yields of the existing chemical routes but also to test and program the synthesis of new compounds. This work is an example of this fact.

At the beginning of the present work, it was thought that the final products of the reactions that were studied with hard nucleophiles were lactams and with soft nucleophile lactones. All the experimental work that was carried out, NMR and mass spectroscopy were pointing to this end but, the final X-ray structure of one of the "lactams" revealed that it was instead a five–membered ring lactone.

During this period of the time, we were also focused on studying the reaction of the hard nucleophiles with the lactone **15** and from which the lactam would be generated (Figure 43).

We found that this reaction only occurs after the formation of the reaction intermediate **17** that adopts an open configuration and the cleavage of the acetal under acid catalysis. The formation of the lactams would be complete through the intramolecular cyclization, in which there would be a nucleophilic attack of nitrogen N6 to carbon C4a, instead of the nucleophilic attack of oxygen Oa to carbon C2 from which the five-membered lactones are obtained.

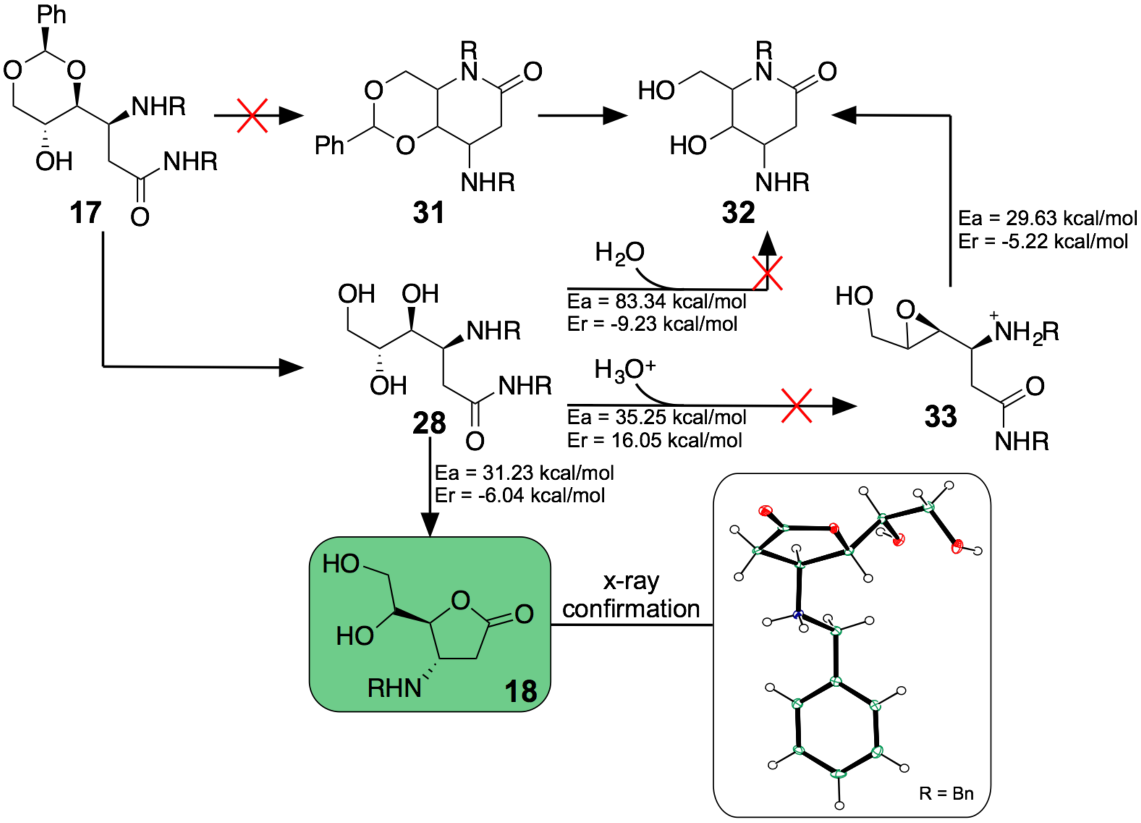

17
31
32
33
28
18
H2O
Ea = 83.34 kcal/mol
Er = -9.23 kcal/mol
H3O+
Ea = 35.25 kcal/mol
Er = 16.05 kcal/mol
Ea = 29.63 kcal/mol
Er = -5.22 kcal/mol
Ea = 31.23 kcal/mol
Er = -6.04 kcal/mol
x-ray
confirmation
R = Bn

*Figure 43 Representation of the studied pathway addressed to the intramolecular cyclization of the product **24**, from the NA@C4 with hard nucleophiles.*

The theoretical and computational results revealed that the formation of the lactam could be done by two mechanisms. The first mechanism, involves the direct cyclization of the compound **28**. The second mechanism involves the formation of epoxide and only afterwards the cyclisation process takes place.

The theoretical and computational results have shown that the energetic profile of the first mechanistically hypothesis is very unfavorable. The direct cyclization of compound **28** in the presence of water requires an activation free energy of 83 kcal/mol and it is exergonic in -9 kcal/mol.

The second mechanism required two steps and it only occur in the presence of acid instead of water. First occurs the formation of an epoxide reaction intermediate. Only afterwards the intramolecular cyclization takes place. The energetic profile of the second step was quite reasonable. It required an activation free energy of 30 kcal/mol and the reaction was exergonic in -5 kcal/mol. The energetic profile of the first step was however very high. The formation of the epoxide required an activation free energy of 35 kcal/mol and the reaction was endergonic in 16 kcal/mol.

Both mechanisms involve very high activation free energies (above 35 kcal/mol) which would turn them impossible to take place even on the most extreme experimental apparatus.

We spend several months trying to find other mechanism that could lead to the formation of the lactams, but with no success. The energetic profiles of these mechanism were very unfavorable and required always activation free energies above 35 kcal/mol.

At some point, we got the X-ray structure of the product of the reaction with hard nucleophiles that showed a five-membered lactone ring instead of a six-membered lactam ring. From this point, we search for all the mechanistic possibilities that would lead to the formation of the five-membered lactone ring. The energetic profile of this reaction was much more favorable than the ones that were obtained for the formation of the lactam, and with lower activation free energies, which would allow them to occur experimentally. In addition, this reaction is much simpler, since it only requires one step instead of two.

This small example illustrates that sometimes we shouldn’t turn away from the evidence that is provided by the theoretical calculations just because it points in a direction that is not observed experimentally. In this case, it was clearly indicating that the formation of lactams would be, experimentally, very difficult to obtain, if not impossible.

However, these results can now be used to see how the formation of lactones can persuade. Different groups can now be introduced in different positions of the reaction intermediate **28** and the calculated activation and reaction free energies can predict if their formation is feasible or not.

As it was initially thought to have lactams as final products, molecular docking studies were conducted for these compounds and the results will be presented below.

2. Molecular docking studies addressed to lactams

The lactams that were studied are illustrated in Figure 44.

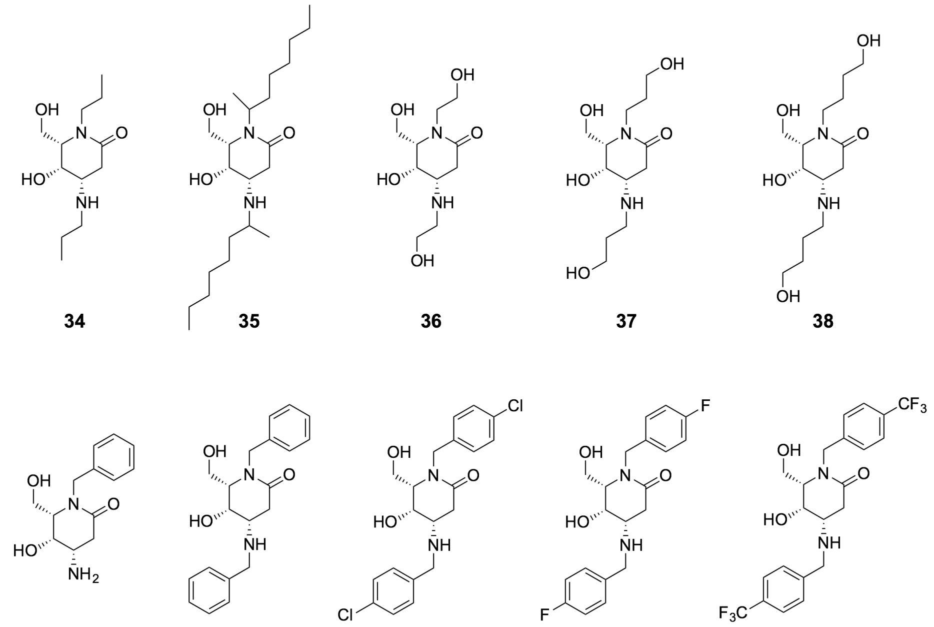


Figure 44 Schematic representation of the studied lactams.

A. α-glucosidase

The molecular docking results have shown that, like in the lactones, the compounds containing polar groups attached to the lactams derivatives (compounds **36**, **37, 38**), are the ones in which the hydrogen bonds with the two Asp residues are the less efficient. This happens because these compounds become folded inside the binding site. This last feature compromises the correct alignment of the compounds inside the active site, in particular with the two catalytic Asp residues. The binding pose of the compounds with flexible hydrophobic substituents (compound **34** and **35**) also present a similar behavior and this may explain their lack of inhibitory activity (Figure 45).

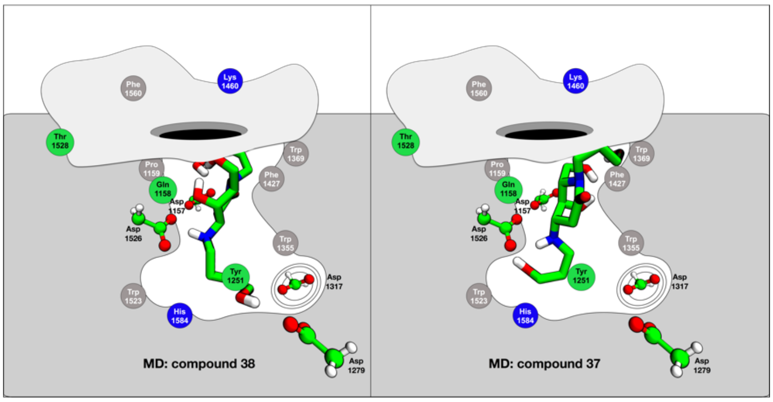
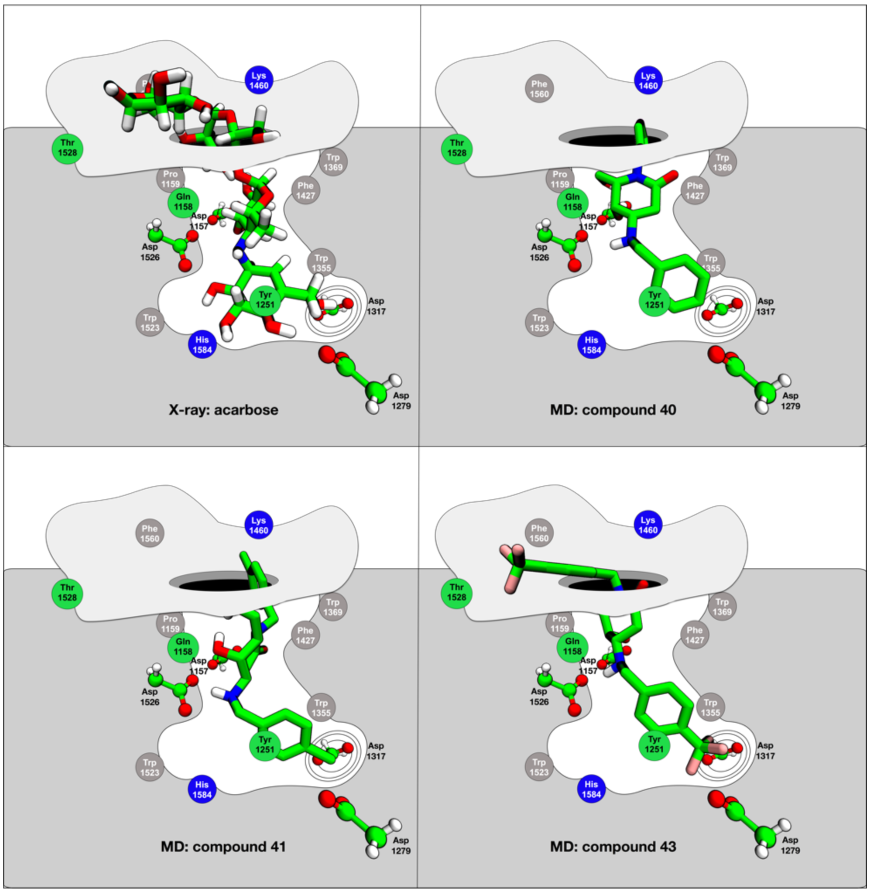

Phe 1560
Lys 1460
Thr 1528
Trp 1369
Pro 1159
Gln 1158
Phe 1427
Asp 1157
Asp 1526
Trp 1355
Tyr 1251
Asp 1317
Trp 1523
His 1584
MD: compound 38
Asp 1279
Phe 1560
Lys 1460
Thr 1528
Trp 1369
Pro 1159
Gln 1158
Phe 1427
Asp 1157
Asp 1526
Trp 1355
Tyr 1251
Asp 1317
Trp 1523
His 1584
MD: compound 37
Asp 1279

Figure 45 Demonstrative image of the two types of interaction between compounds 37 and 38 and the α-glucosidase binding site.

The only compounds that are correctly aligned in the binding site and establish a close interaction with the two Asp residues of the active site are the compounds **40**, **41**, **42** and **43**. All of these compounds present similar binding poses and are able to fulfill almost completely the binding site cavity. The only difference among these compounds is the group that is attached to the *para* position of each phenyl group. Compound **11** has a hydrogen in these positions, compound **12** a chloride, compound **13** a fluorine and compound **14** a CF_3 group. The molecular docking results reveal that the presence of chloride and fluorine in these positions pushes the compounds towards the bottom of the binding site in the direction of the larger pocket where Asp317 is located. This effect turns the interaction between these compounds and the catalytic aspartates less efficient (longer hydrogen bonds), which in turn could compromise their binding affinity. In the case of compound **14**, the volume of the CF_3 group precludes such behavior and becomes caught in a similar position to the one that is occupied by compound **11** (Figure 46).


X-ray: acarbose

MD: compound 40

MD: compound 41

MD: compound 43

Lys 1460
Phe 1560
Thr 1528
Trp 1369
Pro 1159
Phe 1427
Gln 1158
Asp 1157
Asp 1526
Trp 1355
Tyr 1251
Asp 1317
Trp 1523
His 1584
Asp 1279

Figure 46 A: Compounds with the best score against α-glucosidase. (ACR, 40, 41 and 43)

B. Remaining enzymes

Considering the scores obtained by the compounds on the remaining enzymes under study, they will not be promising inhibitors. By the analysis of the molecular docking poses, it is possible to observe that the compounds fail in the essential interactions described in the previous chapter.

C. Conclusion

Molecular docking results of lactams reveal that they can be considered as promising specific inhibitor for α-glucosidase. Since these compounds have shown to have a high degree of interaction with α-glucosidase but, not with the remaining enzymes. With this results in hands, it could be interesting to try to synthetize and test these compounds as specific inhibitors of α-glucosidase.

The molecular docking results also reveal that the presence of bulkier groups attached to the sugar-like moiety of the D-erythrose derivatives is an important feature for the inhibitors of the α-glucosidase. These groups allow the compounds to fulfill almost the entirely cavity of the binding pocket and to align the sugar-like moiety of the tested compounds in relation to the negatively charged residues of the active site. However, they should not have polar substituent attached because, when the compounds have polar groups attached to the bulkier groups, they point towards the cavities of the binding site and interact with the charged residues. This promotes the sliding of the compounds to the bottom region of the binding site and displaces the sugar-like moiety away from the location where the catalytic residues are located. This precludes an efficient interaction of these compounds with the active site of α-glucosidase and therefore should declines their inhibitory potency.

The results can now be used to develop new inhibitors of α-glucosidase with optimized inhibitory activity based on model of Figure 47.

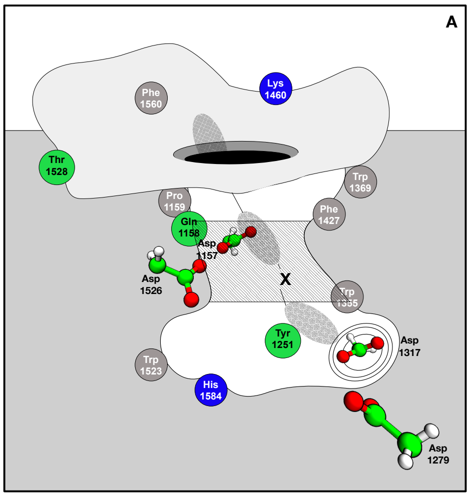


Figure 47 Model of what would, likely, be good inhibitors for α-glucosidase.

IV

CHAPTER IV

Conclusion

> *Science cannot solve the ultimate mystery of nature.*
> *And that is because, in the last analysis, we ourselves*
> *are a part of the mystery that we are trying to solve.*
>
> *Max Planck (1858-1947)*

In the present work, two different types of computational methodologies were employed: quantum mechanics, molecular docking. It was also performed the biological evaluation of the inhibitory activity of several five- and six-membered ring lactones against α-glucosidase, β-glucosidase, α-galactosidase and β-galactosidase.

Quantum mechanical calculations were employed to study the origin of the stereo and regio selectivity of the reactions that lead to the formation of five and six-membered ring lactones. This mechanism brings an advantage over the 1,3-dipolar cycloaddition reactions. 1,3-dipolar cycloaddition were described as *stereo*-selective reactions, but not *regio*-selective. However, in the reaction with hard and soft nucleophiles, it was shown that they are *stereo*- and *regio*-selective (Figure 49). The *stereo*-selectivity of both reactions, 1,3-dipolar cycloaddition and nucleophilic attack by hard and soft nucleophiles, is justified by an intrinsic characteristic of D-erythrosyl δ-lactone **15**, causing the attacks to occur on the upper face. *Regio*-selectivity is justified based on the nature of the attacking molecule. Both, hard and soft nucleophiles follow pathway NA@C4 but, interesting, by different controls. While soft nucleophiles follow this pathway due to thermodynamic control, hard nucleophiles are due to kinetic control. Nucleophiles have an extra difference, while hard nucleophiles attack in C4 followed by the attack in C2, generating an open amide form intermediate **17**, soft nucleophiles attack only in C4 and generate a close lactone form intermediate **19**. After the acetal cleavage, soft nucleophile product **20** is obtained, while the product **18** from hard nucleophiles is obtained after an extra intramolecular cyclization.

The full energetic profile of the reaction that lead to the formation of five and six-membered lactones are depicted in Figure 48.

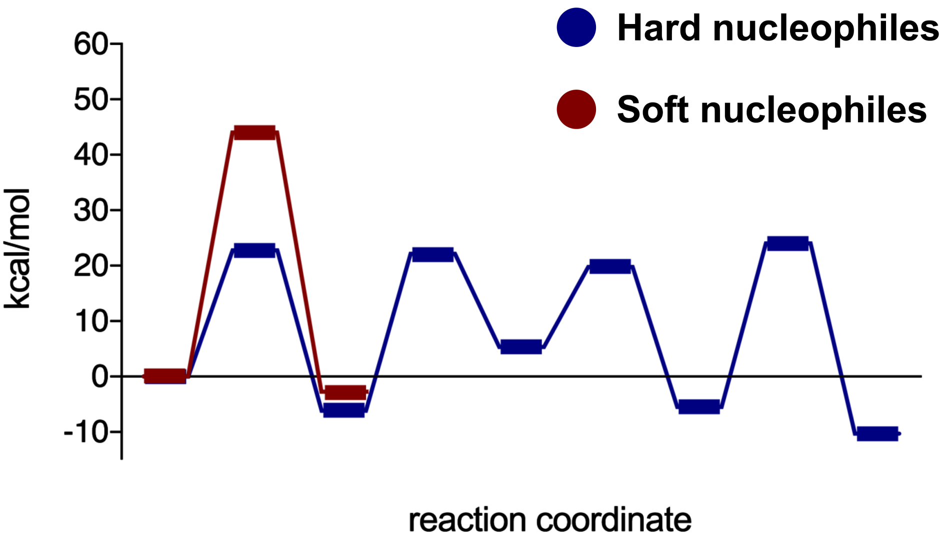


Figure 48 Full energetic profile that lead to six-membered lactones at red and that lead to five-membered lactones at blue.

From Figure 48, it is clear that the synthesis of five-membered lactones involves more steps, being the rate-limiting step the second one that involves the nucleophilic attack at C2. The mechanism of the six-membered ring lactones is simpler and involves only one step. However, when this step is compared with the previous one it is from the energetic point of view more expensive. This goes in line with the experimental observations, since the synthesis of the six-membered lactones requires longer reaction times and takes place at higher temperatures.

At the beginning, there was a big question about the cyclization of the open form intermediate **17**. While the experimental evidences were pointing to a cyclization of a six-membered lactam ring, the theoretical and computational studies were pointing that this was not the path to follow. Later, through x-ray crystallography it was possible to prove that the path was not a six-membered lactam ring but, rather a five-membered lactone ring, demonstrating the strength and increasing utility of theoretical and computational methods.

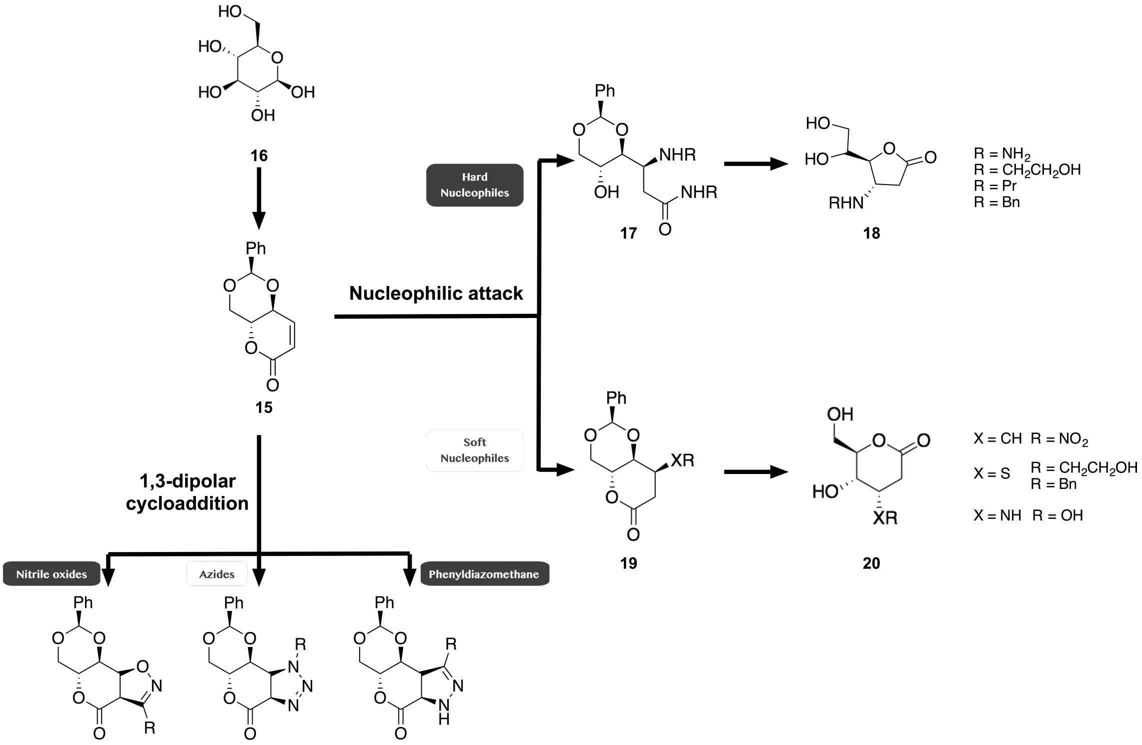

16
15
17
18
19
20
Nucleophilic attack
Hard Nucleophiles
Soft Nucleophiles
1,3-dipolar cycloaddition
Nitrile oxides
Azides
Phenyldiazomethane
R = NH2
R = CH2CH2OH
R = Pr
R = Bn
X = CH R = NO2
X = S R = CH2CH2OH
R = Bn
X = NH R = OH

Figure 49 Schematization of the studied reaction (stereo- and regio-specific) and 1,3-dipolar cycloaddition (stereo-specific) starting from the same compounds.

From the biological and molecular docking results, five inhibitors have been found to be specific for two glycosidases, α-glucosidase and β-galactosidase. Additionally, it was possible to explain the differences in enzyme activity and explain why there is no inhibitors of other glycosidases. All these results provide a better understanding of how to produce optimized, specific and effective inhibitors for certain glycosidases which may be used in the development of new drugs. Moreover, a detailed analysis of possible lactams inhibitors for α- and β-glycosidases was made, and from that results, it was possible to conclude that their synthesis would be advantageous, since they are, probably, potent and specific inhibitor of α- and β-glycosidases.

V

CHAPTER V

References

[1] Cerqueira, N., Brás, N., Ramos, M. J., and Fernandes, P. A. (2012) Glycosidases – A Mechanistic Overview, In *Carbohydrates - Comprehensive Studies on Glycobiology and Glycotechnology* (Chang, C.-F., Ed.), p Ch. 06, InTech, Rijeka.

[2] Bras, N. F., Cerqueira, N. M., Ramos, M. J., and Fernandes, P. A. (2014) Glycosidase inhibitors: a patent review (2008-2013), *Expert Opin Ther Pat 24*, 857-874.

[3] Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases, *Structure 3*, 853-859.

[4] Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013, *Nucleic Acids Res 42*, D490-495.

[5] Hoefsloot, L. H., Hoogeveen-Westerveld, M., Reuser, A. J., and Oostra, B. A. (1990) Characterization of the human lysosomal alpha-glucosidase gene, *Biochem J 272*, 493-497.

[6] Van de Laar, F. A., Lucassen, P. L., Akkermans, R. P., Van de Lisdonk, E. H., Rutten, G. E., and Van Weel, C. (2005) Alpha-glucosidase inhibitors for type 2 diabetes mellitus, *Cochrane Database Syst Rev*, CD003639.

[7] Chinchetru, M. A., Cabezas, J. A., and Calvo, P. (1989) Purification and characterization of a broad specificity beta-glucosidase from sheep liver, *Int J Biochem 21*, 469-476.

[8] Chester, M. A., Hultberg, B., and Ockerman, P. A. (1976) The common identity of five glycosidases in human liver, *Biochim Biophys Acta 429*, 517-526.

[9] Sanchez-Olle, G., Duque, J., Egido-Gabas, M., Casas, J., Lluch, M., Chabas, A., Grinberg, D., and Vilageliu, L. (2009) Promising results of the chaperone effect caused by imino sugars and aminocyclitol derivatives on mutant glucocerebrosidases causing Gaucher disease, *Blood Cells Mol Dis 42*, 159-166.

[10] Dwek, R. A., Platt, F. M., Neises, G. R., and Butters, T. D. (1998) Method for treatment of cns-involved lysosomal storage diseases, Google Patents.

[11] Asano, N., Kato, A., and Watson, A. A. (2001) Therapeutic applications of sugar-mimicking glycosidase inhibitors, *Mini Rev Med Chem 1*, 145-154.

[12] Salehi, S., Eckley, L., Sawyer, G. J., Zhang, X., Dong, X., Freund, J. N., and Fabre, J. W. (2009) Intestinal lactase as an autologous beta-galactosidase reporter gene for in vivo gene expression studies, *Hum Gene Ther 20*, 21-30.

[13] Ishikawa, K., Kataoka, M., Yanamoto, T., Nakabayashi, M., Watanabe, M., Ishihara, S., and Yamaguchi, S. (2015) Crystal structure of beta-galactosidase from Bacillus circulans ATCC 31382 (BgaD) and the construction of the thermophilic mutants, *FEBS J 282*, 2540-2552.

[14] McDonnell, C., Cronin, L., O'Brien, J. L., and Murphy, P. V. (2004) A general synthesis of iminosugars, *J Org Chem 69*, 3565-3568.

[15] E. Sousa, C., R. Mendes, R., T. Costa, F., C.M. Duarte, V., G. Fortes, A., and J. Alves, M. (2014) Synthesis of Iminosugars from Tetroses, *Current Organic Synthesis 11*, 182-203.

[16] Li, H., Liu, T., Zhang, Y., Favre, S., Bello, C., Vogel, P., Butters, T. D., Oikonomakos, N. G., Marrot, J., and Bleriot, Y. (2008) New synthetic seven-membered 1-azasugars displaying potent inhibition towards glycosidases and glucosylceramide transferase, *Chembiochem 9*, 253-260.

[17] Dwek, R. A., Butters, T. D., Platt, F. M., and Zitzmann, N. (2002) Targeting glycosylation as a therapeutic approach, *Nat Rev Drug Discov 1*, 65-75.

[18] Brás, N. F., Coimbra, J. T. S., Neves, R. P. P., Cerqueira, N. M. F. S. A., Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2015) Computational Biochemistry, In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier.

[19] Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2006) Protein-ligand docking: current status and future challenges, *Proteins 65*, 15-26.

[20] Brás, N. F., Cerqueira, N. M. F. S. A., Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2014) Protein Ligand DockingDocking in Drug DiscoveryDrug Discovery, In *Protein Modelling* (Náray-Szabó, G., Ed.), pp 249-286, Springer International Publishing, Cham.

[21] Cerqueira, N. M. F. S. A., Gesto, D., Oliveira, E. F., Santos-Martins, D., Brás, N. F., Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2015) Receptor-based virtual screening protocol for drug discovery, *Archives of Biochemistry and Biophysics 582*, 56-67.

[22] Cerqueira, N. M. F. S. A., Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2010) Virtual Screening of Compound Libraries, In *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols* (Roque, A. C. A., Ed.), pp 57-70, Humana Press, Totowa, NJ.

[23] Sousa, S. F., Cerqueira, N. M., Fernandes, P. A., and Ramos, M. J. (2010) Virtual screening in drug design and development, *Comb Chem High Throughput Screen 13*, 442-453.

[24] Cerqueira, N. M. F. S. A., Bras, N. F., Fernandes, P. A., and Ramos, M. J. (2009) MADAMM: A multistaged docking with an automated molecular modeling protocol, *Proteins: Structure, Function, and Bioinformatics 74*, 192-206.

[25] Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J Comput Chem 19*, 1639-1662.

[26] Taylor, R. D., Jewsbury, P. J., and Essex, J. W. (2002) A review of protein-small molecule docking methods, *J Comput Aided Mol Des 16*, 151-166.

[27] Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications, *Nat Rev Drug Discov 3*, 935-949.

[28] Leach, A. R. (2001) *Molecular Modelling: Principles and Applications (2nd edition)*, Pearson Education EMA.

[29] Cramer, C. J. (2004) *Essentials of Computational Chemistry: Theories and Models* 2nd Edition ed., Wiley.

[30] Jensen, F. (2006) *Introduction to Computational Chemistry*, 2nd Edition ed., Wiley.

[31] Teixeira-Dias, J. J. C. (2017) *Molecular Physical Chemistry: A Computer-based Approach using Mathematica and Gaussian*, Springer International

[32] Hohenberg, P., and Kohn, W. (1964) Inhomogeneous Electron Gas, *Phys Rev B 136*, B864-+.

[33] Kohn, W., and Sham, L. J. (1965) Self-Consistent Equations Including Exchange and Correlation Effects, *Phys Rev 140*, 1133-&.

[34] Lee, C. T., Yang, W. T., and Parr, R. G. (1988) Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density, *Phys Rev B 37*, 785-789.

[35] Tomasi, J., Mennucci, B., and Cammi, R. (2005) Quantum mechanical continuum solvation models, *Chem Rev 105*, 2999-3093.

[36] Sousa, C. E., Ribeiro, A. M., Gil Fortes, A., Cerqueira, N. M., and Alves, M. J. (2017) Total Facial Discrimination of 1,3-Dipolar Cycloadditions in a d-Erythrose 1,3-Dioxane Template: Computational Studies of a Concerted Mechanism, *J Org Chem 82*, 982-991.

[37] Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. . (2009) Gaussian 09, Revision D.01.

[38] Rassolov, V. A., Ratner, M. A., Pople, J. A., Redfern, P. C., and Curtiss, L. A. (2001) 6-31G*basis set for third-row atoms, *J Comput Chem 22*, 976-984.

[39] Scalmani, G., and Frisch, M. J. (2010) Continuous surface charge polarizable continuum models of solvation. I. General formalism, *J Chem Phys 132*.

[40] Chiba, S. (1997) Molecular mechanism in alpha-glucosidase and glucoamylase, *Biosci Biotechnol Biochem 61*, 1233-1239.

[41] Guce, A. I., Clark, N. E., Salgado, E. N., Ivanen, D. R., Kulminskaya, A. A., Brumer, H., 3rd, and Garman, S. C. (2010) Catalytic mechanism of human alpha-galactosidase, *J Biol Chem 285*, 3625-3632.

[42] Rye, C. S., and Withers, S. G. (2000) Glycosidase mechanisms, *Curr Opin Chem Biol 4*, 573-580.

[43] Tom D. Heightman, A. T. V. (1999) Recent Insights into Inhibition, Structure, and Mechanism of Configuration-Retaining Glycosidases, *Angewandte Chemie International Edition 38*, 1521-3773.

[44] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J Comput Chem 30*, 2785-2791.

[45] Cerqueira, N. M. F. S. A., Ribeiro, J., Fernandes, P. A., and Ramos, M. J. (2010) vsLab-An implementation for virtual high-throughput screening using AutoDock and VMD, *International Journal of Quantum Chemistry (Silvi, B., and Eriksson, L., Eds.) 111*, 1208–1212.

[46] Ribeiro, J. V., Tamames, J. A. C., Cerqueira, N. M. F. S. A., Fernandes, P. A., and Ramos, M. J. (2013) Volarea – A Bioinformatics Tool to Calculate the Surface Area and the Volume of Molecular Systems, *Chemical Biology & Drug Design 82*, 743–755.