COMPUTATIONAL IDENTIFICATION OF NOVEL ORGANOCATALYTIC STRUCTURES AS ENZYME MIMICS

by
Alara Öner

Submitted to Graduate School of Natural and Applied Sciences
in Partial Fulfillment of the Requirements
for the Degree of Master of Science in
Chemical Engineering

Yeditepe University
2022

COMPUTATIONAL IDENTIFICATION OF NOVEL ORGANOCATALYTIC STRUCTURES AS ENZYME MIMIC

APPROVED BY:

Prof. Dr. Nihan Çelebi-Ölçüm ……
(Thesis Supervisor)
(Yeditepe University)

Prof. Dr. Tuğba Davran Candan ……
(Yeditepe University)

Asst. Prof. Dr. Burcu Dedeoğlu ……
(Gebze Technical University)

DATE OF APPROVAL: ..../..../20....

I hereby declare that this thesis is my own work and that all information in this thesis has been obtained and presented in accordance with academic rules and ethical conduct. I have fully cited and referenced all material and results as required by these rules and conduct, and this thesis study does not contain any plagiarism. If any material used in the thesis requires copyright, the necessary permissions have been obtained. No material from this thesis has been used for the award of another degree.

I accept all kinds of legal liability that may arise in case contrary to these situations.

Name, Last name Alara ÖNER

Signature …..…….……..

ACKNOWLEDGEMENTS

First of all, I would like to serve my thankfulness to my thesis supervisor Nihan Çelebi-Ölçüm for her endless support throughout my education road. It has been a great chance of learning for me to perform research under the experience and knowledge.

This thesis study was supported by TUBITAK 1001 (116Z514). The computational studies all through this study was performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center.

I would like to thank to my family for all kind of their support from the very begging. I could not be in this point of my life if it was not for them.

ABSTRACT

COMPUTATIONAL IDENTIFICATION OF NOVEL ORGANOCATALYTIC STRUCTURES AS ENZYME MIMIC

Spirocyclic oligomers formed by coupling chiral building blocks carrying the catalytic machinery of hydrolases (spiroligozymes) catalyze the transesterification of vinyl trifluoromethylacetate with methanol but suffer from the lack of enzyme-like preorganization of their catalytic functionalities in a well-defined geometry via H-bond networks. In this thesis, a computational protocol that combines different levels of theories to rapidly explore structural modifications for an improved structural preorganization was applied to a bifunctional parent spiroligozyme containing a pyridine-alcohol dyad that mimics the Ser-His-Asp/Glu catalytic triad of hydrolases. Calculations predict that a modification as simple as replacing the five-membered building block holding the alcohol moiety with a six-membered analog in the parent spiroligozyme significantly increases the occupancy of the H-bond between the benzyl alcohol-pyridine nucleophilic dyad. The computed energy profile is indicative of faster acylation of this derivative compared to the parent spiroligozyme and highlights the importance of inclusion of an oxyanion hole motif for efficient catalysis. With the aim of identifying new organocatalyst candidates containing both a nucleophilic dyad (amine-alcohol) and an oxyanion hole motif (urea/thiourea), a computational approach that combines quantum mechanical calculations with drug design tools was used. A pharmacophore query was generated based on a quantum mechanically optimized transition state model including amine-alcohol-urea catalytic motifs and was screened in the ZINC database. The search identified trifunctional hits in the required three-dimensional constellations. The work presented herein constitutes an important step toward the design of multifunctional organocatalysts with high catalytic efficiency.

.

ÖZET

ENZİM MİMİK EDEN YENİ ORGANOKATALİTİK YAPILARIN HESAPSAL TANIMLANMASI

Hidrolazların katalitik elemanlarını taşıyan kiral yapı blokları birleştirilerek oluşturulan spirosiklik oligomerler (spiroligozimler), vinil triflorometilasetatın metanol ile transesterifikasyonunu katalize eder, ancak katalitik fonksiyonel gruplarının H-bağı ağları yoluyla iyi tanımlanmış bir geometride enzim benzeri ön organizasyonuna sahip değildir. Bu tezde, gelişmiş bir yapısal ön organizasyon sağlayabilecek yapısal modifikasyonları hızla taramak amacıyla oluşturulan ve farklı teori seviyelerini birleştiren bir hesaplama protokolü, hidrolazların Ser-His-Asp/Glu katalitik üçlüsünü taklit eden piridin ve alkol içeren iki işlevli bir ana spiroligozime uygulandı. Hesaplamalar, ana spiroligozimde alkol parçasını tutan beş üyeli yapı bloğunu altı üyeli bir analogla değiştirmek kadar basit bir modifikasyonun, benzil alkol-piridin nükleofilik ikilisi arasındaki H-bağının korunma oranını önemli ölçüde arttırdığını öngörmektedir. Hesaplanan enerji profili, ana spiroligozim ile karşılaştırıldığında bu türevin daha hızlı asilasyonunun göstergesidir ve verimli kataliz için bir oksianyon boşluğu motifinin dahil edilmesinin önemini vurgular. Hem nükleofilik ikili (amin-alkol) hem de oksianyon boşluğu motifi (üre/tiyoüre) içeren yeni organokatalizör adaylarını belirlemek amacıyla, kuantum mekaniksel hesaplamalarını ilaç tasarım araçlarıyla birleştiren bir yaklaşım kullanıldı. Amin-alkol-üre katalitik motifleri içeren kuantum mekaniksel olarak optimize edilmiş bir geçiş konumu modeline dayalı olarak bir farmakofor taraması oluşturuldu ve ZINC veri tabanında tarandı. Tarama, kataliz için tasarlanan fonksiyonel grupların gerekli üç boyutlu yerleşimini sağlayan küçük organik molekülleri belirledi. Burada sunulan çalışma, yüksek katalitik verimliliğe sahip çok işlevli organokatalizörlerin tasarımına yönelik önemli bir adım teşkil etmektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

ABSTRACT v

ÖZET vi

LIST OF FIGURES viii

LIST OF TABLES x

LIST OF SYMBOLS/ABBREVIATIONS xi

1. INTRODUCTION 1

2. METHODOLOGY 6

2.1. DERIVATIZATION 7

2.2. MOLECULAR MECHANICS (MM) EVALUATION 8

2.3. QUANTUM MECHANICS (QM) EVALUATION 12

2.4. MOLECULAR DYNAMIC (MD) EVALUATION 14

2.5. PHARMACOPHORE MODELING 16

2.6. SCREENING 17

3. RESULTS AND DISCUSSION 18

3.1. RAPID COMPUTATIONAL EVALUATION PROTOCOL APPLIED TO BIFUNCTIONAL SPIROLIGOZYME 18

3.1.1. Derivatization 19

3.1.2. MM Evaluation 19

3.1.3. QM Evaluation 21

3.1.4. MD Evaluation 24

3.1.5. Energy Profile 28

3.2. SMALL MOLECULE CATALYST SCREENING OF ZINC DATABASE 29

4. CONCLUSION 35

REFERENCES 36

APPENDIX A: 42

APPENDIX B: 46

LIST OF FIGURES

Figure 1.1. General transesterification reaction. 2

Figure 1.2. Transesterification steps of lipase enzyme catalysis. 2

Figure 1.3. Enzyme-catalyzed transesterification reaction. 4

Figure 1.4. Catalytic functional groups of naturally evolved esterases and spiroligozymes developed using the “inside-out” approach. 5

Figure 2.1. Applied computational protocol for identifying structural modifications for improved H-bond occupancy. 6

Figure 2.2. Applied computational protocol for small molecule catalyst screening. 7

Figure 2.3. MedChem Transformations tool window of MOE software. 8

Figure 2.4. Representative conformer structures of spiroligozymes (green lines) in each cluster based on orientation of the substrate (vinyl trifluoromethylacetate, magenta sticks) upon alignment of the catalytic dyad with the quantum mechanically optimized transition state model (QM theozyme, magenta lines with key distances in Å and angles in degrees). 11

Figure 2.5. Conformational Search tool window of MOE software. 12

Figure 2.6. Pharmacophore Editor tool window of MOE software 16

Figure 2.7. Pharmacophore Search tool window of MOE software 17

Figure 3.1. Design of bifunctional spiroligozymes. 19

Figure 3.2. Most promising derivatives according to MM and QM evaluations. The modifications are shown as red. 21

Figure 3.3. Geometries, relative energies (kcal mol^{-1}), and populations of the lowest-energy conformers in each cluster of promising derivatives based on QM evaluation. 23

Figure 3.4. N···H-O distance-angle scatter plots and MD populations from MD trajectories of one of the representative replicas..... 26

Figure 3.5. Time-distance plots and representative snapshots from the MD simulations for alternative intramolecular interactions in **D20**..... 27

Figure 3.6. Time Energy profile for the **D20**-catalyzed acyl transfer reaction. Relative energies for the analogous acylation pathway of the parent..... 27

Figure 3.7. Theozyme structure for a trifunctional catalyst..... 30

Figure 3.8. Pharmacophore model of the theozyme created by pharmacophore editor in MOE software..... 31

Figure 3.9. Pharmacophore search screen used for the match search of the model..... 32

Figure 3.10. High anodyne in-stock library promising match..... 33

Figure 3.11. High bother agent library promising match..... 33

Figure 3.12. Low clean wait-ok library promising matc..... 34

LIST OF TABLES

Table 3.1. MM populations of selected derivatives 20

Table 3.2. QM populations of selected derivatives 22

Table 3.3. MD populations of selected derivatives 25

Table 3.4. ZINC reference information of purchasability 30

Table 3.5. ZINC reference information of reactivity 30

Table 3.6. ZINC reference information of pH 31

Table A.1. The information of 116 derivative database of BF3 spiroligozyme 42

Table B.1. ZINC Database match library information used during the study 46

Table B.2. The pharmacophore model match result of the ZINC Database library search 48

LIST OF SYMBOLS/ABBREVIATIONS

Å	Angstorm
$\hat{H}$	Hamiltonian operator
E	Eigenvalue
T	Temperature
k_B	Boltzmann constant
V	Potential energy
r	Positions of particles
N	Number of particles
l	Length
Σ	Summation symbol
θ	Angle
Δ	Difference of two variables
ψ	Eigenvalue
Asp	Aspartate
C	Carbon
DFT	Density functional theory
E	Energy
GAFF	General amber force field
Glu	Glutamate
H	Hydrogen
HF	Hartree-Fock
His	Histidine
IEF	Integral equation formalism
IEF-PCM	Integral equation formalism polarizable continuum model
M06	Minnesota 06 functional
MedChem	Medicinal chemistry transformations
MD	Molecular dynamics
MM	Molecular mechanics

MMFF	Merck molecular force field
MOE	Molecular operating environment
N	Nitrogen
O	Oxygen
PCM	Polarizable continuum model
R	Functional group
S	Sulphur
Ser	Serine
SMD	Solvation model based on electron density
TS	Transition state
QM	Quantum mechanics

1. INTRODUCTION

Catalysts, which are chemical substances, are used to speed the chemical reactions while increasing the yield of the desired product. They accelerate the chemical reaction without being involved in the reaction. Catalysts are commonly used in the industry since the time of the manufacturing process is reduced thus money and time is being saved. Although catalysts are known with their benefits, they also have harmful aspects such as polluting the environment. Therefore, development of environmentally friendly catalytic systems is an active area of interest [1]

Enzymes are macromolecular biocatalysts that accelerate chemical reactions, allowing metabolic processes to occur at rates fast enough to sustain life [2]. Each enzyme has its own optimal environmental conditions to take part as a catalyst. The most important conditions are pH, temperature and pressure. To give an example, enzymes which take part in the reactions in a human body is likely to work at its maximum at body temperature, pH and pressure [3].

Besides carrying out reactions to maintain our metabolism, enzymes are increasingly used as environmentally friendly catalysts in the chemical industry. As important green alternatives to biocatalysts, small organic molecules can effectively mimic functional group richness of enzymes within a much simpler molecular framework [4-17]. These low-cost small organic catalysts (organocatalysts) contain essentially carbon (C), hydrogen (H), oxygen (O), nitrogen (N), sulphur (S), and no metal atoms, therefore they display low toxicity [18].

This thesis involves investigation of new organocatalyst candidates for transesterification reactions.

The reversible equilibrium reaction between an alcohol and an ester to form a different ester molecule in an acid or base environment is known as transesterification reaction. The general transesterification reaction is given in the Figure 1.1, where $R^{'}$ group of the ester and $R^{''}$ of the alcohol exchanges in an acid or base environment [19].

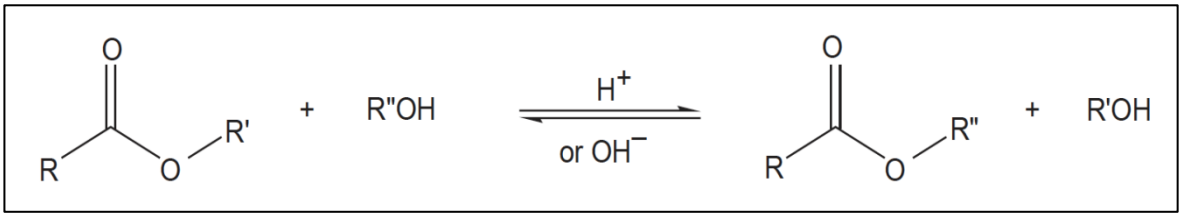

Figure 1.1. General transesterification reaction [19].

In the acid catalyzed transesterification reaction, an acid donates a proton to the carbonyl group of the ester. The advantage of acid catalyzed transesterification reactions is that the yield of the desired product is high. However, the reaction time is long and high temperature environment is required [20].

In the base catalyzed transesterification reaction, a base abstracts the proton from the alcohol group. Base catalyzed transesterification reactions are rapid compared to the acid catalyzed reactions. In industry, base catalyzed transesterification reactions are preferred since the reaction is faster and the base catalysts are less toxic [20].

Enzymes also catalyze transesterification reactions. Essentially, enzymatic transesterification by lipases has been widely explored for biodiesel production, presenting several advantages over chemical catalysts, one of the most important being mild reaction conditions [21]. Lipases are known as catalysts of hydrolysis of triglycerides. In the lipase catalyzed transesterification reaction, triglyceride is reduced to glycerol forming 3 equivalents of methyl ester (Figure 1.2) in three steps. Lipases are part of a class of enzymes known as hydrolyses [22].

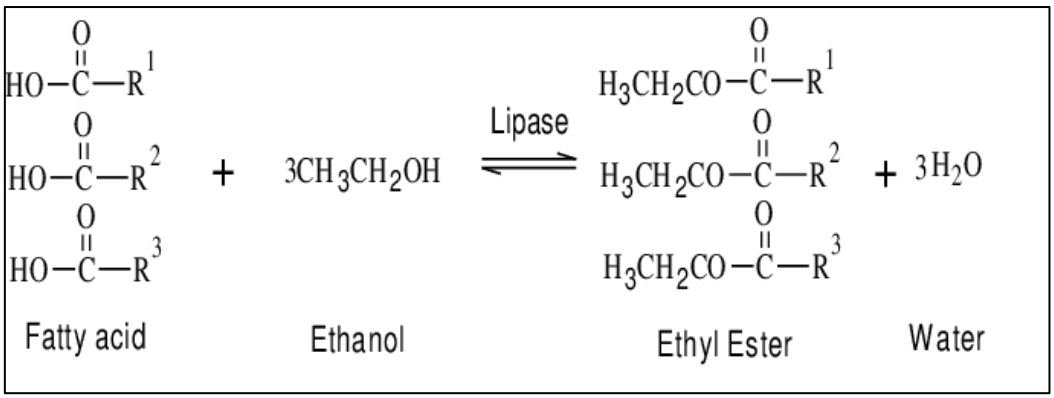

Figure 1.2. Transesterification steps of lipase enzyme catalysis [22].

Hydrolases are enzymes that breakdown bigger molecules into their smaller building blocks. Esterases split esters into an acid and an alcohol. Proteases cleave peptide bonds to form amino acids. Lipases break down lipids to glycerol [23].

Figure 1.3 summarizes the widely accepted mechanism of an exemplary enzyme-catalyzed transesterification reaction [24-31]. Hydrolases function via formation of a covalent bond between the enzyme and the substrate through a nucleophilic catalysis mechanism. The catalysis is initiated by activation of serine as a nucleophile in the enzyme active site. This is achieved via a proton shuttle mechanism in a so-called “catalytic triad” or “nucleophilic triad”, which is composed of Ser-His-Asp/Glu residues interacting with each other via a well-maintained H-bond network. Owing to its increased basicity in the presence of aspartate, histidine activates serine as a nucleophile by abstracting the proton from serine’s hydroxyl group. The activated serine through this proton shuttle mechanism attacks the ester bond of the substrate forming a tetrahedral oxyanionic intermediate. This acyl-enzyme intermediate is stabilized by a number of H-bonding interactions in a region called the “oxyanion hole”. Protonated His-Asp/Glu dyad now acts as the acid and activates the -OR group of ester as a leaving group. Exchange of the leaving group with another alcohol molecule bound in the active site is followed by the attack of the alcohol to the ester bond between serine and the substrate breaking the covalent bond. As a result, product is released and free enzyme is regenerated.

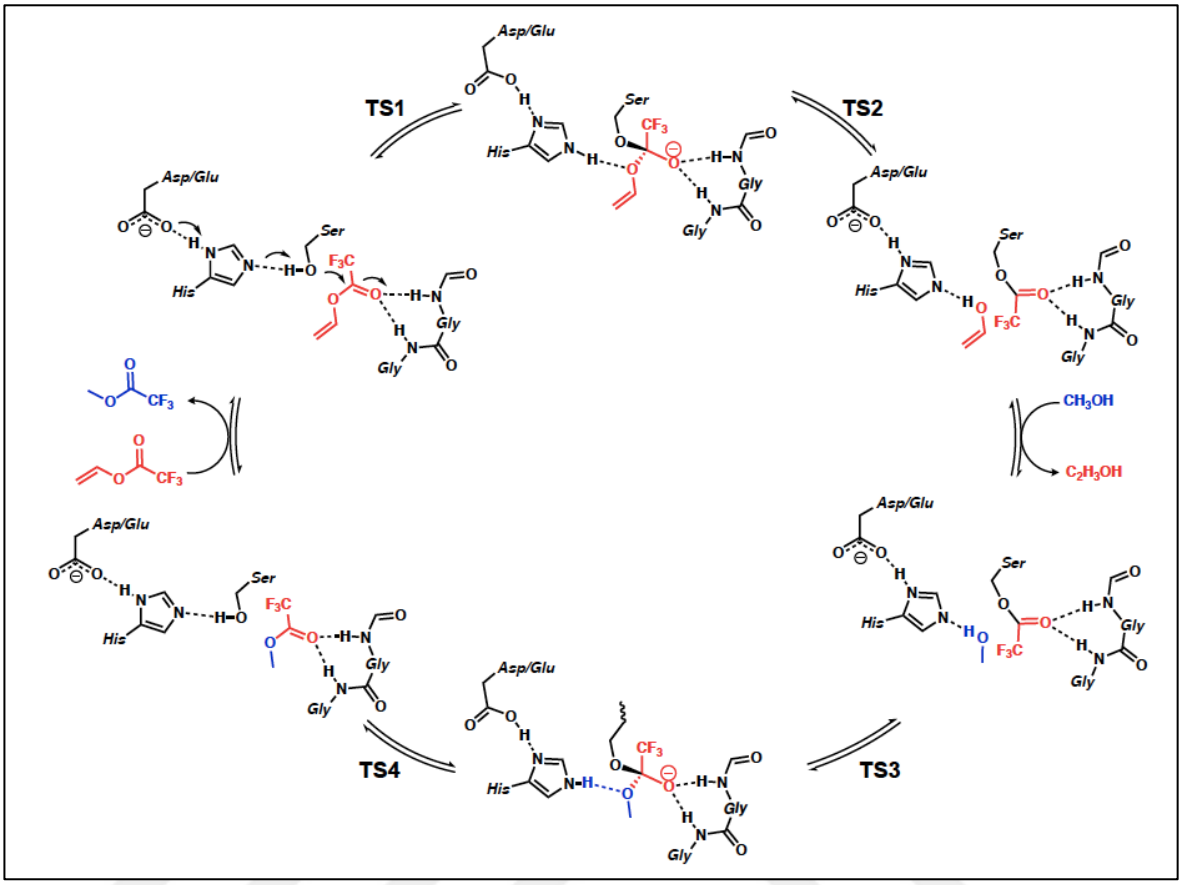

Figure 1.3. Enzyme-catalyzed transesterification reaction [32].

The interest in grafting catalytic active sites of enzymes onto organocatalytic scaffolds led to the development of new transesterification catalysts, called "spiroligozymes", by placing the catalytic machinery of hydrolase enzymes onto modular spiro-fused bispeptides with the help of quantum mechanical transition state (TS) calculations using the "inside-out" approach [17]. The proton shuttle mechanism between serine, histidine, and aspartate/glutamate residues is mimicked with an alcohol-pyridine dyad in spiroligozymes (Figure 1.4) [24]. The best spiroligozyme increased the rate of the transesterification reaction between vinyl trifluoromethylacetate and methanol by 2200-fold for the acylation step and 130-fold for the deacylation step [17]. A tightly maintained H-bond network between Ser-His-Asp/Glu nucleophilic triad is shown to be critical for the activity of hydrolases [25-27]. For spiroligozymes, however, in contrast to the high-level structural and electrostatic preorganization observed in naturally evolved enzymes [33], the catalytic groups in spiroligozymes sample numerous alternative conformations beside the active one, and the H-bond network required for efficient catalysis is not maintained [17].

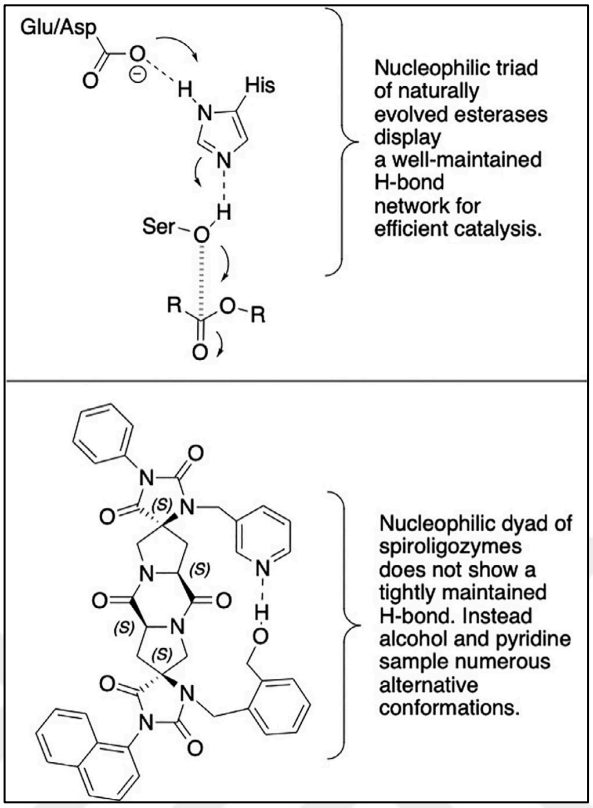

Figure 1.4. Catalytic functional groups of naturally evolved esterases and spiroligozymes developed using the "inside-out" approach [32].

The purpose of this project is to develop a rapid computational evaluation protocol to assess the structural modifications that may result in rate enhancements due to an increased H-bond network stability in small-molecule catalysts, which constitutes an important step toward the design of organocatalysts presenting catalytic functionalities in a well-defined arrangement forming an enzyme-like active site for efficient catalysis. Application of the proposed protocol to the parent bifunctional spiroligozyme, identified derivatives with an improved occupancy of the H-bond between benzyl alcohol and pyridine essential for catalysis. In addition, small molecule databases were screened against theoretical active site (theozyme) models to identify and evaluate novel organocatalytic structures as alternative transesterification catalysts.

2. METHODOLOGY

Figure 2.1 summarizes the applied computational protocol for the identification of structural modifications that will result in increased H-bond occupancy in an organocatalytic framework.

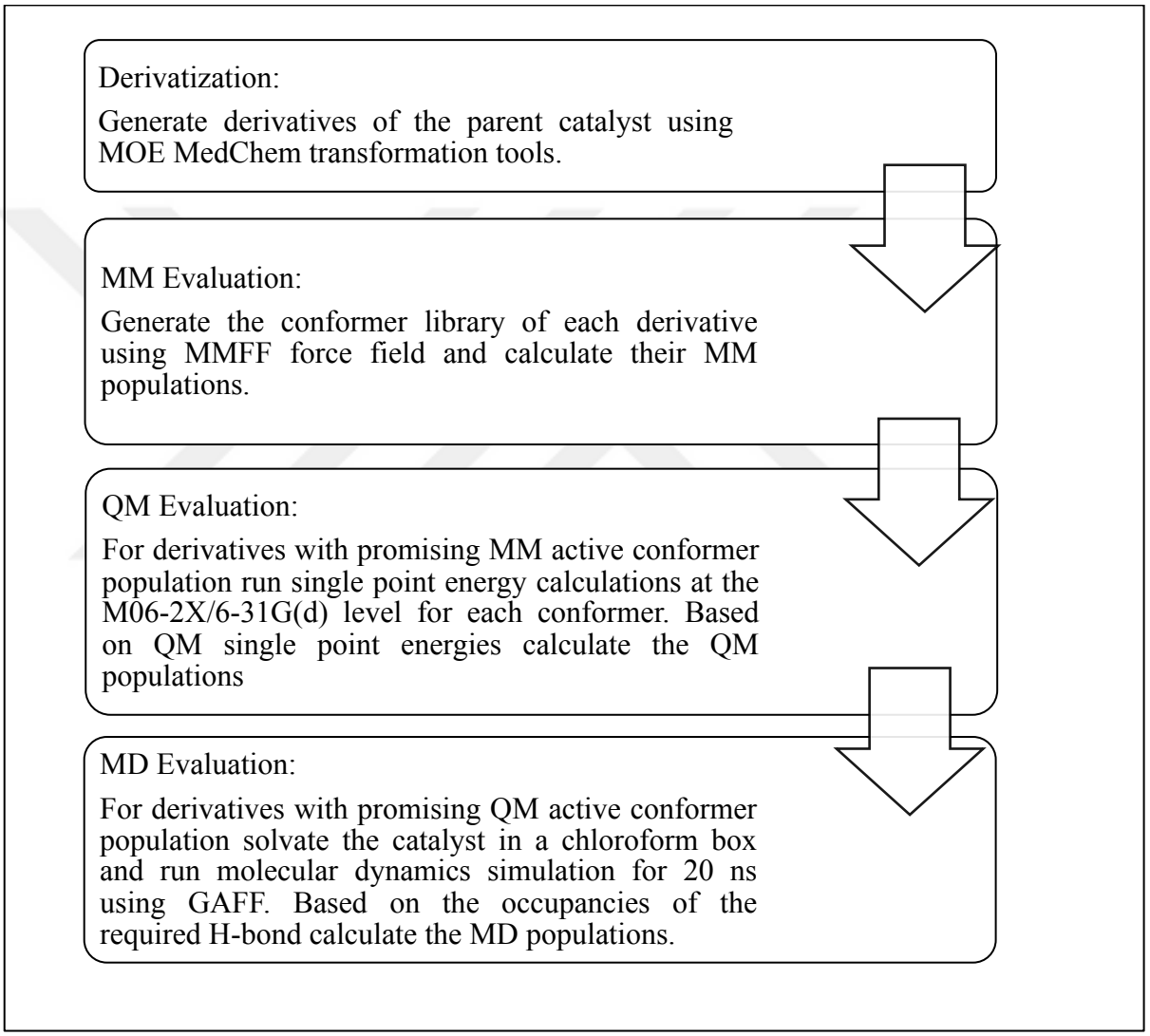

Figure 2.1. Applied computational protocol for identifying structural modifications for improved H-bond occupancy.

Figure 2.2 describes the applied computational protocol for small molecule catalyst screening in the ZINC database.

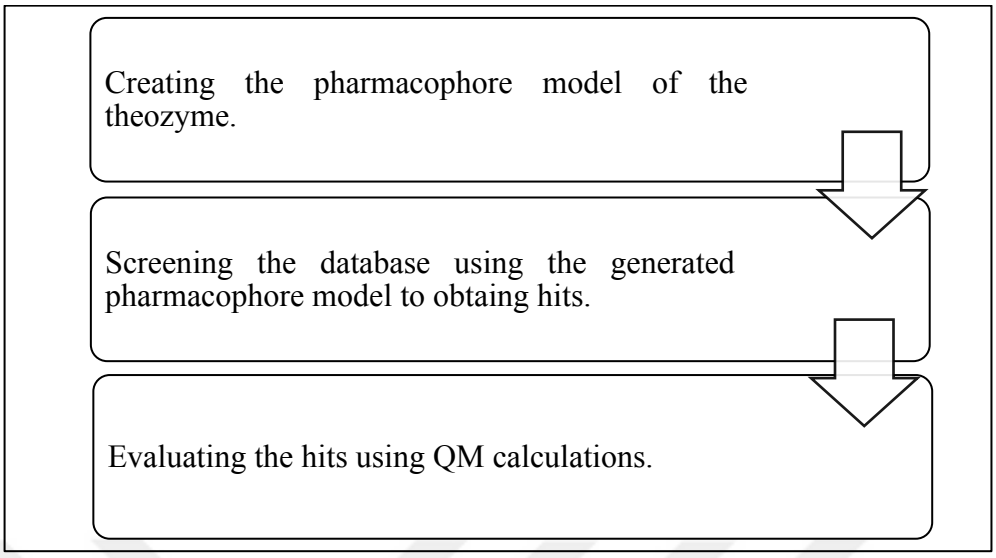

Figure 2.2. Applied computational protocol for small molecule catalyst screening.

2.1. DERIVATIZATION

The protocol starts with the generation of synthesizable derivatives of the parent molecule. For this purpose, MedChem Transformations, an approach that applies predefined transformation rules by a match-and-replace algorithm to existing ligands to obtain new chemical structures, was used as implemented in MOE [31]. A total of 175 reactions is defined in the default library.

The MedChem Transformations tool window of MOE is given in Figure 2.3. This tool is accessed under Compute> Fragments> MedChem Transformations [31].

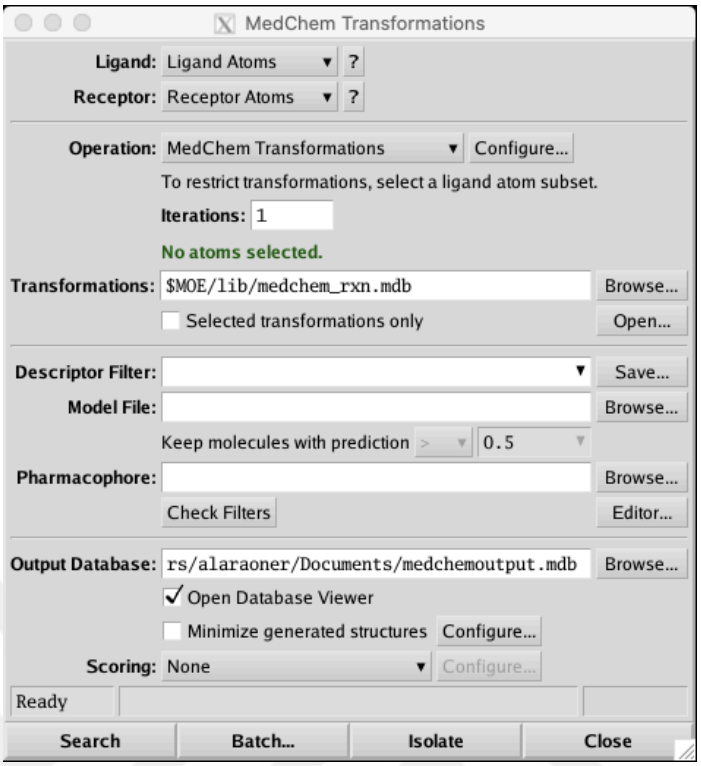

Figure 2.3. MedChem Transformations tool window of MOE software [31].

2.2. MOLECULAR MECHANICS (MM) EVALUATION

Molecular mechanics (MM) is based on a rather simple model of interactions between atoms considered as hard spheres. MM calculations are fast, making them preferable for studying large molecules [34].

MM methods (also known as force field methods) ignore the electronic motions and calculate the energy as a sum of the total bonded energy terms (such as harmonic stretching, bending potentials and periodic torsion potentials), nonbonded electrostatic terms (such as Coulomb interactions between fixed point charges) and Van der Waals interactions [35]. Therefore, relative positions of the nuclei of the atoms are assumed as a function of operating attractive and repulsive forces forming a structure. Force fields are interactions and energies of the computed physical functions with the above-mentioned terms (Equation 2.1) [36].

$$V(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + cos(n\omega - \gamma)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \quad (2.1)$$

The second step of the protocol involves a fast conformational analysis for each generated derivative unless the transformation changed the catalytic functional groups (pyridine and alcohol) in the parent molecule. The conformational search was carried out using MOE with MMFF94x, setting the energy convergence criterion to a root mean square gradient smaller than 0.001 kcal mol^{-1} Å^{-2}. MMFF94x is an all-atom forcefield, parametrized for small organic molecules, modified from MMFF94s to force conjugated nitrogens planar [37]. The solvent effects were implicitly taken into account using the Generalized Born solvation model using an exterior dielectric constant, ε = 4.81, representing chloroform [38]. A LowMode MD sampling procedure was used that performs molecular dynamics (MD) perturbations along low-frequency vibrational modes [39] [40]. A LowMode MD search was selected for its ability to explore conformations near the relevant and realistic low-energy states and for its efficiency in locating most of the local minima of arbitrarily complex multicomponent systems [40]. An iteration limit of 1000 with a rejection limit of 100 was used in order to obtain a sufficient sample of conformational minima for initial evaluation in a reasonable amount of time. Two conformations having a root mean square distance less than 0.25 were considered duplicates. Conformations with ΔE > 7 kcal mol^{-1} relative to the lowest-energy conformer (ΔE_{MM}) were discarded. The Boltzmann populations were calculated according to the formula.

$$MM\ Population = \frac{e^{-\Delta E_{MM}/k_B T}}{\sum(e^{-\Delta E_{MM}/k_B T})} \times 100 \quad (2.2)$$

where ΔE_{MM} is the difference between the energy of the conformer of interest and the energy of the global minimum conformation obtained from molecular mechanics (MM) force field calculations, T is the absolute temperature, and k_B is the Boltzmann constant. The resulting conformer libraries were first examined for the population of the conformers with the required H-bond between the catalytic dyad. An N···H-O H-bond distance cut-off

of 3 Å was applied as the selection criterion for an initial filtering. Conformers with $d_{N\cdots H\text{-}O} < 3$ Å were considered in group H-Bond, as the others were grouped as No-H-Bond. Derivates with H-Bond MM population higher than 50 percent were then clustered into three groups based on the N···H-O distance and the orientation of the substrate upon alignment of the catalytic dyad with the quantum mechanically optimized active site model (QM theozyme) (Figure 2.4).

The first cluster, called **Active-H-Bond**, is composed of conformers in which the arrangement of the pyridine-alcohol hydrogen bond with respect to the chiral backbone of the spiroligozyme resulted in an exposed oxygen nucleophile. The second cluster, called **Inactive-H-Bond**, includes conformers with a buried oxygen nucleophile that requires the approach of the substrate from the hindered face of the dyad. Conformers with $d_{N\cdots H\text{-}O} < 3$ Å, yet having an angle less than 120° between the plane of pyridine and N···H-O, were also considered in this group. Cluster **No-H-Bond** includes all other conformers, in which the arrangement of the two catalytic functional groups does not allow a H-bonding interaction between pyridine and alcohol. The population of each cluster was deduced as the sum of the individual conformer populations in the cluster.

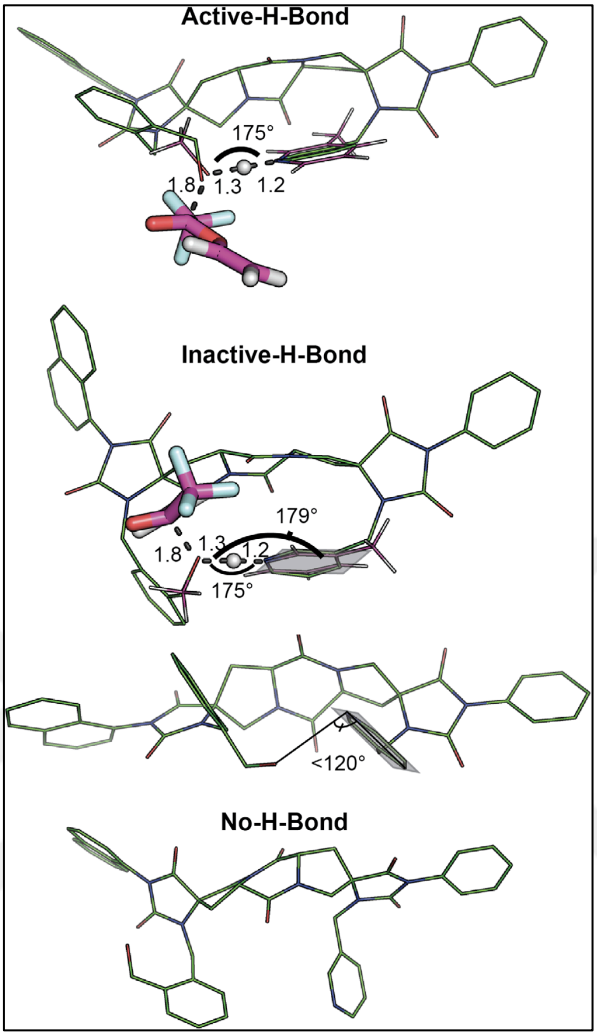

Figure 2.4. Representative conformer structures of spiroligozymes (green lines) in each cluster based on orientation of the substrate (vinyl trifluoromethylacetate, magenta sticks) upon alignment of the catalytic dyad with the quantum mechanically optimized transition state model (QM theozyme, magenta lines with key distances in Å and angles in degrees) [32].

The Conformational Search tool window of MOE is shown in Figure 2.5. Conformational analysis for each derivative was performed with this tool. This tool is accessed under Compute> Conformations> Conformational Search [31].

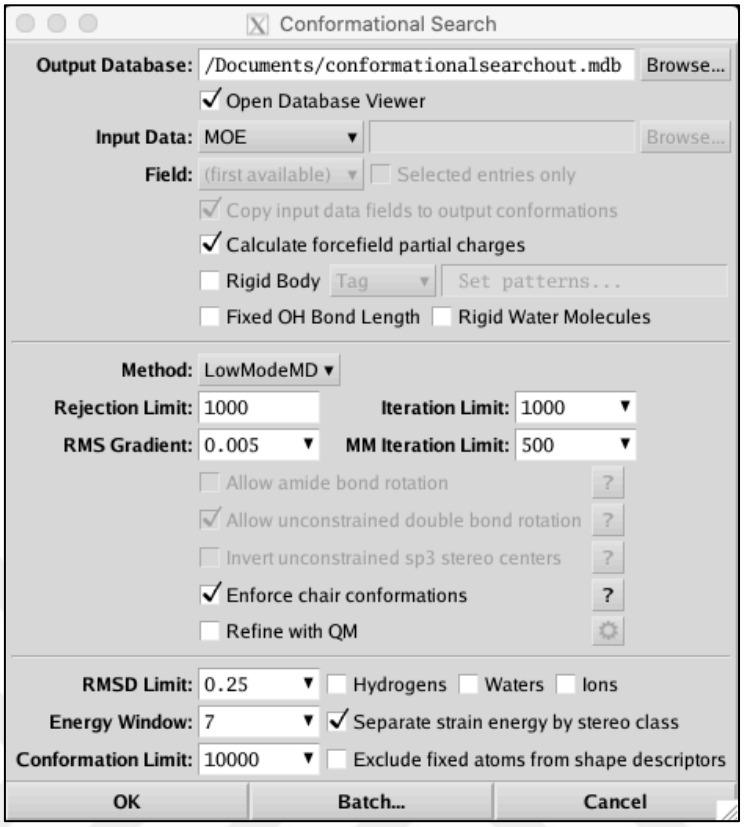

Figure 2.5. Conformational Search tool window of MOE software [31].

2.3. QUANTUM MECHANICS (QM) EVALUATION

The basis of quantum mechanical calculations is Schrödinger equation. The time dependent Schrödinger equation represents a single particle which is moving in space with a definite mass and time with the effect of an external field. By reducing to time independent equation, equation 2.3 is obtained [36].

$$\hat{H}\psi = E\psi \tag{2.3}$$

where $\hat{H}$ is Hamiltonian operator, ψ eigenfunction and E is the eigenvalue. The Schrödinger equation has exact solution for some definite systems such as particle in a box, particle on a ring and hydrogen atom. Thus, necessary assumptions and approximations are applied to obtain solutions of more complex multi electron multi atom systems. By making the approximation that at a minimum energy state the wavefunction is the most favorable causing the first derivative of the energy in the Schrödinger equation zero, the Hartree-Fock (HF) equations are obtained [36]. HF is an *ab initio* method in

which the calculation is performed with the first principle by use of no empirical data; where the average effect of the Coulombic electron-electron repulsion is considered [41]. However, HF equation fails in the treatment of the electron correlation as an electron is considered to be moving in the average field created by all other electrons [36]. Density functional theory (DFT) calculations are preferred to investigate the electronic structure by replacing the electronic wave function with electronic density [41]. In DFT, energy functional is expresses as a sum of this terms, where the first term represents the Coulomb interaction of electrons with the nuclei, and the second term is the sum of kinetic energy of electrons and the contribution from electron-electron interactions (Equation 2.4) [36].

$$E[\rho(r)] = \int V_{ext}(r)\rho(r)dr + F[\rho(r)] \tag{2.4}$$

$F[\rho(r)]$ is the sum of three terms:

$$E[\rho(r)] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)] \tag{2.5}$$

where the first term is the kinetic energy of electrons, the second term is the electron-electron Coulomb potential, and the last term includes contributions from exchange and correlation [36].

The major limitation is that exact functional for exchange and correlation is not known, yet approximations employed in various DFT functionals allow quite accurate calculations. Among them, M0X suite of functionals are known for their good performance in treating dispersion interactions. DFT methods are more preferable due to higher accuracy of the calculations.

The derivatives with **Active-H-Bond** MM population greater than the population of other clusters were subjected to quantum mechanical (QM) evaluation. For each selected derivative, energies of all conformers, regardless of their population, were refined using single-point energy calculations with density functional theory at the M06-2X/6-31G(d) level [42]. The calculations were performed in a dielectric continuum representing chloroform as the solvent using the integral-equation-formalism polarizable continuum model (IEF-PCM) with radii and non-electrostatic terms for the SMD solvation model [43-46]. All quantum mechanical calculations were performed using Gaussian 09 [47]. QM

populations of conformers were determined using the Boltzmann distribution (Equation 2.6).

$$QM\ Population = \frac{e^{-\Delta E_{QM}/k_B T}}{\sum\left(e^{-\Delta E_{QM}/k_B T}\right)} \times 100 \quad (2.6)$$

where ΔE_{QM} is the relative energy of the conformer of interest and the minimum energy conformer obtained from QM calculations, T is the absolute temperature, and k_B is the Boltzmann constant. The populations of the clusters were updated with the QM populations. Derivatives with **Active-H-Bond** QM population greater than the population of other clusters were taken into consideration for molecular dynamics (MD) evaluation.

2.4. MOLECULAR DYNAMIC (MD) EVALUATION

Molecular dynamics (MD) simulations are used to simulate the effect of methods on the movements of molecules. Configuration of the systems are created with the MD simulations. The movement of the atoms are simulated for a defined time as a result defining the trajectory of the molecules where the variation of the positions of the atoms are discovered within time. Newton's equation of motion is the logic during the simulations. The equation of the Newton's second law is solved to obtain the trajectory given in equation 2.5 which the motion of a particle with a mass along a coordinate with a force on the particle direction [36].

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \quad (2.7)$$

MD simulations were performed using AMBER 12 to evaluate the occupancy of the H-bond between the catalytic dyad in a dynamic explicitly solvated environment [48]. The antechamber module of AMBER 12 was used to generate the spiroligozyme parameters using the general Amber forcefield (GAFF), with partial charges set to fit the electrostatic potential computed at HF/6-31G(d) by RESP [49-51]. The Merz-Singh-Kollman scheme was used to calculate the charges using Gaussian 09 [47,52-53]. Spiroligozymes were placed in a chloroform box with a solvent layer of at least 10 Å around the solute [54]. Initially, only the solvent molecules were relaxed, followed by an unrestrained minimization of all atoms. The systems were then heated in six 50 K, 50 ps steps to 300 K

with a 1-fs time step at constant-volume periodic boundary conditions using a harmonic restraint of 10 kcal mol^{-1} to the spiroligozyme. The temperature was controlled and equalized using the Langevin equilibration scheme. Each system was equilibrated for 2 ns with a 2-fs time step at constant volume and another 2 ns at a constant pressure of 1 atm. A 20-ns production run was performed using the isothermal-isobaric ensemble (NPT). Long-range electrostatic interactions were treated with the particle mesh-Ewald method [55]. Trajectories were saved at every 10 ps, resulting in a total of 2000 frames. The trajectories were analyzed using the ptraj module of AMBER 12. H-bond occupancy defines the cumulation of MD snapshots in the H-bond ($d_{N\cdots H\text{-}O}$ < 3.5 Å and 120° < $\theta_{N\cdots H\text{-}O}$< 180°). This region is indicated with a blue rectangle on the corresponding N⋯H-O distance-angle scatter plots, and MD population is calculated as the percentage of the number of frames in the H-bond region with respect to the total number of frames in the simulation. This MD evaluation protocol successfully predicted the active and inactive enzyme designs for Kemp elimination on the basis of the stability of the required H-bond networks in the active site and was previously used for the parent spiroligozymes in this study to assess whether the H-bond between the dyad was maintained in a solvated dynamic environment [17-20, 22-33]. In order to avoid false-positive conclusions, six replicas for each system were performed [56].

Finally, the energy profile of the acyl transfer reaction from vinyl trifluoromethylacetate to methanol was constructed using density functional theory in the presence of the most promising spiroligozyme derivative as a catalyst. All stationary points were located using Gaussian 09 at the M06-2X/6-31G(d) level of theory in a dielectric continuum representing chloroform using the IEF-PCM with radii and nonelectrostatic terms for the SMD solvation model [42-47]. The M06-2X density functional method was chosen for its good performance for thermochemical kinetics and for the treatment of noncovalent interactions [42]. Frequency calculations were used to characterize the stationary points as first-order saddle points or minima. Corrections in enthalpy and free energy, including zero-point vibrational energies, were calculated using the harmonic oscillator approximation at 298.15 K for 1 M standard state. Minimum energy paths were followed from the transition states both in the reverse and forward directions using the intrinsic reaction coordinate (IRC) method [57-58].

2.5. PHARMACOPHORE MODELING

Pharmacophore modeling is three dimensional (3D) arrangements of functional groups. By selecting and preserving the necessary groups of the parent molecule, search is made to find similar molecules [59]. This type of modeling is used to find new compounds or finding molecules which has the necessary features in a library/database [60].

Pharmacophore models of the theozymes used to develop the spiroligozymes, was screened in ZINC database [61]. The matched molecules were considered as catalytic hits. The Pharmacophore Editor tool window of MOE can be seen from the figure (Figure 2.6). This tool is accessed under Compute> Pharmacophore> Query Editor. The chosen atom or ring is added to the pharmacophore model, the radius is determined and the atom or ring is chosen as donor or acceptor with the editor. The smaller the radius, the similar are the matched molecules to the parent molecule [31].

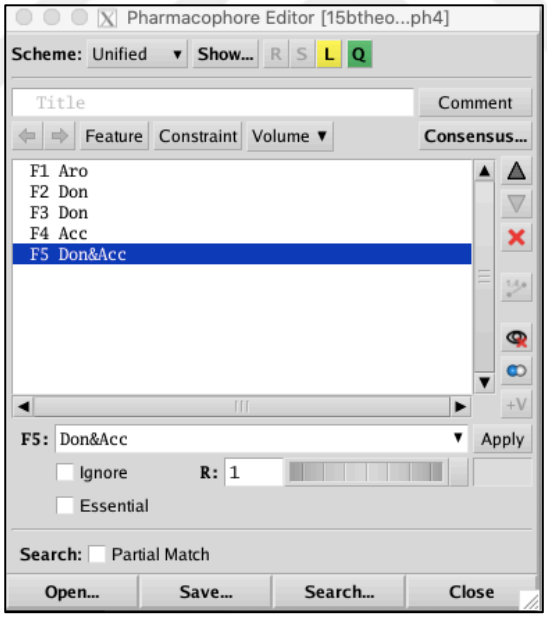

Figure 2.6. Pharmacophore Editor tool window of MOE software [31]

After the determination of pharmacophore model, with Pharmacophore Search tool (Figure 2.7) of MOE software, a pharmacophore query can be generated with the necessary features. Partial match of the model can be allowed [31]

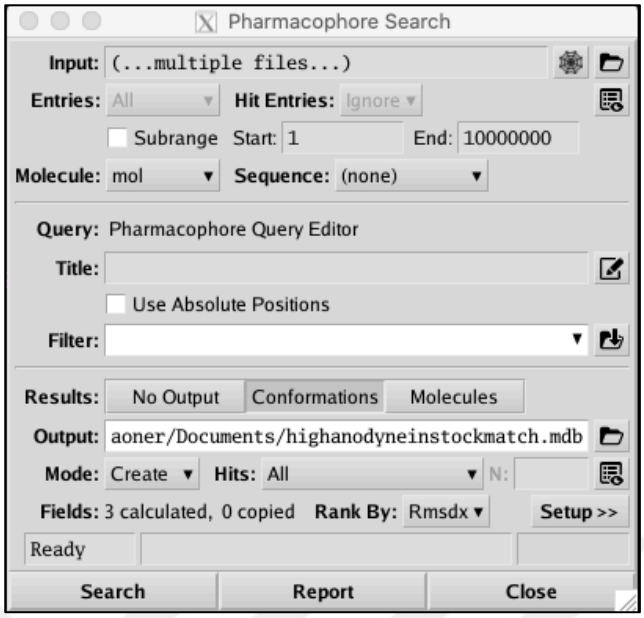

Figure 2.7. Pharmacophore Search tool window of MOE software [31]

2.6. SCREENING

In this thesis, the pharmacophore model is obtained with the required features and search is made in the ZINC database library.

ZINC database is a commercial public access database developed for virtual and pharmacophore screening. Zinc reference information of purchasability, reactivity and pH ranges can be obtained. By choosing the reference information the database can be downloaded for the research studies such as grouped as short delivery times with high molecular diversity or longer delivery times with specific molecular structures [61-63].

ZINC database offers users the purchasability of the molecules and the vendor references. Furthermore, the reactivity and pH ranges are offered to the users so that the reduction in the database can be made according to the research area of the user [61-63].

3. RESULTS AND DISCUSSION

3.1. RAPID COMPUTATIONAL EVALUATION PROTOCOL APPLIED TO BIFUNCTIONAL SPIROLIGOZYME

Esterases catalyze the hydrolysis of carboxyl esters using a proton shuttle mechanism between Ser-His-Asp/Glu catalytic triad for the activation of an alcohol nucleophile that initiates the catalytic cycle by covalently binding to the substrate (Figure 1.4). Their high efficiency is attributed to the high-level structural and electrostatic preorganization with tightly maintained H-bonds in their active sites [25-27]. Catalytic spiroligomers (called spiroligozymes) were constructed by coupling cyclic, sterochemically pure building blocks carrying alcohol and pyridine functional groups to mimic the Ser-His-Asp/Glu nucleophilic triad of natural hydrolases so that the resulting spiroligomer scaffolds with fused ring structures are capable of presenting these catalytic functional groups in a three-dimensional arrangement as suggested by quantum mechanically optimized transition state models (QM theozyme) (Figure 3.1) [17]. Modeling based on the **TS dyad** suggested that spiroligozyme A could approximate the three-dimensional (3D) arrangement of catalytic functional groups depicted in the theozyme (Figure 3.1). Indeed, kinetic experiments using vinyl trifluoromethylacetate as the substrate showed that the rate of acylation of **A** is 250-fold faster than the rate of acylation of the control spiroligomer, displaying a phenyl group instead of a pyridyl group [17]. However, high-level QM calculations and analysis of MD trajectories revealed that the required H-bonding interaction between the catalytic dyad for the proton shuttle mechanism for activating the alcohol as nucleophile is not maintained in designed spiroligozymes. Based on these computational observations, we applied the proposed computational protocol to spiroligozyme **A**, with the aim of identifying structural modifications that could increase the H-bond occupancy of the catalytic dyad by possibly destabilizing alternative unproductive conformations [17]. We selected bifunctional spiroligozyme **A** as the parent compound to be able to focus on the preorganization of the nucleophilic dyad and to understand the importance of the presence of the oxyanion hole, another catalytic motif present in hydrolases previously modeled in trifunctional spiroligozymes [17, 28-30]. All steps of the same computational protocol were applied to

the parent compound A for comparison, and the results are tabulated in the corresponding tables.

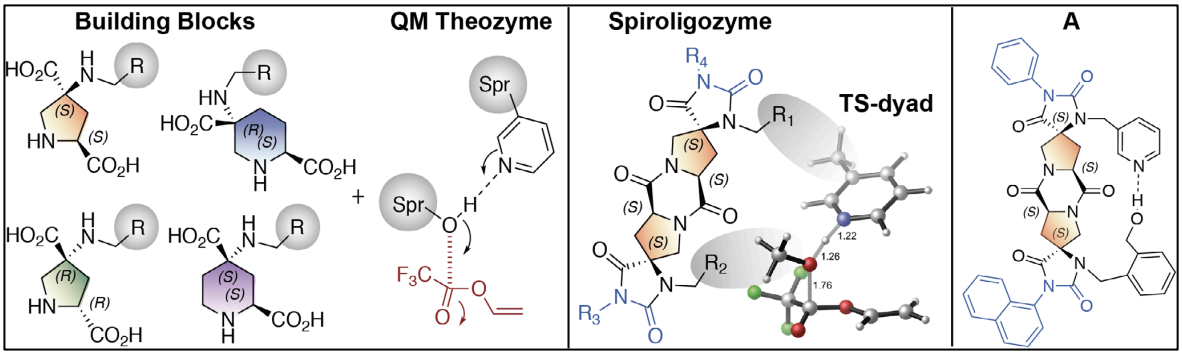

Figure 3.1. Design of bifunctional spiroligozymes [17].

3.1.1. Derivatization

Most of the computational design approaches based on existing molecular structures are faced with significant synthetic challenges as synthetic routes to predicted favorable modifications usually are not known. To overcome this challenge, a derivatization tool that uses a library of 175 transformations commonly applied in medicinal chemistry was used. The derivatization of the parent spiroligozyme **A** using the MedChem transformation tool in MOE gave 116 distinct derivatives. A complete list of derivatives and the corresponding transformation rules are given in Table A1. A total of 42 derivatives involve modifications in the catalytic functional groups (alcohol or pyridine) and are therefore not considered for further evaluation.

3.1.2. MM Evaluation

A total of 74 derivatives were subjected to MM evaluation. The number of conformers obtained for each derivative upon fast initial conformational analysis varies between 150 and 45. The conformer library of each derivative was sorted in the order of increasing N···H-O distance between the dyad. As an initial filter, conformers with $d_{N\cdots H\text{-}O} < 3$ Å were considered H-bonded (**H-Bond**), whereas all other conformers were listed as **No-H-Bond**. A total of 10 derivatives with **H-Bond** MM population greater 50 percent were selected for extended analysis (Table 3.1).

Table 3.1. MM populations of selected derivatives

Derivative Number	Number of Conformer	Active-H-bond (%)	Inactive-H-bond (%)	No-H-bond (%)
A	236	13	6	81
4*	341	35	35	30
6	268	20	41	39
20*	256	84	0	16
21	182	16	68	16
22	152	25	59	16
31*	362	52	2	45
34*	251	73	1	26
35*	204	88	2	9
56	195	5	59	36
59	305	25	25	49

The angle between the plane of pyridine and N···H-O was determined for all individual conformers with $d_{N\cdots H\text{-}O} < 3$ Å for the selected derivatives. As the lone pair of nitrogen is in the plane of pyridine, a planar geometry is indicative of a stronger H-bonding interaction and is essential for proton transfer. A 120° cut-off was applied, and conformers with an N···H-O out-of-plane interaction of pyridine were considered in the **Inactive-H-Bond** cluster. Other conformers with $d_{N\cdots H\text{-}O} < 3$ Å were aligned with the theozyme to assess the extent to which the oxygen nucleophile is exposed to the substrate in the current conformational arrangement and the agreement with the optimal transition state geometry. Conformers with a buried oxygen nucleophile that requires the approach of the substrate from the hindered face of the dyad are in the **Inactive-H-Bond** cluster. The **Active-H-Bond** cluster includes the conformers with an exposed oxygen nucleophile. Table 3.1 summarizes the MM populations of the selected 10 derivatives.

Although 9 of 10 derivatives have an **Active-H-Bond** MM population greater than that of the parent compound **A**, for 5 of the selected derivatives, the conformations that do not allow the optimum geometry required for the acylation step are energetically favored, and the population of the **Inactive-H-Bond** cluster is higher than the population of other

clusters. The derivatives with a higher **Active-H-Bond** MM population compared to the other clusters (Table 3.1, denoted by *) are selected to proceed to the QM evaluation step.

Figure 3.2 shows how the structures of these derivatives differ from that of the parent molecule. **A** is composed of two five-membered building blocks with (S),(S) configuration, each holding one of the catalytic functional groups of the nucleophilic dyad. **D20** results from a ring expansion in the five-membered building block carrying the alcohol moiety to a six-membered building block with the same stereochemical information. **D34** and **D35** are obtained by the amide-to-ketone transformation rule and involve the saturation of the ring nitrogen in each of the building blocks, holding pyridine and alcohol, respectively. This results in a new stereocenter in place of the peptide bond formed from their coupling. The nature of these transformations suggests a possible relief in the backbone strain caused by the arrangement of the catalytic functionalities upon formation of the H-bond. MM evaluation predicts derivatives **D20**, **D34**, and **D35** as the most promising, having an Active-H-Bond MM population greater than 70 percent **Inactive-H-Bond** conformations are significantly disfavored in these derivatives, and the lowest-energy **No-H-Bond** conformer is 1.3 to 1.8 kcal mol^{-1} higher in energy than the global minimum, which is an **Active-H-Bond** conformer. The individual populations of the lowest-energy conformers are 41, 46, and 42 percent for **D20**, **D34**, and **D35**, respectively.

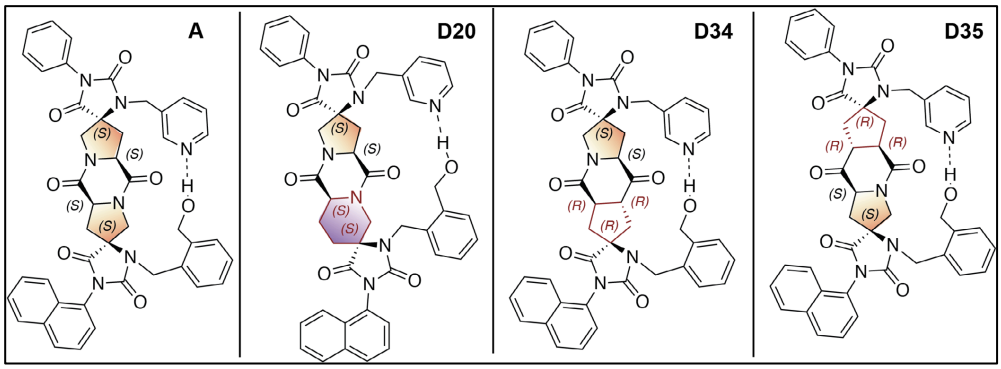

Figure 3.2. Most promising derivatives according to MM and QM evaluations. The modifications are shown as red.

3.1.3. QM Evaluation

Based on the MM evaluation results, five derivatives were subjected to QM evaluation. QM populations of the selected derivatives are summarized in Table 3.2 (Derivatives denoted by * are selected for MD evaluation). The geometries of the lowest-energy conformer of each cluster and their energies with respect to the global minimum are given in Figure 3.3.

Even though assessing the strength of an H-bond is difficult and depends on many parameters, it is generally accepted that the shorter the distance and closer the angle to linear, the stronger the interaction. The aromatic nature of the catalytic groups, benzyl alcohol and pyridine, also suggest the importance of the π-π and C-H···π types of interactions in stabilizing spiroligozyme conformers; yet, it is difficult to quantify and comment on the energetic stability provided by these interactions.

Table 3.2. QM populations of selected derivatives

Derivative Number	**Number of Conformer**	**Active-H-bond (%)**	**Inactive-H-bond (%)**	**No-H-bond (%)**
A	236	13	6	81
4	341	35	35	30
20*	256	84	0	16
31	362	52	2	45
34*	251	73	1	26
35*	204	88	2	9

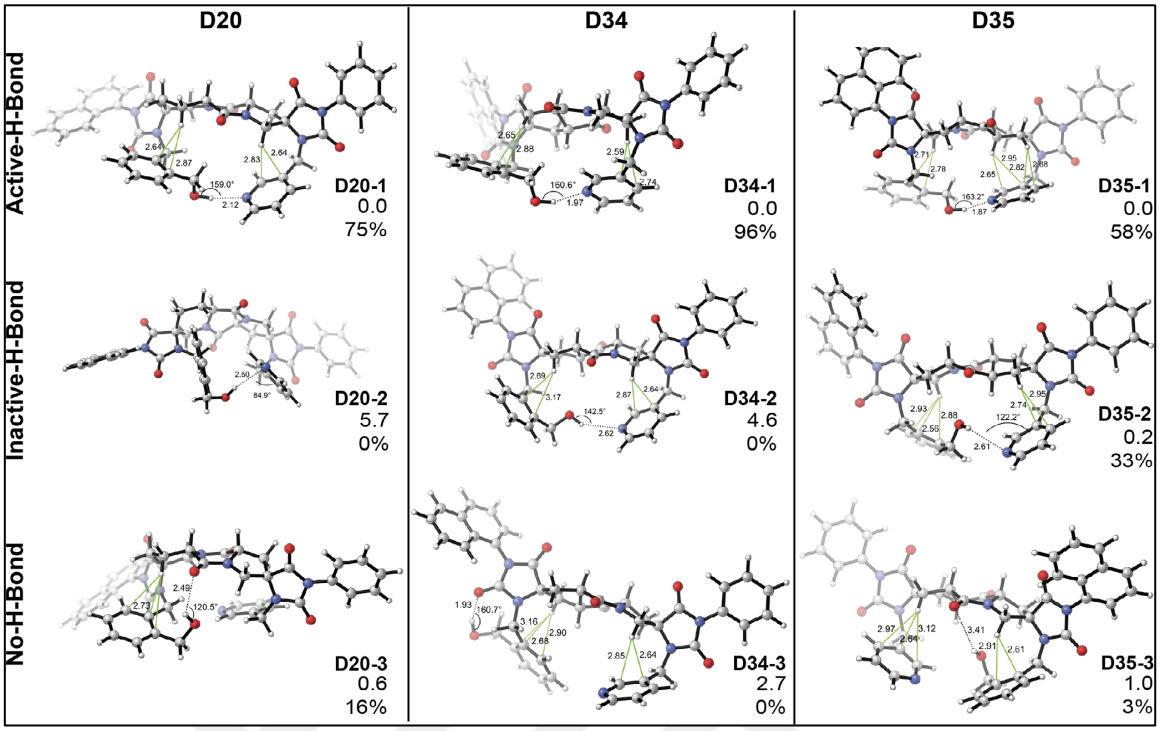

Figure 3.3. Geometries, relative energies (kcal mol^{-1}), and populations of the lowest-energy conformers in each cluster of promising derivatives based on QM evaluation.

The three best conformers based on MM evaluation (**D20**, **D34**, and **D35**) have an **Active-H-Bond** QM population above 60 percent, much higher than that of the parent molecule (8 percent). For derivatives **D4** and **D31**, on the other hand, density functional theory strongly stabilized the **Inactive-H-Bond** conformers, which constitute more than 90 percent of the total population.

For **D20**, 24 of 256 conformers have an N···H-O distance less than 3 Å. Of these, 19 overlays excellently with the theozyme and allow a favorable geometry for the acylation reaction. The lowest-energy conformer **D20-1** (Figure 3.3) constitutes 75 percent of the total population. **D20-1** is 0.6 kcal mol^{-1} more stable than the lowest-energy **No-H-Bond** conformer (Figure 3.3, **D20-3**, 16 percent individual QM population) and 5.7 kcal mol^{-1} compared to the lowest-energy **Inactive-H-Bond** conformer (Figure 3.3, **D20-2**, 0 percent individual QM population).

Density functional theory predicts **D34** as the most promising derivative. Of 251 conformers, 13 display $d_{N\cdots H\text{-}O} < 3$ Å, 6 of which have an exposed oxygen nucleophile. **D34-1** is strongly stabilized, having an individual QM population of 96 percent. The

lowest-energy **Inactive-H-Bond** conformer (Figure 3.3, **D34-2,** 0 percent individual QM population) and the lowest-energy **No-H-Bond** conformer (Figure 3.3, **D34-3**, 0% individual QM population) are 4.6 kcal mol^{-1} and 2.7 kcal mol^{-1} higher in energy, respectively.

For **D35**, of 204 conformers, 18 have an N···H-O distance below 3 Å, and 12 present a nicely oriented catalytic dyad for acylation. Even though the lowest-energy **Active-H-Bond** conformer, **D35-1**, has 58 percent individual QM population, **D35-2**, the **Inactive-H-Bond** conformer, is strong competition, being only 0.2 kcal mol^{-1} higher in energy with an individual QM population of 33 percent. The energy difference between the lowest-energy **Active-H-Bond** and **No-H-Bond** conformers is also as low as 1 kcal mol^{-1}.

3.1.4. MD Evaluation

In the last step, three of the derivatives were evaluated using MD simulations. The H-bond occupancies calculated from MD trajectories are previously shown to be good indicators of activity of protein catalysts involving H-bond networks [33]. Six replicas were run for each system. To calculate the occupancy of the intramolecular H-bonding interactions in selected spiroligozymes, a distance cut-off of 3.5 Å and an angle cut-off of 120° were applied as suggested by the ptraj module of AMBER 12. That is, MD population of each spiroligozyme describes the percent ratio of frames with $d_{N\cdots H\text{-}O} < 3.5$ Å and $120° < \theta_{N\cdots H\text{-}O} < 180°$ to the total number of frames in the simulation and represented by the cumulation of points in the above-mentioned region in the N···H-O distance-angle scatter plot given for a representative run (Figure 3.4). MD populations obtained as averages of six replicas are tabulated in Table 3.3.

D34 and **D35** show only slight improvement of the MD population (5 and 6 percent, respectively) compared to the parent molecule **A**, whereas **D20** displays a much stronger tendency toward the required H-bonding interaction between the catalytic dyad, with a significant increase in the MD population (28 percent). Time-distance graphs and the snapshot geometries given in Figure 3.5 indicate that the H-bond between the catalytic dyad in **D20** is disrupted from time to time, mostly due to competing alternative interactions of the alcohol with the backbone amide oxygens; yet, it is reestablished, and the catalytic groups remain in close proximity throughout the simulation.

All calculations using different methodologies suggest **D20** as a promising spiroligozyme derivative with a better-maintained H-bond between the catalytic dyad compared to the parent molecule. **D20** is also attractive due to its ease of accessibility by simply exchanging a five-membered cyclic building block with an analogous six-membered one in synthesis. Finally, to evaluate the catalytic performance of **D20** in an acyl transfer reaction, we calculated the energy profile of the **D20**-catalyzed transesterification reaction between vinyl trifluoromethylacetate and methanol.

Table 3.3. MD populations of selected derivatives

Derivative Number	**MD Population (%)**
A	4.27 ± 0.85
20	27.3 ± 3.33
34	5.01 ± 1.24
35	5.93 ± 0.76

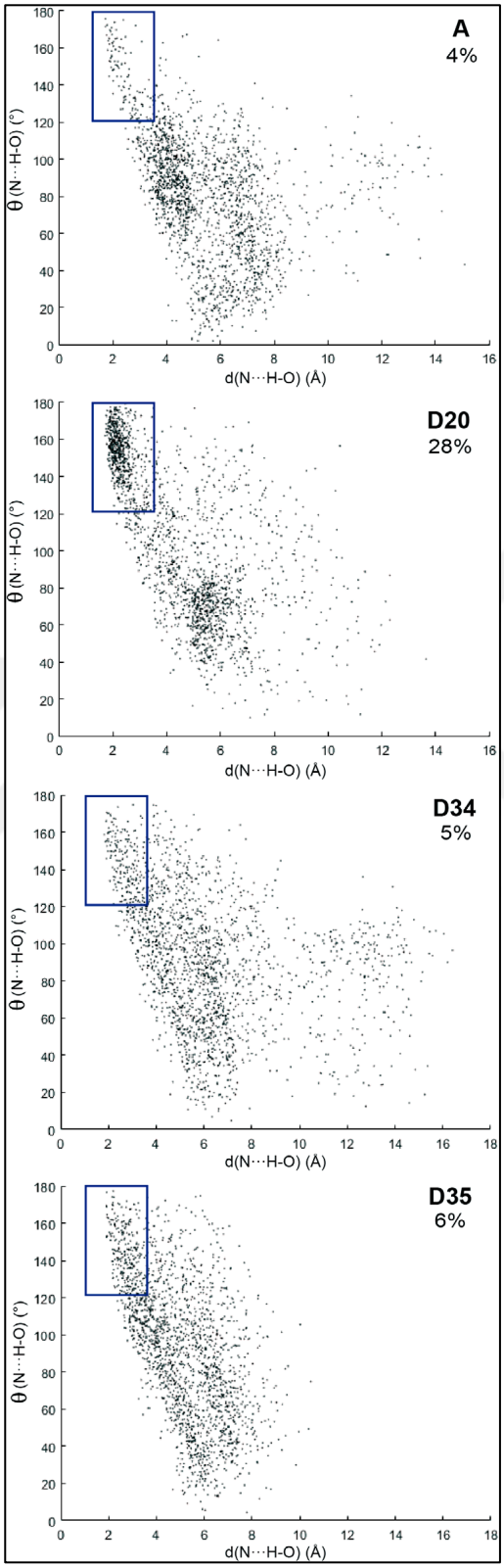

Figure 3.4. N···H-O distance-angle scatter plots and MD populations from MD trajectories of one of the representative replicas.

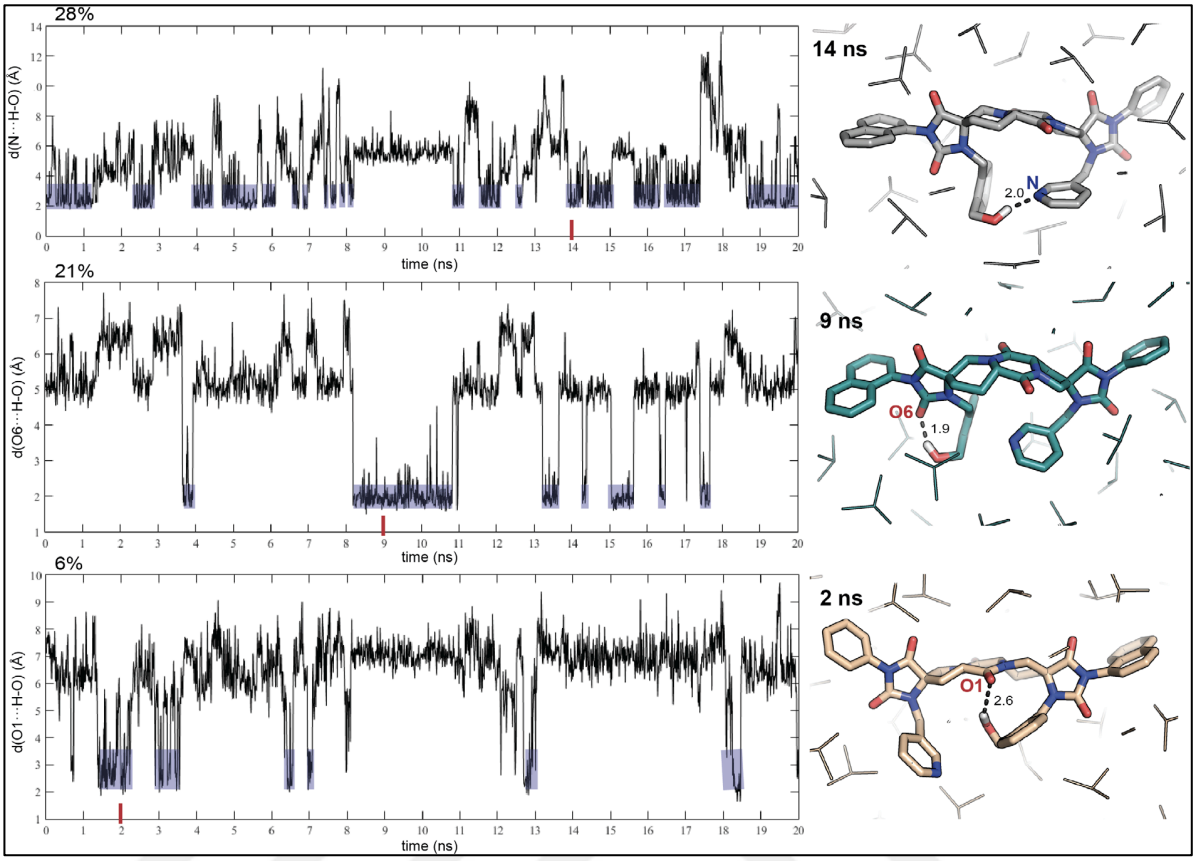

Figure 3.5. Time-distance plots and representative snapshots from the MD simulations for alternative intramolecular interactions in **D20**.

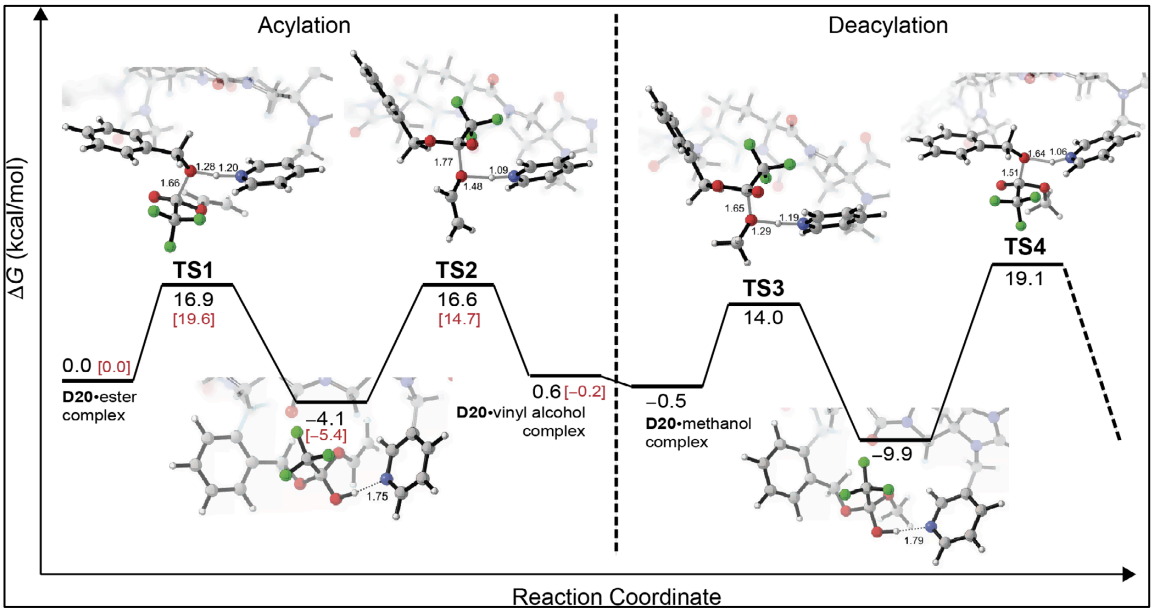

Figure 3.6. Time Energy profile for the **D20**-catalyzed acyl transfer reaction. Relative energies for the analogous acylation pathway of the parent.

3.1.5. Energy Profile

In order to assess whether the spiroligozyme derivative predicted to have an improved H-bond occupancy could function as designed, we computed the stationary points along the transesterification pathway of vinyltrifluoromethylacetate catalyzed by **D20**.

The geometries and relative energies of the optimized structures for the **D20**-catalyzed acyl transfer reaction, together with the analogous acylation pathway calculated with the parent spiroligozyme **A** (relative energies in square brackets), are given in Figure 3.6. The transesterification mechanism starts with the activation of the alcohol nucleophile by abstraction of its proton by pyridine through a proton-shuttle mechanism and the subsequent attack of the nucleophile to the substrate. **TS1** shows a concerted transition state, in which the alcohol is partially deprotonated by pyridine (d_{O-H} = 1.28 Å, d_{N-H} = 1.20 Å) and attacks trifluoromethylacetate (d_{C-O} = 1.66 Å) with an activation barrier of 16.9 kcal mol^{-1}. The activation energy of the analogous transition state located for the acylation of the parent compound **A** is 2.7 kcal mol^{-1} higher ($\Delta G^{\ddagger}$ = 19.6 kcal mol^{-1}), indicative of more facile acylation of **D20** compared to **A**. This transition state leads to a tetrahedral oxyanionic intermediate, which is stabilized by multiple H-bonding interactions in a so-called oxyanion hole in natural enzymes [28-30]. Yet, as **D20** is a bifunctional spiroligozyme with no oxyanion hole motif, in a rather nonpolar solvent, the calculations predict that this oxyanionic tetrahedral intermediate is very unstable and abstract the hydrogen from the protonated pyridine to form a very stable intermediate, which is 4.1 kcal mol^{-1} lower in energy than the starting complex. Even though the solvent effects are included in the geometry optimizations that are essential for stabilizing the zwitterions to prevent a strong artificial tendency for neutralization, chloroform is a solvent with rather low polarity with a dielectric constant of 4.81, and this behavior may well be an artifact of calculations involving zwitterionic intermediates. Yet it is not possible to rule out the aging of the spiroligozyme as conformational flexibility of the tetrahedral species facilitates its protonation in a low-polarity environment. Proton shuffle and conformational reorganization of pyridine enables the activation of vinyl alcohol as the leaving group. **TS2** shows general-acid-assisted loss of vinyl alcohol with a breaking C-O bond of 1.77 Å, with a concurrent proton transfer from pyridine (d_{O-H} = 1.48 Å, d_{N-H} = 1.09 Å). **TS2** is 20.7 kcal mol^{-1} uphill from the tetrahedral intermediate (20.1 kcal mol^{-1} uphill for the parent

spiroligozyme **A**). The exchange of vinyl alcohol with methanol in the generated acylated spiroligozyme vinyl alcohol complex (ΔG = 0.6 kcal mol^{-1}) gives a slightly more stable species (ΔG = −0.5 kcal mol^{-1}), initiating the deacylation step. **TS3** depicts the addition of methanol to the acylated spiroligozyme as methanol is deprotonated by pyridine with an activation barrier of 14.0 kcal mol^{-1}. Similar to **TS2**, following the minimum energy path from **TS3** leads directly to protonation of the oxyanionic tetrahedral intermediate by pyridine, forming a very stable intermediate, which is −9.9 kcal mol^{-1} lower in energy than the starting complex. Release of the transesterification product via **TS4** requires a conformational rearrangement in which the protonated pyridine interacts with the benzyl alcohol of the spiroligozyme for its activation as a leaving group (ΔG = 19.1 kcal mol^{-1}), from which the cleavage of the covalent bond between the substrate and spiroligozyme releasing the transesterification product is a downhill process. These calculations suggest that the transesterification mechanism catalyzed by a bifunctional spiroligozyme in a nonpolar environment in the absence of an oxyanion hole motif involves a high degree of proton shuffling in a catalytic cycle with a free energy span of 29.0 kcal mol^{-1}. This clearly indicates the necessity of inclusion of an oxyanion hole motif in the spiroligozyme to attain reasonable catalytic efficiencies.

3.2. SMALL MOLECULE CATALYST SCREENING OF ZINC DATABASE

The ZINC database is used for pharmacophore search of theozyme model given in Figure 3.7. ZINC database offers users the information of the molecules such as purchasability, molecule chemical information and vendor references. ZINC database is a large database containing over 750 million purchasable compounds. To reduce the scanning time, the database is filtered according to purchasability (Table 3.4), reactivity (Table 3.5) and pH ranges (Table 3.6). By taking in consideration the fact that finding a good match which can be obtained effortlessly is important, wait ok and in stock ZINC molecules with different pH range were chosen and a new search library was created which is given in Table B.1 [61].

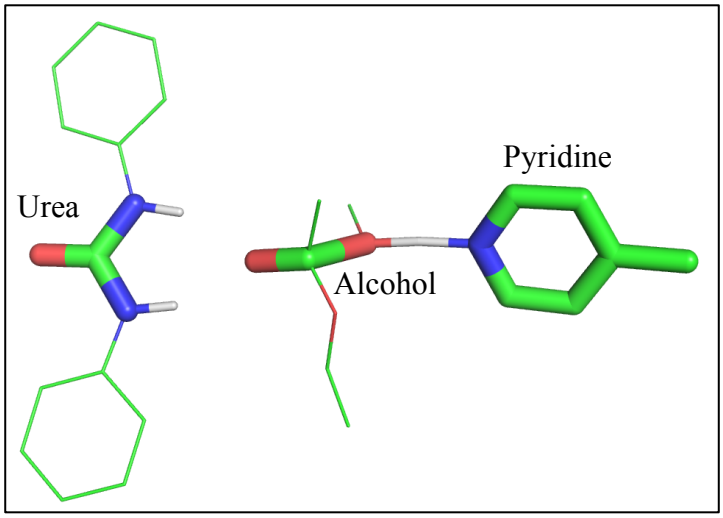

Figure 3.7. Theozyme structure for a trifunctional catalyst.

Table 3.4. ZINC reference information of purchasability [61]

Name	Description
In-stock	2 weeks, direct from manufacturer
On-demand	8-10 weeks, 65% success
Boutique	As on-demand and may be expensive
Annotated	Not currently for sale
Benched	Not in any catalog
Now	In-stock + agent
Wait-ok	Now + on-demand
For-sale	Wait-ok + boutique

Table 3.5. ZINC reference information of reactivity [61]

Name	Description
Anodyne	Compound matches no pattern
Pains-ok	PAINS sans mechanism ok
Clean	No reactive group allowed
Standard	Weakly reactive permitted
Reactive	Reactive allowed
Unstable	Even boronic acids ok

Table 3.6. ZINC reference information of pH [61]

Name	Description
Reference (ref)	7.4
Middle (mid)	6.4 to 8.4
Low pH (lo)	5.4 to 6.4
High pH (hi)	8.4 to 9.4

By using the pharmacophore editor of MOE software (Figure 3.8), the pharmacophore model of the theozyme is created from the parent molecule to search for matches of similar molecules in the molecule library created by filtering the ZINC database. For the desired pharmacophore model, the atom and/or ring is added, the radius is determined and the atom and/or ring is chosen as donor, acceptor or other definitions which fits the modeling with the editor. In Figure 3.8, the pharmacophore model of the theozyme is created by selecting urea, pyridine and alcohol.

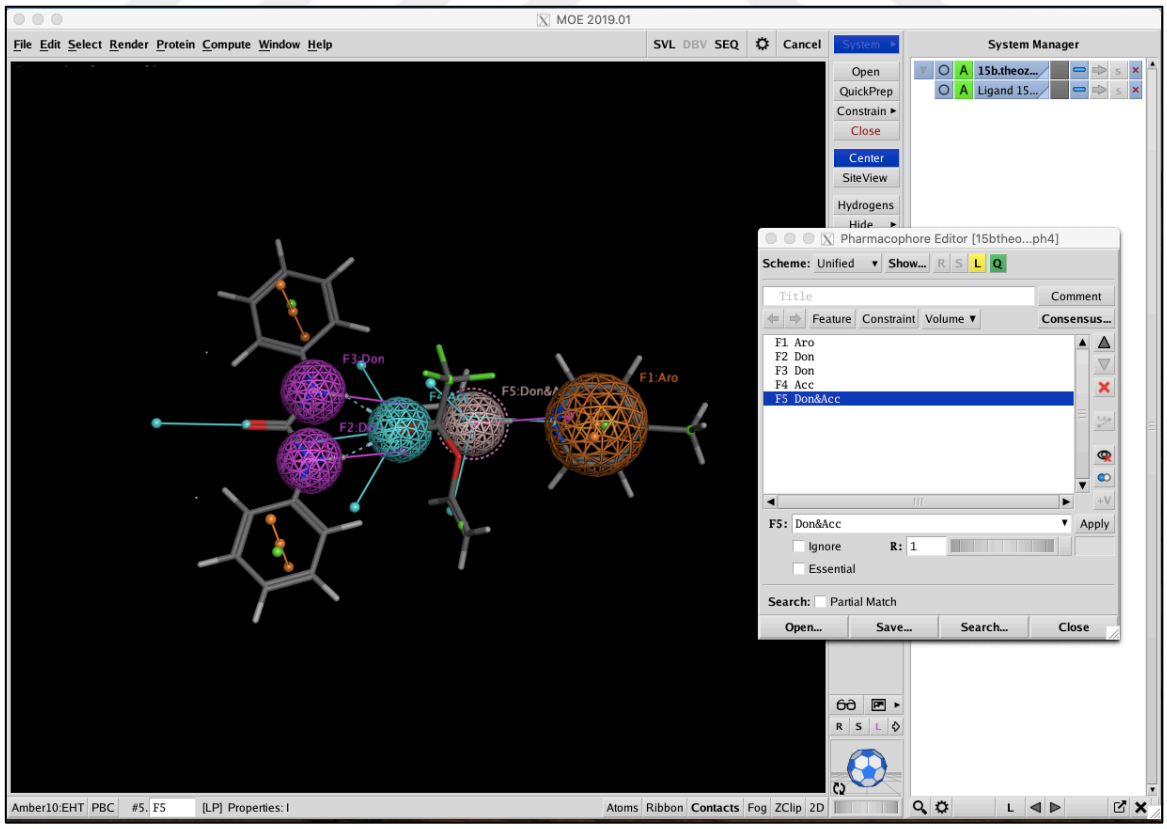

Figure 3.8. Pharmacophore model of the theozyme created by pharmacophore editor in MOE software [31]

The pharmacophore editor offers a feature that the search can be made for partial match. In other words, in this project 5 features were selected for the molecule match search. The search was made for no partial match (exact match) and at least 4 matches (Figure 3.9) since the aim is to find the potential best match. The number of hits is given in Table B.2 for the match search. To begin with, the hit molecules from the no partial match search of each library was examined.

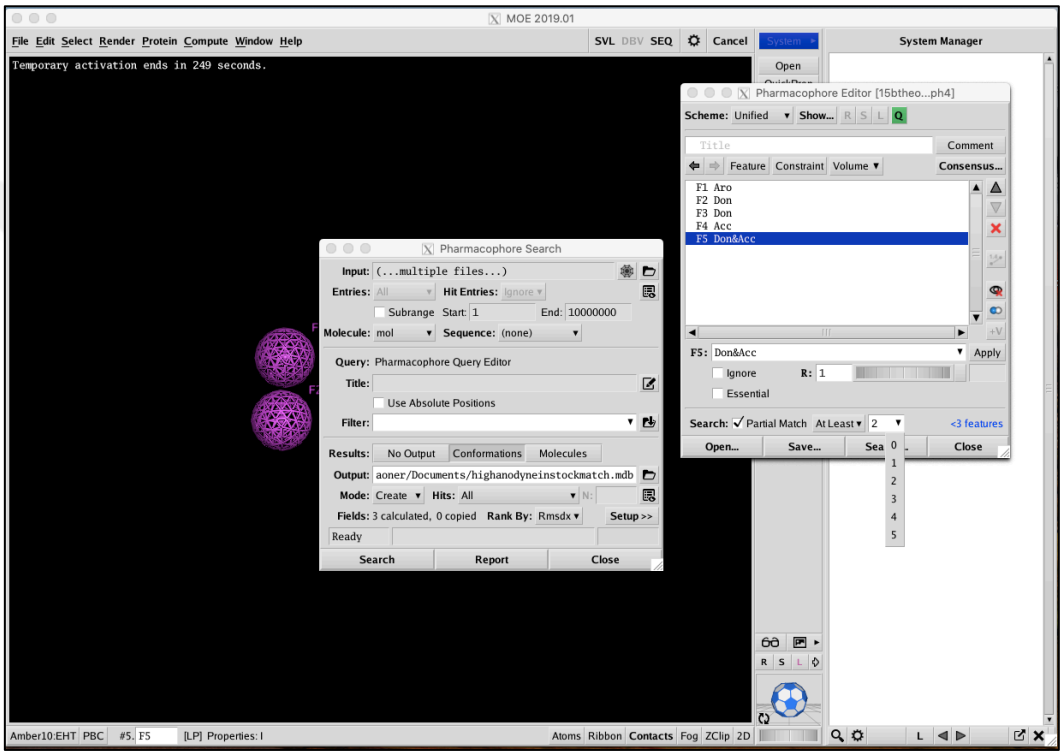

Figure 3.9. Pharmacophore search screen used for the match search of the model [31]

The exact match of each library was examined and 3 promising molecules were determined. In Figure 3.10, the promising match of high anodyne in-stock library match; in Figure 3.11, the promising match of high bother agent library match and in Figure 3.12 the promising match of low clean wait-ok library match can be seen.

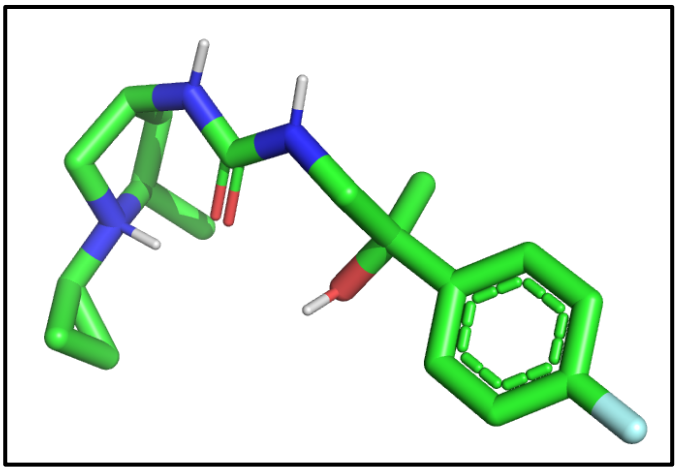

Figure 3.10. High anodyne in-stock library promising match.

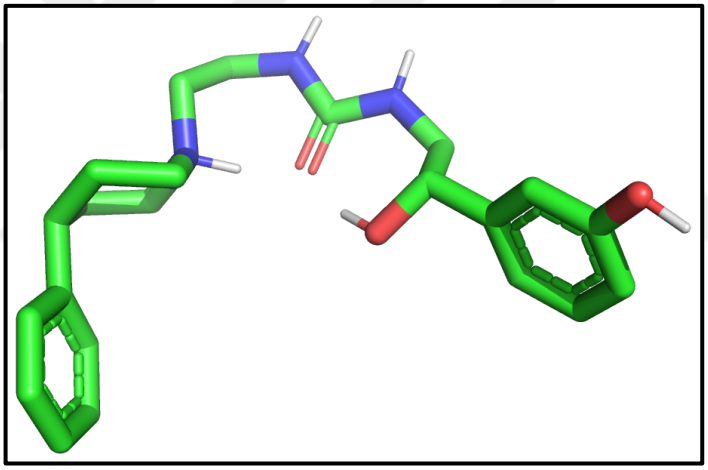

Figure 3.11. High bother agent library promising match

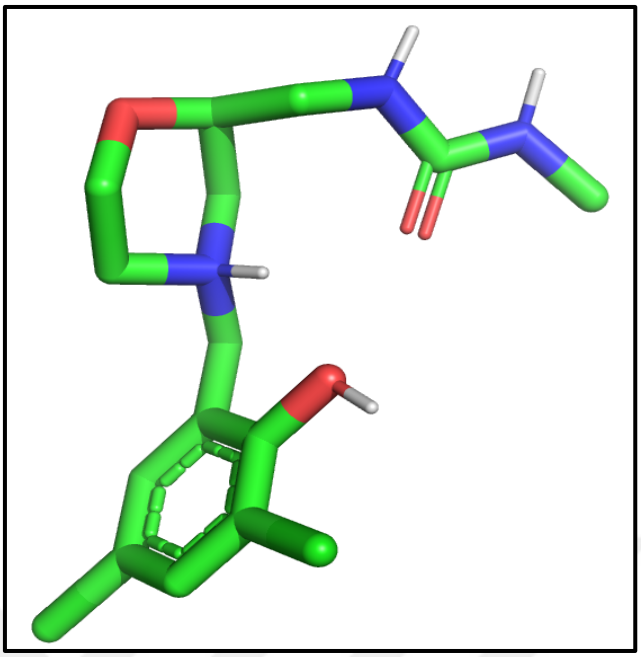

Figure 3.12. Low clean wait-ok library promising match

4. CONCLUSION

Derivatization of the parent bifunctional spiroligozyme A yielded 116 compounds, which are evaluated by the proposed computational protocol using a ladder of state-of-the-art methods of increasing complexity. Calculations predict that the H-bond between the catalytic dyad is significantly better maintained in **D20** compared to the parent compound, yet suggest the importance of the inclusion of an oxyanion hole motif to reach reasonable catalytic activities for the transesterification of vinyl trifluoromethylacetate. In order to enhance the catalytic efficiency of this bifunctional spiroligozyme, expanding **D20** scaffold to involve a urea/thiourea motif can be considered as a future work.

With the aim of identifying new organocatalyst candidates that involve an oxyanion hole motif beside a catalytic dyad, a computational approach that combines the predictive power of quantum mechanical calculations with drug design tools was employed. The pharmacophore search of a trifunctional theozyme was performed in ZINC database which was narrowed down to wait-ok and in-stock molecules to find catalytic hits easily accessible.

Partial and exact match queries of the generated pharmacophore models returned catalytic hits with amine, alcohol and urea motifs in 3D constellations depicted in catalytic atom map based on the QM optimized theozyme for a trifunctional catalyst.

Promising molecules can further be evaluated and modified with the above-mentioned computational protocol to identify catalysts with higher efficiency.

REFERENCES

1. Rahman R , Uahengo V , Daniel LS. Green chemistry concept: Applications of catalysis in pharmaceutical insdustry. Global Drugs and Therapeutics. 2017; 2(4): 3-6.
2. Radzicka A , Wolfenden R. A proficient enzyme. Science. 1995; 267(5194): 90-3.
3. Blanco A , Blanco G. Enzymes. Medical Biochemistry. 2017;: 153-157.
4. Sinibaldi A , Nori V , Baschieri A , Fini F , Arcadi A , Carlone A. Organocatalysis and beyond: activating reactions with two catalytic species. Catalysts. 2019; 9(11): 928.
5. van Der Helm MP , Klemm B , Eelkema R. Organocatalysis in aqueous media. Nature Reviews Chemistry. 2019; 3: 491-508.
6. Zabka M , Sebesta R. Experimental and theoretical studies in hydrogen-bonding organocatalysis. Molecules. 2015; 20(9): 15500-15524.
7. Auvil TJ , Schafer AG , Mattson AE. Design strategies for enhanced hydrogen-bond donor catalysts. European Journal of Enhanced Hydrogen-Bond Donor Catalysts. 2014; 2014(13): 2633-2646.
8. Serdyuk OV , Heckel CM , Tsogoeva SB. Bifunctional primary amine-thioureas in asymmetric organocatalysis. Organic and Biomolecular Chemistry. 2013; 11(41): 7051-7071.
9. Marchetti L , Levine M. Biomimetic catalysis. ACS Catalysis. 2011; 1(9): 1090-1118.
10. Larionov E , Zipse H. Organocatalysis: acylation catalysts. Wiley Interdisciplinary Reviews Computational Moleculer Science. 2011; 1(4): 601-619.
11. Cheong PHY , Legault CY , Um JM , Celebi-Olcum N , Hou KN. Quantum mechanical investigations of organocatalysis: mechanisms, reactivities, and selectivities. Chemical Reviews. 2011; 111(8): 5042-5137.
12. Knowles RR , Jacobsen EN. Attractive noncovalent interactions in asymmetric catalysis: Links between enzymes and small molecule catalysts. Proceedings of the Natural Academy of Sciences of the United States of America. 2010; 107(48): 20678-20685.
13. MacMillan DWC. The advent and development of organocatalysis. Nature. 2008; 455: 304-308.

14. Houk KN , Cheong PHY. Computational prediction of small-molecule catalysts. Nature. 2008; 455(7211): 309-313.
15. Ema T , Tanida D , Matsukawa T , Sakai T. Biomimetic trifunctional organocatalyst showing a great acceleration for the transesterification between vinyl ester and alcohol. Chemical Communications. 2008; 8: 957-959.
16. Parker MFL , Osuna S , Bollot G , Vaddypally S , Zdilla MJ , Houk KN , et al. Acceleration of an aromatic claisen rearrangement via a designed spiroligozyme catalyst that mimics the ketosteroid isomerase catalytic dyad. Journal of American Chemical. 2014; 136(10): 3817-3827.
17. Kheirabadi M , Celebi-Olcum N , Parker MFL , Zhao Q , Kiss G , Houk KN , et al. Spiroligozymes for transesterifications: design and relationship of structure to activity. Journal of American Chemical Society. 2012; 134(44): 18345-18353.
18. Leadbeater NE. Organic synthesis using microwave heating. Comprehensive Organic Synthesis (Second Edition). 2014; 9: 234-286.
19. McCarthy SM , Melman JH , Reffel OK , Gordon-Wylie SW. Synthesis and partial characterization of biodiesel via base-catalyzed transesterification. Bioenergy. 2015;: 361-365.
20. Mumtaz MW , Adnan A , Mukhtar H , Rashid U , Danish M. Biodiesel production through chemical and biochemical transesterification: trends, technicalities and future perspectives. Clean Energy for Sustainable Development. 2017;: 465-485.
21. Norjannah B , Ong HC , Masjuki HH , Juan JC , Chong WT. Enzymatic transesterification for biodiesel production: a comprehensive review. RSC Advances. 2016; 6(65): 60034-60055.
22. Ponnarasy G , Khan MR , Kalam A , Mahmud MS. Light induced esterification of oleic acid catalyzed by pseudomonas cepacia lipase. International Journal of Environmental Science and Development. 2014; 5(4): 344-346.
23. Hou CT , Shimada Y. Lipases. Encyclopedia of Microbiology (Third Edition). 2009;: 385-392.
24. Hedstrom L. Serine protease mechanism and specifity. Chemical Reviews. 2002; 102(12): 4501-4524.
25. Kirby A. Efficiency of proton transfer catalysis in models and enzymes. Accounts of

Chemical Research. 1997; 30(7): 290.

26. Warshel A , Sharma PK , Kato M , Xiang Y , Liu H , Olsson MHM. Electrostatic basis for enzyme catalysis. Chemical Reviews. 2006; 106(8): 3210-3235.
27. Smith AJT , Muller R , Toscano MD , Kast P , Hellinga HW , Hilvert D , et al. Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. Journal of the American Chemical Society. 2008; 130(46): 15361-153753.
28. Menard R , Storer AC. Oxyanion hole interactions in serine and cysteine proteases. Biological Chemistry Hoppe-Seyler. 1992; 373(7): 393-400.
29. Kirby A. Enzyme mechanisms, models and mimics. Angewandte Chemie International Edition England. 1996; 35(7): 706-724.
30. Nothling MD , Xiao Z , Bhaskaran A , Blyth MT , Bennett CW , Coote ML , et al. Synthetic Catalysts Inspired by Hydrolytic Enzymes. ACS Catalysis. 2019; 9(1): 168-187.
31. Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2022. .
32. Oner A , Celebi-Olcum N. Rapid computational evaluation of small-molecule hydrolase mimics for preorganized H-bond networks. International Journal of Quantum Chemistry. 2021; 121(2).
33. Kiss G , Rothlisberger D , Baker D , Houk KN. Evaluation and ranking of enzyme designs. Protein Science. 2010; 19(9): 1760-1773.
34. Marques HM , Brown KL. Molecular mechanics and molecular dynamics simulations of porphyrins, metalloporphyrins, heme proteins and cobalt corrinoids. Coordination Chemistry Reviews. 2002; 225(1-2): 123-158.
35. Danovich D , Shaik S , Chen H. Theoretical toolkits for inorganic and bioinorganic complexes: their applications and insights. rehensive Inorganic Chemistry II (Second Edition). 2013; 9: 1-57.
36. Leach AR. Molecular Modelling Principles and Applications: Pearson Education Limited; 2001.
37. Halgren TA. MMFF VI. MMFF94s option for energy minimization studies. Journal of

Computational Chemistry. 1999; 20(7): 720-729.

38. Wojciechowski M , Lesyng B. Generalized born model: analysis, refinement, and applications to proteins. The Journal of Physical Chemistry B. 2004; 108(47): 18368-18376.
39. Kolossvary I , Guida WC. Low mode search. an efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides. Journal of American Chemical Society. 1996; 118(21): 5011-5019.
40. Labute P. LowModeMD--implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. Journal of Chemical Information and Modeling. 2010; 50(5): 792-800.
41. Gece G. The use of quantum chemical methods in corrosion inhibitor studies. Corrosion Science. 2008; 50(11): 2981-2992.
42. Zhao Y , Truhlar DG. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06 functionals and 12 other functionals. Theoretical Chemistry Accounts. 2008; 120: 215-241.
43. Cossi M , Barone V , Cammi R , Tomasi J. Ab initio study of solvated molecules: a new implementation of the polarizable continuum model. Chemical Physics Letters. 1996; 255(4-6): 327-335.
44. Cances E , Mennucci B , Tomasi J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. The Journal of Chemical Physics. 1997; 107(8): 3022-3041.
45. Tomasi J , Mennucci B , Cammi R. Quantum mechanical continuum solvation models. Chemical Reviews. 2005; 105(8): 2999-094.
46. Marenich AV , Cramer CJ , Truhlar DG. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. The Journal of Physical Chemistry B. 2009; 113(18): 6378-6396.
47. Frisch MJ , Trucks GW , Schlegel HB , Scuseria GE , Robb MA , Cheeseman JR , et al. Gaussian, Inc., Wallingford CT, 2009. .

48. Case DA , Darden TA , Cheatham TE , III CLS , Wang J , Duke RE , et al. AMBER, Vol. 12, University of California, San Francisco, CA 2012. .
49. Wang J , Wolf RM , Caldwell JW , Kollman PA , Case DA. Development and testing of a general amber force field. Journal of Computational Chemistry. 2004; 25(9): 1157-1174.
50. Wang JM , Cieplak P , Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry. 2000; 21(12): 1049-1074.
51. Bayly CI , Cieplak P , Cornell W , Kollman PA. well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. Journal of Physical Cehmistry. 1993; 97(40): 10269-10280.
52. Besler BH , Merz KM , Kollman PA. Atomic charges derived from semiempirical methods. Journal of Computational Chemistry. 1990; 11(4): 431-439.
53. Singh UC , Kollman PA. An approach to computing electrostatic charges for molecules. Journal of Computational Chemistry. 1984; 5(2): 129-145.
54. Cieplak P , Caldwell J , Kollman P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of nucleic acid bases. Journal of Computational Chemistry. 2001; 22(10): 1048-1057.
55. Darden T , York D , Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. Journal of Computational Chemistry. 1993; 98(12): 10089-10092.
56. Knapp B , Ospina L , Deane CM. Avoiding false positive conclusions in molecular simulation: the importance of replicas. Journal of Chemical Theory and Computation. 2018; 14(12): 6127-6138.
57. Gonzalez C , Schlegel HB. An improved algorithm for reaction path following. Journal of Chemical Physics. 1989; 90(4): 2154-2161.
58. Gonzalez C , Schlegel HB. Reaction path following in mass-weighted internal coordinates. Journal of Chemical Physics. 1990; 94(14): 5523-5527.
59. Afshar M , Lanoue A , Sallantin J. Multiobjective/multicriteria optimization and

decision support in drug discovery. Comprehensive Medicinal Chemistry II. 2007; 4: 767-774.

60. Martin YC. Pharmacophore modeling:1 – methods. Comprehensive Medicinal Chemistry II. 2007; 4: 119-147.
61. Sterling T , Irwin JJ. ZINC 15 – Ligand discovery for everyone. Journal of Chemical Information and Modeling. 2015; 55(11): 2324-2337.
62. Irwin JJ , Sterling T , Mysinger MM , Bolstad ES , Coleman RG. ZINC: A free tool to discover chemistry for biology. Journal of Chemical Information and Modeling. 2012; 52(7): 1757-1768.
63. Irwin JJ , Shoichet BK. ZINC – A free database for commercially available compounds for virtual screening. Journal of Chemical Information and Modeling. 2005; 45(1): 177-182.
64. Cieplak P , Caldwell J , Kollman P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/. Journal of Computational Chemistry. 2001; 22(10): 1048-1057.

APPENDIX A: BIFUNCTIONAL SPIROLIGOZYME DATABASE

Table A.1. The information of 116 derivative database of BF3 spiroligozyme

Derivative Number	Transformation Reaction	Number of Conformers	MM Population (%)	
			H-Bond $d_{N\cdots H\text{-}O} < 3$ Å	No-H-Bond $d_{N\cdots H\text{-}O} > 3$ Å
A	Parent Molecule	236	19	81
1	C=O to C=S	325	34	66
2	C=O to C=S	283	23	77
3	C=O to C=S	337	26	4
4	C=O to C=S	341	70	30
5	C=O to C=S	184	3	97
6	C=O to C=S	268	60	40
7	C=O to SO2	250	13	87
8	C=O to SO2	169	0	100
9	C=O to SO2	239	27	73
10	C=O to SO2	209	3	97
11	C=O to SO2	236	10	90
12	C=O to SO2	162	0	100
13*	C[CNOSFClBrI] to C(OMe)	-	-	-
14*	C[CNOSFClBrI] to C(OMe)	-	-	-
15*	C[CNOSFClBrI] to C(OMe)	-	-	-
16*	C[COSFClBrI] to C(N)	-	-	-
17*	C[NOSFClBrI] to C(C)	-	-	-
18*	C[NOSFClBrI] to C(cyano)	-	-	-
19	Cyclo56	272	0	100
20	Cyclo56	256	84	16
21	Cyclo56	187	84	16
22	Cyclo56	152	84	16
23	Cyclo56	192	1	99
24	Cyclo56	190	0	100

Derivative	Transformation Reaction	Number of	MM Population (%)	
25*	ace [NOS]	-	-	-
26*	amide to amine	-	-	-
27*	amide to amine	-	-	-
28	amide to amine	422	4	96
29	amide to amine	177	3	97
30	amide to hydroxyethyl	446	6	94
31	amide to hydroxyethyl	362	55	45
32	amide to hydroxyethyl	146	17	83
33	amide to hydroxyethyl	448	4	96
34	amide to ketone	251	74	26
35	amide to ketone	204	91	9
36	amide to ketone	280	47	53
37	amide to ketone	180	42	58
38	amide to ketone	365	39	61
39	amide to ketone	324	7	93
40	amine to amide	427	47	53
41*	amine to amide	-	-	-
42	amine to amide	272	1	99
43*	amine to amide	-	-	-
44	amine to amide	334	4	96
45	aro6 N 12 switch	449	8	92
46	aro6 N 12 switch	438	0	100
47	benzene to pyridine	338	25	75
48	benzene to pyridine	300	12	88
49	benzene to pyridine	255	8	92
50*	benzene to pyridine	-	-	-
51*	benzene to pyridine	-	-	-
52*	benzene to pyridine	-	-	-
53*	benzene to pyridine	-	-	-
54	benzene to pyridine	183	3	97
55	benzene to pyridine	279	33	67

Derivative	Transformation Reaction	Number of	MM Population (%)	
56	benzene to pyridine	195	64	36
57	benzene to pyridine	257	43	57
58	benzene to pyridine	261	20	80
59	benzene to pyridine	305	51	49
60	benzene to pyridine	202	22	78
61	hom [NOS]CH2	239	1	99
62*	hom [NOS]CH2	-	-	-
63	hom [NOS]CH2	175	1	99
64*	phenyl to indole	-	-	-
65*	phenyl to indole	-	-	-
66*	phenyl to indole	-	-	-
67	phenyl to indole	248	5	95
68	phenyl to indole	220	12	88
69	phenyl to indole	240	6	94
70*	phenyl to indole	-	-	-
71*	phenyl to indole	-	-	-
72*	phenyl to indole	-	-	-
73*	phenyl to indole	-	-	-
74*	phenyl to indole	-	-	-
75*	phenyl to indole	-	-	-
76*	phenyl to indole	-	-	-
77*	phenyl to indole	-	-	-
78*	phenyl to indole	-	-	-
79*	phenyl to indole	-	-	-
80*	phenyl to indole	-	-	-
81*	phenyl to indole	-	-	-
82	phenyl to indole	213	41	59
83	phenyl to thiophene 2	287	20	80
84	phenyl to thiophene 2	250	11	89
85	phenyl to thiophene 2	343	8	92
86	phenyl to thiophene 2	270	13	87

Derivative	Transformation Reaction	Number of	MM Population (%)	
87	phenyl to thiophene 2	274	13	87
88	phenyl to thiophene 2	308	23	77
89*	phenyl to thiophene 2	-	-	-
90*	phenyl to thiophene 2	-	-	-
91*	phenyl to thiophene 2	-	-	-
92*	phenyl to thiophene 2	-	-	-
93*	phenyl to thiophene 2	-	-	-
94*	phenyl to thiophene 2	-	-	-
95*	phenyl to thiophene 2	-	-	-
96*	phenyl to thiophene 2	-	-	-
97*	phenyl to thiophene 2	-	-	-
98	phenyl to thiophene 2	291	42	58
99	phenyl to thiophene 2	302	33	67
100	phenyl to thiophene 2	275	45	55
101	phenyl to thiophene 2	317	26	74
102	phenyl to thiophene 2	239	37	63
103	phenyl to thiophene 2	313	42	58
104	phenyl to thiophene 2	360	36	64
105	phenyl to thiophene 2	289	13	87
106	phenyl to thiophene 2	325	36	64
107	phenyl to thiophene 2	252	11	89
108	phenyl to thiophene 2	314	15	85
109	phenyl to thiophene 2	324	43	57
110	phenyl to thiophene 2	259	22	78
111	phenyl to thiophene 2	318	26	74
112	phenyl to thiophene 2	208	20	80
113*	pyridine to benzene	-	-	-
114*	pyridine to pyrimidine	-	-	-
115	urea to guanidine	205	37	63
116	urea to guanidine	212	32	68

* The derivatives not considered since change occurred in catalytic functional groups.

APPENDIX B: ZINC DATABASE

Table B.1. ZINC Database match library information used during the study

File Number	pH	Reactivity	Purchasability
1	High	Anodyne	Agent
2	High	Anodyne	In-stock
3	High	Bother	Agent
4	High	Bother	In-stock
5	High	Clean	In-stock
6	High	Clean	Wait-ok
7	High	Hot	In-stock
8	High	Reactive	In-stock
9	High	Reactive	Wait-ok
10	High	Standard	In-stock
11	High	Standard	Wait-ok
12	Low	Anodyne	In-stock
13	Low	Bother	In-stock
14	Low	Bother	Wait-ok
15	Low	Clean	In-stock
16	Low	Clean	Wait-ok
17	Low	Hot	In-stock
18	Low	Hot	Wait-ok
19	Low	Reactive	In-stock
20	Low	Standard	In-stock
21	Low	Standard	Wait-ok
22	Reference-Middle	Anodyne	In-stock
23	Reference-Middle	Bother	In-stock
24	Reference-Middle	Clean	In-stock
25	Reference-Middle	Hot	Wait-ok
26	Reference-Middle	Reactive	In-stock

File Number	pH	Reactivity	Purchasability
27	Reference-Middle	Reactive	Wait-ok
28	Reference-Middle	Standard	In-stock
29	Reference-Middle	Standard	Wait-ok

Table B.2. The pharmacophore model match result of the ZINC Database library search

File No.	pH / Reactivity / Purchasability	Number of Hits	
		No Partial Match	At Least 4 Match
1	High / Anodyne / Agent	2	170
2	High / Anodyne / In-stock	9	183
3	High / Bother / Agent	2	98
4	High / Bother / In-stock	0	35
5	High / Clean / In-stock	0	8
6	High / Clean / Wait-ok	0	25
7	High / Hot / In-stock	0	7
8	High / Reactive / In-stock	0	12
9	High / Reactive / Wait-ok	2	38
10	High / Standard / In-stock	0	29
11	High / Standard / Wait-ok	0	29
12	Low / Anodyne / In-stock	1	244
13	Low / Bother / In-stock	0	51
14	Low / Bother / Wait-ok	4	142
15	Low / Clean / In-stock	1	34
16	Low / Clean / Wait-ok	1	53
17	Low / Hot / In-stock	6	89
18	Low / Hot / Wait-ok	1	110
19	Low / Reactive / In-stock	8	59
20	Low / Standard / In-stock	10	261
21	Low / Standard / Wait-ok	8	384
22	Reference-Middle / Anodyne / In-stock	0	25
23	Reference-Middle / Bother / In-stock	0	8
24	Reference-Middle / Clean / In-stock	0	20
25	Reference-Middle / Hot / Wait-ok	0	0
26	Reference-Middle / Reactive / In-stock	1	11
27	Reference-Middle / Reactive / Wait-ok	2	19

File No.	pH / Reactivity / Purchasability	Number of Hits	
		No Partial Match	At Least 4 Match
28	Reference-Middle / Standard / In-stock	1	15
29	Reference-Middle / Standard / Wait-ok	0	8