

Shape Theory Applied to Molecular Docking and Automatic Localization of Ligand Binding Pockets in Large Proteins

Iliana Ramírez-Velásquez,* Álvaro H. Bedoya-Calle, Ederley Vélez, and Francisco J. Caro-Lopera*



Cite This: *ACS Omega* 2022, 7, 45991–46002



Read Online

ACCESS |



Metrics & More

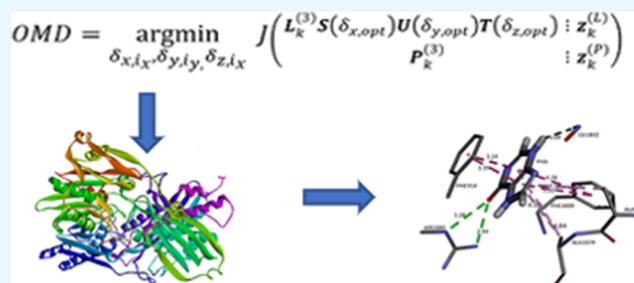


Article Recommendations



Supporting Information

ABSTRACT: Automatic search of cavities and binding mode analysis between a ligand and a 3D protein receptor are challenging problems in drug design or repositioning. We propose a solution based on a shape theory theorem for an invariant coupled system of ligand–protein. The theorem provides a matrix representation with the exact formulas to be implemented in an algorithm. The method involves the following results: (1) exact formulae for the shape coordinates of a located-rotated invariant coupled system; (2) a parameterized search based on a suitable domain of van der Waals radii; (3) a scoring function for the discrimination of sites by measuring the distance between two invariant coupled systems including the atomic mass; (4) a matrix representation of the Lennard-Jones potential type 6–12 and 6–10 as the punctuation function of the algorithm for a molecular docking; and (5) the optimal molecular docking as a solution of an optimization problem based on the exploration of an exhaustive set of rotations. We apply the method in the xanthine oxidase protein with the following ligands: hypoxanthine, febuxostat, and chlorogenic acid. The results show automatic cavity detection and molecular docking not assisted by experts with meaningful amino acid interactions. The method finds better affinities than the expert software for known published cavities.



1. INTRODUCTION

There is currently a growing interest in the search for therapeutic targets and the development of different computational techniques that support research in the field of pure and applied sciences, such as the area of pharmaceutical chemistry. There is also a marked advance in machine learning techniques, where the shape theory has emerged as a powerful tool in this area.^{1–3} The union between a molecule and a receptor is the starting point for the design of drugs and their subsequent synthesis.⁴

The 3D ligand-based models analyze the structure–activity relationship. Namely, the molecular docking explores the conformational space and uses a scoring function to classify the poses of the ligand in the active site of the protein.^{5,6} Additionally, it is based on the geometric complementarity between the ligand and the receptor.^{7–10} For this reason, the positions of atoms are essential for detecting a protein region characterized by favorable ligand–receptor interactions.¹¹

The literature reports several studies on the detection and descriptions of protein cavities. These works include geometric methods,^{8,12} energy field methods,^{10,13–15} molecular dynamics calculations,^{16,17} residues of conserved sequences,^{18–20} and physicochemical characteristics of the system.¹²

Another approach considers the evaluation of molecular similarity based on equivalent structures and properties. These methods are based on the interatomic distance, atom-centered Gaussian overlay-based representations, Zernike descriptors,

and spherical harmonics surfaces.^{9,11,21–26} Additionally, these approaches have been combined with molecular docking, generating successful results in discovering inhibitors.^{27,28} The classical methods explore the chemical and biological spaces; however, there is an interest in developing alternative methods centered on the geometric invariances in non-Euclidean spaces. The shape complementarity can be seen as the Riemann space information after filtering some Euclidean noise.^{29,30}

A shape theory technique can score the interaction between molecules by equivalence classes. They can be seen as partitions of a quotient space quoting the differences by some Riemannian geometric descriptors.^{29,31} Some automatic scanning methods use the Lennard-Jones (L-J) potential^{32,33} to score the receptor–ligand complex's geometry. It allows the conformational space exploration without limitations in the number of atoms of the interacting molecules. The incorporation of the shape theory^{29,31,34} in the context of molecular docking appeared for the first time in a recent work.² Multiple applications in chemistry, physics, and engineering

Received: April 10, 2022

Accepted: October 11, 2022

Published: December 6, 2022



have been reported. For example, ref 3 applied the shape theory for spectroscopic analysis and ref 35 studied rocking curves, both in the context of experimental physics; in engineering, ref 36 implemented a remote sensing analysis of burning satellite products; finally, in the area that concerns us, ref 37 set the shape theory for the automatic search of large clusters not assisted by experts.

In general, the shape theory is a broad branch of statistics. Initially, models based on the normal distribution received wide diffusion, for example, refs 3038, and 39. Then, robust models were included in several numerical fields: real normalized division algebras (real, complex, quaternion, and octonion); multiple transformations such as SVD, QR, pseudo-Wishart, polar, and so forth.^{40–43}

Some tools such as those defined in refs 1044, and 45 were used to perform an automatic search for cavities based on multiple classification criteria. However, they require a priori energy calculations or physicochemical information of the studied system. With the correct parameters, the Riemannian geometry can perform automatic exploration of cavities in large proteins, avoiding the usual manual intervention of an expert. In particular, we can search cavities and perform molecular docking by finding new shape coordinates of a coupled ligand–protein system. The invariance property means that only the significantly different scenarios are studied. The shape theory usually deals with separated clusters, but in this case, the required ligand–protein complementarity claims the setting of an invariant couple system. For doing so, this paper establishes a new theorem with the exact formulas in a matrix form describing the automatic search and the molecular docking. The proposed method involves the following parts:

1. Exact expressions with the shape's coordinates of the invariant coupled ligand–protein.
2. A parametric search based on the exploration of a suitable domain of van der Waals radii.
3. A shape and atomic mass descriptor for similarity or dissimilarity between two any coupled systems.
4. An interaction energy score criterion based on the L-J 6–12 and 6–10 type potentials.³³
5. Geometry optimization routine for reaching the optimal molecular docking (OMD) via the L-J potential.

Once the OMDs are obtained, profuse interaction studies can be performed using expert molecular docking software.⁴⁵

We organize the document as follows: In Section 2, we present the case study. Section 3 describes the methods and algorithms. Section 4 gives the results and discussion. The following applications of the algorithm are fully detailed:

Application 1. A known system: XO–HPA.

Application 2. A forbidden pocket (an apparently narrow cavity for the ligand): XO–chlorogenic Acid.

Application 3. A new cavity: XO–FBX.

Application 4. A new cavity: XO–chlorogenic Acid.

Application 5. A new cavity: XO–HPA and complementary analysis using AUTODOCK VINA.

2. CASE STUDY

In a previous work,² we used the enzyme xanthine oxidase (XO), which is distributed in various species, from bacteria to humans, and is present in mammalian tissues. XO contains the cofactor FAD (flavin adenine dinucleotide), a molybdenum cofactor, and [2Fe–2S] centers. Mammalian enzymes may have xanthine dehydrogenase (XDH) activity. They exist primarily in the cell, but they are readily converted to the XO form by oxidation of sulfhydryl residues. XO uses dioxygen as a substrate, leading to the formation of a superoxide anion, hydrogen peroxide, and urate.^{46,47} The inhibition of its activity is the objective in treating diseases such as gout caused by hyperuricemia and vascular inflammation, among others. For this reason, the importance of XO lies in its affinity for producing reactive oxygen species (ROS), a situation implicated in several pathological states.^{48–51} Consequently, other investigations seek to minimize ROS formation through various approaches, including molecular docking with the XO enzyme. XO hydroxylates HPA, and then, the xanthine is obtained, which is then converted into uric acid, generating ROS.⁴⁷ The inhibition of XO is relevant in treatments to prevent the accumulation of uric acid and, therefore, ROS.^{52,53} In our case, we selected the XO enzyme in complex with hypoxanthine (HPA) and the ligands febuxostat (FBX) and chlorogenic acid,⁵⁴ which are XO inhibitors.

3. METHODS AND ALGORITHMS

In this section, we propose and prove the paper's main result about the automatic search of cavities in large proteins and molecular docking via the L-J potential.

The algorithm is based on the following result:

Theorem. Consider a protein of n atoms and a ligand with $m \ll n$ atoms. $\mathbf{p}_j = (p_{j,1}, p_{j,2}, p_{j,3})$, $j = 1, \dots, m$, denotes the original Euclidean coordinates of the atoms in the protein $\mathbf{P} = (\mathbf{p}_j)$. Also let $\mathbf{l}_i = (l_{i,1}, l_{i,2}, l_{i,3})$, $i = 1, \dots, m$, be the original location of the atoms in the ligand $\mathbf{L} = (\mathbf{l}_i)$. Assume that $2b$ (eq 2) is the largest Euclidean distance in the ligand, which is reached between the points $\mathbf{l}_u = (l_{u,1}, l_{u,2}, l_{u,3})$ and $\mathbf{l}_v = (l_{v,1}, l_{v,2}, l_{v,3})$. In some symmetric clusters, there could exist more than one pair of points associated to such a maximal distance, in that case, we just take any of them. In the same way, consider that the pair $\mathbf{p}_w = (p_{w,1}, p_{w,2}, p_{w,3})$ and $\mathbf{p}_t = (p_{t,1}, p_{t,2}, p_{t,3})$ of the protein attain the maximal distance $2d$, (eq 3).

Then:

Part (1): A located–rotated invariant coupled system of shape coordinates for ligand–protein is given by $\mathbf{C}^{(3)} = \begin{pmatrix} \mathbf{L}^{(3)} \\ \mathbf{P}^{(3)} \end{pmatrix}$.

The invariant ligand is the $m \times 3$ matrix $\mathbf{L}^{(3)} = (\mathbf{l}_i^{(3)})$, where $\mathbf{l}_i^{(3)} = (l_{i,1}^{(3)}, l_{i,2}^{(3)}, l_{i,3}^{(3)})$, $i = 1, \dots, m$; with

$$\begin{aligned}
 l_{i,1}^{(3)} &= \frac{-l_{u,1}^2 - l_{u,2}^2 + 2l_{i,1}(l_{u,1} - l_{v,1}) + l_{v,1}^2 + 2l_{i,2}(l_{u,2} - l_{v,2}) + l_{v,2}^2 + (2l_{i,3} - l_{u,3} - l_{v,3})(l_{u,3} - l_{v,3})}{2\sqrt{(l_{u,1} - l_{v,1})^2 + (l_{u,2} - l_{v,2})^2 + (l_{u,3} - l_{v,3})^2}}, \\
 l_{i,2}^{(3)} &= \frac{-(l_{i,3}l_{u,1}l_{v,3} + l_{u,2}l_{v,1} - l_{i,3}l_{u,3}l_{v,1} + l_{u,3}l_{v,1} + l_{i,2}(l_{u,1} - l_{v,1})(l_{u,2} - l_{v,2}) - l_{u,1}l_{u,2}l_{v,2} - l_{u,2}l_{v,1}l_{v,2} + l_{u,1}l_{v,2}^2 - l_{i,1}((l_{u,2} - l_{v,2})^2 + (l_{u,3} - l_{v,3})^2) - (l_{u,1}(l_{i,3} + l_{u,3}) + (-l_{i,3} + l_{u,3})l_{v,1})l_{v,3} + l_{u,1}l_{v,3}^2)/(\sqrt{(l_{u,2} - l_{v,2})^2 + (l_{u,3} - l_{v,3})^2})\sqrt{(l_{u,1} - l_{v,1})^2 + (l_{u,2} - l_{v,2})^2 + (l_{u,3} - l_{v,3})^2})}{}, \\
 l_{i,3}^{(3)} &= -\frac{l_{i,3}(l_{u,2} - l_{v,2}) + l_{u,3}(-l_{i,2} + l_{v,2}) + (l_{i,2} - l_{u,2})l_{v,3}}{\sqrt{(l_{u,2} - l_{v,2})^2 + (l_{u,3} - l_{v,3})^2}}
 \end{aligned} \quad (1)$$

Also, the invariant protein is the $n \times 3$ matrix $\mathbf{P}^{(3)} = (p_{j,1}^{(3)}, p_{j,2}^{(3)}, p_{j,3}^{(3)})$, $j = 1, \dots, n$, where the formulae for $p_{j,1}^{(3)}$, $p_{j,2}^{(3)}$, and $p_{j,3}^{(3)}$ are just the expressions of $l_{i,1}^{(3)}$, $l_{i,2}^{(3)}$, and $l_{i,3}^{(3)}$, respectively, substituting $l \rightarrow p$, $i \rightarrow j$, $m \rightarrow n$, $u \rightarrow w$, and $v \rightarrow t$.

Part (2): When $i = u$, the above rigid transformations send atom \mathbf{l}_u to $\mathbf{l}_u^{(3)} = (b, 0, 0)$, with

$$b = \frac{1}{2}\sqrt{(l_{u,1} - l_{v,1})^2 + (l_{u,2} - l_{v,2})^2 + (l_{u,3} - l_{v,3})^2} > 0 \quad (2)$$

Meanwhile, atom \mathbf{l}_v is placed at $\mathbf{l}_v^{(3)} = (-b, 0, 0)$. For the other values of i , the remaining $m - 2$ atoms \mathbf{l}_i , $i = 1, \dots, m$; $i \neq u, v$ in the original ligand are sent to the given expressions (1) in the three Euclidean space. For the protein, the original atoms at $\mathbf{p}_w, \mathbf{p}_t$ are, respectively, sent to $\mathbf{p}_w^{(3)} = (d, 0, 0)$ and $\mathbf{p}_t^{(3)} = (-d, 0, 0)$, where

$$d = \frac{1}{2}\sqrt{(p_{w,1} - p_{t,1})^2 + (p_{w,2} - p_{t,2})^2 + (p_{w,3} - p_{t,3})^2} > 0 \quad (3)$$

and the remaining $n - 2$ atoms are rigidly translated to the addressed coordinates $\mathbf{p}_j^{(3)}$, $j = 1, \dots, n$; $j \neq w, t$ from (1).

Part (3): For a more meaningful docking along the segment of atoms $\mathbf{p}_w^{(3)}$ and $\mathbf{p}_t^{(3)}$, (3) is restricted such that

$$d = b + r \quad (4)$$

with $b > 0$ (eq 2). $r > 0$ is a real parameter in a certain appropriate interval, (g, h) , varying in a suitable domain of the van der Waals radius, according to the atoms located at $\mathbf{l}_u^{(3)}, \mathbf{l}_v^{(3)}, \mathbf{p}_w^{(3)}, \mathbf{p}_t^{(3)}$. The q possible ligand–protein invariant

coupled system $\mathbf{C}_k^{(3)} = \begin{pmatrix} \mathbf{L}_k^{(3)} \\ \mathbf{P}_k^{(3)} \end{pmatrix}$, $k = 1, \dots, q$ is considered,

satisfying the restriction for d . The atomic mass of the system is added in a fourth column, and the new coupled system is

denoted by $\mathbf{M}_k^{(3)} = \begin{pmatrix} \mathbf{L}_k^{(3)} : \mathbf{z}_k^{(L)} \\ \mathbf{P}_k^{(3)} : \mathbf{z}_k^{(P)} \end{pmatrix}$, where $\mathbf{z}_k^{(L)}$ is a $m \times 1$ vector

containing the atomic mass of the k -th ligand $\mathbf{L}_k^{(3)}$. Similarly, $\mathbf{z}_k^{(P)}$ is a $n \times 1$ vector with the atomic mass of the k -th protein $\mathbf{P}_k^{(3)}$.

A shape and atomic mass descriptor for similarity or dissimilarity between any two coupled systems $\mathbf{M}_e^{(3)}$ and $\mathbf{M}_f^{(3)}$ is given by the following distance

$$\text{dist}(\mathbf{M}_e^{(3)}, \mathbf{M}_f^{(3)}) = \sqrt{T_e^2 + T_f^2 + 2T_e T_f \cos R} \quad (5)$$

where

$$\begin{aligned}
 T_o &= \sqrt{\text{trace}\{[\mathbf{D}\mathbf{M}_o^{(3)}]'\mathbf{D}\mathbf{M}_o^{(3)}\}}, \quad o = e, f, \\
 \mathbf{D} &= \begin{pmatrix} \mathbf{I}_{m+n} - \frac{1}{m+n}\mathbf{1}_{m+n}\mathbf{1}'_{m+n} \end{pmatrix}
 \end{aligned}$$

\mathbf{I} is the identity matrix, $\mathbf{1}$ is a vector of ones, and $'$ denotes the transpose. Finally, R is the Riemannian distance of $\mathbf{M}_e^{(3)}$ and $\mathbf{M}_f^{(3)}$, namely,

$$R = \text{Riemannian distance}(\mathbf{M}_e^{(3)}, \mathbf{M}_f^{(3)}) = \cos^{-1} \sum_{s=1}^4 \alpha_s$$

where $0 \leq R \leq \pi/2$ and α_s are the square roots of the eigenvalues of $\mathbf{A}_e' \mathbf{A}_e \mathbf{A}_f' \mathbf{A}_f$. Here, $\mathbf{A}_o = \frac{1}{T_o} \mathbf{B}\mathbf{M}_o^{(3)}$, $\mathbf{o} = e, f$;

$\mathbf{B} = (\mathbf{b}_i)$, is $(m + n - 1) \times (m + n)$ matrix such that

$$\mathbf{b}_i = \left(\begin{array}{c} -\frac{1}{\sqrt{i(i+1)}}, \dots, -\frac{1}{\sqrt{i(i+1)}}, -\frac{i}{\sqrt{i(i+1)}}, 0, \\ \dots, 0 \end{array} \right)$$

For details about the Riemannian geometry, see, for example, Kendall (1984).²⁹

Part (4): Instead of studying the complete protein, consider the subset of atoms inside a sphere of radius $2d$, eq 4 centered at the origin. It is possible that the sphere cuts some amino acids of the protein; thus, we must include the necessary atoms outside the sphere. Let n be the number of such confined atoms with complete amino acids. Let $\boldsymbol{\varepsilon} = (\varepsilon_{\tau, \varphi})$, $\boldsymbol{\sigma} = (\sigma_{\tau, \varphi})$, $\tau, \varphi = 1, \dots, \omega$; the matrices of interactions among the total ω classes of elements present in the system ligand–protein $\mathbf{M}_k^{(3)}$, $k = 1, \dots, q$. The $(mn) \times 2$ matrix is defined as $\boldsymbol{\varepsilon} = (\varepsilon_{\beta, \theta})$, $\theta = 1, 2$. The index β recovers all the interactions of the indexes $i = 1, \dots, m$ in the ligand and the indexes $j = 1, \dots, n$ of the protein. It is listed as follows: $\beta = ((1, 1), (1, 2), \dots, (1, n), (2, 1), (2, 2), \dots, (2, n), \dots, (m, 1), (m, 2), \dots, (m, n))$.

Then, the entries of $\boldsymbol{\vartheta}$ are given by

$$\varepsilon_{\beta, 1} = -\frac{4\varepsilon_{\tau(i,j), \varphi(i,j)}\sigma_{\tau(i,j), \varphi(i,j)}^6}{\rho_{i,j}^6} + \frac{4\varepsilon_{\tau(i,j), \varphi(i,j)}\sigma_{\tau(i,j), \varphi(i,j)}^{12}}{\rho_{i,j}^{12}}$$

and

$$\varepsilon_{\beta, 2} = -\frac{\varepsilon_{\tau(i,j), \varphi(i,j)}\sigma_{\tau(i,j), \varphi(i,j)}^{10}}{\rho_{i,j}^{10}} + \frac{\varepsilon_{\tau(i,j), \varphi(i,j)}\sigma_{\tau(i,j), \varphi(i,j)}^{12}}{\rho_{i,j}^{12}}$$

Here, $\tau(i, j)$ and $\varphi(i, j)$ correspond to the entries $\varepsilon_{\tau, \varphi}$, $\sigma_{\tau, \varphi}$ matrices $\boldsymbol{\varepsilon}, \boldsymbol{\sigma}$ according to the i -th atom of the ligand and the j -th atom of the protein. Thus, the L-J energy of the k -th invariant coupled system is given by

$$J(\mathbf{M}_k^{(3)}) = \text{mn}(\overline{\epsilon}_{\cdot,1} + \overline{\epsilon}_{\cdot,2}) \quad (6)$$

where $\overline{\epsilon}_{\cdot,1}$, $\overline{\epsilon}_{\cdot,2}$ are the means of the corresponding columns of the matrix ϵ .

Part (5): The rotation matrix is defined about the y -axis

$$\mathbf{U}(a) = \begin{pmatrix} \cos a & 0 & -\sin a \\ 0 & 1 & 0 \\ \sin a & 0 & \cos a \end{pmatrix}$$

and the rotation matrices $\mathbf{S}(a)$, $\mathbf{T}(a)$ are considered with respect to the x and z axes, given in eqs 7 and 8 of the Supporting Information. A sequence of a_x , a_y , a_z angles are set such that $0 \leq \delta_{x,i_x} < 2\pi$, $i_x = 1, \dots, a_x$; $0 \leq \delta_{y,i_y} < 2\pi$, $i_y = 1, \dots, a_y$; $0 \leq \delta_{z,i_z} < 2\pi$, $i_z = 1, \dots, a_z$. The OMD via the L-J potential of eq 6 in a certain cavity $k = 1, \dots, q$ can be seen as the coupled ligand–protein

$$\mathbf{O}_k^{(3)} = \begin{pmatrix} \mathbf{L}_k^{(3)} \mathbf{S}(\delta_{x,\text{opt}}) \mathbf{U}(\delta_{y,\text{opt}}) \mathbf{T}(\delta_{z,\text{opt}}) & : \mathbf{z}_k^{(L)} \\ & \mathbf{P}_k^{(3)} & : \mathbf{z}_k^{(P)} \end{pmatrix}$$

satisfying the following optimization problem

$$\text{OMD} = \underset{\delta_{x,i_x}, \delta_{y,i_y}, \delta_{z,i_z}}{\text{argmin}} J(\mathbf{O}_k^{(3)}) \quad (9)$$

where $i_x = 1, \dots, a_x$, $i_y = 1, \dots, a_y$, $i_z = 1, \dots, a_z$.

Proof: See the Supporting Information.

Remark: The distance (see eq 5) can be seen as a new punctuation function in molecular docking because it involves an atomic mass feature to the classical geometrical invariant Riemannian distances. We have used it here only to avoid a repetition of searching the same cavity with different symmetrical atoms associated to the same maximal distance $2d$. An explicit relation of eqs 5 and 6 could lead a more robust optimization solution, eq 9. This problem will be part of a future research.

Algorithm. In the following steps, we use the notations and the context given in the Theorem subsection.

Section A: Steps 1 to 13.

1. Set the original protein $\mathbf{P} = (\mathbf{p}_j)$, $j = 1, \dots, n$.
2. Set the original ligand $\mathbf{L} = (\mathbf{l}_i)$, $i = 1, \dots, m$.
3. Find \mathbf{l}_u and \mathbf{l}_v with the largest Euclidean distance $2b$, (eq 2).
4. Set g and h explained in part (3), and provide a sequence r_k , $k = 1, \dots, q$ such that $g \leq r_1 < r_2 < \dots < r_q \leq h$.
5. Set $k = 1$.
6. Find \mathbf{p}_w and \mathbf{p}_t with the largest Euclidean distance $2d = 2b + 2r_k$, (eq 4).
7. Instead of studying the complete protein, consider the subset of atoms inside a sphere of radius $2d$, eq 4, centered at the origin. It is possible that the sphere cuts some amino acids of the protein; thus, the necessary atoms outside the sphere are included. Let n be the number of such confined atoms with complete amino acids. Update the new protein $\mathbf{P}_k = (\mathbf{p}_j)$, $j = 1, \dots, n$, under this restriction.
8. Compute the Riemannian invariant system

$$\mathbf{M}_k^{(3)} = \begin{pmatrix} \mathbf{L}_k^{(3)} : \mathbf{z}_k^{(L)} \\ \mathbf{P}_k^{(3)} : \mathbf{z}_k^{(P)} \end{pmatrix} \text{ with the exact formulae of eq 1.}$$

9. Repeat steps 5 to 8 for $k = 2, \dots, q$.

10. Compute the Riemannian distance of eq 5 between all possible pairs indexed by $e, f \in \{1, \dots, q\}$ of coupled systems $\mathbf{M}_e^{(3)}$ and $\mathbf{M}_f^{(3)}$.
11. Order the systems according to the largest distances found in step 10. Assume that the resulting order is indexed by k_1, k_2, \dots, k_q . Delete from the list those systems that are similar according to a near zero distance in eq 5. Let k_1, k_2, \dots, k_Q the new updated list with the $Q \leq q$ most discrepant systems $\mathbf{M}_{k_i}^{(3)}$, $i = 1, \dots, Q$.
12. Find the L-J energies $\text{LJ}_{k_i} = J(\mathbf{M}_{k_i}^{(3)})$, $i = 1, \dots, Q$ of eq 6.
13. Select the minimum L-J energy of step 12.

Section B: Steps 14 to 16.

14. Set the required parameters in part (5) in order to refine the system of step 13.
15. Finally, the OMD of eq 9 provides an optimal cavity and ligand in the context of the theory level given in the Theorem subsection.
16. For further applications in molecular biology, complete the amino acids of the OMD in step 15 by adding the atoms out of the corresponding sphere in step 7.

The algorithm can be divided in two sections: steps 1 to 13, the search for the best invariant coupled ligand–protein, and steps 14 to 16, giving rise to the OMD suitable for expert analysis in AutoDock Vina or similar software.

Getting Started. We use the free software R⁵⁵ to implement the cavity search, optimization, and classification algorithms under the potential criteria of L-J 6–10 and 6–12 potentials. We downloaded the HPA–XO complex (PDB ID: 3NRZ) from the Protein Data Bank (<http://www.rcsb.org/pdb>), and the compounds FBX (Pubchem CID: 134018) and chlorogenic acid (Pubchem CID: 1794427) from the PubChem repository (<https://pubchem.ncbi.nlm.nih.gov>). We prepare the files via AutoDockTools version 1.5.6.⁴⁵ We then apply the method to the search and optimization process in a known XO region⁵² to obtain the new coordinates of the protein and ligand pockets. Molecular docking is calculated for the protein and ligands in the new coordinate system using the Vina AutoDock tool.⁵⁶ A search of the entire receptor conformational space was then performed for each of the compounds. For the analyses, we used the VMD—Visual Molecular Dynamics software.⁵⁷ Finally, we apply molecular docking to the original coordinates to compare these results with those obtained in the previous step.

4. RESULTS AND DISCUSSION

In this section, we applied the proposed algorithm in the XO protein with three different ligands: HPA, FBX, and chlorogenic acid. The challenge of the method consists of finding three non-reported cavities for the ligands and showing that the L-J potential behaves well in the corresponding molecular docking. The source code used in the applications of this paper is provided in the Supporting Information.

The examples work well for the case study; however, we do not provide a generalization about the performance of the method for any protein. The most important result is the algorithm's OMD output (eq 9), which is based on the proposed theorem contextualized in the L-J formulation.

Each protein deserves a detailed study, involving a number of issues about computation and molecular docking. The Supporting Information provides an additional complete

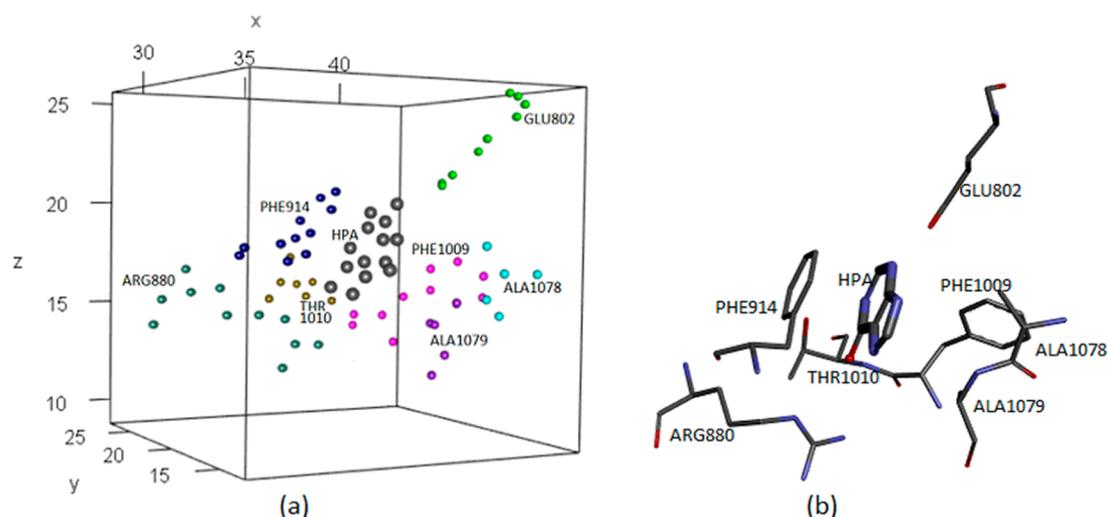


Figure 1. (a) Ligand (HPA) atoms in gray and XO protein pocket atoms in other colors. (b) Interactions of our method obtained by using a visualization software and without any assistance of an expert docking software.

Table 1. Parameterization in Terms of the Pocket Radius, the Number of Atoms, and the L-J Potential

| position | radius (Å) | no. atoms | L-J(kcal/mol) | position | radius (Å) | no. atoms | L-J(kcal/mol) |
|----------|------------|-----------|---------------|----------|------------|-----------|---------------|
| 1 | 0.70 | 24 | -8.21 | 16 | 0.91 | 42 | -10.54 |
| 2 | 0.71 | 26 | -8.62 | 17 | 0.92 | 42 | -10.54 |
| 3 | 0.73 | 29 | -9.82 | 18 | 0.93 | 43 | -10.60 |
| 4 | 0.74 | 30 | -10.52 | 19 | 0.95 | 44 | -10.75 |
| 5 | 0.76 | 32 | -11.90 | 20 | 0.96 | 46 | -8.53 |
| 6 | 0.77 | 34 | -13.04 | 21 | 0.98 | 49 | -8.92 |
| 7 | 0.78 | 34 | -13.04 | 22 | 0.99 | 50 | -9.08 |
| 8 | 0.80 | 35 | -13.30 | 23 | 1.02 | 50 | -9.08 |
| 9 | 0.81 | 36 | -13.52 | 24 | 1.03 | 53 | -9.14 |
| 10 | 0.82 | 37 | -10.20 | 25 | 1.03 | 54 | -9.17 |
| 11 | 0.84 | 39 | -10.25 | 26 | 1.04 | 54 | -9.17 |
| 12 | 0.85 | 41 | -10.40 | 27 | 1.06 | 54 | -9.17 |
| 13 | 0.87 | 41 | -10.40 | 28 | 1.07 | 54 | -9.17 |
| 14 | 0.88 | 42 | -10.54 | 29 | 1.09 | 54 | -9.17 |
| 15 | 0.89 | 42 | -10.54 | 30 | 1.10 | 54 | -9.17 |

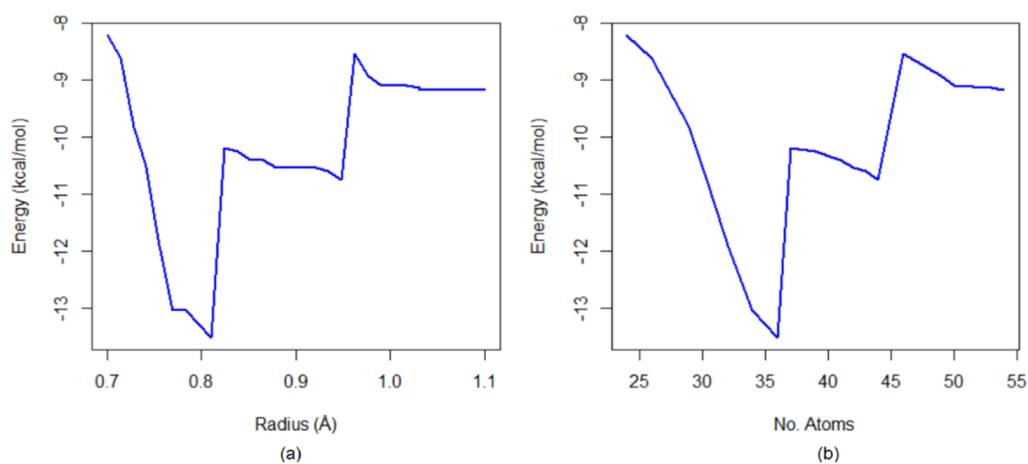


Figure 2. (a) Relationship between the energy and the radius of the pocket and (b) relationship between the energy and the number of atoms in the pocket.

example (user guide) with a low computational time application by using a segment of the Aldose Reductase.

For a complete validation of the method, we will consider a data base with hundreds of proteins and an updated version of the theorem. This topic is under research.

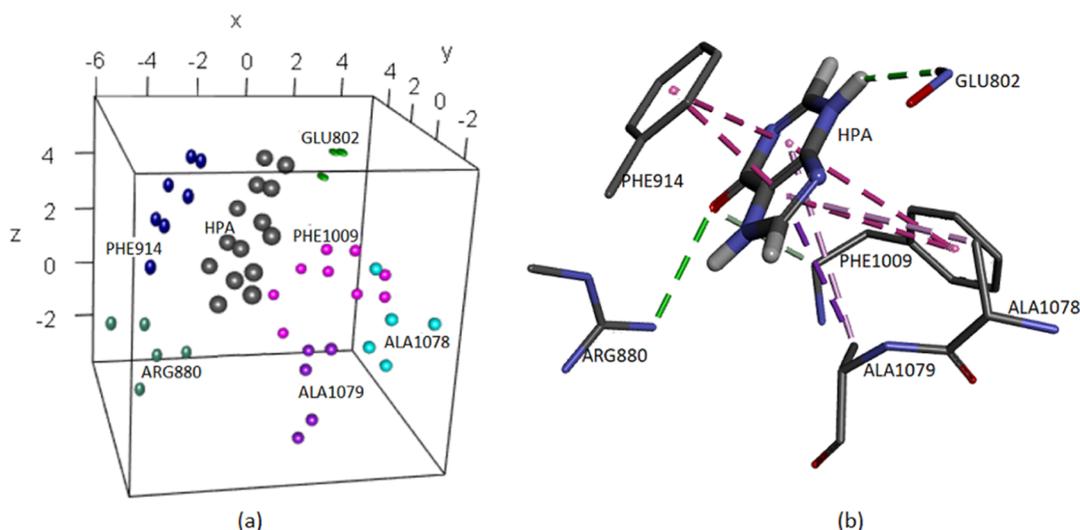


Figure 3. (a) Representation of active sites of the XO–HPA complex found using the new method (Section A of the algorithm). (b) Interactions of our method obtained by using a visualization software and without any assistance of an expert docking software.

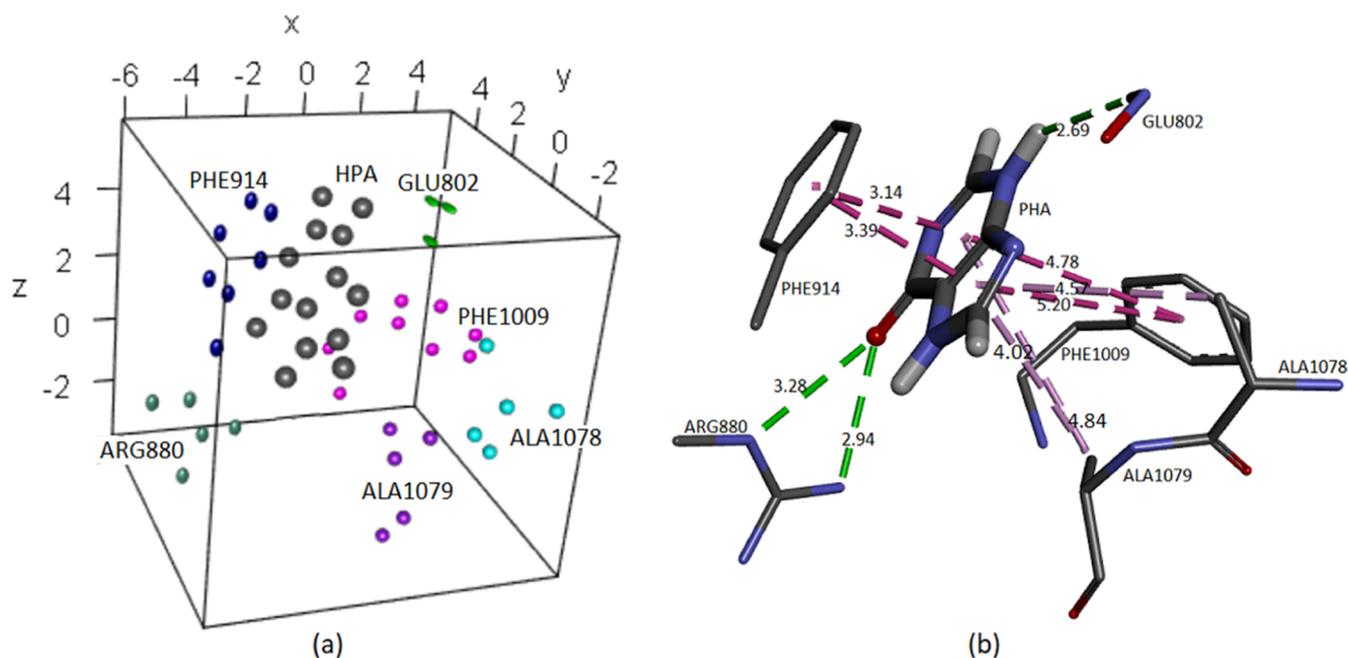


Figure 4. (a) Representation of active sites of the XO–HPA complex found using the new method (Section B of the Algorithm section). (b) Interactions of our method obtained by using a visualization software and without any assistance of an expert docking software.

Application 1. A Known Cavity: XO–HPA. We start with a simple application of docking for the correctness of our method. We take a cavity reported in the literature by an expert. In this case, the input consists of the Euclidean coordinates of the pocket and the ligand. Then, the algorithm will find the invariant coupled system and perform the corresponding docking via the L-J potential.

We start with the crystal structure of the bovine XO enzyme in a complex with HPA at a 1.8 Å resolution. Figure 1 presents two groups, the ligand atoms and the protein pocket, according to ref 47, see the coordinates in Tables S1 and S2. Note that the theorem accepts isolated reference systems for the ligand and the protein, but in this trivial case, the reported coordinates correspond to the optimized docking given by ref 47. In any case, our algorithm considers the worst situation,

and the components are separated in matrices assuming that they are located initially in any Euclidean region.

Figure 1a shows the atoms that make up each amino acid in the active site differentiated by colors, and Figure 1b represents the interactions of the ligand with the amino acids of the protein pocket. The performance of steps 1 to 13 of our algorithm is summarized in Table 1. The method's convergence depends on the cavity radius, the number of atoms in the pocket, and the corresponding L-J potential.

Figure 2a shows the relationship between the energy and the radius of the pocket. Figure 2b shows the energy and the number of atoms in the pocket.

Figure 2 shows a similar energy behavior between the radii and the number of atoms in the spheres.

Thus, the first part of the algorithm (Section 3.2.1) provides an invariant coupled system with 36 atoms in the pocket and

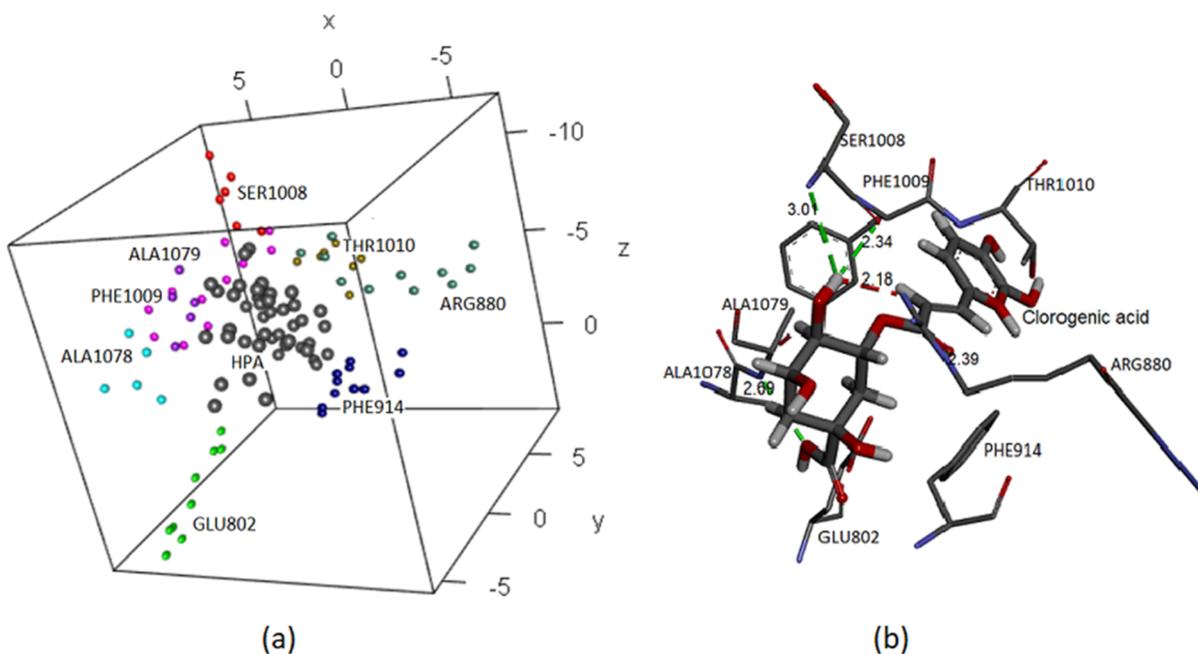


Figure 5. (a) Representation of active sites of the XO–chlorogenic acid complex found using the new method in a prohibited cavity. (b) Interactions of our method obtained by using a visualization software and without any assistance of an expert docking software.

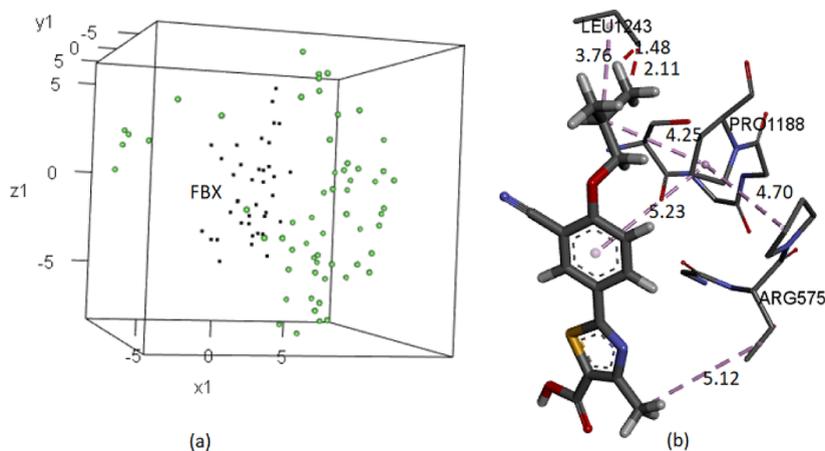


Figure 6. (a) XO–FBX complex corresponding to the best local optimum found for a new cavity. (b) Representation of the interactions between the ligand and pocket residues. (b) shows the interactions of our method obtained by using a visualization software and without any assistance of an expert docking software.

an optimal energy of -13.52 kcal/mol. Figure 3 shows this system.

The second part of the algorithm involves steps 14 to 16. In this case, we consider step 14. Finally, eq 9 reaches an energy of -14.59 kcal/mol with the angles 177.31 , 176.29 , and 184.72° . Similar neighborhoods cannot improve the addressed energy. Figure 4a depicts the amino acids surrounding the ligand, and Figure 4b shows the interactions between HPA and protein residues. The distance between interacting atoms is measured in angstroms. HPA exhibits hydrogen bonding with ARG880 (3.28 and 2.94 Å) and GLU802 (2.69 Å) and hydrophobic interactions with PHE914 (3.14 and 3.39 Å), PHE1009 (4.78 and 5.20 Å), ALA1078 (4.57 Å), and ALA1079 (4.02 and 4.84 Å). The results are in the same line as those consulted in refs 47, 54, and 58.

Investigations that include molecular docking of XO with other small molecules^{59,60} have shown interactions of the ligands with the amino acids that we report here.

Application 2. A Forbidden Pocket: XO–Chlorogenic Acid. In this application, we check the performance of the theorem with a larger ligand that cannot be docked in a small cavity. We test a narrow cavity that does not achieve significant energy at the end of step 13, but the final step 16 reaches a satisfactory reduction. We set the same pocket with 65 atoms of Application 1 (Table S1) and consider a new ligand. In this case, we select the 43 atoms of chlorogenic acid (Table S3), a ligand reported as an antioxidant in ref 54. Steps 1–13 register a very high energy of $230,403,762,231$ kcal/mol. However, the energy drops substantially to 1581.62 kcal/mol when we perform the geometry optimization in a small neighborhood (steps 14–16). The interaction is still repulsive but accommodates the ligand in a forbidden cavity as well as

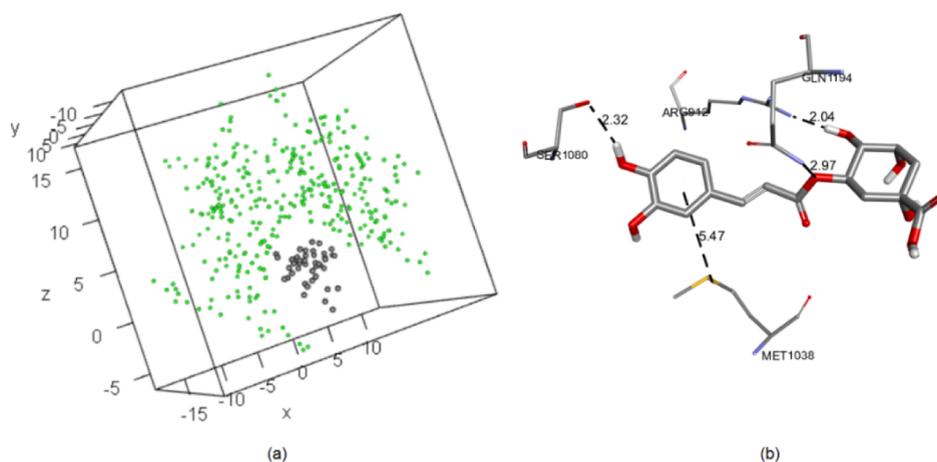


Figure 7. (a) XO–chlorogenic acid complex corresponding to the best local optimum found for a new cavity. (b) Representation of the complex considering with the amino acids. (b) shows the interactions of our method obtained by using a visualization software and without any assistance of an expert docking software.

possible; in this case, the minimum value is obtained with the corresponding rotations 196.01, 248.12, and 257.18°. Figure 5 shows the complex, although it does not present favorable interactions energetically because the ligand cannot enter in such a small and narrow cavity.

Figure 5a shows the ligand surrounded by amino acids, and in Figure 5b, we can observe the interaction between the ligand and the pocket residues in more detail. Thus, hydrogen bonds ARG880 (2.39 Å), ALA1079 (2.69 Å), and SER1008 (3.01 Å, 2.34 Å) and unfavorable bonds ALA79 and ARG880 affected the value of the interaction energy between the ligand and the pocket.

Application 3. A New Cavity: XO–FBX. We consider the most difficult challenge for the method: to find a new possible cavity automatically by exploring a large protein. We tested the method on the XO (C chain) in a complex with FBX. This chain has 5832 atoms and has been motivated by the previous section, “Case Study.”

We search for a new cavity in XO for a possible docking with FBX, a ligand with 38 atoms. In this case, the resulting cavity reaches a high energy of 5484.73 kcal/mol in a concave pocket of 60 atoms. Nevertheless, step 14, through the subspace associated with y -axis rotation, finds the minimum of -9.64 kcal/mol with a rotation of 0, 138, and 0°. Then, we apply step 15, and a search around this local minimum results in the OMD with the angles -49.45 , 213.56, and 171.83° and a potential of -17.49 kcal/mol.

The coordinates of the ligand and the cavity are given in Tables S4 and S5, respectively. Figure 6 shows the corresponding OMD.

In Figure 6a, the ligand is enveloped by amino acid atoms. Figure 6b shows the amino acids interacting with the ligand (distances in Å).

Application 4. A New Cavity: XO–Chlorogenic Acid. We study other cavities in the C chain by a larger ligand. Steps 1–13 find a possible cavity for docking with chlorogenic acid. As before, step 13 gives a high energy of 59,735.92 kcal/mol. Optimization of step 14 with 360 x -axis rotations of 1° results in the energy of -7.07 kcal/mol. Taking this result as a seed of a specific application of step 15, we obtain the OMD with angles 170.03, 126.33, and 110.37° and an energy of -27.81 kcal/mol.

We remark that our OMD is not reported in the literature; it guides interesting future work for the associated biological activity.

The positions of the OMD are in given in Tables S6 and S7 (see Figure 7a). Figure 7b shows the interactions.

In Figure 7 a, the XO–chlorogenic acid complex is surrounded by specific amino acids.

The interactions present distances between 2.04 and 5.47 Å and hydrogen bonds such as SER1080 (2.32 Å), ARG912 (2.04 Å), and GLN1194 (2.97 Å). They also show van der Waals interactions at MET1038 (5.47 Å).

Application 5. A New Cavity: XO–HPA and Complementary Analysis Using AUTODOCK VINA. In this example, we provide a complementary analysis using AutoDock Vina. Our method gives the automatic docking with the L-J potential.

First, we consider the automatic search of a cavity in OX for HPA.

Steps 1–13 select a cavity with 206 atoms. In the following lines, we describe step 14 (see also Part 5) of the theorem. Figure 8a shows the energy versus 360 ligand rotations of 1° around the invariant axis given in step 13. The minimum energy is reached at 358°. A neighborhood exploration by subspaces of rotations is shown in Figure 8b–d.

The local minima in the three previous subspaces suggest the starting point for a more robust optimization. Finally, step 15 reaches the energy of -13.27 kcal/mol and occurs in the x , y , and z rotations of 4.48, 64.47, and 31.67°, respectively. Figure 9a,b represents the OMD with the best local optimum (see Tables S8 and S9).

This new cavity presents hydrophobic interactions with LEU1243 (3.70 Å), PRO1188 (5.23 Å), and ARG575 (5.12 Å) and other interactions with LEU1243 (2.11 Å and 1.48 Å).

Complementary Analysis Using the Expert Software in Molecular Docking. Once the OMD was obtained, a further analysis can be implemented using expert software in molecular docking.

To carry out the docking simulations, we use Autodock Vina.⁵⁶ We provide the original coordinates of the suggested site (step 1) and the isolated coordinates of the ligand taken from a database. Furthermore, we compare the results of AutoDock with the OMD of the theorem (step 16).

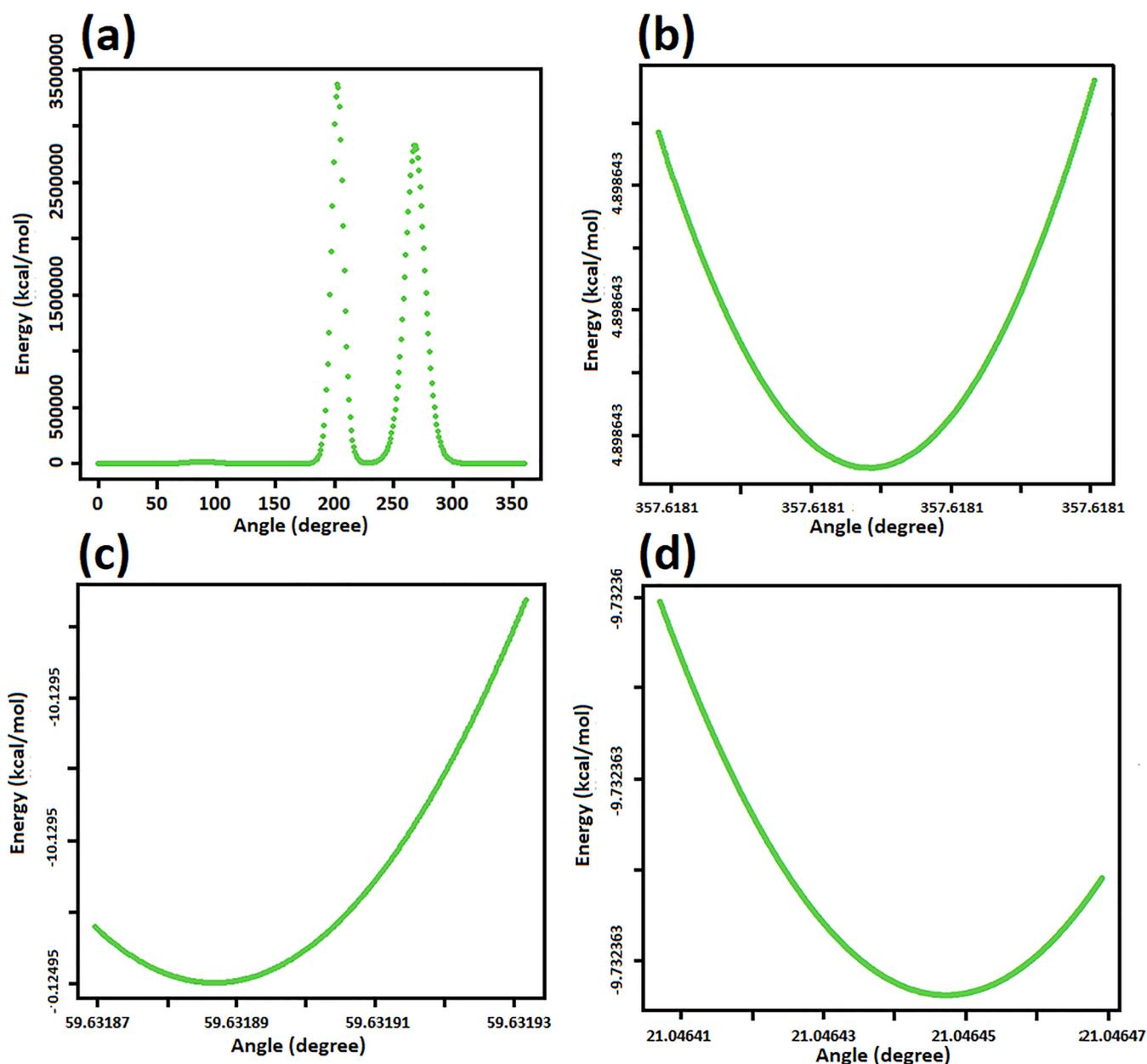


Figure 8. (a) Energy in kilocalories per mole vs 1° rotations for the ligand around the invariant axis of symmetry. In this figure, we show only the wide range of energy in terms of the angle exploration. A local minimum is explored in the remaining figures. (b) Path with a local minimum potential of around 357.61° with respect to the z -axis. The x -axis cannot provide a higher resolution because the differences require a large number of decimals. This figure shows the zoomed-in view of the local minimum not observed in Figure 8a. (c) Path with a local minimum potential of around 60° with respect to the y -axis. (d) Path with a local minimum potential of around 21° with respect to the x -axis.

In Table 2, we report the results for the XO–HPA complex. The original repository coordinates are indexed by the superscript 2, and the OMDs are indexed by 1.

The order of the results agrees with our expectations: (1) The best system is the OMD(pocket)–OMD(ligand) of our method, with an affinity of -4.4 kcal/mol. (2) The second system is the OMD(pocket)–repository(ligand), with an affinity of -4.3 kcal/mol. Both results explain that our OMD cavity performs better than the repository pocket. (3) In the third place, the repository(pocket)–OMD (ligand) appears with an affinity of -4.2 kcal/mol. (4) Finally, the worst system (-4.0 kcal/mol) is repository(pocket)–repository(ligand).

Figure 9 shows a number of amino acids not reported before; thus, a new possibility for applications is still to come.

5. CONCLUSIONS AND FUTURE WORK

In this work, we set the automatic search for cavities and docking in large protein in the mathematical framework of the shape theory.

The method is summarized in the formulation and demonstration of a theorem involving five main parts:

- (1) The exact formulas for the shape coordinates of the docking protein–ligand are provided; the coupled system is invariant under location and rotation.
- (2) An automatic search for cavities in a large protein is parameterized by a sequence of possible radii in a suitable van der Waals interval.

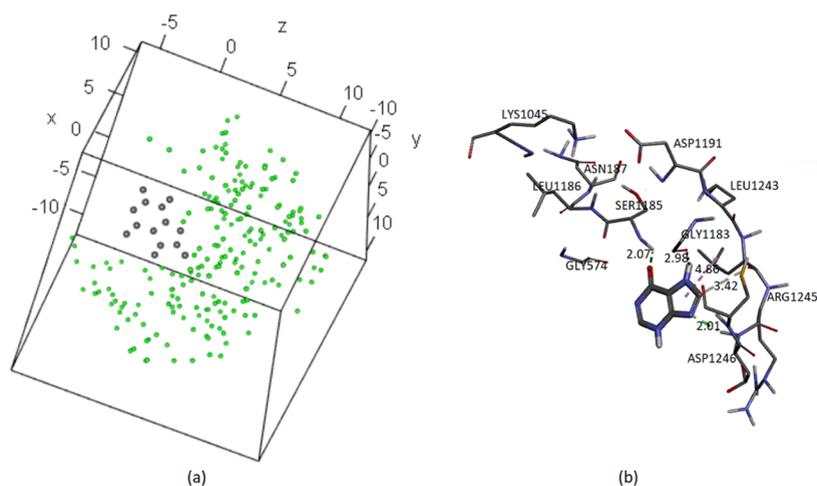


Figure 9. (a) XO–HPA complex in gray corresponding to the best local optimum found and representation of the amino acids that surround the HPA ligand. (b) shows the interactions and distances (Å) of our method obtained by using a visualization software and without any assistance of an expert docking software.

Table 2. Affinity of the XO–HPA Complex

| XO ¹ –HPA ¹ | affinity (kcal/mol) | | |
|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| | XO ¹ –HPA ² | XO ² –HPA ¹ | XO ² –HPA ² |
| –4.4 | –4.3 | –4.2 | –4.0 |
| –4.4 | –4.3 | –4.2 | –3.9 |
| –4.4 | –4.2 | –4.1 | –3.9 |
| –4.0 | –4.1 | –3.8 | –3.8 |
| –4.0 | –4.1 | –3.8 | –3.8 |
| –3.9 | –3.9 | –3.8 | –3.8 |
| –3.9 | –3.9 | –3.7 | –3.5 |
| –3.9 | –3.9 | –3.7 | –3.5 |
| –3.8 | –3.9 | –3.6 | –3.5 |

- (3) A four-dimensional space for the invariant coupled system is proposed in terms of the atom shape coordinates and the atomic mass. Then a non-scaled Riemannian distance between coupled systems can be included in the geometry optimization process via the L–J potential types 6–12 and 6–10.
- (4) A matrix representation of the algorithm in terms of the potential scoring function.
- (5) Geometry optimization routine for reaching the OMD via the L–J potential.

The theorem leads to the algorithm written in a matrix form, which can be easily programed on a personal computer. The method contains two sections: (A) Steps 1 to 13. They find a plausible cavity and the best docking via the L–J potential in the corresponding invariant coupled system.

The algorithm was applied in three different ligands (FBX, HPA, and chlorogenic acid), and three unpublished receptors were found.

The method is self-contained, and the expert analysis using AutoDock Vina is not necessary. However, we included the analysis for a comparison of our OMDs and the public database repositories. The interactions in the receptor can be shown using a visualization software, without requiring an additional expert optimization.

The algorithm can be applied in any virus chain with unknown active site. Then, docking with a plausible drug can be proposed by the method. The OMDs of different cavities can be used for comparisons of vaccines via the Riemannian

distance of eq 5. In a future work, we will apply the method in the SARS-CoV-2 virus and Monkeypox virus.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.2c02227>.

Methods and algorithms: proof of the theorem (parts 1 and 2); Application 1: a known cavity: XO–HPA, coordinates of the ligand; Application 1: a known cavity: XO–HPA, coordinates of the cavity; Application 2: a forbidden pocket: XO–chlorogenic acid, coordinates of the ligand; Application 3: a new cavity: XO–FBX, coordinates of the ligand; Application 3: a new cavity: XO–FBX, coordinates of the cavity; Application 4: a new cavity: XO–chlorogenic acid, coordinates of the ligand; Application 4: a new cavity: XO–chlorogenic acid, coordinates of the cavity; Application 5: a new cavity: XO–HPA, coordinates of the ligand; application 5: a new cavity: XO–HPA, coordinates of the cavity; and a complete example with the source code in R software (licensed in Colombia): user competences, general procedure of the example, and the required sources files in R for the example (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Iliana Ramírez-Velásquez – Faculty of Exact and Applied Sciences, Instituto Tecnológico Metropolitano ITM, Medellín 050034, Colombia; Doctorate in Modeling and Scientific Computing, Faculty of Basic Sciences, University of Medellín, Medellín 050026, Colombia; orcid.org/0000-0001-9779-4019; Email: ilianaramirez@itm.edu.co

Francisco J. Caro-Lopera – Faculty of Basic Sciences, University of Medellín, Medellín 050026, Colombia; orcid.org/0000-0002-4295-3120; Email: fjcaro@udemedellin.edu.co

Authors

Alvaro H. Bedoya-Calle – Faculty of Basic Sciences, University of Medellín, Medellín 050026, Colombia; orcid.org/0000-0001-7133-2553

Ederley Vélez — Faculty of Basic Sciences, University of Medellín, Medellín 050026, Colombia

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acsomega.2c02227>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by Ministerio de Ciencia, Tecnología e Innovación (MINCIENCIAS) in Colombia through grant no. 120680762977 (contrato: 795–2018). I.R. was supported by the Instituto Tecnológico Metropolitano (ITM), Colombia. The authors also thank Doctorate in Modeling and Scientific Computing at University of Medellín.

REFERENCES

- (1) Luo, H.; Li, M.; Yang, M.; Wu, F.-X.; Li, Y.; Wang, J. Biomedical Data and Computational Models for Drug Repositioning: A Comprehensive Review. *Briefings Bioinf.* **2021**, *22*, 1604–1619.
- (2) Ramirez-Velasquez, I. M.; Velez, E.; Bedoya-Calle, A.; Caro-Lopera, F. J. Mechanism of Antioxidant Activity of Betanin, Betanidin and Respective C15-Epimers via Shape Theory, Molecular Dynamics, Density Functional Theory and Infrared Spectroscopy. *Molecules* **2022**, *27*, 2003.
- (3) Villarreal-Rios, A. L.; Bedoya-Calle, A. H.; Caro-Lopera, F. J.; Ortiz-Méndez, U.; García-Méndez, M.; Pérez-Ramírez, F. O. Ultrathin Tunable Conducting Oxide Films for Near-IR Applications: An Introduction to Spectroscopy Shape Theory. *SN Appl. Sci.* **2019**, *1*, 1553.
- (4) Lange, J. H. M.; Coolen, H. K. A. C.; van der Neut, M. A. W.; Borst, A. J. M.; Stork, B.; Verveer, P. C.; Kruse, C. G. Design, Synthesis, Biological Properties, and Molecular Modeling Investigations of Novel Tacrine Derivatives with a Combination of Acetylcholinesterase Inhibition and Cannabinoid CB1 Receptor Antagonism. *J. Med. Chem.* **2010**, *53*, 1338–1346.
- (5) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput.-Aided Drug Des.* **2011**, *7*, 146–157.
- (6) Morris, G. M.; Lim-Wilby, M. Molecular Docking. In *Molecular Modeling of Proteins*; Kukol, A., Ed.; Methods Molecular Biology; Humana Press: Totowa, NJ, 2008; pp 365–382.
- (7) Diago, L. A.; Moreno, E. Evaluation of Geometric Complementarity between Molecular Surfaces Using Compactly Supported Radial Basis Functions. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2009**, *6*, 689–694.
- (8) Mylonas, S. K.; Axenopoulos, A.; Daras, P. DeepSurf: A Surface-Based Deep Learning Approach for the Prediction of Ligand Binding Sites on Proteins. *Bioinformatics* **2021**, *37*, 1681–1690.
- (9) Gupta, A.; Mukherjee, A. Capturing Surface Complementarity in Proteins Using Unsupervised Learning and Robust Curvature Measure. *Proteins* **2022**, *90*, 1669.
- (10) Harris, R.; Olson, A. J.; Goodsell, D. S. Automated Prediction of Ligand-Binding Sites in Proteins. *Proteins* **2008**, *70*, 1506–1517.
- (11) Di Rienzo, L.; Milanetti, E.; Alba, J.; D'Abramo, M. Quantitative Characterization of Binding Pockets and Binding Complementarity by Means of Zernike Descriptors. *J. Chem. Inf. Model.* **2020**, *60*, 1390–1398.
- (12) Morris, R. J.; Najmanovich, R. J.; Kahraman, A.; Thornton, J. M. Real Spherical Harmonic Expansion Coefficients as 3D Shape Descriptors for Protein Binding Pocket and Ligand Comparisons. *Bioinformatics* **2005**, *21*, 2347–2355.
- (13) Laurie, A. T. R.; Jackson, R. M. Q-SiteFinder: An Energy-Based Method for the Prediction of Protein-Ligand Binding Sites. *Bioinformatics* **2005**, *21*, 1908–1916.
- (14) Weisel, M.; Proschak, E.; Schneider, G. PocketPicker: Analysis of Ligand Binding-Sites with Shape Descriptors. *Chem. Cent. J.* **2007**, *1*, 7.
- (15) An, J.; Totrov, M.; Abagyan, R. Pocketome via Comprehensive Identification and Classification of Ligand Binding Envelopes. *Mol. Cell. Proteomics* **2005**, *4*, 752–761.
- (16) Ivetac, A.; McCammon, J. A. A Molecular Dynamics Ensemble-Based Approach for the Mapping of Druggable Binding Sites. *Methods Mol. Biol.* **2012**, *819*, 3–12.
- (17) Tan, Y. S.; Reeks, J.; Brown, C. J.; Thean, D.; Ferrer Gago, F. J.; Yuen, T. Y.; Goh, E. T. L.; Lee, X. E. C.; Jennings, C. E.; Joseph, T. L.; Lakshminarayanan, R.; Lane, D. P.; Noble, M. E. M.; Verma, C. S. Benzene Probes in Molecular Dynamics Simulations Reveal Novel Binding Sites for Ligand Design. *J. Phys. Chem. Lett.* **2016**, *7*, 3452–3457.
- (18) Glaser, F.; Morris, R.; Najmanovich, R.; Laskowski, R.; Thornton, J. A Method for Localizing Ligand Binding Pockets in Protein Structures. *Proteins* **2006**, *62*, 479.
- (19) Guo, F.; Wang, L. Computing the Protein Binding Sites. *BMC Bioinf.* **2012**, *13*, S2.
- (20) Capra, J. A.; Laskowski, R. A.; Thornton, J. M.; Singh, M.; Funkhouser, T. A. Predicting Protein Ligand Binding Sites by Combining Evolutionary Sequence Conservation and 3D Structure. *PLoS Comput. Biol.* **2009**, *5*, No. e1000585.
- (21) Ballester, P. J.; Westwood, I.; Laurieri, N.; Sim, E.; Richards, W. G. Prospective Virtual Screening with Ultrafast Shape Recognition: The Identification of Novel Inhibitors of Arylamine N-Acetyltransferases. *J. R. Soc., Interface* **2010**, *7*, 335–342.
- (22) Berenger, F.; Voet, A.; Lee, X. Y.; Zhang, K. Y. A Rotation-Translation Invariant Molecular Descriptor of Partial Charges and Its Use in Ligand-Based Virtual Screening. *J. Cheminf.* **2014**, *6*, 23.
- (23) Kumar, A.; Zhang, K. Y. J. Prospective Evaluation of Shape Similarity Based Pose Prediction Method in D3R Grand Challenge 2015. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 685–693.
- (24) Liu, H.; Lin, F.; Qian, K.; Seah, H. S.; Lee, Y. T. Visual Analysis with Dynamic Geometric Complementarity and Physicochemical Matching in Protein Docking. In *2014 18th International Conference on Information Visualisation*, 2014; pp 315–320.
- (25) Kumar, A.; Zhang, K. Y. J. Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery. *Front. Chem.* **2018**, *6*, 315.
- (26) Di Rienzo, L.; Milanetti, E.; Alba, J.; D'Abramo, M. Quantitative Characterization of Binding Pockets and Binding Complementarity by Means of Zernike Descriptors. *J. Chem. Inf. Model.* **2020**, *60*, 1390–1398.
- (27) Das, D.; Maeda, K.; Hayashi, Y.; Gavande, N.; Desai, D. V.; Chang, S. B.; Ghosh, A. K.; Mitsuya, H. Insights into the Mechanism of Inhibition of CXCR4: Identification of Piperidinyethanamine Analogs as Anti-HIV-1 Inhibitors. *Antimicrob. Agents Chemother.* **2015**, *59*, 1895–1904.
- (28) Salo, H. S.; Laitinen, T.; Poso, A.; Jarho, E.; Lahtela-Kakkonen, M. Identification of Novel SIRT3 Inhibitor Scaffolds by Virtual Screening. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 2990–2995.
- (29) Kendall, D. G. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bull. Lond. Math. Soc.* **1984**, *16*, 81–121.
- (30) Goodall, C.; Mardia, K. Multivariate Aspects of Shape Theory. *Ann. Stat.* **1993**, *21*, 848.
- (31) Berger, M. Riemannian Manifolds as Metric Spaces and the Geometric Meaning of Sectional and Ricci Curvature. In *A Panoramic View of Riemannian Geometry*; Berger, M., Ed.; Springer: Berlin, Heidelberg, 2003; pp 221–297.
- (32) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

- (33) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230–252.
- (34) Goldstein, M.; Wooff, D. *Bayes Linear Statistics*; Wiley Series in Probability and Statistics; John Wiley & Sons, Ltd, 2007; pp 509–516.
- (35) Quintero, J. H.; Mariño, A.; Siller, L.; Restrepo-Parra, E.; Caro-Lopera, F. J. Rocking Curves of Gold Nitride Species Prepared by Arc Pulsed - Physical Assisted Plasma Vapor Deposition. *Surf. Coat. Technol.* **2017**, *309*, 249–257.
- (36) Valencia, G. M.; Anaya, J. A.; Velásquez, É. A.; Ramo, R.; Caro-Lopera, F. J. About Validation-Comparison of Burned Area Products. *Remote Sens.* **2020**, *12*, 3972.
- (37) Arias, E.; Jose Caro-Lopera, F. J.; Flórez, E.; Fredy Pérez-Torres, J. Two Novel Approaches Based on the Thompson Theory and Shape Analysis for Determination of Equilibrium Structures of Nanoclusters: Cu₈, Ag₈ and Ag₁₈ as Study Cases. **2019**, *1247*, 012008. DOI: 10.1088/1742-6596/1247/1/012008.
- (38) Klingenberg, C. P. Walking on Kendall's Shape Space: Understanding Shape Spaces and Their Coordinate Systems. *Evol. Biol.* **2020**, *47*, 334–352.
- (39) Rudemo, M. Statistical Shape Analysis. I. L. Dryden and K. V. Mardia, Wiley, Chichester 1998. No. of Pages: xvii+347. Price: £60.00. ISBN 0-471-95816-6. *Stat. Med.* **2000**, *19*, 2716–2717.
- (40) Caro-Lopera, F.; Díaz-García, J. A.; González-Farías, G. Noncentral Elliptical Configuration Density. *J. Multivariate Anal.* **2010**, *101*, 32.
- (41) Díaz-García, J.; Caro-Lopera, F. Estimation of Mean Form and Mean Form Difference under Elliptical Laws. *Electron. J. Stat.* **2017**, *11*, 2424.
- (42) Díaz-García, J. A.; Caro-Lopera, F. J. Generalised Shape Theory Via Pseudo-Wishart Distribution. *Sankhya, Ser. A* **2013**, *75*, 253–276.
- (43) Díaz-García, J. A.; Caro-Lopera, F. J. Statistical Theory of Shape under Elliptical Models and Singular Value Decompositions. *J. Multivariate Anal.* **2012**, *103*, 77–92.
- (44) Tsujikawa, H.; Sato, K.; Wei, C.; Saad, G.; Sumikoshi, K.; Nakamura, S.; Terada, T.; Shimizu, K. Development of a Protein-Ligand-Binding Site Prediction Method Based on Interaction Energy and Sequence Conservation. *J. Struct. Funct. Genomics* **2016**, *17*, 39–49.
- (45) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- (46) Enroth, C.; Eger, B. T.; Okamoto, K.; Nishino, T.; Nishino, T.; Pai, E. F. Crystal Structures of Bovine Milk Xanthine Dehydrogenase and Xanthine Oxidase: Structure-Based Mechanism of Conversion. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10723–10728.
- (47) Cao, H.; Pauff, J. M.; Hille, R. Substrate Orientation and Catalytic Specificity in the Action of Xanthine Oxidase. *J. Biol. Chem.* **2010**, *285*, 28044–28053.
- (48) Schmidt, H. M.; Kelley, E. E.; Straub, A. C. The Impact of Xanthine Oxidase (XO) on Hemolytic Diseases. *Redox Biol.* **2019**, *21*, 101072.
- (49) Sekizuka, H. Uric Acid, Xanthine Oxidase, and Vascular Damage: Potential of Xanthine Oxidoreductase Inhibitors to Prevent Cardiovascular Diseases. *Hypertens. Res.* **2022**, *45*, 772.
- (50) Fernandez, M. L.; Upton, Z.; Shooter, G. K. Uric Acid and Xanthine Oxidoreductase in Wound Healing. *Curr. Rheumatol. Rep.* **2013**, *16*, 396.
- (51) Roy, J.; Galano, J.-M.; Durand, T.; Le Guennec, J.-Y.; Chung-Yung Lee, J. C.-Y. Physiological Role of Reactive Oxygen Species as Promoters of Natural Defenses. *FASEB J.* **2017**, *31*, 3729–3745.
- (52) Malik, U. Z.; Hundley, N. J.; Romero, G.; Radi, R.; Freeman, B. A.; Tarpey, M. M.; Kelley, E. E. Febuxostat Inhibition of Endothelial-Bound XO: Implications for Targeting Vascular ROS Production. *Free Radicals Biol. Med.* **2011**, *51*, 179–184.
- (53) Okamoto, K.; Eger, B. T.; Nishino, T.; Kondo, S.; Pai, E. F.; Nishino, T. An Extremely Potent Inhibitor of Xanthine Oxidoreductase. Crystal Structure of the Enzyme-Inhibitor Complex and Mechanism of Inhibition. *J. Biol. Chem.* **2003**, *278*, 1848–1855.
- (54) Yusuff, O. K.; Abdul Raheem, M. A. O.; Mukadam, A. A.; Sulaimon, R. O. Kinetics and Mechanism of the Antioxidant Activities of C. Olitorius and V. Amygdalina by Spectrophotometric and DFT Methods. *ACS Omega* **2019**, *4*, 13671–13680.
- (55) R Team. *Core. R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.
- (56) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, *31*, 455.
- (57) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (58) Teles Fujishima, M. A.; Silva, N. D. S. R. d.; Ramos, R. D. S.; Batista Ferreira, E. F.; Santos, K. L. B. d.; Silva, C. H. T. d. P. d.; Silva, J. O. d.; Campos Rosa, J. M.; Santos, C. B. R. d. An Antioxidant Potential, Quantum-Chemical and Molecular Docking Study of the Major Chemical Constituents Present in the Leaves of *Curatella Americana* Linn. *Pharmaceuticals* **2018**, *11*, 72.
- (59) Lin, H.-C.; Tsai, S.-H.; Chen, C.-S.; Chang, Y.-C.; Lee, C.-M.; Lai, Z.-Y.; Lin, C.-M. Structure-Activity Relationship of Coumarin Derivatives on Xanthine Oxidase-Inhibiting and Free Radical-Scavenging Activities. *Biochem. Pharmacol.* **2008**, *75*, 1416–1425.
- (60) Ardjani, A. T. E.; Mekelleche, S. Analysis of the Antioxidant Activity of 4-(5-Chloro-2-Hydroxyphenylamino)-4-Oxobut-2-Enoic Acid Derivatives Using Quantum-Chemistry Descriptors and Molecular Docking. *J. Mol. Model.* **2016**, *22*, 302.