2020-04-18

A computer-based approach for developing linamarase inhibitory agents

Paul, Lucas

Walter de Gruyter GmbH

https://doi.org/10.1515/psr-2019-0098

Lucas Paul[1,6] / Celestin N. Mudogo[2,7] / Kelvin M. Mtei[3] / Revocatus L. Machunda[3] / Fidele Ntie-Kang[4,5,8]

A computer-based approach for developing linamarase inhibitory agents

[1] The Department of Materials and Energy Science & Engineering, The Nelson Mandela African Institution of Science and Technology, P.O. Box 447 Arusha, Tanzania, E-mail: lucaspaul33@gmail.com, lucasp@nm-aist.ac.tz
[2] Biochemistry and Molecularbiology, University of Hamburg Institute of Biochemistry and Molecularbiology, Hamburg, Germany, E-mail: cmudogo@gmail.com
[3] The Department of Water and Environmental Science and Engineering, The Nelson Mandela African Institution of Science and Technology, P.O. Box 447 Arusha, Tanzania, E-mail: kelvin.mtei@nm-aist.ac.tz, revocatus.machunda@nm-aist.ac.tz
[4] Department of Pharmaceutical Chemistry, Martin-Luther University Halle-Wittenberg, Wolfgang-Langenbeck Str. 4, Halle (Saale) 06120, Germany, E-mail: ntiekfidele@gmail.com
[5] Department of Informatics and Chemistry, University of Chemistry and Technology Prague, Technická 5, Prague 6, Dejvice 166 28, Czech Republic, E-mail: ntiekfidele@gmail.com
[6] Department of Chemistry, Dar es Salaam University College of Education, P.O. Box 2329, 255 Dar es Salaam, Tanzania, E-mail: lucaspaul33@gmail.com, lucasp@nm-aist.ac.tz
[7] Department of Basic Sciences, School of Medicine, University of Kinshasa, Kinshasa, Congo (Democratic Republic of the), E-mail: cmudogo@gmail.com
[8] Department of Chemistry, University of Buea, P. O. Box 63 Buea, Cameroon, E-mail: ntiekfidele@gmail.com

Abstract:
Cassava is a strategic crop, especially for developing countries. However, the presence of cyanogenic compounds in cassava products limits the proper nutrients utilization. Due to the poor availability of structure discovery and elucidation in the Protein Data Bank is limiting the full understanding of the enzyme, how to inhibit it and applications in different fields. There is a need to solve the three-dimensional structure (3-D) of linamarase from cassava. The structural elucidation will allow the development of a competitive inhibitor and various industrial applications of the enzyme. The goal of this review is to summarize and present the available 3-D modeling structure of linamarase enzyme using different computational strategies. This approach could help in determining the structure of linamarase and later guide the structure elucidation *in silico* and experimentally.

Keywords: linamarase, cassava, structural determination, computational strategies
DOI: 10.1515/psr-2019-0098

1 Introduction

Cassava (*Manihot esculenta* Crantz) (Euphorbiaceae), also known as mandioca, yucca, tapioca or manioc. It is the leading supplier of energy ranked after rice and corn [1]. It is the most grown crop in the tropics and subtropics regions. The tuber is the primary source of carbohydrate while leaves provide protein as a vegetable. About 105 countries grow cassava as a strategical crop against famine since it can sustain and produces in drought and poor soil, can stay within the farm and be harvested at the time of demand [2]. The roots are the primary source of carbohydrate while leaves provide vitamins, minerals, and protein, as well as a vegetable that is available throughout the year. The leaves have high crude protein, and the amino acids, which are well balanced and its amount is beyond the minimal amount recommended by the Food and Agriculture Organization [2]. Leaves also have various minerals like iron, zinc, manganese, magnesium and calcium, vitamin B1, B2, C and carotenoids [2, 3], The combination of cassava roots and leaves can provide a meal with almost all essential dietary needs.

All body parts of the cassava except seeds contain cyanogenic glucosides compounds known as linamarin and lotaustralin. Linamarase hydrolyses these compounds to hydrogen cyanide as the main product [4]. There are about 5,000 varieties of cassava; all of them are known to contain cyanogenic glucosides which range from 10 to 500 mg HCN/kg. Based on the amount of hydrogen cyanide (HCN) released, cassavas are classified into three groups: group one those with greater than 100 mg HCN/kg are called very bitter and very toxic, group

Lucas Paul is the corresponding author.

two those with between 50 to 100 mg HCN/kg are regarded as moderate bitter and moderate toxic while those with less than 50 mg HCN/kg are sweet cassava [5].

Cyanogenic glucosides are mainly composed of Linamarin (95%) and Lotaustralin (5%) [6], enzymatic hydrolysis of cyanogenic glucosides by linamarase is initiated by any physical damage of cassava tissue this allows interaction between enzyme from cell wall which is physically separated from substrate found in cell vacuole so are not compartmentalized [7]. Interaction between enzymes and substrate (mainly Linamarin) start by the release of glucose and acetone cyanohydrin at pH > 4 and temperature >35 °C is converted to hydrogen cyanide [8], Figure 1. The presence of hydroxynitrile lyase (HNL) helps to complete the reaction of acetone cyanohydrin to cyanide. This enzyme is highly available in cassava leaves and very little in the roots [9]. The study by [10] has reported that processed cassava flours contain high levels of acetone cyanohydrin but little linamarin or HCN, this is due to the little amount of HNL in the roots, which brings about this accumulation.

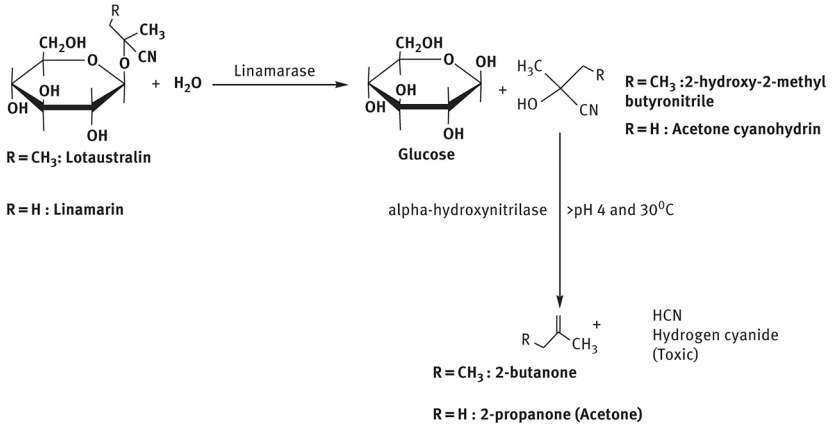

Figure 1: A complete enzymatic reaction between linamarase and linamarin.

Linamarin is a chemically stable compound, soluble in water and resists boiling in acid. Acetone cyanohydrin is also soluble in water and has a boiling point of 82 °C. But, HCN is a volatile compound, which evaporates only at 27 °C so volatilize at ambient temperature. So for effective processing techniques to be useful, we should reduce cyanogens to a safe level. The methods should maximize cassava tissue rupture to ensure effective enzyme-substrate interaction, to release acetone cyanohydrin and finally, volatile cyanide [11, 12].

The potential concentration of HCN determines the toxicity of the cyanogenic product consumed. If inadequately processed food is ingested, the HCN concentration is expected to be high within the body. For the toxicity of cyanogens depends on the following factors:

a. Unsufficient processing of plant which causes linamarin or HCN to remain in the food.

b. When raw cassava is consumed or insufficiently processed cassava product.

HCN, when released continuously until when low pH value from the stomach, deactivates the enzyme (linamarase). Cyanide ingested into the body always follows the metabolic pathway of detoxification, whereby rhodanese works by converting it to thiocyanide which later excreted in the urine [13]. When HCN ingested in the body gets absorbed quickly into the blood and combines with all forms of iron (methemoglobin and hemoglobin) which are present in erythrocytes [14]. The body eliminates the toxic cyanide by using the enzyme rhodanese, which contains an active disulfide group. It works by reacting with thiosulphate and cyanide, which converts cyanide to excretable thiocyanate, for this process to be complete sulfur donors that usually is provided by dietary sulfur amino acids are highly required [15, 16]. HCN binds to the Fe^{3+}/Fe^{2+} present in cytochrome and inactivates its activity.

This HCN inhibits the oxygen uptake and then causes glucose and lactic acid accumulation and deficiency of Adenosine triphosphate (ATP)/adenosine diphosphate (ADP), which brings the body to anaerobic instead of aerobic respiration [17]. HCN in the body inhibits many enzymatic reactions; if they contain iron, copper or molybdenum and its effect is highly and immediately appreciated in the respiratory system and heart. The amount recommended for cassava products, should not exceed 10 mg HCN/kg [4]. Any consumption of cassava products beyond the recommended amount can cause the following health problems vomiting, nausea, diarrhoea, dizziness, headache, stomach pains and sometimes death.

Linamarase is among the β-glucosidase belonging to the GH1 family which can convert glucosyl group from a glycoside (nonreducing) or carbohydrate by hydrolysis resulting in water or by transglycosylation gives alcohol. It has the $(\beta/\alpha)_8$ barrel structure, the properties of acid-base catalysis and the nucleophilic are contributed by the two carboxylic acid residues at β-strands 4 and 7 [18]. One most crucial property of linamarase which

differentiate it from other GH1 family is the ability to effectively catalyze the transglycosylation using primary, secondary and tertiary alcohol as acceptors [19]. However, cannot synthesize oligosaccharides and glycosides by reverse hydrolysis [20].

The detailed crystal structure of enzyme linamarase is still lacking [21]. The only effort has been done is to obtain function active-site amino acid residues, which has been modeled using homology modeling by the MODELLER9v4 program (Figure 2) [22]. In the study by [21], they modeled the residues which are likely to be involved in the activity of dalcochinase in an effort to identify the amino acids which bring about enzyme specificity as compared to linamarase which have 47% sequence similarities.

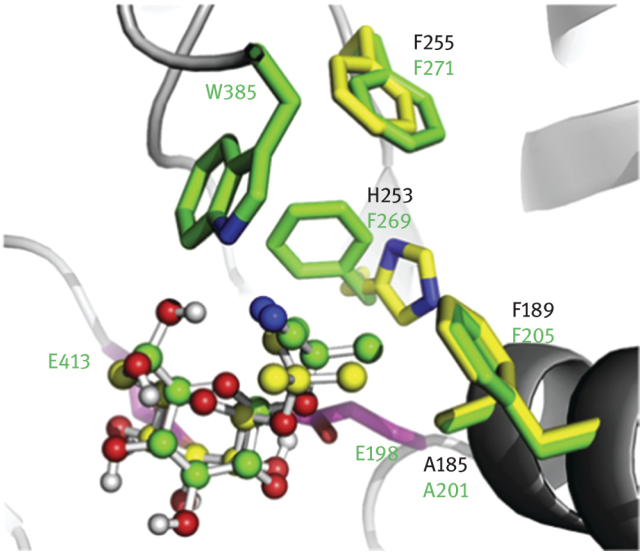

Figure 2: The generated three-dimensional models by MODELLER 9v4 of dalcochinase and Wild-type linamarase mutant's (1185A/N189F/V255F) (adapted from reference [22]).

There is the inclusion of linamarin in the 1185A/N189F/V255F mutant of dalcochinase as well as linamarase models. The catalytic acid/base (E198) and nucleophile (E413) of linamarase are shown in pink its docked linamarin is in yellow. The residues of linamarase and its docked linamarin are shown in green adopted from [22].

This review describes the computational approach toward the development of linamarase inhibitor, which is more competitive than natural substrate linamarin. For the development of linamarase inhibitor, we need first to develop *in silico* model structure of the enzyme (linamarase) by homology modeling. Then steps to discover the best and stable inhibitor, this will involve virtual screening (VS), molecular docking and finally, molecular dynamics. The details of these steps are analyzed below.

2 Determination of the three-dimension linamarase structure prediction

Generally, there are several computational methodologies and algorithms which are currently used to solve the problems of three-dimensional (3-D) structure of the protein which have not yet experimentally determined. The only available information is the sequence of amino acid, these provide essential information that relates to the 3-D macromolecules structures which are obtained by the experimental method like Protein Crystallography (X-ray diffraction), electron microscopy or nuclear magnetic resonance (NMR) [23]. There are four main methods that can be used as an approach of obtaining the linamarase 3-D structure these include the following.

2.1 Method without database information

The method uses the *ab initio* method which uses the concept from thermodynamics assuming all native protein structure always corresponds to the global minimum free energy [24–26]. Here, it does not use the structural templates from a database like Protein Data Bank (PDB). It mainly considers the potential energy functions in integrating the parameters of all atoms. The general goal is to obtain a global minimum free energy that corresponds to the native protein [26–28]. Using this approach, we can predict the new folds, since it is not limited to template from PDB. This principle uses the following simulation package; AMBER (Assisted Model Building

with Energy Refinement) [29, 30], CHARMM (Chemistry at HARvard Molecular Mechanics) [31], UNRES [32], GROMACS (Groningen Machine for Chemical Simulation) [33], TINKER (Software tools for Molecular Design) [34].

2.2 Methods with database information

The starting point here is the 3-D protein structure is obtained from the database, and it compares the fragments of the target sequences to that of known protein. The short amino acid sub-sequence of the target structure against the known protein's structure fragments [35]. The newly discovered protein structure will be composed of a similar structure of motifs like the known protein. Therefore, this method is based on fragments of amino acid sequence with a different motif which, when combined, they form the 3-D protein structure [26]. The homology fragments are used for finding the structures which are achieved through scoring functions and algorithm optimization to get the structure with the lowest potential energy [36]. The fragment-based approach always needs to look for a criterion that exists between the fragments so that the final fragment will have a high chance of being inserted at the final structure predicted as summarized in illustration (Figure 3) [37]. This method is similar to *ab initio* when it finds polypeptide structures with the lowest energy, but the main difference is that it uses the database to predict the structure of polypeptide [38].

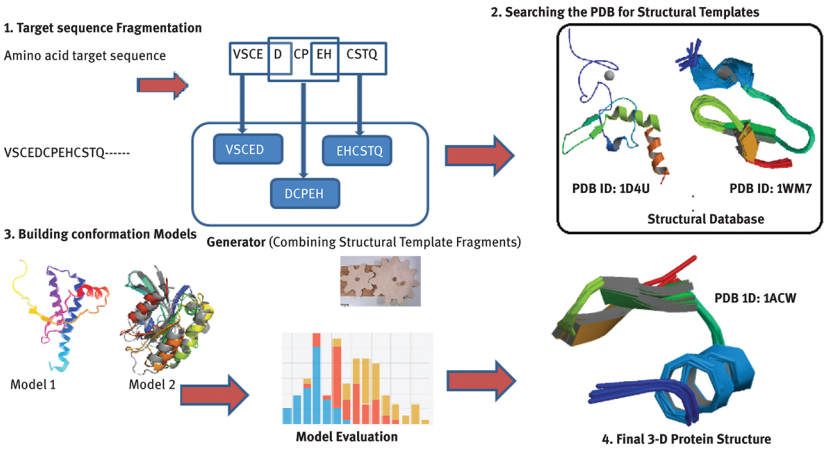

Figure 3: Schematic representation of the method based on fragments (adapted from reference [23]).

2.3 Fold recognition and threading methods

The method involves using the amino acid sequence and evaluates how well it fits the 3-D structure of the protein of the known. This approach is used because structures are more evolutionary preserved than the sequence [35, 39–41]. The sequential order is followed in placing the target amino acid sequences and is governed by two procedures; first searching the correct replacement between the target sequence versus the model which is in the space of possible sequence-structure alignment. The threads tried to find the templates with a similar fold that have or don't have direct evolutionary relations (analogue).

2.4 Comparative modeling method and sequence alignment strategies

This method use target protein's amino acid sequences to align against know protein's amino acid sequence (used as a template). The information of a known protein is experimentally determined and deposited in PDB [42]. If there are high similarities between the two amino acid sequences, then the structural information of the known (template) can be used to modal the target protein of interest [43, 44]. The homology protein with full information obtained experimentally are the ones to be used to model the target protein, and their amino acid residue is similar as they occupy the same position in the homology protein and have similar physic-chemical properties. Currently, comparative modeling is highly used because it is useful in protein structure prediction, which has more impact in the field of drug discovery [45]. The sequence alignment can either be pairwise which is the sequence–sequence comparison or multiple sequence comparison. The first approach uses the target sequence to compare with sequences in the database independently [46]. It uses methods like FASTA [47, 48], PSI-BLAST [49] and BLAST [50], while multiple sequence comparison allows multiple sequence alignments

whereby the sensitivity of the search is maximized [51–53]. The methods used here include CLUSTALW [54], PSI-BLAST [49] and T-COFFEE [55].

In the study by [21]identify the specific residues with similarities but have different catalytic properties. So eight amino acid residues in the glycine binding pocket of dalcochinase were replaced with respective residues of linamarase. Since the crystal structure of both enzymes is unavailable, then homology modeling of dalcochinase which has 47% amino acid sequence identity with linamarase, were performed by using ClustalW 2.0 with similar procedures as reported by [54]. The 3-D structure of wild-type dalcochinase was obtained using the template with a 45% identity to dalcochinase from Maize β-glucosidase 1 (Figure 4) with its substrate DIMBOA-β-D-glucoside with PDB code 1E56A as also reported by [56].

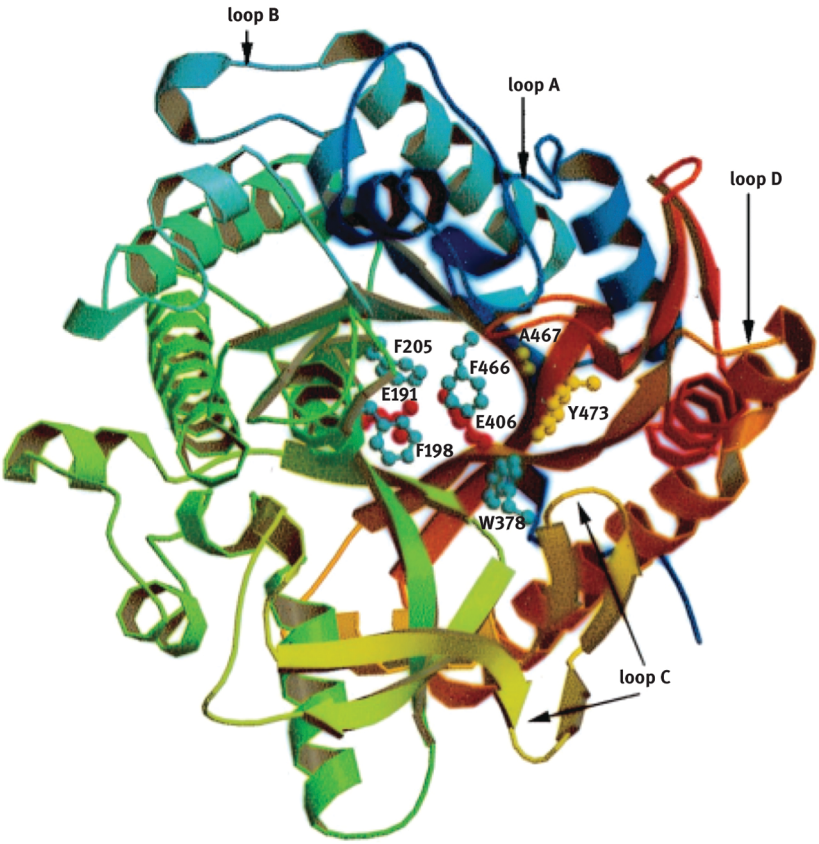

Figure 4: The three-dimensional structure of Maize β-glucosidase (from N terminus dark blue to C terminus dark red) (adapted from reference [56]).

The structure was built using the Sybyl 7.2 molecular modeling package, the overall structure of the model was checked by the PROCHECK, ProSA, Verify-3D and WHATIF programs [57–60]. Whereby PROCHECK showed 97% of the residues in the homology were located in the most favorable regions of the Ramachandran plot, whereby only 1.5% was in the rejected region when compared with the template with 98.5% and 0%. PROCHECK was also used to obtain G-factor, which was 0.22 above −0.5 which shows the model was reliable as compared to 0.34 of a template.

The ProSA Z-scores is explained well by its application to check the errors of the 3-D models. It can be used to determine errors of the experimentally and theoretically determined structures. It uses the coordinates then the structure's energy is evaluated by using a distance-based pair potential as well as the potential that relates to solvent exposure. The Z-scores validates the quality of the model and measures the total energy of the structure with regards to energy distribution derived from random conformations. The obtained Z-score which appears outside a range property of native proteins verifies erroneous structure [58]. In general positive value correspond to problematic or erroneous parts of a model. For example the model showedProSA Z-score of −8.15 while template −9.68 this is within the acceptable range [58]. The compatibility of the residues with the surrounding environment was done by verify-3D which scored above 0.2 (91.2%) shows reliable as compared with 93.9% for the template. WHATIF program managed to bring the confidence of the packing quality which scored above −5.0 and the picture of the 3-D model of the active site pocket of dalcochinase was generated by using PyMOL version 0.99.

Another study by [22] used the report of single mutation brought by replacing eight amino acid residues in the dalcochinase's binding pocket using residues from linamarase. The mutants namely 1185A, N189F, V255F

which have been identified to contribute to the hydrolytic and transglycosylation specificity of dalcochinase. The 1185A and V255F mutants have a low contribution to natural (Dal-Glc) substrate while all three mutants have more significant transglycosylation activities by using primary and secondary alcohol as acceptors, but none of these mutants demonstrated linamarase activities of transglycosylation; glucose to tertial alcohol and hydrolyzing linamarin. So in the standing study by [22], it brought an intention of further mutating the residues of dalcochinase in order to attain the linamarase specificities (transglycosylation and hydrolysis reaction). So the 3-D model of the previously reported mutants (1185A, N189F, V255F) of dalcochinase [61] and wild-type linamarase [18] were created using MODELLER9v4 [62], using the template of cyanogenic β-glucosidase from *Trifolium repens* L with (PDB code 1CBG), which have a similarity of 60% to dalcochinase and 51% to wild-type linamarase [56]. The model's quality was checked using PROCHECK, ProSA and Verify-3D programs [58, 59] where the active site was defined as 15 Å which is at the center of residues E182 and E396 of dalcochinase.

3 Identification and validation of linamarase inhibitors

The approach of identifying the inhibitor of a specific enzyme uses the same approach of drug discovery which targets to obtain a small molecule, known as the entity that can preferentially interact with the valid target. The target which has identified to cause or have a link with the disease or biological effect and need to be inhibited [63].

3.1 Virtual Screening for the identification of linamarase inhibitors

This is described as the step by step approach of searching novel compounds that referred to as hit and led with potential biological effects is achieved by filtering and narrowing down until the lead is obtained as an alternative to the natural ligand. Depending on the intended application the databases for VS consist of up to about 10 million compounds and they can be obtained from compound libraries that are provided from commercial venders, public and commercial databases. The application of VS depends mainly on the availability of the validated structural target (3-D structure). VS is categorized into either structural-based virtual screening (SBVS) or ligand-based Virtual Screening (LBVS).

3.1.1 Structure-based virtual screening

This approach is used in identifying the best ligand through searching to the chemical library for identifying its interaction with drug target, and it uses the 3-D structure of the protein which obtained either experimentally using X-ray crystallography, NMR or computational modeling [64]. Where the candidate is docked then ranked based on the binding affinity to the binding site [65].

3.1.2 Ligand-based virtual screening

The approach uses information obtained from the known ligand rather than structural protein for led identification as well as optimization, and it typically applied when there is no 3-D structure of the protein. It depends on the pharmacophores and relies on the knowledge of the ligand that will bind the active site of the biological target. The primary goal is to come up with the structure which retains the physicochemical properties. The approach is based on the principle that structurally similar molecules will always have similar properties [66].

3.2 Molecular docking

A method is an essential tool in the field of drug discovery and design. The main objective is to predict the best ligand's conformation in a target binding site/protein of known 3-D structure [67]. It concentrates much on either accuracy of the structure or correct prediction activity. The algorithms and scoring function allows the evaluation of the interaction of compounds and potential targets. It starts from simple, then advances to its complicated stage of the scoring function. It depends mainly on electrostatic and van der Waals of the interaction of solvation or entropic effects.

There are basic ways of representing the protein and ligands during the docking process these ensure evaluation of their methods used, which include; **atomic, surface, and grid** [68]**. Atomic representation** is mainly for evaluating pair-wise atomic interaction which brings the complexity so it uses potential energy function [69]. **Surface-based** is used to minimize the angles of the opposite or different interacting molecules [70, 71]. **Rigid body** access the energy contribution of receptor specifically on grid points which are used in ligand scoring and stores electrostatic and van der Waals (potentials) [72].

3.2.1 Searching methods for ligands

These are methods that allow molecular flexibility by focusing on the algorithms which treat ligands flexible but in a few cases, protein. These methods are divided into (a) systematic approach, (b) random or stochastic method and (c) shape matching.

3.2.1.1 Systematic searching

A method is used for flexible ligand docking whereby all number of possible conformations for ligand binding to the active site is measured, visiting the degrees of freedom of ligand. It can be considered in three approaches; Exhaustive, incremental/fragment, and assembling of conformation. The exhaustive is done by rotating all bonds of the ligand at a specific time allocated. They generate good conformation. To avoid the exhaustive search, the screening is done initially for different poses, filtered and then optimized. It uses Glide [73, 74] and FRED [75] for the sampling method. The fragment method is used to avoid the combinatorial explosion. So the algorithm uses the fragments which may be generated by three steps.

(i) Core fragment selection (ii) Core fragment ligand placement (iii) Incremental ligand placement.

By incrementally growing of the ligand into the binding site at a specific time, it is covalently linked, the programs used include Dock [76], LUDI [77], FlexX [77], ADAM [78] and eHiTs [79]. The conformational ensemble methods search for the pre-generated ligand confirmation with the libraries. The binding mode is compared by ranking them by considering the scoring energy. Programs used include FLOG [80], DOCK 3.5 [81], PhDock [82], MS-DOCK [83], MDock [84, 85] and Q-Dock [86].

3.2.1.2 Random search (stochastic algorithms)

The algorithms consider the conformation of ligands at the active site, whether individually or populated. It examines the translational/rotational space and conformational space of ligand. The approach includes;

The first is the Monte Carlo method (MC); the algorithms allows the generation of random conformation with translation and rotation at an initial stage of ligands docking at the active site. The new configuration is generated after scoring of the initial one and the probability of being accepted is achieved by considering the Boltzmann probability function,

$$P \sim exp\left[\frac{-(E_1 - E_0)}{k_B T}\right] \tag{1}$$

where E_0 and E_1 are the energy scores of the ligand before and after the random change.

Respectively, k_B is the Boltzmann constant, and T is the absolute temperature of the system. MC uses programs like DockVision [87], ICM [88], Prodock [89] and MCDOCK [90].

The second approach is genetic/evolutionary algorithms, and this uses the approach of biological competition and population dynamics. The varying parameters are included in the chromosome and randomly varied. The result produced is evaluated by its fitness. The chromosomes that produce optimal characteristics are crossed to produce the next generation. This uses GOLD [91, 92], AutoDock [93], DIVALI [94], DARWIN [95], MolDock [96], PSI-DOCK [97], FLIPDock [98], Lead finder [99] and EADock [100].

The third approach is Tabu search algorithms, and this considers the Tabu (those rejected conformation). It operates by making random changes on available conformations then each change is ranked. The Tabu is determined, and their changes that have the lower value are going to be accepted even if it was in tabu otherwise the non-tabu change is accepted. The program uses PRO_LEADS [101] and PSI-DOCK [97].

3.2.1.3 Shape matching

This method is used at the initial stage of docking, and it is the simplest approach. It places the ligand in the position that its molecular surface is in complementary with the surface of the binding site of the protein

involves translation and rotational which allow different orientations of ligands at the binding site. So, it mainly looks at which the ligands will be easily placed at the binding site as quickly as possible. Programs used include DOCK [102], FRED [103], FLOG [104], LigandFit [105], Surflex [106] MS-Dock [67] and MDock [107, 108].

3.2.2 Scoring functions

For determining the accuracy of the docking algorithm, it is important to consider the scoring function. So it is the fundamental element for the protein-ligand algorithm. It looks at the reliability and efficiency of any algorithm. The scoring functions can be grouped into three categories: Force field, empirical and knowledge-based scoring functions.

3.2.2.1 Force field scoring function

It uses the molecular mechanics' force field to consider the interaction between the ligand and receptor. Basically, the Van der Waals energies, the bond bending/stretching/torsional energies, are used at this strategy. The program which can be used here include AMBER [109] or CHARMM [110, 111]. The effect of water for the force field is accounted for by the inclusion of distance-dependent dielectric constant $\varepsilon(r_{ij})$, which uses a program like DOCK [112] and implements eq. (2) below.

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon\left(r_{ij}\right) r_{ij}} \right) \tag{2}$$

where r_{ij} stands for the distance between protein atom i and ligand atom j, A_{ij} and B_{ij} are the VDW parameters, and q_i and q_j are the atomic charges. $\varepsilon(r_{ij})$ is usually set to $4r_{ij}$, reflecting the screening effect of water on electrostatic interactions.

3.2.2.2 Empirical scoring function

This method is useful for reproducing the experimental data by summation of the binding energy of ligand to a receptor such as VDW, electrostatic, hydrogen bonding, desolvation, entropy and hydrophobicity represented by eq. (3) below:

$$\Delta G = \sum_i W_i \cdot \Delta G_i \tag{3}$$

where $\{\Delta G_i\}$ is for individual empirical energy and $\{W_i\}$ produced by the binding affinity after training the complex ligand-protein from the known 3-D structures. So, this method mainly depends on the information from the crystal structure of different protein-ligand complex whose binding affinities are known. It uses programs like Glide Score [74, 113], LUDI [114–117], SCORE [118–120] X-SCORE [121], ChemScore [122, 123], SFscore [124, 125].

3.2.2.3 Knowledge-based scoring function

These are scoring functions which aim at reproducing the experimental structures of protein-ligand complexes. It uses the potential of mean force which is given by inverse Boltzmann eq. (4).

$$w(r) = -k_B T In\left[\rho(r)/\rho^*(r)\right] \tag{4}$$

where k_B is the Boltzmann constant, T is the absolute temperature of the system, $\rho(r)$ is the number density of the protein-ligand atom pair at distance r in the training set, and $\rho^*(r)$ is the pair density in a reference state where the interatomic interactions are zero. Different from the field and empirical scoring function have a right balance between the accuracy and speed because the potentials from eq. (4) above are obtained from a larger number of structures rather than generating the known affinity, it uses a programs like PMF [126, 127], DrugScore [128], SMOG [129].

3.2.2.4 Consensus scoring function

The approach is not a specific type of scoring function; it combines and balances the scoring information by removing/minimizing errors of each scoring function. So it makes the true binder to be distinguished from the others [130, 131]. Programs used include X-Score and MultiScore [132, 133].

3.2.2.5 Clustering and entropy-based scoring methods

This method also used to improve the performance of different scoring functions. It includes the entropic effect by taking the ligand binding modes then divided into various clusters [134, 135] The impact of entropic contribution in each of the clusters is considered, and it uses AutoDock [135, 136].

A case study for docking was reported by [21], after homology modeling and identification of the structure of Thai rosewood dalcochinase which have high similarity to linamarase. The natural substrate (Dal-Glc) of dalcochinase were docked into the active site pocket of dalcochinase obtained using homology modeling, the docking program used was GOLD 3.1 [137]. Maize β-glucosidase was used by docking with DIMBOA-β-D-glucoside, which achieved a fitness score of 69.15, and when as compared with the fitness of the score for docking Dal-Glc into the dalcochinase it gave score 61.41. The interaction of the substrate positioned it by stacking π–π interaction between the phenyl ring of substrate and indole ring of the conserved Trp368. The distance between amino acid residues and Dal-Glc (substrate) specifically in the binding pocket of dalcochinase model was predicted by molecular docking as summarized in Table 1.

Table 1: The binding pocket of dalcochinase model showing the distances between Dal-Glc and amino acid residues obtained by molecular docking, (adapted from reference [21]).

Dal-Glc position	Amino acid position	Distance (Å)
Sugar ring		
O2	Y325-HH	2.49
O3	N181-HD22	2.89
O4	W453-He1	1.48
O6	W445-He1	2.16
H at O2	E396-Oe1	1.72
C		
Methylene	H235-He1	2.34
Methylene	W368-Hz3	2.51
Ring-A		
Ring-B		π–π (3.52–4.09)
Ring-C		
OH	W368-He1	2.35

In different studies [22] and [21], the authors considered the single mutation, whereby residues Ile185, Asn189, and Val255 in dalcochinase binding pocket were replaced by Ala201, Phe205, and Phe271 from linamarase consider Figure 2. This affected dalcochinase by increasing transglycosylation, mainly primary and secondary alcohol. However, this replacement of single mutation did not change dalcochinase activities to be similar to that of linamarase specifically transglycosylation to tertiary alcohol.

To bring dalcochinase specificity to linamarase multiple mutations of the identified key amino acid residues. The homology models of the 1185A/N189F/V255F which considered a triple mutation of the respective enzyme were obtained and docked to linamarin which its structure was driven from PubChem (CID 11128) using the AutoDock version 4.2 [138]. The analysis of the docked conformation was analyzed by Accelrys DsVisualizer 3.0 consider Figure 2.

The results of docking of these multiple mutations show that for both enzymes when were docked with linamarin the covalent bond glycosidic and the C1 atom are proposed to be at the same location.

4 Conclusion

The enzymatic reaction of linamarase is associated with application in biotechnology like transglycosylation and hydrolysis by using acceptors. The most important and which has more concern is in cassava which the hydrolysis of cyanogenic glucoside and ultimately leads to the release of toxic product hydrogen cyanide. To get the insight of linamarase's potential and how can be used to different productive reaction, there is a need to elucidate the structure experimentally. However, due to the improvement of computational approaches

in carrying out structural modeling, this work has reviewed and analyzed the approaches to be used predict the structure of linamarase. Homology modeling is mainly used then other techniques can be applied to identify the competitive inhibitor against the natural substrate and inhibits the enzymatic reaction. These are mainly molecular docking and molecular dynamics. In homology modeling of dalcochinase, it has proved that the enzyme shares 47% amino acid sequence with dalcochinase, which its structure is now available. So it can be used as a template to generate the 3-D structure of linamarase.

Regardless, of these similarities the enzymes they have the different catalytic ability as linamarase works by hydrolyzing the natural substrate linamarin and dalcochinase the substrate dalcochinin-8-*O*-β-*D*-glucoside, but not the reverse. Their distinct also is that dalcochinase can catalyze the transglycosylation of primary and secondary alcohol but not tertiary while linamarse can work effectively on both. Additionally, linamarase cannot synthesize oligosaccharides and glycosides by reverse hydrolysis while dalcochinase can do.

Therefore, regardless of the availability of amino acid sequences similar to linamarase, there is a great need to find out the crystal structure experimentally. This will unlock more scientific understanding, research, and application of linamarase especially in food processing like products with glycosides residues retained during the processing.

Acknowledgements

The authors are indebted to the Nelson Mandela African Institutional of Science and Technology (NM-AIST) through African Development Bank (AfDB) project for financial support. FNK acknowledges a return fellowship and equipment subsidy from the Alexander von Humboldt foundation, Germany.

References

[1] Andama M, Oloya B. Effectiveness of traditional processing techniques in reducing cyanide levels in selected cassava varieties in Zombo District, Uganda. Int J Food Sci Biotechnol. 2017;2:121–5.

[2] Montagnac JA, Davis CR, Tanumihardjo SA. Processing techniques to reduce toxicity and antinutrients of Cassava for use as a staple food. Compr Rev Food Sci Food Saf. 2009;8:17–27.

[3] Burns AE, Bradbury JH, Cavagnaro TR, Gleadow RM. Journal of food composition and analysis total cyanide content of cassava food products in Australia. J Food Compos Anal. 2012;25:79–82.

[4] Nicolau AI. Safety of fermented cassava products. In: Prakash V, Martín-Belloso O, Keener L, Astley S, Braun S, McMahon H, Lelieveld H, editors. Regulating safety of traditional and ethnic foods. London, Oxford, Boston, New York und San Diego: Academic Press, 2016:319–35.

[5] AttahDaniel BE, Ebisike K, Adeeyinwo CE, Ojumu TV, Olusunle SO, Adewoye OO. Towards arresting the harmful effect of cyanogenic potential of cassava to man in the environment. Int J Eng Sci. 2013;2:100–4.

[6] Nartey F. Manihot Esculemta (Cassava): cyanogenesis, ultrastructure and seed germination. Australia: Copenhagen, Munksgaard, 1978.

[7] Mcmahon JM, White WL, Sayre RT. Review article: cyanogenesis in cassava (manihot esculenta crantz). J Exp Bot. 1995;46:731–41.

[8] Mkpong OE, Yan H, Chism G, Sayre RT. Purification, characterization, and localization of linamarase in cassava. Plant Physiol. 1990;93:176–81.

[9] White WL, Arias-Garzon DI, McMahon JM, Sayre RT. Cyanogenesis in cassava: the role of hydroxynitrile lyase in root cyanide production. Plant Physiol. 1998;116:1219–25.

[10] Tylleskar T, Banea M, Bikangi N, Cooke RD, Poulter NH, Rosling H. Cassava cyanogens and konzo, an upper motoneuron disease found in Africa. Lancet. 1992;339:208–11.

[11] Cooke RD. An enzymatic assay for the total cyanide content of cassava (manihot esculenta crantz). J Sci Food Agric. 1978;29:345–52.

[12] Brien GM, Taylor AJ, Poulter NH. Improved enzymic assay for cyanogens in fresh and processed cassava. J Sci Food Agric. 1991;56:277–89.

[13] Frankenberg L. Enzyme therapy in cyanide poisoning: effect of rhodanese and sulfur compounds. Arch Toxicol. 1980;45:315–23.

[14] Food Standards Australia New Zealand, Cassava and bamboo shoots; a human health risk assessment. 2005.

[15] Cardoso AP, Ernesto M, Cliff J, Egan SV, Bradbury JH. Cyanogenic potential of cassava flour: field trial in Mozambique of a simple kit. Int J Food Sci Nutr. 1998;49:93–9.

[16] Rosling H. Measuring effects in humans of dietary cyanide exposure from cassava. International Society for Horticultural Science (ISHS), Nov. 1994.

[17] Solomonson LP. Cyanide as a metabolic inhibitor. Cyanide Biol. 1981;2013:11–28.

[18] Keresztessy Z, Kiss L, Hughes MA. Investigation of the active site of the cyanogenic β-D-glucosidase (linamarase) from Manihot esculenta Crantz (cassava). II. Identification of Glu-198 as an active site carboxylate group with acid catalytic function. Arch Biochem Biophys. 1994;315:323–30.

[19] Svasti J, Phongsak T, Sarnthima R. Transglucosylation of tertiary alcohols using cassava β-glucosidase. Biochem Biophys Res Commun. 2003;305:470–5.

[20] Srisomsap C, Subhasitanont P, Techasakul S, Surarit R, Svasti J. Synthesis of homo- and hetero-oligosaccharides by Thai rosewood β-glucosidase. Biotechnol Lett. 1999;21:947–51.

[21] Kongsaeree PT, Ratananikom K, Choengpanya K, Tongtubtima N, Sujiwattanarat P, Porncharoennop C, et al. Substrate specificity in hydrolysis and transglucosylation by family 1 β-glucosidases from cassava and Thai rosewood. J Mol Catal B Enzym. 2010;67:257–65.
[22] Tongtubtim N, Thenchartanan P, Ratananikom K, Choengpanya K, Svasti J, Kongsaeree PT. Multiple mutations in the aglycone binding pocket to convert the substrate specificity of dalcochinase to linamarase. Biochem Biophys Res Commun. 2018;504:647–53.
[23] Dorn M, Silva MB, Buriol LS, Lamb LC. Three-dimensional protein structure prediction: methods and computational strategies. Comput Biol Chem. 2014;53:251–76.
[24] Anfinsen CB, Haber E, Sela M, White FH. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci USA. 1961;47:1309–14.
[25] Anfinsen CB. Principles that govern protein folding Publication;. Science. 1973;181:223–30.
[26] Büssow KA. Protein structure prediction. Concepts and applications. Anal Bioanal Chem. 2006;386:1579–80.
[27] Bujnicki JM. Protein-structure prediction by recombination of fragments. ChemBioChem. 2006;7:19–27 1579–80.
[28] Osguthorpe DJ. Ab initio protein folding. Curr Opin Struct Biol. 2000;10:146–52.
[29] Case DA, Cheatham III TE, Darden T, Gohlke H, Luo R, Merz Jr KM, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005;26:1668–88.
[30] Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham III TE, DeBolt S, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun. 1995;91:1–41.
[31] Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral Angles. J Chem Theory Comput. 2012;8:3257–3273.
[32] Liwo A, Kaźmierkiewicz R, Czaplewski C, Groth M, Ołdziej S, Wawak RJ, et al. United-residue force field for off-lattice protein-structure simulations: III. Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. J Comput Chem. 1998;19:259–76.
[33] Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013;29:845–54.
[34] Kundrot CE, Ponder JW, Richards FM. Algorithms for calculating excluded volume and its derivatives as a function of molecular conformation and their use in energy minimization. J Comput Chem. 1991;12:402–9.
[35] Floudas CA, Fung HK, McAllister SR, Mönnigmann M, Rajgaria R. Advances in protein structure prediction and de novo protein design: a review. Chem Eng Sci. 2006;61:966–88.
[36] Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol. 1997;268:209–25.
[37] Zhang Y, Skolnick J. SPICKER: A clustering approach to identify near-native protein folds. J Comput Chem. 2004;25:865–71.
[38] Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—Round XII. Proteins Struct Funct Bioinforma. 2018;86:7–15.
[39] Finkelstein AV, Ptitsyn OB. Why do globular proteins fit the limited set of folding patterns? Prog Biophys Mol Biol. 1987;50:171–90.
[40] Levitt M, Chothia C. Structural patterns in globular proteins. Nature. 1976;261:552–8.
[41] Setubal JC, Meidanis J. Introduction to computational molecular biology. Boston: PWS Pub, 1997.
[42] Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, et al. The protein data bank. Acta Crystallogr Sect D Biol Crystallogr. 2002;58:899–907.
[43] Bajorath J, Stenkamp R, Aruffo A. Knowledge-based model building of proteins: concepts and examples. Protein Sci. 1993;2:1798–810.
[44] Sternberg MJ, Thornton JM, Blundell TL, Sibanda BL. Knowledge-based prediction of protein structures and the design of novel molecules. Nat Int J Sci. 1987;326:347–52.
[45] Kopp J, Schwede T. Automated protein structure homology modeling: a progress report. Pharmacogenomics. 2004;5:405–16.
[46] Apostolico A, Giancarlo R. Sequence alignment in molecular biology. J Comput Biol. 1998;5:173–96.
[47] Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA. 1988;85:2444–8.
[48] Lipman DJ, Pearson WR. Rapid and sensitive protein similarity searches. Science (80-.). 1985;227:1435–41.
[49] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
[50] Altschup SF, Gish W, Pennsylvania T, Park U. Basic local alignmen. J Mol Biol. 1990;215:403–10.
[51] Notredame C. Recent progresses in multiple sequence alignment: a survey. Pharmacogenomics. 2002;3:131–44.
[52] Notredame C. Recent evolutions of multiple sequence alignment algorithms. PLoS Comput Biol. 2007;3:1405–8.
[53] Thompson JD, Plewniak F, Poch O. A comprehensive comparison of multiple sequence alignment programs. Nucleic Acids Res. 1999;27:2682–90.
[54] Thompson JD, Higgins DG. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–80.
[55] Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000;302:205–17.
[56] Czjzek M, Cicek M, Zamboni V, Bevan DR, Henrissat B, Esen A. The mechanism of substrate (aglycone) specificity in β-glucosidases is revealed by crystal structures of mutant maize β-glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. Proc Natl Acad Sci USA. 2000;97:13555–60.
[57] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993;26:283–91.
[58] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35:407–10.
[59] Kresge JS, Leonowicz CT, Roth ME, Vartuli WJ, Beck JC. Assessment of protein models with three-dimensional profiles. Nature. 1992;359:710–13.
[60] Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph. 1990;8:52–6.

[61] Ketudat Cairns JR, Champattanachai V, Srisomsap C, Wittman-Liebold B, Thiede B, Svasti J. Sequence and expression of Thai Rosewood beta-glucosidase/beta-fucosidase, a family 1 glycosyl hydrolase glycoprotein. J Biochem. 2000;128:999–1008.
[62] Eswar N, Webb B, Marti-Renom MA, Madhusudan MS, Eramian D, Shen M, et al. Comparative protein structure modeling using Modeller. Curr Protoc BioinformaticsCurr. 2006;15:5.6.1–5.6.30.
[63] Lavecchia A, Giovanni C. Virtual screening strategies in drug discovery: a critical review. Curr Med Chem. 2013;20:2839–60.
[64] Li Q, Shah S. Structure-based virtual screening. Methods Mol Biol. 2017;1558:111–24.
[65] Dror O, Shulman-peleg A, Nussinov R, Wolfson HJ. Predicting molecular interactions. Curr Med Chem. Curr. 2004;11:71–90.
[66] Hamza A, Wei NN, Zhan CG. Ligand-based virtual screening approach using a new scoring function. J Chem Inf Model. 2012;52:963–74.
[67] Kukol A. Molecular docking. In: Kahl G, editor(s). The dictionary of genomics, transcriptomics and proteomics, 5th ed. Vol. 443. Mannheim: Wiley-Blackwell, 2015:1–1.978-3-527-32852-9.
[68] Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins Struct Funct Genet. 2002;47:409–43.
[69] Taylor JS, Burnett RM. DARWIN: A program for docking flexible molecules. Proteins Struct Funct Genet. 2000;41:173–91.
[70] Norel R, Lin SL, Wolfson HJ, Nussinov R. Shape complementarity at protein–protein interfaces. Biopolymers. 1994;34:933–40.
[71] Norel R, Petrey D, Wolfson HJ, Nussinov R. Examination of shape complementarity in docking of unbound proteins. Proteins. 1999;36:307–17.
[72] Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem. 1985;28:849–57.
[73] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004;47:1739–49.
[74] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem. 2004;47:1750–9.
[75] McGann MR, Almond HR, Nicholls A, Grant JA, Brown FK. Gaussian docking functions. Biopolymers. 2003;68:76–90.
[76] Ewing TJ, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem. 1997;18:1175–89.
[77] Bohm HJ. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des. 1992;6:61–78.
[78] Mizutani MY, Tomioka N, Itai A. Rational automatic search method for stable docking models of protein and ligand. J Mol Biol. 1994;243:310–26.
[79] Zsoldos Z, Reid D, Simon A, Sadjad BS, Johnson AP. eHiTS: An innovative approach to the docking and scoring function problems. Curr Protein Pept Sci. 2006;7:421–35.
[80] Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. J Comput Aided Mol Des. 1994;8:153–74.
[81] Lorber DM, Shoichet BK. Flexible ligand docking using conformational ensembles Despite important successes. Protein Sci. 1998;7:938–50.
[82] Joseph-McCarthy D, Thomas BE, Belmarsh M, Moustakas D, Alvarez JC. Pharmacophore-based molecular docking to account for ligand flexibility. Proteins Struct Funct Genet. 2003;51:172–88.
[83] Sauton N, Lagorce D, Villoutreix BO, Miteva MA. MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. BMC Bioinformatics. 2008;9:1–12.
[84] Di Costanzo L, Jr LV, Christianson DW. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins. 2006;642:637–42.
[85] Huang SY, Zou X. Efficient molecular docking of NMR structures: application to HIV-1 protease. Protein Sci. 2006;16:43–51.
[86] Brylinski M, Skolnick J. Q-dock: low-resolution flexible ligand docking with pocket-specific threading restraints. J Comput Chem. 2008;29:1574–88.
[87] Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. J Med Chem. 2006;49:3315–21.
[88] Benigni R, Bossa C. Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. Chem Rev. 2011;111:2507–36.
[89] McGann M. FRED and HYBRID docking performance on standardized datasets. J Comput Aided Mol Des. 2012;26:897–906.
[90] Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins Struct Funct Genet. 2003;52:80–7.
[91] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004;3:935–49.
[92] Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct. 2003;32:335–73.
[93] Rishton GM. Reactive compounds and in vitro false positives in HTS. Drug Discov Today. 1997;2:382–4.
[94] Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol. 2008;153:7–26.
[95] Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. Phys Chem Chem Phys. 2010;12:12899–908.
[96] Krovat E, Steindl T, Langer T. Recent advances in docking and scoring. Curr Comput Aided-Drug Des. 2006;1:93–102.
[97] Jain AN. Scoring functions for protein-ligand docking. Curr Protein Pept Sci. 2006;7:407–20.
[98] Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, Norton S, et al. ChemBank: a small-molecule screening and cheminformatics resource database. Nucleic Acids Res. 2008;36:351–9.
[99] Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. J Comput Chem. 1992;13:505–24.
[100] Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. Proteins Struct Funct Bioinforma. 1990;8:195–202.
[101] Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. Proteins. 1999;34:4–16.

[102] Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. Brief Bioinform. 2009;10:579–91.
[103] Knox C, Law W, Jewison T, Liu P, Ly S, Frokis A, et al. DrugBank 3.0: A comprehensive resource for 'Omics' research on drugs. Nucleic Acids Res. 2011;39:1035–41.
[104] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012;40:1100–7.
[105] Del Rio A, Barbosa AJ, Caporuscio F, Mangiatordi GF. CoCoCo: a free suite of multiconformational chemical databases for high-throughput virtual screening purposes. Mol Biosyst. 2010;6:2122–8.
[106] Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem. 2000;43:4759–67.
[107] Baell JB. Observations on screening-based research and some concerning trends in the literature. Future Med Chem. 2010;2:1529–46.
[108] Lagorce D, Maupetit J, Baell J, Sperandio O, Tufféry P, Miteva MA, et al. The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections. Bioinformatics. 2011;27:2018–20.
[109] Seifert MH. Robust optimization of scoring functions for a target class. J Comput Aided Mol Des. 2009;23:633–44.
[110] Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J Med Chem. 1999;42:5100–9.
[111] Feher M. Consensus scoring for protein-ligand interactions. Drug Discov Today. 2006;11:421–8.
[112] Raub S, Steffen A, Kämper A, Marian CM. AIScore - Chemically diverse empirical scoring function employing quantum chemical binding energies of hydrogen-bonded complexes. J Chem Inf Model. 2008;48:1492–510.
[113] Warren GL, Andrews CW, Capelli AM, Clarke M, LaLonde J, Lambert MH, et al. A critical assessment of docking programs and scoring functions. J Med Chem. 2006;49:5912–31.
[114] Willett P, Barnard JM, Downs GM. Chemical similarity searching. J Chem Inf Comput Sci. 1998;38:983–96.
[115] Ma X, Jia J, Zhu F, Xue Y, Li Z, Chen Y. Comparative analysis of machine learning methods in ligand-based virtual screening of large compound libraries. Comb Chem High Throughput Screen. 2009;12:344–57.
[116] Böhm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des. 1994;8:243–56.
[117] Boehm HJ. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. J Comput Aided Mol Des. 1998;12:309–23.
[118] Melville JL, Burke EK, Hirst JD. Machine learning in virtual screening. Comb Chem High Throughput Screen. 2009;12:332–43.
[119] Wang R, Liu L, Lai L, Tang Y. SCORE: a new empirical method for estimating the binding affinity of a protein-ligand complex. J Mol Model. 1998;4:379–94.
[120] Tao P, Lai L. Protein ligand docking based on empirical method for binding affinity estimation. J Comput Aided Mol Des. 2001;15:429–46.
[121] Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des. 2002;16:11–26.
[122] Eckert H, Vogt I, Bajorath J. Mapping algorithms for molecular similarity analysis and ligand-based virtual screening: design of DynaMAD and comparison with MAD and DMC. J Chem Inf Model. 2006;46:1623–34.
[123] Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des. 1997;11:425–45.
[124] Barreiro G, Guimarães CR, Tubert-Brohman I, Lyons TM, Tirado-Rives J, Jorgensen WL. Search for non-nucleoside inhibitors of HIV-1 reverse transcriptase using chemical similarity, molecular docking, and MM-GB/SA scoring. J Chem Inf Model. 2007;47:2416–28.
[125] Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol. 1996;261:470–89.
[126] Muegge I. A knowledge-based scoring function for protein-ligand interactions: probing the reference state. Perspect Drug Discov Des. 2000;20:99–114.
[127] Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem. 1999;42:791–804.
[128] Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol. 2000;295:337–56.
[129] DeWitte RS, Ishchenko AV, Shakhnovich EI. SMoG: de novo design method based on simple, fast, and accurate free energy estimates. 2. Case studies in molecular design. J Am Chem Soc. 1997;119:4608–17.
[130] Smith RD, Hu L, Falkner JA, Benson ML, Nerothin JP, Carlson HA. Exploring protein-ligand recognition with binding MOAD. J Mol Graph Model. 2006;24:414–25.
[131] Okuno Y, Tamon A, Yabuuchi H, Niijima S, Minowa Y, Tonomura K, et al. GLIDA: GPCR - Ligand database for chemical genomics drug discovery - Database and tools update. Nucleic Acids Res. 2008;36:907–12.
[132] Mcgaughey GB, Sheridan RP, Bayly CI, Culberson JC, Kreatsoulas C, Lindsley S, et al. Comparison of topological, shape, and docking methods in virtual screening. J Chem Inf ModelJ . 2007;47:1504–19.
[133] Krejsa CM, Horvath D, Rogalski SL, Penzotti JE, Mao B, Barbosa F, et al. Predicting ADME properties and side effects: the BioPrint approach. Curr Opin Drug Discov Devel. 2003;6:470–80.
[134] Lin X, Huang XP, Chen G, Whaley R, Peng S, Wang Y, et al. Life beyond kinases: structure-based discovery of sorafenib as nanomolar antagonist of 5-HT receptors. J Med Chem. 2012;55:5749–59.
[135] Corbeil CR, Englebienne P, Moitessier N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. J Chem Inf Model. 2007;47:435–49.
[136] Wang R, Fang X, Lu Y, Yang CY, Wang S. The PDBbind database: methodologies and updates. J Med Chem. 2005;48:4111–19.
[137] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003;52:609–23.
[138] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem. 2009;30:2785–91.