AN ASSESSMENT OF POTENTIAL THERMOSTABLE D-ALLULOSE EPIMERASES OBTAINED FROM GENOME MINING USING A COMPUTATIONAL SIMULATION APPROACH

RUDI NIRWANTONO[1], ALAM AHMAD HIDAYAT[1,2], JOKO P. TRINUGROHO[1,*], DIGDO SUDIGYO[1], BENS PARDAMEAN[1,3]

[1]Bioinformatics & Data Science Research Center, Bina Nusantara University, Jakarta 11480, Indonesia

[2]Department of Mathematics, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

[3]Computer Science Department, BINUS Graduate Program - Master of Computer Science Program, Bina Nusantara University, Jakarta 11480, Indonesia

Abstract: D-Allulose is a type of rare sugar with low-calorie and many health benefits, which shows a great prospect to be utilized for the food industries. The production of D-Allulose can be achieved biotechnologically using D-Allulose 3-Epimerase (DAEase) enzyme, which is an environmentally friendly approach compared to the chemical method. However, for the sustainable and large-scale production of D-Allulose, it is necessary to use the DAEase enzyme that has several characteristics, such as high thermal activity, high stability, and high specificity. The present work aims to discover new potential DAEase from thermophilic bacteria via a computational biology approach. Genome mining was first performed to obtain the relevant sequences from the database. The sequences were subsequently examined in order to ascertain their phylogenetic placements and identify the conserved motifs. Additionally, 3D modeling and Molecular Dynamic (MD) simulation of the protein were undertaken to explore the

---

*Corresponding author

E-mail address: joko.trinugroho@binus.edu

potential of the presumed enzyme using computational analysis. Two sequences originating from the species *C. composti* were uncovered through the process of genome mining. These sequences were subsequently confirmed to be homologous and displayed a comparable structure to the already established DAEases, including the presence of active site conformations and metal binding sites. Based on molecular dynamics (MD) simulation, it was determined that CcDAE1 exhibited relatively greater structural stability in comparison to CcDAE2, although still lower than that observed in NtDAE. Overall, this work has shown the applicability of computational biology to find the new DAEase with the desired characteristics.

Keywords: D-Allulose; DAEase; computational biology; thermophilic bacteria.

2020 AMS Subject Classification: 92-08.

1. INTRODUCTION

The altering of daily activity into a sedentary lifestyle along with excessive calorie and fat consumption leads to the elevating number of people suffering from metabolic syndrome [1]. Therefore, obesity, diabetes, hypertension, and hyperlipidemia have gained the attention of experts to seek solutions. Correspondingly, D-Allulose successfully attracted the interest of researchers as a novel supplement that contains low calories without altering the taste and flavor of the food [2]. In addition, D-allulose has now been received in many regions including South America, Asia Pacific Region, the Middle East, and Africa. D-Allulose is currently designated as a new food in the United States and is recognized as a safe food. [3,4]. Therefore, the growing market is also still established with a market value of about US\$ 232.91 million by 2021 and is projected to rise to US\$ 362.24 million in 2028 [5].

Until now, the production of D-Allulose has relied heavily on biotechnology production using a recombinant enzyme named D-Allulose 3-Epimerase (DAEase) because the chemical production approach is labor intensive and harmful to the environment. Several bacteria have been known to be sources of enzymes including DaeM [6], *Clostridium bolteae* [7], *Dorea* sp. CAG317 [8], *Ruminococcus* sp. [9], *Treponema primitia* ZAS-1 [10], *Thermoclostridium caenicola* [11], *Arthrobacter globiformis* M30 [12], *A. tumefaciens* [13], *Flavonifractor plautii* [14], *Pirellula* sp. SH-Sr6A [15], *C. cellulolyticum* H10 [16], *Novibacillus thermophilus* [17].

However, D-allulose production commonly undergoes under harsh settings, including high

thermal treatment. Thus, the industry tends to employ an enzyme with a thermal-resistant feature. Even though the engineering approach can improve thermal stability, it can also sometimes reduce the activity instead. Due to this industrial requirement, the exploration of a novel enzyme with intrinsic properties of high thermal activity and stability besides high enzyme activity is still necessary to discover more potential enzymes.

Genome mining highly depends on computational technology and bioinformatics methods. This approach was regularly used to discover the biosynthetic gene clusters involved in the synthesis of natural bioproducts or to reveal the gene involved in biological pathways [18–21]. This approach utilized a combination of statistical and alignment-based methods to find the similarity motif shared among homologous sequences to determine whether a stand of DNA is the target or not. Due to its versatile usage, genome mining was used to discover a potential DAEase gene within thermophilic bacteria sourced from Indonesia.

In this study, database genome mining to discover new potential DAEase was conducted against a list of genomes from thermophilic bacteria obtained from the literature study. The sequences retrieved from the gene mining were then analyzed to determine the phylogenetic positions and to find conserved motifs. The 3D modelling and Molecular Dynamic (MD) simulation of the protein were also performed to study the potential of the putative enzyme based on *in-silico* analysis.

2. METHODS

2.1. Listing thermophilic bacteria and gene mining from NCBI. The thermophilic bacteria were listed in the literature related to the isolation and identification of thermophilic bacteria from Indonesia. The screened bacteria were listed and grouped according to their Phylum and Class to understand their taxonomic status. Then, the genome mining was conducted by using the Biopython package in Python 3.10 framework targeting DAEase genes in the genome of given bacteria. The Bio.Blast module was applied to retrieve the protein enzyme from a non-redundant sequence database in NCBI and screened the sequence encoding DAEase using a sequence of DAEase from *Novibacillus thermophilus* (Accession Number: WP_077721022.1) as the query. Similarity-scoring matrix of BLOSSUM80 was also set to ensure the filtering of the sequences with higher sequence similarity. The sequences with a length between 280-300 and a query cover

of more than 90% were selected for the following steps.

2.2. Homology modelling of DAEase Proteins. In this work, comparative/homology modeling was performed to determine the three-dimensional (3D) protein structures at the tertiary level. The method is widely used for predicting the 3D structure of a protein, given some known homologous template structures are found a priori. The method relies on the observation that the 3D structures are more conserved than the sequences. In this work, we adopted a homology modeling technique proposed by Shen and Sali that utilizes a scoring function called Discrete Optimized Protein Energy (DOPE) [22]. DOPE is defined as the negative logarithm of a joint probability function of a given protein that depends on the atomic distances of the protein. The distant-dependent probability function is derived from a number of native structures (i.e., known structures retrieved from the Protein Data Bank) from different proteins using a probability theory. The objective is to find the global minimum of the DOPE function that can be physically interpreted as the free energy of the native structure of a given protein.
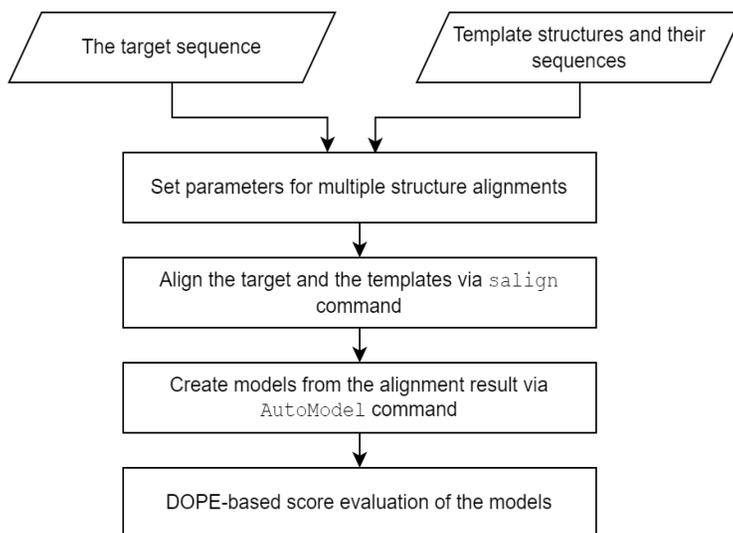
```
┌─────────────────────────┐      ┌─────────────────────────────┐
│   The target sequence   /      /  Template structures and their │
└─────────────────────────┘      /          sequences            /
                                 └─────────────────────────────┘
```

```
┌──────────────────────────────────────────────┐
│   Set parameters for multiple structure alignments │
└──────────────────────────────────────────────┘
```

```
┌──────────────────────────────────────────────┐
│  Align the target and the templates via salign │
│                   command                      │
└──────────────────────────────────────────────┘
```

```
┌──────────────────────────────────────────────┐
│  Create models from the alignment result via  │
│             AutoModel command                  │
└──────────────────────────────────────────────┘
```

```
┌──────────────────────────────────────────────┐
│      DOPE-based score evaluation of the models │
└──────────────────────────────────────────────┘
```

FIGURE 1. The workflow of the comparative modeling used in this work.

The workflow of the comparative modeling in our work, summarized in Figure 1, was implemented using Modeller 10.1 software with Python 3.10 framework. More specifically, three crystal structures with the highest sequence similarity from the PDB database were used as the templates. The templates were determined by conducting template selection using BLAST-p search in NCBI against the PDB database with the CcDAE1 and CcDAE2 sequences as the query

to calculate the three most similar templates [2,23]. The *satisfaction of spatial restraints* method was used for the model building, which is the default algorithm in Modeller via AutoModel command. Guided by the above alignment result, the method constructs homology-based restraints on the geometry of the modeled protein (angles and distances) from the templates, considering physicochemical properties derived from the CHARMM22 force field [24]. An optimization process is then carried out to minimize the deviation of the restraints (the objective function) until the desired structure candidates are obtained according to the DOPE assessment score.

In practice, the evaluation of the model was conducted by seeking the minimum DOPE score and also plotting the DOPE score profile for each residual amino acid from both the model and the template. Afterward, the PyMol software was used to visualize the final model. A superimposition of the model over the templates was also conducted to compare the secondary structure and the amino acid positioning of the model over the template in a more specific manner.

2.3. Molecular dynamics. We conducted the structural stability analysis of the resulting proteins by performing Molecular Dynamics (MD) simulations (Figure 2). The setup of the simulations requires the proteins to be dissolved in an optimal water system, allowing interactions between atoms in the protein as well as interactions with the surrounding water molecules. The simulation in this work was conducted using Gromacs 2023 with NVIDIA-CUDA support on Ubuntu 18.04.

First, a topology describing the geometry and interactions of atoms (bond lengths, bond angles, dihedrals, etc) in all residues of a "clean" protein (i.e., the structure file with no water molecules and ligand) was built with CHARMM36 force field parameterization. The magnesium ion topology in the binding pocket was also parametrized by the force field, assuming non-bonded interactions (van der Waals and Coulomb forces) with nearby atoms. We utilized CHARMM TIP3P as the water molecule model. Further, the solvation is based on the solvent model spc216. The volume of the simulated solvation system is required to be as optimal as possible. This was achieved using a water box in the form of a rhombic dodecahedron where the protein was placed in the center of the box with a minimal distance to each edge to be 1.01-1.05 nanometers. In order to neutralize the system, a number of charged ions were added to the solvent (it was either Na+ or Cl-).
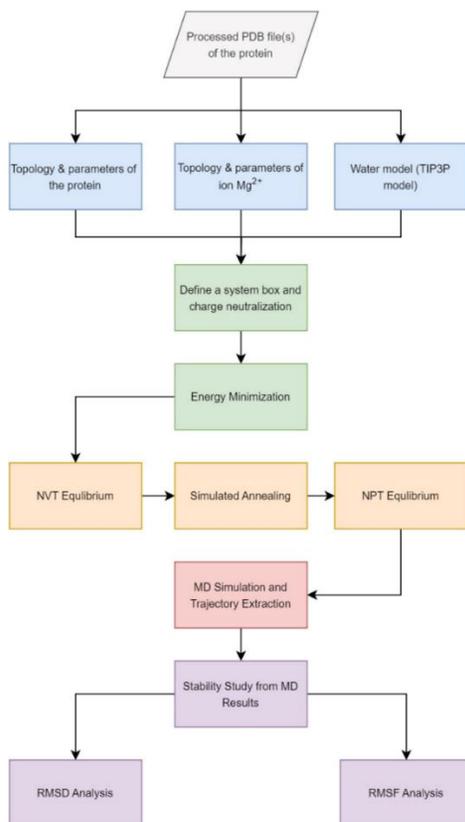
FIGURE 2. The protocol of the MD-based analysis of the modelled structures.

We employed the standard periodic boundary conditions for the system box. The particle mesh Ewald was deployed for the long-range Coulomb interactions with a grid size of 1.2 Angstrom, which was also the same cut-off for van der Waals forces. LINCS was used as the constraint algorithm. Meanwhile, during the simulation V-rescale algorithm was utilized for the constant temperature coupling and Parinello-Rahman was for the pressure coupling.

The system was then relaxed via energy minimization using maximum iterations of 50000 until the total potential energy of the system was less than 10 kJ/mol. The relaxed system was then brought into three phases of equilibration: NVT ensemble, simulated annealing, and NPT ensemble. The NVT equilibration was to bring the system into an initial reference temperature of 5 K with a period of 250 picoseconds (ps). Simulated annealing was done to slowly increase the temperature from 5 K to desired temperature (300 K or 338 K) for 1000 ps. Lastly, the NPT equilibration for 250 ps was to stabilize the pressure at 1 bar.

The MD simulation was finally conducted for the equilibrated system for 30000 ps (30 ns) with a time step of 0.002 ps to produce sufficient trajectory data for analyzing the stability of the protein structure. In this work, we employed the Root Mean Square Deviation (RMSD) of the structural

backbone to examine the convergence of the stability during the simulation. If the obtained RMSD is properly observed, the Root Mean Square Fluctuation (RMSF) of alpha carbons of the protein is computed to identify the most dynamic regions along the residues of the protein.

3. RESULTS

3.1. DAEase gene mining of thermophilic bacteria. The literature study led to the finding of 32 bacteria: *Paenibacillus barengoltzii, Paenibacillus cisolokensis, Chelatococcus composti, Geobacillus thermoleovorans, Geobacillus stearothermophilus, Geobacillus kaustophilus, Anoxybacillus flavithermus, Anoxybacillus rupiensis, Paenibacillus cisolokensis, Brevibacillus borstelensis, Gandjariella thermophila, Aneurinibacillus migulanus, Pseudomonas balearica, Bacillus halotolerans, Bacillus megaterium, Bacillus subtilis subsp. stercoris strain N12, Bacillus licheniformis,* and *Bacillus caldotenax.* The genome mining of those bacteria using the BLAST-p search module (BLOSSUM80) with WP_077721022.1 as the query and based on their query cover (>90%), which resulted in a species containing putative DAEase. The species was *Chelatococcus composti*, which contained two sequences, WP_183335203.1 (next called CcDAE1) and WP_183336036.1 (next called CcDAE2) (Table 1). Sequence CcDAE1 and CcDAE2 were retrieved from Whole Genome Sequencing data of *Chelatococcus composti* strain CGMCC 1.15283 (Accession Number: NZ_BMHX01000008.1).

By possessing 282 amino acids, CcDAE1 was slightly shorter than CcDAE2, compared to 296 amino acids in CcDAE2. The CDS region of CcDAE1 was located at nucleotide 269568-270416 in the forward direction, while the CDS of the sequence CcDAE2 was at 36757-37647 in the reverse direction. Both sequences were also known to bear a conserved *AP_endonuclease_2* domain in the sequence.

TABLE 1. Two best sequences candidates of DAEase.

Protein Accession Number	Species	Strain	WGS Accession Number	Length	Region name	E	Similarity	Query cover
WP_183335203.1	*Chelatococcus composti*	CGMCC 1.15283	NZ_BMHX0 1000008.1	282 aa	AP_end onuc_2	1e-31	29.26%	92%
WP_183336036.1	*Chelatococcus composti*	CGMCC 1.15283	NZ_BMHX0 1000008.1	296 aa	AP_end onuc_2	8e-33	30.31%	95%

3.2. Protein modelling. According to the BLAST-p search over the PDB database, the three crystal structures with the highest identity score with CcDAE1 were from *Agrobacterium* sp. SUL3 (PDB accession: 7erm), *Methylomonas* sp. DH-1 (PDB accession: 7cj4), and *Arthrobacter globiformis* (PDB accession: 5zfs) by 65.59%, 35.67%, and 33.94%, respectively. On the other hand, CcDAE2 revealed the highest identity with *Mesorizhobium loti* (PDB accession: 3vyl), *Methylomonas* sp. DH-1 (PDB accession: 7cj4), and *Pseudomonas chicorii* (PDB accession: 4xsl) by 34.44%, 31.35%, and 31.09%. Therefore, those 3D structures were used for the 3D modeling of each sequence.

Two 3D models (CcDAE1 and CcDAE2) obtained from the homology modelling were evaluated to determine the backbone dihedral angles ψ against φ of amino acid residues in each protein structure and plot each residue in the Ramachandran plot (Figure 3). The evaluation showed that 98.814% and 1.186% residues in CcDAE1 (Figure 3A) were observed in highly preferred and preferred areas. Meanwhile, 97.015% and 1.493% residues in CcDAE2 (Figure 3B) were laid in highly preferred and preferred areas. The results indicated that both models were accurate since the amino acid residues in preferred were more than 90%.
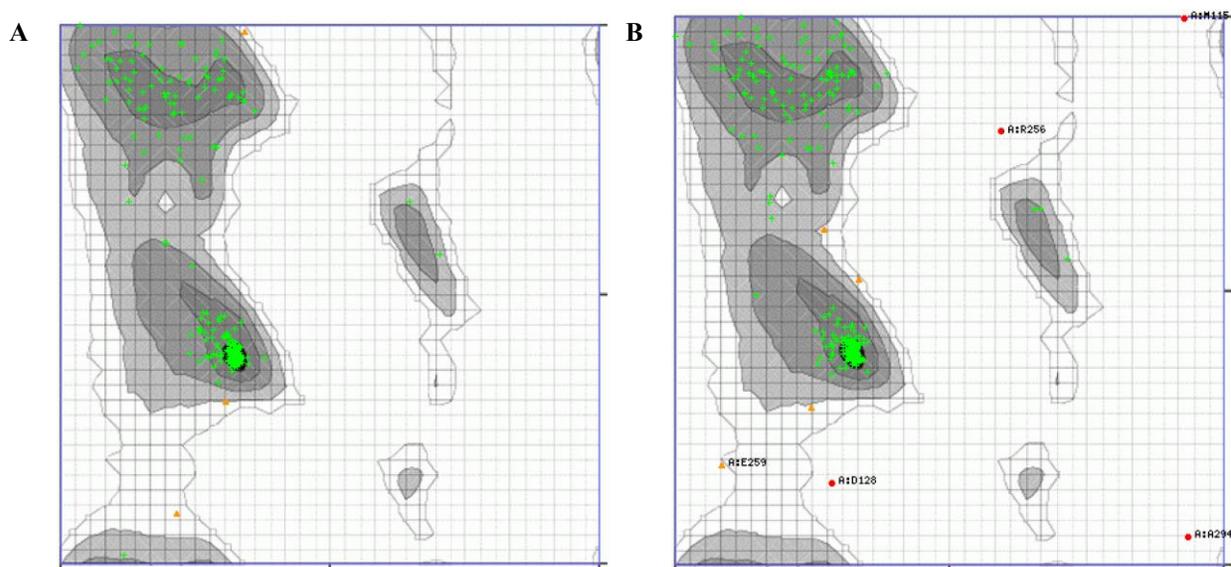

FIGURE 3. Ramachandran plot of protein modeling. Figure 3A and 3B show the position of each amino acid residue of CcDAE1 and CcDAE2, respectively.

According to the tertiary structure analysis, the 3D structure model of both sequences from *Chelatococcus composti* follows (β/α)8 TIM barrel fold, similar to the existing DAEases (Figure 4A). The eight β-strands inside the barrel were wrapped over by the eight α-helices outside. A number of irregular segments called loops, with which a single segment connects two adjacent regular secondary structures, were also observed in our models. Most loops were fully and partially buried within the barrel structure, but some were also clearly visible (see Figure 4B of loop4 for a clear example). In addition, we assigned numbers from 1-7 to these loops according to the order of appearance in their primary sequences. Several loops were found in a pore region situated in the center of the barrel, in which a metal ion binding and the reaction took place. This pore was covered by loops 5-7 on which some substrate binding residues were located. The pore existed in the center of the barrel as the location at which a metal ion took place and the reaction occurred. This pore was covered by several loops on which some substrate binding residues were located (loop5-7). Despite the inherent structural similarity between CcDAE1 (Figure 4A, blue) and CcDAE2 (Figure 4A, red), there were some differences between the two models. Some differences were also disclosed on their α-helix, β-strands, and loops after superimposition between the two of them was performed.

The superimposition of both models over template DAEases from *Mesorizhobium loti*, *Methylomonas* sp. DH-1, and *Arthrobacter globiformis* (Figure 4B) also showed apparent differences on some α-helixes, even though both models were almost similar in structure with the templates. However, some structural differences were also monitored. It includes the irregular structure of loop4 that extended differently compared to the templates in the model of CcDAE2, even though this loop does not bear any significant residue neither metal binding nor substrate binding. Figure 4C also shows the different lengths and trajectories of β4 and β7 in both CcDAE1 and CcDAE2 compared to the DAEase crystal structure from A. *globiformis*. This was correlated with the amino acids variation that was monitored to build-up the base of those β-strands. Besides, Figure 4D also reveals variations in amino acids and the position of some important residues. The metal binding at β7 in CcDAE2 was Glutamine (Q217) instead of Histidine like in CcDAE1 (H203) and in A. *globiformis* (H205). The residue that existed in loop7 (R209 in CcDAE1 and R223 in CcDAE2) was positioned differently compared to residue R211 in DAEase from A. *globiformis*.

The R-group of Arginine in CcDAE2 positioned inverted compared to that in CcDAE1 and DAEase from A. *globiformis*. The metal binding of E252 at β8 in CcDAE2 was slightly rotated compared to it in CcDAE1 (E238) and DAEase from A. *globiformis* (E240). On the other hand, the rests were positioned with almost similar position and trajectories.
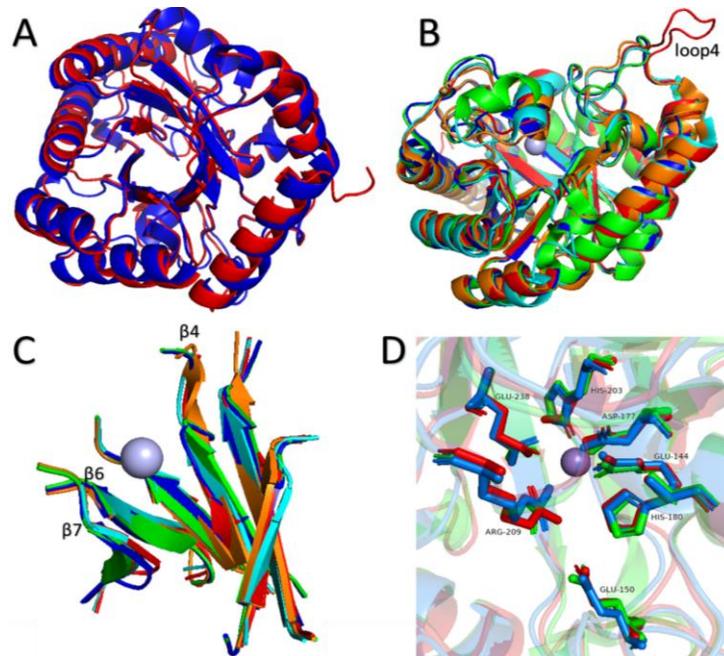

FIGURE 4. The 3D visualization of ApDAE. (A) The ribbon representation of CcDAE1 and CcDAE2 structure models. (B) The superimposition of models with DAEases from various organisms. (C) The comparison of β-strands "core" structure. (D) Variations in amino acids and the position of important residues.

3.3. MD Simulation. The MD simulation was performed at temperatures 27°C and 65°C to examine the structural stability of CcDAE1 and CcDAE2 following the protocol described previously. The RMSD trajectory of both models was also compared with the RMSD trajectory of the control model from *Novibacillus thermophilus* (NtDAE). NtDAE was chosen due to its outstanding thermal stability and superior enzyme activity over other DAEases. It optimally converts D-fructose to D-allulose at 70°C, so the Tm should be higher than it. The RMSD trajectories between the relaxed structure and the simulated ones were examined and recorded every 0.01 ns to ensure the convergence of the stability. The calculation excluded the N- and C-terminal tails due to their high flexibility and less importance in thermal stability.
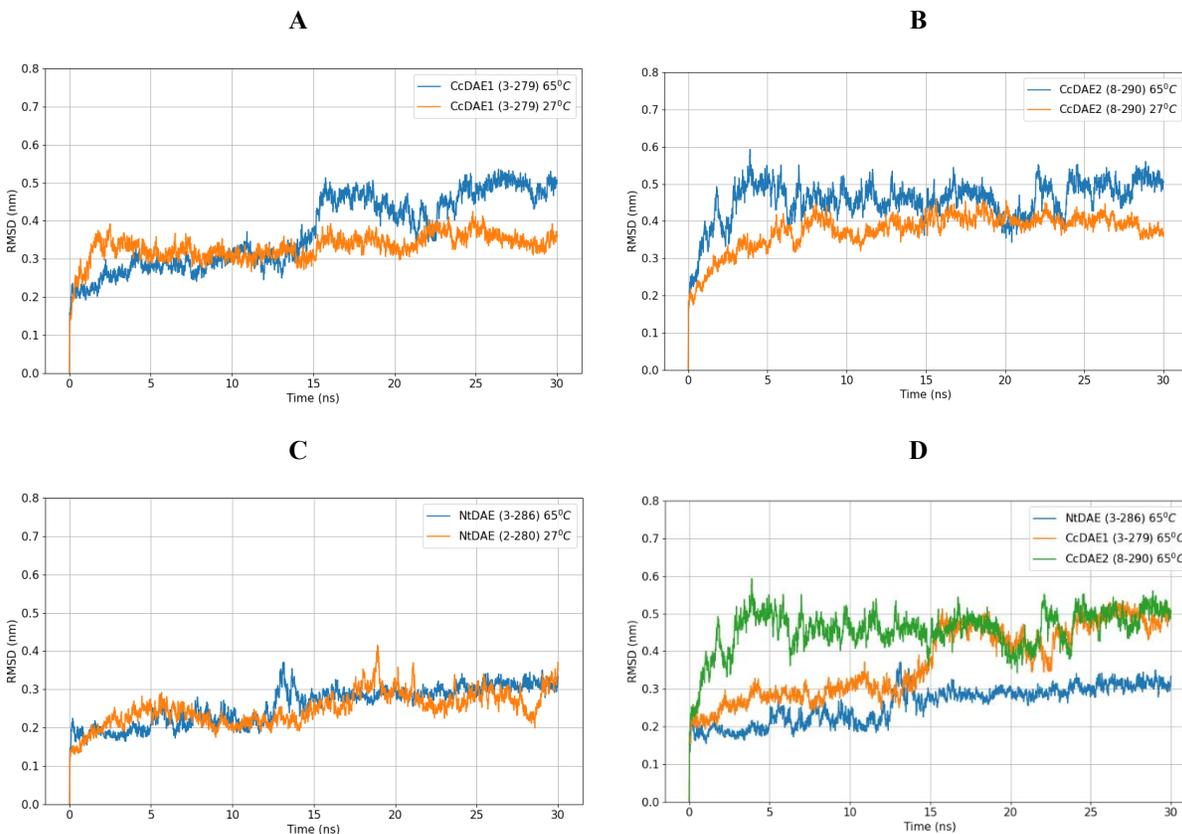
FIGURE 5. The MD simulation results. Figure 5A, 5B, and 5C show the RMSD trajectories at both temperatures of CcDAE1, CcDAE2 and NtDAE, respectively. Meanwhile, Figure 5D shows the RMSD trajectories of all proteins at 65°C.

According to Figure 5A, the RMSD trajectory of CcDAE1 at 27°C was apparently stable over the time frames with an average of 3.333 Å. The peak of the RMSD trajectory was at 4.245 Å and the lowest at 2.659 Å, or 1.586 Å in difference. At a temperature of 65°C, the RMSD trajectory of CcDAE1 gradually inclined to 15 ns, and it started to flatten but still fluctuated afterward with an average of 4.591 Å. It recorded the highest RMSD trajectory at 5.354 Å and the bottom of 3.444 Å of which the difference was 1.910 Å. Then Figure 5B revealed that CcDAE2 at both 27°C and 65°C reached their stability after 5 ns and recorded the average of 3.918 Å and 4.650 Å, subsequently. They recorded the highest RMDS trajectory at 4.594 Å and 5.604 Å, while the minimum was at 3.125 Å and 3.428 Å. Moreover, the deviation between the maximum and the minimum value for the two structures held at temperatures 27°C and 65°C were 1.469 Å and 2.1777 Å, respectively. As a comparison, the RMSD trajectory of NtDAE at 27°C and 65°C apparently behaved almost similarly where they reported averages of 2.584 and 2.773, subsequently (Figure 5C). The Figure 5D shows the comparison between the RMSD trajectory of

CcDAE1, CcDAE, and NtDAE at 65°C. Uniquely, CcDAE1 started off with the same trajectory trend as NtDAE but after 10 ns the value began to hop and then be stagnant with slight fluctuations around the same RMSD level as CcDAE2.
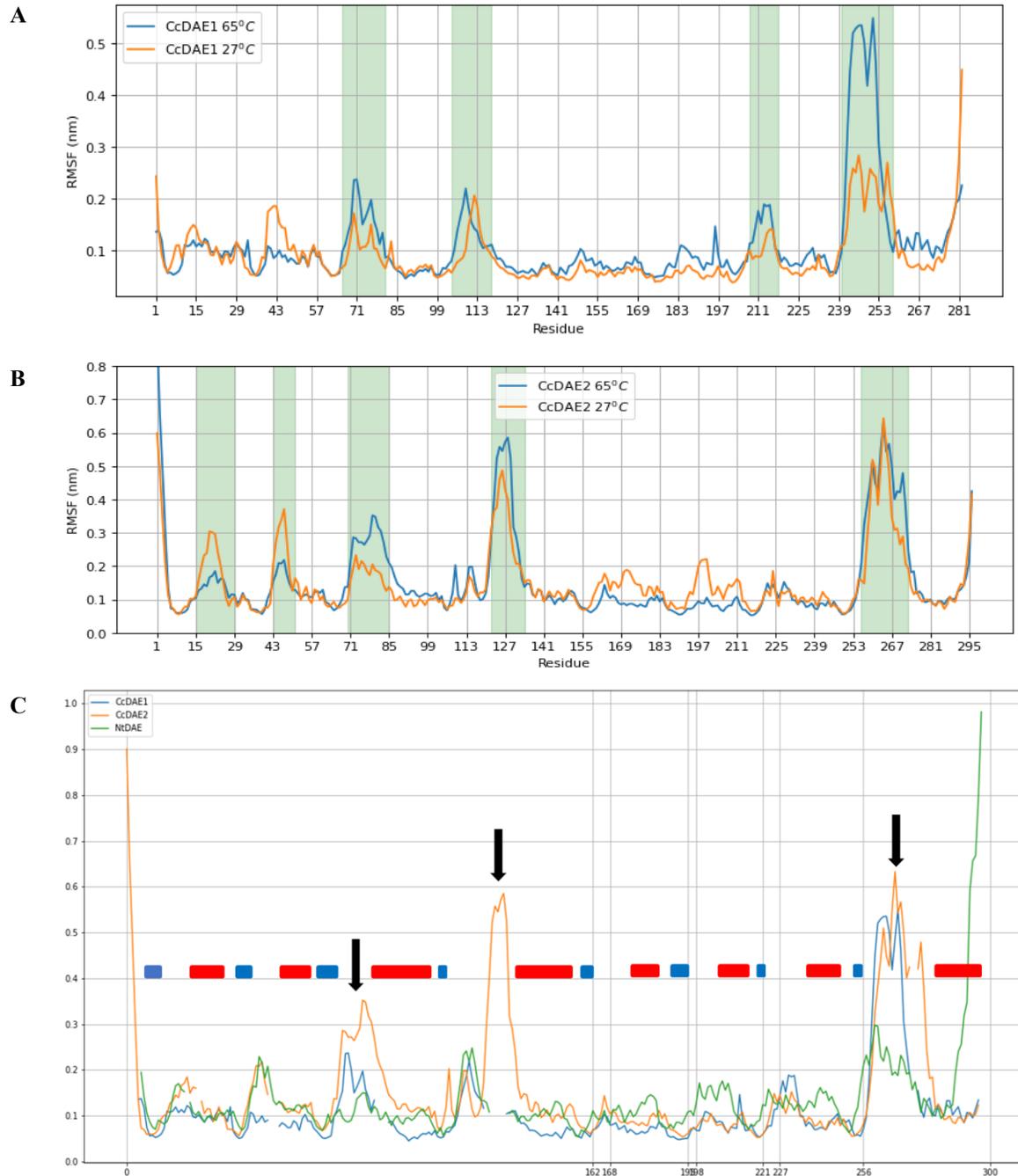

FIGURE 6. RMSF comparison of protein residues. (A) The RMSF trajectories of CcDAE1. (B) The RMSF trajectories of CcDAE2. (C) The RMSF comparison of CcDAE1, CcDAE2, and NtDAE.

In addition, the RMSF trajectories of CcDAE1 and CcDAE2 at both 27°C and 65°C were compared in Figure 6A and 6B. From the trajectory, it can be identified several prominent energetic regions in CcDAE1 & CcDAE2 (highlighted in green). In CcDAE1 and CcDAE2, the dynamic elevations were undergone regularly at their loops. The prominent different elevations in CcDAE1 and CcDAE2 were monitored almost at all loops. However, loop 8 in CcDAE1 showed an outstanding difference in dynamics in which it increased from below 3 Å at 27°C to exceed 5 Å at 65°C. Similar patterns were also shown by loops 3 and 7 but with a smaller difference. In CcDAE2, loop 8 was almost similar for both temperatures where the height of trajectory at the region elevated significantly to more than 5 Å and even reached 6 Å. A similar pattern was also shown by loops 3 and 4, but it was higher at 65°C than at 27°C. Uniquely, loops 1 and 2 in both models behaved inversely where the trajectory was higher at 27°C than at 65°C.

Figure 6C shows the comparison of the RSMF trajectory of CcDAE1, CcDAE2, and NtDAE at 65°C. Referring to the loop numbering from the 3D modeling results, the RMSF plot shows obvious distinctions in loops 3, 4, and 8. The trajectory at those regions was noticeably higher in CcDAE2 than at CcDAE1 and NtDAE indicating the higher dynamics movement at 65°C over the regions. The extended loop 4 of CcDAE2 shown in the 3D model also generated an excessive elevation of energetic dynamics (middle arrow), while the others were absent since they did not have such an extended loop. Loop 3 also showed an almost similar pattern in which the RMSF trajectory of CcDAE2 over the region elevated, and it was followed by CcDAE1 at the second and NtDAE (the lowest trajectory). At loop 8, the dynamic movements in CcDAE1 and CcDAE2 formed an almost similar pattern, while in NtDAE it was significantly lower. Then, unlike CcDAE1 and CcDAE2, NtDAE retains relatively steady dynamics along all of its residues as no residue shifting beyond 3 Å is observed during the simulation (except for the loop region in the C-terminal).

In contrast, the energetic movement at the **α**-helices (red bar) and the **β**-strands (blue bar) regions tended to be stable at around 1 Å and never surpassed 2 Å. The residues forming metal binding sites and active sites (vertical line) hardly moved (i.e., the respective alpha carbons only deviate around 1 Å from their initial positions) over the course of the simulations with almost indiscernible atom displacements at 27°C and 65°C. It indicated that the stability of the structures was relatively higher than the loop regions.

4. DISCUSSION

The importance of D-allulose epimerase (DAEase) in D-allulose conversion from D-fructose has fostered the discovery of enzymes from various species of bacteria. This focuses on the exploration of stable enzymes that support the production of sugar under a harsh setting in industrial applications [2,25]. In this study, two DAEase were successfully found from a single strain of bacteria *Chelatococcus composti*. This species of bacteria was first found in the left-over of penicillin fermentation [26], but researchers in Indonesia also identified this species from spring water [27]. According to the data from Zhang et al. [26], the GC content of this species was up to 70.9 %, meaning the genome of the species is stable enough to stay in an environment with a high temperature [26]. Therefore, the finding of two putative DAEases from this species is promising for industrial application.

Interestingly, the discovery of two homologous DAEase from a single genome of *Chelatococcus composti* is the first report since all of the previously studied species only contain a single DAEase in their genome [2,28]. The two non-allelic genes with a similar catalytic reaction are called isoenzymes. It commonly could arise from the gene duplication event in the past (divergent evolution), and the amino acid sequence was then altered over time due to mutations. The further sequence analysis of the two enzymes and the sequences from relative species also strengthen this hypothesis that the enzymes are likely to arise from a gene duplication event that occurred in the ancestor of *Chelatococcus composti*.

The results of the sequence analysis and the 3D structures indicate that both sequences are highly possible to have an epimerase activity to convert D-fructose to D-allulose. The active sites of both models are located exactly similar to other known DAEases. The only difference is in CcDAE2 which uses Glutamine amino acid (Q217) at metal binding sites instead of Histidine. While almost all DAEases used Histidine in a similar site, the usage of Glutamine at the region was only monitored at *Cereibacter sphaeroides* (previously *Rhodobacter sphaeroides*) [29]. Although the RMSF trajectory of CcDAE2 at the region was not apparently different to it from NtDAE (control), this difference could affect the binding to the metal ion. According to Zhang et al. [29], the Allulose production from Fructose using DAEase from *Cereibacter sphaeroides* was relatively low in conversion ratio and only occurred at a low temperature (optimum at 40°C). Thus, CcDAE2 could probably have a similar tendency to be active at a low temperature like it

Cereibacter sphaeroides. The 3D structure supported this since the loop regions, especially at loop 4, was excessively extended and reported high energetic movement reflected by elevated RMSF at the region. Nagi and Regan [30] also said that the loop length has an inverse correlation with its thermostability. A protein engineering conducted by Liu et al [31] on trypsin also indicated an improvement in thermal stability after truncating the loop.

On the other hand, CcDAE1 could be adequately confirmed to be more stable than CcDAE2. However, it is less stable than NtDAE at a higher temperature based on the RMSF behavior. Its pattern of the RMSF trajectory was almost similar to NtDAE at most regions, except at loops 3 and 8 where they were significantly higher in CcDAE1 that most likely to reduce the stability at a higher temperature.

5. CONCLUSION

This study has revealed the potential enzymes for D-Allulose production by employing computational biology and simulation techniques. Two sequences originating from the species *C. composti* were discovered through genome mining, namely CcDAE1 and CcDAE2. These sequences have been confirmed to be homologous and exhibit a similar structure to existing DAEases, including the conformation of active sites and metal binding sites. In addition, based on the findings of molecular dynamics (MD) simulations, it was observed that CcDAE1 possesses a relatively higher level of thermal stability compared to CcDAE2, although it is still lower than that of NtDAE. Future studies involving experimental work should be undertaken to elucidate the characteristics of the two enzymes.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] A. Bankoski, T.B. Harris, J.J. McClain, et al. Sedentary activity associated with metabolic syndrome independent of physical activity, Diabetes Care. 34 (2011), 497–503. https://doi.org/10.2337/dc10-0987.

[2] F.A. Laksmi, R. Nirwantono, I. Nuryana, et al. Expression and characterization of thermostable D-allulose 3-epimerase from Arthrobacter psychrolactophilus (Ap DAEase) with potential catalytic activity for bioconversion of D-allulose from d-fructose, Int. J. Biol. Macromol. 214 (2022), 426–438. https://doi.org/10.1016/j.ijbiomac.2022.06.117.

[3] C. Lambré, J.M. Barat Baviera, C. Bolognesi, et al. Safety evaluation of the food enzyme d-psicose 3-epimerase from the genetically modified Corynebacterium glutamicum strain FIS002, FSA J. 21 (2023), 8084. https://doi.org/10.2903/j.efsa.2021.6870.

[4] W. Zhang, S. Yu, T. Zhang, et al. Recent advances in d-allulose: Physiological functionalities, applications, and biological production, Trends Food Sci. Technol. 54 (2016), 127-137. https://doi.org/10.1016/j.tifs.2016.06.004.

[5] The Insight Partners. Allulose market size worth US$ 362.24Mn, globally, by 2028 at 6.5% CAGR - exclusive report by the insight partners 2022. https://www.globenewswire.com/en/news-release/2022/04/25/2428178/0/en/Allulose-Market-Size-Worth-US-362-24Mn-Globally-by-2028-at-6-5-CAGR-Exclusive-Report-by-The-Insight-Partners.html.

[6] S.N. Patel, G. Kaushal, S.P. Singh, d-Allulose 3-epimerase of Bacillus sp. origin manifests profuse heat-stability and noteworthy potential of d-fructose epimerization, Microbial Cell Fact. 20 (2021), 60. https://doi.org/10.1186/s12934-021-01550-1.

[7] M. Jia, W. Mu, F. Chu, et al. A d-psicose 3-epimerase with neutral pH optimum from Clostridium bolteae for d-psicose production: cloning, expression, purification, and characterization, Appl. Microbiol. Biotechnol. 98 (2013), 717–725. https://doi.org/10.1007/s00253-013-4924-8.

[8] W. Zhang, H. Li, T. Zhang, et al. Characterization of a d-psicose 3-epimerase from Dorea sp. CAG317 with an acidic pH optimum and a high specific activity, J. Mol. Catalysis B: Enzymatic. 120 (2015), 68-74. https://doi.org/10.1016/j.molcatb.2015.05.018.

[9] Y. Zhu, Y. Men, W. Bai, X. Li, L. Zhang, Y. Sun, Y. Ma, Overexpression of d-psicose 3-epimerase from Ruminococcus sp. in Escherichia coli and its potential application in d-psicose production, Biotechnol. Lett. 34 (2012), 1901-1906. https://doi.org/10.1007/s10529-012-0986-4.

[10] W. Zhang, T. Zhang, B. Jiang, et al. Biochemical characterization of a d-psicose 3-epimerase from *Treponema*

primitia ZAS-1 and its application on enzymatic production of d-psicose, J. Sci. Food Agric. 96 (2015), 49-56. https://doi.org/10.1002/jsfa.7187.

[11] J. Chen, D. Chen, M. Ke, S. Ye, X. Wang, W. Zhang, W. Mu, Characterization of a Recombinant d-Allulose 3-epimerase from Thermoclostridium caenicola with Potential Application in d-Allulose Production, Mol Biotechnol. 63 (2021) 534–543. https://doi.org/10.1007/s12033-021-00320-z.

[12] A. Yoshihara, T. Kozakai, T. Shintani, et al. Purification and characterization of d-allulose 3-epimerase derived from Arthrobacter globiformis M30, a GRAS microorganism, J. Biosci. Bioeng. 123 (2017), 170-176. https://doi.org/10.1016/j.jbiosc.2016.09.004.

[13] H.J. Kim, E.K. Hyun, Y.S. Kim, et al. Characterization of an Agrobacterium tumefaciens d-Psicose 3-Epimerase That Converts d-Fructose to d-Psicose, Appl. Environ. Microbiol. 72 (2006), 981–985. https://doi.org/10.1128/aem.72.2.981-985.2006.

[14] C.S. Park, T. Kim, S.H. Hong, et al. D-Allulose production from d-fructose by permeabilized recombinant cells of corynebacterium glutamicum cells expressing d-allulose 3-epimerase flavonifractor plautii, PLoS ONE. 11 (2016), e0160044. https://doi.org/10.1371/journal.pone.0160044.

[15] C. Li, L. Li, Z. Feng, et al. Two-step biosynthesis of d-allulose via a multienzyme cascade for the bioconversion of fruit juices, Food Chem. 357 (2021), 129746. https://doi.org/10.1016/j.foodchem.2021.129746.

[16] W. Mu, F. Chu, Q. Xing, et al. Cloning, Expression, and Characterization of a d-Psicose 3-Epimerase from Clostridium cellulolyticum H10, J. Agric. Food Chem. 59 (2011), 7785-7792. https://doi.org/10.1021/jf201356q.

[17] D.X. Jia, C.Y. Sun, Y.T. Jin, et al. Properties of d-allulose 3-epimerase mined from Novibacillus thermophilus and its application to synthesis of d-allulose, Enzyme Microbial Technol. 148 (2021), 109816. https://doi.org/10.1016/j.enzmictec.2021.109816.

[18] G.D. Hannigan, D. Prihoda, A. Palicka, et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction, Nucleic Acids Resea. 47 (2019), e110-e110. https://doi.org/10.1093/nar/gkz654.

[19] A.H. Russell, A.W. Truman, Genome mining strategies for ribosomally synthesised and post-translationally modified peptides, Comput. Struct. Biotechnol. J. 18 (2020), 1838-1851. https://doi.org/10.1016/j.csbj.2020.06.032.

[20] M.H. Medema, T. de Rond, B.S. Moore, Mining genomes to illuminate the specialized chemistry of life, Nat. Rev. Genet. 22 (2021), 553-571. https://doi.org/10.1038/s41576-021-00363-7.

[21] J.P. Trinugroho, F. Asadi, B. Pardamean, Identifying cyanobacteria through next-generation sequencing

technology for modern agriculture, E3S Web Conf. 388 (2023), 01007. https://doi.org/10.1051/e3sconf/202338801007.

[22] M. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, Protein Sci. 15 (2006), 2507-2524. https://doi.org/10.1110/ps.062416606.

[23] R. Sánchez, A. Šali, Advances in comparative protein-structure modelling, Current Opinion Struct. Biol. 7 (1997), 206-214. https://doi.org/10.1016/s0959-440x(97)80027-9.

[24] B. Webb, A. Sali, Comparative protein structure modeling using MODELLER, Curr. Protoc. Bioinformatics. 2006 Oct; Chapter 5: Unit-5.6. https://doi.org/10.1002/cpbi.3.

[25] J. Chapman, A. Ismail, C. Dinu, Industrial applications of enzymes: recent advances, techniques, and outlooks, Catalysts. 8 (2018), 238. https://doi.org/10.3390/catal8060238.

[26] Z. Zhang, J. Zhao, C. Yu, et al. Chelatococcus composti sp. nov., isolated from penicillin fermentation fungi residue with pig manure co-compost, Int. J. Syst. Evol. Microbiol. 67 (2017), 565-569. https://doi.org/10.1099/ijsem.0.001651.

[27] F. Ningsih, D.G. Nurlaili, A. Yokota, et al. Isolation and molecular identification of thermophilic bacteria from litter of mount Galunggung hot spring, Tasikmalaya, Indonesia, J. Phys.: Conf. Ser. 1943 (2021), 012097. https://doi.org/10.1088/1742-6596/1943/1/012097.

[28] R. Nirwantono, J.P. Trinugroho, D. Sudigyo, et al. Genome mining of potential enzyme from Chelatococcus composti for sustainable production of D-Allulose, IOP Conf. Ser.: Earth Environ. Sci. 1169 (2023), 012083. https://doi.org/10.1088/1755-1315/1169/1/012083.

[29] L. Zhang, W. Mu, B. Jiang, et al. Characterization of d-tagatose-3-epimerase from Rhodobacter sphaeroides that converts d-fructose into d-psicose, Biotechnol. Lett. 31 (2009), 857-862. https://doi.org/10.1007/s10529-009-9942-3.

[30] A.D. Nagi, L. Regan, An inverse correlation between loop length and stability in a four-helix-bundle protein, Folding Design. 2 (1997), 67-75. https://doi.org/10.1016/s1359-0278(97)00007-2.

[31] L. Liu, H. Yu, K. Du, et al. Enhanced trypsin thermostability in Pichia pastoris through truncating the flexible region, Microbial Cell Fact. 17 (2018), 165. https://doi.org/10.1186/s12934-018-1012-x.