Journal Editorial Board

ISSN: 2165-3445 (Print)  ISSN: 2165-3453 (Online)
http://www.scirp.org/journal/cmb

Editor-in-Chief
Dr. Christo Z. Christov Northumbria University, UK

Editorial Board
Dr. David R. Bevan Virginia Polytechnic Institute and State University, USA
Dr. Nicolay Ivanov Dodoff Bulgarian Academy of Sciences, Bulgaria
Prof. Leif A. Eriksson University of Gothenburg, Sweden
Prof. Emilio Gallicchio Rutgers University, USA
Prof. Juraj Gregan University of Vienna, Austria
Dr. Ian S. Haworth University of Southern California, USA
Prof. Srividhya Jeyaraman Indiana University, USA
Prof. Cizhong Jiang Tongji University, China
Prof. Tatyana Karabencheva-Christova Northumbria University, UK
Dr. Daisuke Kihara Purdue University, USA
Prof. Jianyong Li Virginia Polytechnic Institute and State University, USA
Dr. Jose L. Medina-Franco Florida Atlantic University, USA
Dr. Sihua Peng Shanghai Ocean University, China
Dr. Olli T. Pentikäinen University of Jyväskylä, Finland
Prof. Giulio Rastelli University of Modena and Reggio Emilia, Italy
Prof. Igor A. Topol Frederick National Laboratory for Cancer Research, USA
Dr. Ivanka Tsakovska Institute of Biophysics and Biomedical Engineering, Bulgaria
Prof. Nagarajan Vaidehi City of Hope National Medical Center, USA
Dr. Jinhua Wang New York University Langone Medical Center, USA
Dr. Yanggan Wang Emory University, USA
Prof. Arieh Warshel University of Southern California, USA
Prof. Dongqing Wei Shanghai Jiao Tong University, China
Dr. Yasushige Yonezawa Kinki University, Japan
Table of Contents

Volume 6  Number 2  June 2016

In Silico Evaluation of Anti-Malarial Agents from Hoslundia opposita as Inhibitors of Plasmodium falciparum Lactate Dehydrogenase (PfLDH) Enzyme
D. M. Shadrack, S. S. Nyandoro, J. J. E. Munissi, E. B. Mubofu .......................................................... 23

Use of FFT in Protein Sequence Comparison under Their Binary Representations
J. Pal, S. Ghosh, B. Maji, D. K. Bhattacharya ...................................................................................... 33

The figure on the front cover is from the article published in Computational Molecular Bioscience, 2016, Vol. 6, No. 2, pp. 23-32 by Daniel M. Shadrack, Stephen S. Nyandoro, Joan J. E. Munissi and Egid B. Mubofu.
Computational Molecular Bioscience (CMB)

Journal Information

SUBSCRIPTIONS


Subscription rates:
Print: $79 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
E-mail: sub@scirp.org

COPYRIGHT

COPYRIGHT AND REUSE RIGHTS FOR THE FRONT MATTER OF THE JOURNAL:
Copyright © 2016 by Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

COPYRIGHT FOR INDIVIDUAL PAPERS OF THE JOURNAL:
Copyright © 2016 by author(s) and Scientific Research Publishing Inc.

REUSE RIGHTS FOR INDIVIDUAL PAPERS:
Note: At SCIRP authors can choose between CC BY and CC BY-NC. Please consult each paper for its reuse rights.

DISCLAIMER OF LIABILITY

Statements and opinions expressed in the articles and communications are those of the individual contributors and not the statements and opinion of Scientific Research Publishing, Inc. We assume no responsibility or liability for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained herein. We expressly disclaim any implied warranties of merchantability or fitness for a particular purpose. If expert assistance is required, the services of a competent professional person should be sought.

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: cmb@scirp.org
In Silico Evaluation of Anti-Malarial Agents from Hoslundia opposita as Inhibitors of Plasmodium falciparum Lactate Dehydrogenase (PfLDH) Enzyme

Daniel M. Shadrack1*, Stephen S. Nyandoro2, Joan J. E. Munissi2, Egid B. Mubofu2

1Chemistry Department, Faculty of Natural and Applied Sciences, St. John's University of Tanzania, Dodoma, Tanzania
2Chemistry Department, College of Natural and Applied Sciences, University of Dar es Salaam, Dar es Salaam, Tanzania

Received 9 March 2016; accepted 14 June 2016; published 17 June 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

Abstract

Malaria has continued to be a health and economic problem in Africa and the world at large. Many anti-malarial drugs have been rendered ineffective due to the emergence of resistant strains of Plasmodium falciparum. A key malaria parasite enzyme in glycolytic pathway, P. falciparum lactate dehydrogenase (PfLDH) is specially targeted for anti-malarial drugs development. Thus, the aim of this investigation was to determine the in silico inhibition effects of antimalarial compounds from Hoslundia opposita Vahl. namely hoslundin, hoslundal and hoslunddiol on PfLDH enzyme. The compounds were docked to the three-dimensional structure of PfLDH as enzyme using AutoDock Vina in PyRx virtual screening software. Binding affinity and position of the inhibitors were evaluated using PyMol software. The PfLDH enzyme showed two binding sites: the cofactors binding site (Site A) and secondary binding site (Site B). In the absence of the cofactor all ligands showed higher affinity than NADH, and were bound to the cofactors binding site (Site A). When docked in the presence of the cofactor, site B was the preferred binding site. Binding to cofactor site with higher binding energy than NADH suggests that these ligands could act as preferential competitive inhibitors of PfLDH. However, the binding to site B also suggests that they may be non-competitive allosteric inhibitors. Amino acid residues Gly99, Asn140, Phe100 and Thr97 were indicated to form hydrogen bonds with Hoslundin. Hoslunddiol showed hydrogen bonding with Thr97 and Met30, while Hoslundal formed hydrogen bond with Thr101 and Asn140.

*Corresponding author.

Keywords

PfLDH, Hoslundin, Hoslundal, Hoslunddiol, Anti-Malarial, Docking

1. Introduction

Malaria remains to be a health and economic problem in the tropical Africa and the world at large. In Africa malaria is responsible for over 430,000 children deaths every year [1]. In the year 2013, the World Health Organization (WHO) estimated 198 million cases of malaria leading to 584,000 deaths worldwide [1]. The disease mostly prevails in the poor countries of tropical and sub-tropical regions of Sub-Saharan Africa, America and Asia [2]-[7]. The Plasmodium species namely P. falciparum, P. vivax, P. malariae and P. ovale, and the lately reported species P. knowlesi are known to cause malaria [4] [5] [8] [9]. Of the five species P. falciparum is the most responsible for causing the disease [2] [3] [5] [7].

Over the recent past years, chloroquine was used to treat malaria [4] [8]. Artemisinin and its derivatives were later discovered and took over to treat the disease [3]. However, the occurrence of resistant strains of Plasmodium to chloroquine and artemisinin based drugs [3] [4] [8], has necessitated the search and development of new drug targets to combat the disease. Development of resistance to antimalarial drugs is associated to mutations in the parasite active site for the drug target [3] [4] [8]. Thus, identifying new drug targets with new mechanisms of action of the drug may help in fighting the disease [4] [8]. Today, many other natural products and synthetic anti-malaria agents have been designed to target different enzymes involved in parasitic life cycle [6] [10]-[15].

Enzymes of the glycolytic pathways are thought to be an important drug target due to parasitic dependence on glycolysis for energy production [3] [4]. Plasmodium falciparum Lactate Dehydrogenase (PfLDH) enzyme is involved in the final step of glycolysis and catalyzes the interconversion of pyruvate to lactate [3] [4]. The enzyme LDH is further involved in the formation of NAD+ which is required for the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase [4]. P. falciparum parasites depend on this enzyme for energy production needed for biochemical process, growth and development. Thus, this enzyme is thought to be an important drug target in malaria treatment. Inhibition of this enzyme activities results in death of the parasite. Thus, some anti-malarial drugs have been designed to target this enzyme. Thus, efforts to search for new potent anti-malarial drugs have been a constant ambition to scientists in this field of research.

In a search for potent anti-malarial compounds, Ngadjui and co-workers isolated anti-malarial compounds namely hoslundin, hoslundal and hoslunddiol from Hoslundia opposita (Figure 1) [16]. The compounds showed anti-malarial activity against multi-drug resistant K1 strain of P. falciparum, hoslundal being the most active [16]. An interesting anti-malarial activity of hoslundal suggested it to have a different mode of action from chloroquine and other anti-malaria drugs [16]. H. opposita is traditionally used in the treatment of malaria and other diseases in East and West Africa [16] [17]. Several research groups have also reported the anti-malarial activities of the compounds from H. opposita [17]-[21]. However, the in silico assessment of molecular interaction of these compounds with PfLDH enzyme remains unreported. In this work, a molecular docking study was carried on hoslundin, hoslundal and hoslunddiol to evaluate their potential inhibitory activity against PfLDH enzyme. The binding interactions were investigated and are reported herein.

2. In Silico Experimental Procedures

The three dimensional protein crystal structures of PfLDH (PDB ID: 1T2C) with resolution of 2.01 Å, (Figure 2(a)), [22] was obtained from the RCSB Protein Data Bank and used for docking studies. Docking experiments
Figure 2. Structure of PfLDH enzyme (PDB ID: IT2C) retrieved from the RCSB Protein Data bank as shown in (a); (b) shows the binding surface pockets present in PfLDH enzyme. Site A is the cofactor binding pocket as viewed in PyMol software; (c) shows the amino acids binding pocket in blue colour near the surface of the enzyme.

were performed following the procedure reported in the literature by [23] [24]. Briefly, Protein binding pockets sites were analysed by using of 3D Ligand Binding Site Predication Server. Possible proteins clashes and amino acid from the active sites were checked by PyMol software v 0.9 (DeLano Scientific LLC). The protein crystal structure was cleaned by using Arguslab software by deleting all heteroatoms, cofactor and water in the protein active sites. Hydrogen atom was added to the geometry. The crystal structure was converted to PDBQT, then re-fined and geometries optimized using AutoDock Vina in PyRx tools to generate atomic coordinates. Ligands used in this study are chemical structures of hoslundin, hoslundal and hoslunddiol which are anti-plasmodial compounds isolated *Hoslundia opposita* [16]. A list of ligands was generated and optimized using ChemDraw Ultra 12.0 and ChemDraw 3D Pro. The optimized geometry structures were served in .pdb file format. Ligands were then converted to PDBQT in PyRx. Finally, ligands were automatically docked to PfLDH enzymes using AutoDock Vina in PyRx (v 0.8) virtual screening tool with 8 ligand exhaustiveness. PyRx employs Lamarckian Geometric Algorithm (LGA) in docking processes; ligands were docked using flexible conformation. Ligands were re-docked three times in the active sites. Docking grid from autogrid with dimensions $25 \times 25 \times 25 \, \text{Å}$ size was used. A default vina search space grid dimensions of 26.09, 26.99 and 9.106 for x, y, z, coordinates was used, respectively. PyMol v 0.9 software (DeLano Scientific LLC,) was used to visualize the binding sites and orientations of the complex.
3. Results and Discussion

3.1. Binding Pockets Analysis

The PfLDH enzyme possesses two important binding pockets, the cofactors binding pocket (Site A) and Site B ([Figure 2(b)] [4]). Site A, identified as NADH binding pocket comprised of amino acid residues: Gly29, Met30, Ile31, Phe52, Asp53, Ile54, Tyr85, Thr97, Ala98, Thr101, Val138, Thr139, Asn140, Val142, Leu163, Leu167 and Pro250 (Appendix). Site B comprised of amino acid residues: Asp230, Lys198, Val233, Lys314, Glu317, Asp230, Leu201, Glu226, Phe229, Val200, Leu237, and Asn241. These amino acid residues were near the surface end of enzyme ([Figure 2(c)] [4]). The presence of NADH cofactor affected the distribution and conformation of the docked ligands. It was observed that, in the presence of the cofactors all ligands bound to site B, while in the absence of the cofactor all docked ligands showed a stable conformation at site A.

3.2. Molecular Docking

All ligands were successfully docked to binding sites of PfLDH enzyme. All ligands showed better docking score with stable conformation than the crystalized NADH cofactor when docked in the absence of the cofactor (in the cofactor binding site). In the absence of cofactor, all ligands bound to site A which is the binding site of the cofactor. When ligands were docked in the presence of cofactor, site B was the preferred binding site for all ligands and possessed lower docking scores (Table 1).

In the present study, the anti-malarial compound hoslunddiol from Hoslundia opposita had a good docking score (−8.0 kcal/mol) higher than all the other ligands to PfLDH enzyme. The best docking score was obtained in the absence of the cofactor (Table 1). When re-docked in the absence of the cofactor, similar binding affinities were obtained (Table 2). The interaction of hoslunddiol with amino acid residues of PfLDH involved four hydrogen bonds which were formed as follows: The first hydrogen bond was formed by the oxygen atom of the ligand ([Figure 3(d)]). The second hydrogen bond was formed by the hydroxyl group bonding with -NH2 of Met30 (HO----Met30, 2.38 Å). The third and fourth hydrogen bonds involved oxygen of methoxy group (MeO ----Thr97, 3.22 Å), carbonyl group (C=O----Thr97, 3.54 Å) and oxygen atom of the pyran ring (-O----Thr97, 3.30 Å). [Figure 3(e)] shows the binding of hoslunddiol in the PfLDH active site A. The docking results showed hoslundin to have the second docking score rank with −7.8 kcal/mol. Docking of hoslundin involved four hydrogen bonds between PfLDH and oxygen of the ligand ([Figure 3(b) and Figure 3(c)]). The first hydrogen bond formed between C=O of Gly99 (C=O---- Gly99, 3.06 Å) while the second hydrogen bond was formed between -NH2 of Asn140 (C=O----Asn140, 3.05 Å). The hydroxyl (OH) group in the ligand formed the third hydrogen bond with -NH2 group of Phe100 (HO----Phe100, 3.02 Å) whereas the fourth bond was formed between amine group of Thr97 with the carbonyl group in the ligand (C=O----Thr97, 2.35 Å). Molecular docking studies of hoslundal showed only two hydrogen bonds, which involved Thr101 (3.08 Å) and Asn140 (3.26 Å) ([Figure 3(a)]). Other amino acid residues which interacted with hoslundal in the PfLDH binding pocket were; Gly99, Thr97, Ala98, Gly27, Lys102, Ala103, Phe101, Met30, Ile31, Gly29, Gly32, Pro246, Asn116, Thr139, Ser28, Met36, Tyr247, Gly33, Ala251, Asp110, Leu112, Ser245 and Val248.

Docked ligands showed higher binding affinity than NADH in absence of the cofactor, implying a possible competitive inhibition. However, on closer examination, the interaction of NADH and PfLDH indicated many more hydrogen bonds than the ligands (Figure 4). Concomitantly, other studies have reported many hydrogen bonds formed between NADH and PfLDH, regardless of the lower binding energy obtained for NADH in the

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Absence of cofactor (NADH)</th>
<th>Presence of cofactor (NADH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoslundal</td>
<td>−7.3</td>
<td>−5.9</td>
</tr>
<tr>
<td>Hoslundin</td>
<td>−7.8</td>
<td>−6.1</td>
</tr>
<tr>
<td>Hoslunddiol</td>
<td>−8.0</td>
<td>−6.4</td>
</tr>
<tr>
<td>NADH</td>
<td>−6.8</td>
<td>−6.8</td>
</tr>
</tbody>
</table>

Table 1. Best binding energy (kcal/mol) of the favourable conformation based on PyRx
Figure 3. (a) shows the interactions of hoslundal with PfLDH enzyme, two hydrogen bonds were involved, the H-bond formed involved Thr101 (3.08 Å) and Asn140 (3.26 Å); (b) shows the interaction of hoslundin with PfLDH enzyme, four hydrogen bonds were involved (indicated by yellow dotted lines). The amino residues forming hydrogen bond with hoslundin were: Thr97 (2.35 Å), Asn140 (3.05 Å), Phe100 (3.02 Å) and Gly100 (3.06 Å); (c) shows the binding position of hoslundin in the cavity of PfLDH in the absence of cofactor; (d) indicates the PfLDH-hoslunddiol complex. The yellow dotted lines indicates the hydrogen bonds formed between Met30 (2.38 Å), Thr97 (3.22 Å), Thr97 (3.54 Å) and Thr97 (3.30 Å); (e) shows the binding conformation of hoslunddiol in the cavity of PfLDH at site A.
Figure 4. Interaction of NADH to PfLDH (left) and binding position in the pocket (right).

Table 2. Best binding affinity on re-docking experiments

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Docking run (kcal/mol)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Mean ± SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoslunndiol</td>
<td>−8.0</td>
<td>−8.1</td>
<td>−8.1</td>
<td>−8.1</td>
<td>−8.06 ± 0.03</td>
</tr>
<tr>
<td>Hoslunndin</td>
<td>−7.8</td>
<td>−7.8</td>
<td>−7.8</td>
<td>−7.8</td>
<td>−7.8 ± 0.0</td>
</tr>
<tr>
<td>Hoslunndal</td>
<td>−7.3</td>
<td>−7.0</td>
<td>−7.0</td>
<td>−7.0</td>
<td>−7.1 ± 0.1</td>
</tr>
</tbody>
</table>

PfLDH enzyme active site [3]. The presence of a high number of hydrogen bonds may reduce the binding affinity of other ligands in the cofactor binding sites. Hydrogen bonds stabilize the ligand interaction in the active site [4] [25]. Docking of ligands in the presence of NADH reduced the affinity and frequency of binding to the cofactor active site. Such phenomenon has also been reported when ligands were docked in the presence of NADH [4], the frequency and affinity of binding to the cofactors binding site was found to be low. In the present study, all ligands showed lower binding affinity than NADH when they were docked in the presence of the NADH (Table 1). However, in the absence of NADH all ligands showed higher docking score than NADH. This suggests competitive inhibition in the NADH cofactor binding site, thus, inhibiting PfLDH enzyme from producing energy [3]. Binding of ligands to site B may suggest allosteric inhibition [4]. The amino acid residues forming H-bonds and the bond length in the active site with hoslunndal, hoslunndin, hoslunndioll and NADH are presented in Table 3.

4. Conclusion

*Plasmodium falciparum* is an important malaria parasite in Africa and other parts of the world. The glycolytic PfLDH enzyme is thought to be one of the important anti-malarial drug targets due to the parasite dependence on glycolysis for energy production. The present study investigated the *in silico* inhibition effects of the three *Hoslundia opposita* compounds (hoslunndiol, hoslunndin and hoslunndal) to PfLDH enzyme. Results indicated that hoslunndiol (−8.0 kcal/mol) and hoslunndin (−7.8 kcal/mol) had better interaction with the target. Hoslunndal showed less binding energy (−7.3 kcal/mol) and had few interactions. The former two compounds showed four hydrogen bonds while the latter showed only two hydrogen bonds. Since hoslunndal was reported to be the most active anti-malarial compound against the multi-drug resistant K1 strain yet weakly active against the chloroquine sensitive strain NF54 *in vitro*, the present findings further affirm that hoslunndal possesses antimalarial activities with probable different modes of action. Within the limitations of the docking procedure, these findings suggest that the studied compounds could act as competitive inhibitors as they had higher binding energy than the cofactor. The pharmacokinetics and *in vivo* studies of these promising anti-malarial compounds are recommended. Furthermore, studies on the interaction and mode of action with other *Plasmodium falciparum* enzymes...
Table 3. Amino acid residues forming H-bonds with their bond length in PfLDH active site

<table>
<thead>
<tr>
<th>H-bond forming residues</th>
<th>Bond length (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoslundal</td>
<td></td>
</tr>
<tr>
<td>Thr101</td>
<td>3.08</td>
</tr>
<tr>
<td>Asn140</td>
<td>3.26</td>
</tr>
<tr>
<td>Thr97</td>
<td>2.35</td>
</tr>
<tr>
<td>Phe100</td>
<td>3.02</td>
</tr>
<tr>
<td>Hoslundin</td>
<td></td>
</tr>
<tr>
<td>Thr97</td>
<td>2.35</td>
</tr>
<tr>
<td>Phe100</td>
<td>3.02</td>
</tr>
<tr>
<td>Asn140</td>
<td>3.05</td>
</tr>
<tr>
<td>Gly99</td>
<td>3.06</td>
</tr>
<tr>
<td>Met30</td>
<td>2.38</td>
</tr>
<tr>
<td>Hoslunddiol</td>
<td></td>
</tr>
<tr>
<td>Thr97</td>
<td>3.22</td>
</tr>
<tr>
<td>Thr97</td>
<td>3.54</td>
</tr>
<tr>
<td>Thr101</td>
<td>3.29</td>
</tr>
<tr>
<td>Gly164</td>
<td>2.99</td>
</tr>
<tr>
<td>NADH</td>
<td></td>
</tr>
<tr>
<td>(2) Thr101</td>
<td>3.03 and 3.18</td>
</tr>
<tr>
<td>(2) Thr97</td>
<td></td>
</tr>
<tr>
<td>(2) His195</td>
<td>2.81 and 3.38</td>
</tr>
<tr>
<td>(2) Ile31</td>
<td>3.12</td>
</tr>
<tr>
<td>Ala236</td>
<td>2.8</td>
</tr>
<tr>
<td>Val233</td>
<td>2.27</td>
</tr>
<tr>
<td>Asn234</td>
<td>2.81</td>
</tr>
</tbody>
</table>

and any other life cycle target points are recommended for the studied and other anti-plasmodial compounds from *H. opposita*.

**Acknowledgements**

DMS express his sincerely thanks to the NAPRECA (T) for sponsorship and facilitation on *In Silico* studies during the 16th NAPRECA Pre-Symposium workshop. JJEM is grateful to the International Foundation for Sciences (IFS) for financial support.

**References**

D. M. Shadrack et al.


Appendix

Predicated binding active site in the *P*/*LDH* enzyme by 3D ligand site prediction server.
Selected letter amino acid code sequence of PfLDH in docking server.

**Single letter amino acid code sequence**

<table>
<thead>
<tr>
<th>Selected amino acids:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A P K A K I V L V G S G M I G G V M A T -26</td>
</tr>
<tr>
<td>L I V Q K N L G D V V L F D I V K N M P -56</td>
</tr>
<tr>
<td>H G K A L D T S H T N V M A S N C K V S -76</td>
</tr>
<tr>
<td>G S N T Y D D L A G A D V V I V T A G F -100</td>
</tr>
<tr>
<td>T K A D E W N R D D L L P L N N K I M I -121</td>
</tr>
<tr>
<td>E I G G H I K K N C P A F I I V V T N P -141</td>
</tr>
<tr>
<td>V D V M V Q L L H Q H S G V P K N K I I -101</td>
</tr>
<tr>
<td>G L G G V L D T S R L K Y Y I S Q K L N -191</td>
</tr>
<tr>
<td>V C P R D V N A H I V G A H G N K M V L -201</td>
</tr>
<tr>
<td>L K R Y I T V G L E F I N N K L I S D A -222</td>
</tr>
<tr>
<td>E L E A I F D R T V N A L E I V N L H -243</td>
</tr>
<tr>
<td>A S P Y V A P A A A I I E M A E S Y L K -293</td>
</tr>
<tr>
<td>D L K K V L I C S T L L E G Q Y G H S D -283</td>
</tr>
<tr>
<td>I F G G T P V L G A N G V E Q V I E L -303</td>
</tr>
<tr>
<td>Q L N S E E K A K F D E A I A E T K R M -325</td>
</tr>
<tr>
<td>K A L A</td>
</tr>
</tbody>
</table>
Use of FFT in Protein Sequence Comparison under Their Binary Representations

Jayanta Pal¹, Soumen Ghosh², Bansibadan Maji³, Dilip Kumar Bhattacharya⁴

¹Department of Computer Science & Engineering, Narula Institute of Technology, Kolkata, India
²Department of Information Technology, Narula Institute of Technology, Kolkata, India
³Department of Electronics & Communication Engineering, National Institute of Technology, Durgapur, India
⁴Department of Pure Mathematics, Calcutta University, Kolkata, India

Email: jayantapal1978@yahoo.com

Received 14 February 2016; accepted 27 June 2016; published 30 June 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

Abstract

The paper considers Voss type representation of amino acids and uses FFT on the represented binary sequences to get the spectrum in the frequency domain. Based on the analysis of this spectrum by using the method of inter coefficient difference (ICD), it compares protein sequences of ND5 and ND6 category. Results obtained agree with the standard ones. The purpose of the paper is to extend the ICD method of comparison of DNA sequences to comparison of protein sequences. The topic of discussion is to develop a novel method of comparing protein sequences. The main achievements of the work are that the method applied is completely new of its kind, so far as protein sequence comparison is concerned and moreover the results of comparison agree with the previous results obtained by other methods for the same category of protein sequences.

Keywords

Voss Type Representation, Inter-Coefficient Difference (ICD) Method, Distance Matrix, Phylogenetic Tree, Fast Fourier Transform (FFT), ND5 and ND6 Category of Protein

1. Introduction

Among the numerous available amino acids only 20 are generally found in living beings and every protein sequence is expressed by these 20 amino acids. The representation of protein in terms of its amino acids is called its primary sequence. Based on this primary sequence representation, protein sequence comparison involves basically two types of methods: 1) Alignment Based Method and 2) Alignment Free Method. Protein sequence comparison was primarily done by different alignment based methods [1]-[3]. But especially due to execution
time and comparatively difficult procedure, alignment free methods were preferred subsequently. So far as alignment free methods are concerned, a good literature up to 2003 is available in [4]. So we start with highlighting some of the most important contribution in protein sequence comparisons by alignment free methods from 2003 onwards [5]-[25]. Obviously in most cases, protein sequence comparison also follows similar approach as is considered in genome sequence analysis, because the role of four nucleotides is the same as the role of 20 amino acids in a protein sequence. In details, first of all, numerical representations of the protein sequences are obtained from the numerical values given to the individual amino acids, then graphical representation of the protein sequences is obtained; from these graphs descriptors are derived. These are finally used in comparing protein sequences. All the papers from [7]-[24] involve graphical representations. But another completely different approach is also followed in protein sequence comparison. These are based on classification of amino acids in different groups with different cardinality [8] [26] [27]. Again application of Discrete Fourier Transform in Bioinformatics is also well known. Discrete Fourier Transform (DFT) is nicely used in signal and image processing [28]-[34]. The main areas of its application in DNA research are found in gene prediction, hierarchical analysis and such others [35]-[40]. It is effectively used in identification of protein coding regions, because a DFT spectrum of a DNA sequence reflects the distribution and periodic pattern of the sequence [41]. Use of DFT on binary sequence is found in [42], where the binary sequence is generated from genome sequences by Voss type of representation. Naturally to find similar use of DFT in protein sequence analysis, corresponding Voss type representation of amino acids is to be known priori. Fortunately Voss representation of DNA sequences involving 4 nucleotides has already been generalized to Voss type representation of 20 amino acids in protein sequences [43]. Such representation of amino acids has already been used in obtaining fuzzy representation of amino acids [43]. These are found to be effective in classification of amino acids in 6 different groups. Finally protein sequence classification has been obtained based on such classified groups of amino acids [8] [25]-[27] [44]. Thus Voss type representation of amino acids is an important contribution in protein sequence analysis. But use of FFT on the binary representations of protein sequences generated by such Voss type representation of amino acids has not yet been attempted in protein sequence comparison. This is the motivation of the paper to consider such binary sequences in comparing protein sequences.

2. Methodology

2.1. Voss Type Binary Representation of Amino Acid

20 amino acids are taken in the following order:

- Alanine (A), Cysteine (C), Aspartic acid (D), Glutamic acid (E), Phenylalanine (F), Glycine (G), Histidine (H), Isoleucine (I), Lysine (K), Leucine (L), Methionine (M), Asparagine (N), Proline (P), Glutamine (Q), Arginine (R), Serine (S), Tyrosine (T), Valine (V), Tryptophan (W) and Threonine (Y).

Each amino acid is represented by a 20 component vector of which one bit is 1 and others are 0. But the representation follows the order of amino acid taken. For example amino acid Alanine(a) is represented by 10000000000000000000. The same rule is applied for other amino acids also, so that the last amino acid Threonine (Y) is represented by 000000000000000001.

From each protein sequence S we get 20 different representations corresponding to 20 different amino acids by putting in the protein sequence 1 for the particular amino acid considered and the rest all 0 for the remaining amino acids. Thus 20 different binary representations viz., U_A, U_C, U_D, U_E, U_F, U_G, U_H, U_I, U_K, U_L, U_M, U_N, U_P, U_Q, U_R, U_S, U_T, U_V, U_W and U_Y are obtained.

2.2. ICD Method for Protein Sequence Analysis

The ICD method of DNA sequence and Protein sequence analysis basically remains the same as both deals with binary sequence only. So we describe ICD method as described in [43]. First of all FFT is applied on the binary represented protein sequences of length N say. In the Fourier spectrum the amplitudes are taken, which are N/2 distinct numbers. We normalize these N/2 components by their lengths. On these N/2 normalized components, we take absolute value of the inter coefficient difference (ICD) by calculating the differences of the succeeding terms from the preceding ones. Thus we get (N/2 – 1) distinct elements corresponding to each amino acid. Now 20 such (N/2) − 1 distinct components are concatenated to give a descriptor of length 20*((N/2) − 1)). From such descriptors distance matrix is formed by considering Euclidian Distance measures as follows.
If \( x = (x_1, x_2, \cdots, x_n) \) and \( y = (y_1, y_2, \cdots, y_n) \) are two sequences for two proteins X and Y, then the distance between X and Y is given by

\[
d(x, y) = \left( \sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}
\]

This is the Euclidean distance between X and Y. The smaller is the distance; more similar are the protein sequences. On the basis of this formula the distances between pair of proteins are calculated and they are used to form the diagonal matrix. Due to similarity, only the lower half of the matrix is taken. Now using the UPGMA software on this matrix the Phylogenetic Tree for all the species is obtained. For comparison of protein sequences of different lengths the question of making all the lengths same does not arise normally in FFT. But if necessary, the length may be manually adjusted by putting additional zeros. For example, suppose two protein sequences are of lengths M and N. Then the descriptors for the first and second sequences are of lengths \( 20^*((N/2) - 1) \) and \( 20^*((M/2) - 1) \) respectively. As the descriptors are of unequal lengths, so comparison becomes infeasible. Hence if \( M = N - 2 \), say, then we first make the lengths of both the sequences equal to \( N \), by putting two additional zeros to the second sequence. But there is no problem in doing so, as the Fourier transform of zeros gives zero spectrum.

3. Sequences for Comparison

We have used the NADH dehydrogenase subunit 5 (ND5) and subunit 6 (ND6) protein sequences of nine species for comparison as shown in Table 1.

4. Results and Discussions

4.1. Results

Distance matrix obtained by applying our method for 9 protein sequences of ND5 and ND6 category have been presented in Table 2 and Table 3 respectively. Phylogenetic tree obtained from these data have been presented in Figure 1 and Figure 2 for ND5 and ND6 category respectively.

4.2. Discussion

- ICD method, which is dependent on Voss type representation of DNA sequences, is already known to be very much successful in comparing DNA sequences. Voss type representation for protein sequences is comparatively a newer concept. As Voss type representation for protein sequences has been applied recently in different areas and found to be very much successful there, so it is expected that this type of representation might be useful in protein sequence comparison also. This is why; in our paper ICD method based on Voss type representation for protein sequences has been developed and used for protein sequence comparison. No doubt that the present method is a new contribution to the literature of protein sequence comparison.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Species</th>
<th>ND5 NCBI Reference</th>
<th>Length</th>
<th>ND6 NCBI Reference</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HUMAN</td>
<td>AP-000649.1</td>
<td>603</td>
<td>AP-000650.1</td>
<td>174</td>
</tr>
<tr>
<td>2</td>
<td>GORILLA</td>
<td>NP-008222.1</td>
<td>603</td>
<td>NP-008223.1</td>
<td>174</td>
</tr>
<tr>
<td>3</td>
<td>COMMON CHIMPANZEE</td>
<td>NP-008196.1</td>
<td>603</td>
<td>NP-008197.1</td>
<td>174</td>
</tr>
<tr>
<td>4</td>
<td>PYGMY CHIMPANZEE</td>
<td>NP-008209.1</td>
<td>603</td>
<td>NP-008210.1</td>
<td>174</td>
</tr>
<tr>
<td>5</td>
<td>FIN WHALE</td>
<td>NP-006899.1</td>
<td>606</td>
<td>NP-006900.1</td>
<td>175</td>
</tr>
<tr>
<td>6</td>
<td>BLUE WHALE</td>
<td>NP-007066.1</td>
<td>606</td>
<td>NP-007067.1</td>
<td>175</td>
</tr>
<tr>
<td>7</td>
<td>RAT</td>
<td>AP-004902.1</td>
<td>610</td>
<td>AP-004903.1</td>
<td>172</td>
</tr>
<tr>
<td>8</td>
<td>MOUSE</td>
<td>NP-904338.1</td>
<td>607</td>
<td>NP-904339.1</td>
<td>172</td>
</tr>
<tr>
<td>9</td>
<td>OPOSSUM</td>
<td>NP-007105.1</td>
<td>602</td>
<td>NP-007106.1</td>
<td>168</td>
</tr>
</tbody>
</table>
Figure 1. Phylogenetic tree obtained for 9 protein sequences of ND5 category.

Figure 2. Phylogenetic tree obtained for 9 protein sequences of ND6 category.

Table 2. Distance matrix (lower triangular) for 9 protein sequences of ND5 category.

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Gorilla</th>
<th>P. Chim</th>
<th>C. Chim</th>
<th>Rat</th>
<th>Mouse</th>
<th>B_Whale</th>
<th>F_Whale</th>
<th>Opossum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gorilla</td>
<td>0.88011</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. Chim</td>
<td>0.78392</td>
<td>0.83493</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Chim</td>
<td>0.80377</td>
<td>0.86392</td>
<td>0.70748</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>1.33343</td>
<td>1.32515</td>
<td>1.32102</td>
<td>1.33353</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>1.31723</td>
<td>1.3021</td>
<td>1.29413</td>
<td>1.31185</td>
<td>1.16305</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B_Whale</td>
<td>1.28719</td>
<td>1.28518</td>
<td>1.27026</td>
<td>1.29161</td>
<td>1.32947</td>
<td>1.3423</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_Whale</td>
<td>1.27388</td>
<td>1.2825</td>
<td>1.26707</td>
<td>1.29284</td>
<td>1.31818</td>
<td>1.32474</td>
<td>0.73081</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Opossum</td>
<td>1.39053</td>
<td>1.38974</td>
<td>1.37859</td>
<td>1.3812</td>
<td>1.39501</td>
<td>1.3896</td>
<td>1.4018</td>
<td>1.40041</td>
<td>0</td>
</tr>
</tbody>
</table>
• Obviously ICD method, may be for DNA sequence comparison or Protein sequence comparison, is comparatively easier and straight forward to apply.
• To compare our results with those obtained earlier by other methods on the same species, we first mention them as far as possible. The phylogenetic tree obtained in [25] for 9 species of ND5 category is given in Figure 3. Similarly the phylogenetic trees obtained in [26] for 9 species of ND5 category and ND6 category are given in Figure 4 and Figure 5 respectively and the phylogenetic trees obtained in [44] for 9 species of ND5 category and ND6 category are given in Figure 6 and Figure 7 respectively.

![Figure 3. Phylogenetic tree obtained in [25] for 9 species of ND5 category.](image)

![Figure 4. Phylogenetic tree obtained in [26] for 9 species of ND5 category.](image)

**Table 3. Distance matrix (lower triangular) for 9 protein sequences of ND6 category.**

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Gorilla</th>
<th>P. Chim</th>
<th>C. Chim</th>
<th>Rat</th>
<th>Mouse</th>
<th>B_Whale</th>
<th>F_Whale</th>
<th>Opossum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>0</td>
<td>0.61527</td>
<td>0.66602</td>
<td>0.65876</td>
<td>1.40056</td>
<td>1.44023</td>
<td>1.29625</td>
<td>1.28281</td>
<td>1.48984</td>
</tr>
<tr>
<td>Gorilla</td>
<td>0.61527</td>
<td>0</td>
<td>0.60947</td>
<td>0.45713</td>
<td>1.3828</td>
<td>1.38812</td>
<td>1.28369</td>
<td>1.28498</td>
<td>1.46302</td>
</tr>
<tr>
<td>P. Chim</td>
<td>0.66602</td>
<td>0.60947</td>
<td>0</td>
<td>1.37694</td>
<td>1.38812</td>
<td>1.44952</td>
<td>1.28337</td>
<td>1.35032</td>
<td>1.4753</td>
</tr>
<tr>
<td>C. Chim</td>
<td>0.65876</td>
<td>0.45713</td>
<td>1.37694</td>
<td>0</td>
<td>1.10948</td>
<td>1.35476</td>
<td>1.28337</td>
<td>1.36975</td>
<td>1.47082</td>
</tr>
<tr>
<td>Rat</td>
<td>1.40056</td>
<td>1.3828</td>
<td>1.37694</td>
<td>1.10948</td>
<td>0</td>
<td>1.35032</td>
<td>1.28337</td>
<td>0.86639</td>
<td>1.44397</td>
</tr>
<tr>
<td>Mouse</td>
<td>1.44023</td>
<td>1.38812</td>
<td>1.44952</td>
<td>1.35476</td>
<td>1.35032</td>
<td>0</td>
<td>1.37165</td>
<td>0.86639</td>
<td>1.43112</td>
</tr>
<tr>
<td>B_Whale</td>
<td>1.29625</td>
<td>1.28369</td>
<td>1.28337</td>
<td>1.35476</td>
<td>1.35032</td>
<td>1.37165</td>
<td>0</td>
<td>1.36975</td>
<td>1.46181</td>
</tr>
<tr>
<td>F_Whale</td>
<td>1.28281</td>
<td>1.28369</td>
<td>1.28337</td>
<td>1.35476</td>
<td>1.35032</td>
<td>0.86639</td>
<td>1.36975</td>
<td>0</td>
<td>1.44397</td>
</tr>
<tr>
<td>Opossum</td>
<td>1.48984</td>
<td>1.46302</td>
<td>1.4753</td>
<td>1.47082</td>
<td>1.37682</td>
<td>1.43112</td>
<td>1.46181</td>
<td>1.44397</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 5. Phylogenetic tree obtained in [26] for 9 species of ND6 category.

Figure 6. Phylogenetic tree obtained in [44] for 9 species of ND5 category.

Figure 7. Phylogenetic tree obtained in [44] for 9 species of ND6 category.
From the above phylogenetic trees obtained for ND5 and ND6 categories of protein, it is revealed that in both the cases the phylogenetic trees obtained by our method almost agree with the earlier phylogenetic trees obtained by other methods.

5. Conclusion

Our method is effective and easier to apply in protein sequence comparison.

References


Computational Molecular Bioscience (CMB) is an international journal dedicated to the latest advancement of Computational Molecular Bioscience. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of Computational Molecular Bioscience. All manuscripts must be prepared in English, and are subject to a rigorous and fair peer-review process. Accepted papers will immediately appear online followed by printed hard copy. The journal publishes original papers including but not limited to the following fields:

◆ Ab Initio and Density Functional Calculations of Biomolecules
◆ Atomistic and Coarse Grained Molecular Dynamics
◆ Combined Computational and Experimental Studies of Biomolecular Interactions
◆ Combined Quantum Mechanical and Molecular Mechanical Methods (QM/MM)
◆ Computational Chemistry of Biomolecules, Ligands and Drugs
◆ Computational Drug Design: Structure-Based; Ligand-Based; Rational; De Novo
◆ Computational Modelling of Biomolecular Structures Interactions and Processes
◆ Computational Systems Biology and Chemistry
◆ Development and Applications of Monte Carlo Methods
◆ Development and Design of New Biological and Chemical Databases and Data Mining Techniques
◆ Development, Testing and Applications to Biomolecular Systems
◆ Enzymatic Reaction Mechanisms and Inhibition
◆ High Performance Computing in Molecular and Biomolecular Sciences
◆ Ligand Binding and Free Energy Calculations
◆ Modelling of Membrane Processes and Protein-Membrane Interactions
◆ Modelling Protein Structure, Conformational Dynamics an Interactions
◆ Molecular Mechanics, Force Field Development and Evaluation
◆ Molecular Visualizations and Data Analysis
◆ Multilevel Computational Simulations
◆ Nucleic Acids Structure, Dynamics and Interactions with Ligands
◆ Protein Folding
◆ Protein Ligand Docking New Algorithm, Codes and Applications
◆ Protein-Nucleic Acids Interactions
◆ Quantitative Structure-Activity Relationships (QSAR)
◆ Semiempirical Electronic Structure Calculations
◆ Structural Bioinformatics and Homology Modelling

We are also interested in: 1) Short Reports—2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book Reviews—Comments and critiques.

Website and E-Mail
http://www.scirp.org/journal/cmb E-mail: cmb@scirp.org
What is SCIRP?
Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

What is Open Access?
All original research papers published by SCIRP are made freely and permanently accessible online immediately upon publication. To be able to provide open access journals, SCIRP defrays operation costs from authors and subscription charges only for its printed version. Open access publishing allows an immediate, worldwide, barrier-free, open access to the full text of research papers, which is in the best interests of the scientific community.

- High visibility for maximum global exposure with open access publishing model
- Rigorous peer review of research papers
- Prompt faster publication with less cost
- Guaranteed targeted, multidisciplinary audience

Website: http://www.scirp.org
Subscription: sub@scirp.org
Advertisement: service@scirp.org