Computational study of Cu$^{2+}$, Fe$^{2+}$, Fe$^{3+}$, Mn$^{2+}$ and Mn$^{3+}$ binding sites identification on HSA 4K2C

Syahputra Wibowo*1, Sutiman B Sumitro1, Sri Widyarti1

1Biology Dept., Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia

*Corresponding authors: wibowo@student.ub.ac.id

Abstract. This research aims to computationally characterize HSA 4K2C protein and describe as well as its ability to bind transition metal ions. Data mining is performed to obtain HSA 4K2C from PDB and transition metal ions such as Cu$^{2+}$ (ID: 27099), Fe$^{2+}$ (ID: 27284), Mn$^{2+}$ (ID: 27854), Mn$^{3+}$ (ID: 105130) and Fe$^{3+}$ (ID: 29936) from PubChem. The analysis consists of ProtParam, Motif Search, CFSSP, DLP-SVM, and docking. Docking used PyRx Autodock Vina. Analysis of receptor-ligand interactions used DS 2016. The results of the ProtParam analysis provide some information on HSA 4K2C, which is it has 585 amino acids with an isoelectric point (pI) of 5.67, an index of protein instability of 38.85, then total amino acids (aa) residues of negatively charged (Asp + Glu) are 98 while the positively charged ones (Arg + Lys) are 83. Motive Search shows that there are three HSA motifs namely motif 1 (aa. 551-575), motif 2 (aa. 353-377), and motif 3 (aa. 61-185). CFSP shows α-helix structure is the dominant structure compared to β-sheets, turn and coil in 4K2C. DLP-SVM shows two domain linkers where DL-1 (aa. 410-451) and DL-2 (aa. 96-122). Docking shows the ability of HSA 4K2C in binding metal ions such as Cu$^{2+}$, Fe$^{2+}$, Mn$^{2+}$, Mn$^{3+}$ and Fe$^{3+}$. 

1. Introduction
Scavenger can be found both endogenous and exogenous [1]. An example of endogenous scavenger is albumin (human serum albumin) which has the largest antioxidant capacity in human serum working well by sequestering transition metal ions or directly by binding and carrying scavenger [2]. Serum albumin is a very abundant protein in human blood [3]. Albumin has a molecular weight of 66.5 kDa which is a single chain with 585 amino acids. Albumin consists of three homologous structures with 67% of helix, each domain (I, II, and III) containing two subdomains i.e. A and B [4]. The ability of albumin as a scavenger is the ability of metal ions binding, which is known before that heme has pro-oxidant properties and albumin is an effective protein in the binding of heme. When it binds to albumin, the pro-oxidant is decreased which means albumin has the function as an antioxidant [5].

Ligands binding such as metal ions on a protein is capable of causing structural changes. The albumin structure can be used as a diagnosis of disease. In diabetes mellitus, it has been found that albumin in the form of HNA (human nonmercaptalbumin) increased while HMA (human mercaptalbumin) decreased. Conversion of the form of albumin affects the ability of albumin as scavenger [6]. Another example is cancer where there are differences in the structure of albumin in patients with cancer, compare with healthy people caused by the bonding of albumin that be the reason for allosteric modification in albumin [7]. The ability of albumin as a scavenger against various metal ions can be done not just with in vivo analysis but also in silico. Research using domain analysis,
motive, and metal ion docking in HSA 4K2C has never been done so that it needs to be done in order to give an overview of the ability of albumin as a scavenger with a focus on metal ions sequestering.

2. Methods

2.1. Data Mining
Human Serum Albumin obtained from PDB database (Protein Data Bank) with ID 4K2C as HSA Ligand Free. Next datamining of metal ions obtained from PubChem such as Cu$^{2+}$ ion (ID: 27099), Fe$^{2+}$ ion (ID: 27284), Mn$^{2+}$ ion (ID: 27854), Mn$^{3+}$ ion (ID: 105130), and Fe$^{3+}$ ion (ID: 29936) with molecular weight 63.546 gr/mol, 55.84 gr/mol, 54.938 gr/mol, 54.938 g/mol, 55.845 gr/mol, respectively. Data from PDB database is a PDB file that is a 3D conformation and a sequence file (.fasta). As for data from Pubchem in the form of 3D conformer in SDF format. Data with formats (.pdb) and (.sdf) are used in docking and RAMPAGE analysis while (.fasta) format is used for ProtParam analysis, MOTIF Search, DLP-SVM, and CFSSP.

2.2. ProtParam and Motif Search
ProtParam is used in conducting computation of physicochemical properties analyzed from the sequence of proteins that we input. Sequence in Fasta or ID format and SwissProt/TrEMBL accession number. ProtParam is one of the tools integrated with SwissProt (web.expasy.org). The parameters are input in the form of molecular weight, theoretical pI, amino acid composition, atomic composition, and instability index [8]. The Motif Search uses PROSITE as part of its system integration. The Motif Search is one of the tools provided by GenomeNet (genome.jp).

PROSITE is a method for determining the function of a protein sequence. The techniques used include alignment to get the pattern, motif, or fingerprint of the input proteins. PROSITE database consists of two ASCII files where the first one is (PROSITE. DAT) which is a file containing all the pattern and matrix information. The second file is (PROSITE. DOC) which contains the text information of each profile and pattern [9].

2.3. CFSSP and DLP-SVM
CFSSP (biogem.org/tool/chou-fasman) stands for (Chou & Fasman Secondary Structure Prediction Server) which is a dedicated Server to predict the secondary structure of the protein amino acid sequence. This server can predict α-helix, β-sheet, turn and coil. The algorithm used is divided into three, where any part consisting of six or more amino acid residues (Pa) ≥ 1.03 and (Pa > Pb) is predicted as helical. Then if the part of the amino acid sequence is five or more residue or (Pb) ≥ 1.05 with (Pb > Pa), it will be predicted as β-sheet. As for the turn is measured from the content of proline or glycine because both amino acids are commonly found in the turn area [10].

DLP-SVM is a server for domains linker predictor which is a specific amino acids sequence that become a barrier as well as interconnection between domains, so the desired interaction does not occur in those proteins. This server uses SVM or support vector machine which is a method for machine-learning that consists of three types namely SVM-long, SVM-Short and SVM-Joint which combines the predicted results of linkers [11].

2.4. Docking
Docking has a function in studying receptor and ligands interactions through the identification of a protein-compatible active site, then obtaining the best geometry of the complex between the ligands and its receptors and conducting calculation of energy interactions with ligands. As expected, the results of the docking analysis can help in designing effective ligands. The component that affects the success of a docking program is the scoring function and search algorithm [12]. The scoring component has a function in estimating the binding affinity between macromolecules and ligands. While the use of algorithms serves to determine the conformation of receptors interacting with ligands [13]. Docking was done with AutoDock Vina where Vina Search Space is used the same for all ligand with Center X (9.3232), Y (-23.360), and Z (5.6878) and Dimensions X (85.7204), Y (109.1736) and Z (79.3440).
3. Results and Discussion

3.1. Characterization of HSA 4K2C

Albumin as one of the proteins synthesized by hepatocytes as well as the continuous circulation has a total amount 60% to 65% of the overall plasma protein. Total amino acids were identified by ProtParam analysis. The results of the ProtParam analysis provided some information on the HSA 4K2C, it has an amount of 585 amino acids with an isoelectric point (pI) 5.67. The isoelectric point is the pH of a molecule when the electrical charge is neutral or uncharged. Further information is obtained by the negative amino acid residue (Asp + Glu) is 98 while the positively charged (Arg + Lys) is 83. Besides, the atomic composition is shown in Table 1. Further data obtained that 4K2C has a protein instability index of 38.85 which according to Walker (2005) if the protein has an index of instability under 40, then the protein is stable while above 40 means unstable. Stability of HSA 4K2C have been proved. Normal albumin before damage such as glycation is a stable protein, it is evidenced by molecular dynamics simulation by YASARA of human serum albumin in the period 15,000 ps [14]. Albumin until the end of the simulation still shows the initial conformation and does not occur unfolding. But when molecular dynamic simulation of glycated albumin (HSA 4K2C + glucose), there are a lot of conformational change happened including unfolding event between sidechain A and B of this glycated albumin model that represent the condition of albumin in diabetes mellitus.

| Atom     | Total |
|----------|-------|
| Carbon (C)| 2936  |
| Hydrogen (H) | 4624 |
| Nitrogen (N)| 786   |
| Oxygen (O) | 889   |
| Sulfur (S)   | 41    |

| Formula | \( C_{2936}H_{4624}N_{786}O_{889}S_{41} \) |

Table 1. Human Serum Albumin’s Atom Compositions (ProtParam).

High solubility of this protein in liquid media is caused by the existence of these amino acid residues. In terms of structure, it is known that human albumin is a protein monomer, whereby sub domains A and B have six and four \( \alpha \)-helix connected with loops that are flexible [15]. The characterization of HSA 4K2C such as the domain linker, pattern of motives and secondary structure in this study showed that albumin is a protein complex with amino acids integrated in order to complete the function. Motif and domain linker are a characteristic of protein so it can be distinguished from other proteins in the database. By knowing the character as well as the position of amino acids from a protein then in the research involving the process of docking, researchers can choose which proteins from databases should be used for next analysis.

The results of Motif Search indicate that there are three patterns of protein HSA (Human Serum Albumin) (Fig. 1A). Albumin consists of two side chains in which the Fig. 1B, side chain A is green, and the side chain B is light blue. The first motif at 551-575 amino acid residual arrangement which colored by magenta (FAAFVEKCKADDKETCFAEQK) while the second motif is in the yellow residue 353-377 (YETLEKKCAAADPHECYAKVFDEF). The last motif colored in red and the amino acid residues at 161-185 (YKAAFECCQAADKAACLLPKDEL). These motives are small sequences of protein fragments with highly conserved properties that have been used for predicting secondary and tertiary structure of a protein. These structure motifs also have application in drug design where it will provide ligand binding sites data or the active sites of protein.
The linker domain analysis using DLP-SVM. The domain linker is a loop area that separates two structural domains. The Linker is divided into two types i.e. short linker (SVM-Short) which contains less than equal to 9 amino acid residues. While long linker (SVM-Long) is more than 9 residues of amino acids. The combination is called SVM-All on this server, but it is used as a linker representation of SVM-Joint where the presentation of accuracy is much higher compared to SVM-Long and SVM-Short. Figure 3 shows the linker probability of the SVM-Long analysis shown in blue, with the SVM-Short (yellow) and SVM-All colored in red.

Figure 1. (A) Motif Search Result (B) Motif Search Human Serum Albumin Visualization with PyMOL.

Figure 2. Secondary Structure of HSA 4K2C.
Figure 3. SVM linkers, Long (Blue), Short (Yellow), All (Red).

Figure 3 on SVM-All has a peak value of 1,835 in the 411-450 and 1,558 regions of the amino acid 90-122. SVM-Long has a 1,966 peak value in the 93-122 region. Then SVM-Short with a peak value of 1,720 in amino acid residue 92-100. Further SVM-Joint will be taken as final data for visualization using PyMOL. Figure 4 is a visualization of the SVM-Joint which is the highest predictor of the linker domain found in the HSA 4K2C. The results of the SVM-Joint analysis consist of two linker domains that are the first DL-1 with 410-451 amino acid residues (RYTKYPQVSTPTLEVSRLNGKGVGSKCCKHPEAKRMPCAED) then the DL-2 is composed of amino acid residue number 96-122 (PERNECFLQHKDDNPNLPLRPEVDV).

Figure 4. HSA 4K2C Domain Linker Visualization (SVM-Joint).

3.2. Docking
Docking analysis is to know the ability of albumin in metal ions sequestering. Before performing the preparation, the first thing to do is Ramachandran Plot (RAMPAGE) analysis (Fig. 5). Ramachandran Plot shows the quality of the proteins to be used based on the location of certain amino acid residues in four quadrants that are representations of proteins to be analyzed.
Figure 5. Ramachandran Plot of HSA 4K2C. RAMPAGE evaluation of residues show that Ser5(A), Glu60(A), Thr79(A), Gly85(A), Pro113(A), Ala364(A), Ser5(B), Arg81(B), Gly85(B), Ala362(B), Ala364(B), Glu400(B), Val469(B), Lys541(B) in outlier region or less than 1.2%.

Data from the Ramachandran Plot analysis indicates that HSA 4K2C can be used for docking because the amount of residue that is in place should be as much as 1061 or 91.9% followed by 6.8% of residues in allowed region. Then it is followed by docking using PyRx. Before the docking, ligand preparation is done by minimizing energy using Open Babel integrated with PyRx software. Open Babel also used for the conversion of ligand format from (. sdf) to (. pdbqt).

Table 2. Binding Energy of Ligands and HSA Protein.

| Metal Ions | Binding Energy (kcal/mol) |
|------------|---------------------------|
| Cu^{2+}    | -1.2                      |
| Fe^{2+}    | -1.2                      |
| Mn^{2+}    | -1.2                      |
| Mn^{3+}    | -1.2                      |
| Fe^{3+}    | -1.2                      |

As for protein did not need preparation because HSA 4K2C has no innate ligand nor water in the file. The docking results show albumin can bind a wide range of metal ions both copper ions, iron ions (II) and (III), as well as manganese ions (II) and (III). Binding Energy (Table 2) of all ligands with HSA 4K2C shows the same value. However, the binding location of each ligand on the HSA 4K2C amino acid residue is different. Analysis of binding site is done using Discovery Studio (Table 3) where the iron ions (II) bind to Gln29 (A), Tyr30 (A), and Gln32 (A). Meanwhile, manganese ion (II) binds to CYS34 (A), Arg144 (A), Leu31 (A), Gln32 (A), and His39 (A).
Copper ions (II) bind to Gln29 (A), Tyr30 (A), and Gln32 (A). Manganese ions (III) interact with amino acid residues Gln29 (A), Tyr30 (A), and Gln32 (A). The last ion is Iron (III) where it binds to Cys34 (B), Arg144 (B), Leu31 (B), and Gln32 (B). In Figure 6 the protein-side chain of human serum albumin is divided into two namely side A (green) and side chain B (light blue). The type of bond between the Cu\(^{2+}\) ion and the amino acid residue in HSA is hydrogen bond acceptor, while the Fe\(^{3+}\) and Mn\(^{2+}\) bond type is divided into two namely hydrogen bond acceptor and metal donor. Then the type of bond between Mn\(^{3+}\) ions and Fe\(^{2+}\) is hydrogen bond acceptor.

The transition metal ions used in this study include iron ions (II) and (III), copper ion (II), manganese ions (II) and (III). Metal ions are often associated with the cause of various diseases in the human body, one of which is atherosclerosis [16]. Some examples of hazards from the accumulation of transition metal ions such as iron ions on human lesions and can contribute to the progression of an illness. Oxidative damage caused by multiple transition metal ions such as iron and copper ions capable of damaging the extracellular matrix components, the oxidation that occurs can even activate enzyme which capable to degrading matrix. One example of this enzyme is metalloproteinase [17].

From the results of this study HSA 4K2C can bind with a wide range of transition metal ions. The 3D visualization of the metal ion docking with HSA 4K2C can be seen in Figure 6. The binding energy of the docking result shows the same number of all simulated results i.e.-1.2 kcal/mol. The binding power of metal ions with proteins cannot only be determined by the binding energy. The interaction of metal and ligands conducted on the model, in this case, docking can weaken the true condition of the binding of metal ions in nature. The effect that occurs is called a shielding effect [18].

The binding energy of the transition metal ion docking results is also a character of the albumin protein. Binding sites of both human and bovine albumin to the metal ions consisted of four locations i.e. N-terminal site, Cys34 residue, side chains A and B [19].

The results of the NMR site A of Human Serum Albumin study constitute the MBS or multi-metal binding site [20]. It can explain an amino acid residue can bind to some metals above. As well as manganese ions (II) and iron (III) that interact with residue Cys34 both in the side chain A and B. Metal ions binding on Cys amino acid residues, can be caused due to cysteine have negative charge so that interactions are easier to occur compared to other metal ions that have a dipole charge interaction [21]. Metal ions binding on the amino acid residue of Gln and His is also due to the hydrophilic and polar properties of these two amino acid residues. The residue capability of certain amino acids as a metal donor as well as the acceptor is determined by the affinity of each amino acid. The amino acids Cys, Asp and Glu can be as donors and acceptors [21]. Site A as a multiple binding site indicated by amino acid residues that interact with metal ions such as Cu\(^{2+}\), Fe\(^{2+}\), and Mn\(^{3+}\).

| Metal Ions | Binding Site                  |
|-----------|-------------------------------|
| Fe\(^{2+}\) | Gln29 (A), Tyr30 (A), Gln32 (A) |
| Fe\(^{3+}\) | Cys34 (B), Arg144 (B), Leu31 (B), Gln32 (B) |
| Mn\(^{2+}\) | Cys34 (A), Arg144 (A), Leu31 (A), Gln32 (A), His39 (A) |
| Mn\(^{3+}\) | Gln29 (A), Tyr30 (A), Gln32 (A) |
| Cu\(^{2+}\) | Gln29 (A), Tyr30 (A), Gln32 (A) |
The ability of albumin in free radical trapping is due to the presence of cysteine (Cys34) residue wherein as much as 70-80% of the residue in an adult contains the sulfhydryl group. The existence of such residue that makes albumin capable of binding to radical hydroxyl [22]. Cys 34 on a subdomain of IA is located in the slit of a protein surface and not related with disulfide bridge, but its sulfhydryl group is able to bind Hg$^{2+}$, Ag$^{2+}$, Au$^+$ and Pt$^{2+}$ [15]. The scavenger activity owned by albumin is able to protect the ligands from oxidation events, an example is the binding of unsaturated fatty acids [23]. Cys34 in albumin is capable to reduce thiol in human blood plasma [24]. In addition to the Cys amino acids there is also the Met in human albumin with the function of Cys residue as a free radical scavenger while the Met functions in a metal chelator [25]. It was found that 80% of the group – SH

**Figure 6.** 3D Visualization of HSA 4K2C docking with metal ions.
which is a strong scavenger of ROS in albumin comes from Cys-34 [26]. The presence of cysteine and methionine residues is essential in the antioxidant ability of human serum albumin. Research shows that when a mutation is performed on Cys34 the antioxidant capabilities of albumin fall by 30% and if free cysteine residues are placed in another position on the HSA 4K2C, the antioxidant capability still remain. Site 2 in HSA 4K2C is the target of oxidative stress, but lost or modified cysteine residue results in more easily albumin to be degraded [24].

4. Conclusion

The characteristic albumin as a stable protein also has a function as an endogenous scavenger that is closely related to sequestering metal ions. Albumin has MBS (Multi-Metal Binding Site) represented from the docking results of several metal ions that bind to the same amino acid residue in the side chain A. The ability of amino acid residue in albumin as free radical trapping like Cys34 has donor and acceptor metal ions binding type. The binding energy of transitional metal ions to HSA 4K2C has the same amount of energy as -1.2 kcal/mol, allegedly shielding effect plays a role in determining binding energy value.

References
[1] Murray R K, Granner D K and Rodwell V W 2006 Harper's illustrated biochemistry 27th Edition (Asia: Mc Graw Hills Companie)
[2] Fasano M, Curry S, Terreno E, Gilliano M, Fanali G, Narciso P, Notari S and Ascenzi P 2005 IUBMB Life 57 787-96
[3] Xu X, Zhang L, Shen D, Wu H and Liu Q 200 J Fluoresc 18 193-201
[4] Stewart A J, Blindauer C A, Berezenko S, Sleep D and Sadler P J 2003 PNAS 100 3701-06.
[5] Quinlan G J, Martin G S and Evans T W 2005 Hepatology 41 1211-19
[6] Suzuki E, Yasuda K, Takeda N, Sakata S, Era S, Kuwata K, Sogami M and Miura K 1992 Diabetes Research and Clinical Practice 18 153-58
[7] Kazmierczak S C, Gurachevsky A, Matthes G and Muravsky V 2006 Clinical Chemistry 52 2129-
[8] Walker, J M 2005 The proteomics Protocols Handbook (USA: Humana Press)
[9] Sigrist C J A, C Gerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A and Bucher P 2002 Brief Bioinform 3 265-74
[10] Kumar T A 2013 Wide Spectrum 1 15-19
[11] Ebina T, Toh H and Kuroda Y 2008 Peptide Science 92 1-8
[12] Mukes B and Rakesh K 2011 International Journal of Research in Ayurveda & Pharmacy 6 1746-51
[13] Serina J C 2013 Master Dissertation Universidade de Madeira, Portugal
[14] Wibowo S, Widyarti S, Sabarudin A, Soe amd Tami S B 2019 Asian Journal of Pharmaceutical and Clinical Research 12 276-82
[15] Otagiri M and Chuang V T G 2016 Albumin in Medicine: Pathological and Clinical Applications (Singapore: Springer)
[16] Horwitz L D and Rosenthal E A 1999 Vasc Med. 4 93–99
[17] Stadler N, Lindner R A and Davies M J 2004 Arterioscler Thromb Vasc Biol 24 949-54
[18] Chen D, Li Y, Guo W, Li Y, Savidge T, Li X and Fan X 2019 Physical Chemistry Chemical Physics 21 205-16
[19] Bal W, Sokolowska M, Kurowska E and Faller P 2013 Biochim Biophys Acta 12 5444-55
[20] Blindauer C A, Lu J, Stewart A J, Sadler P J and Pinheiro T J T 2008 Biochem Soc Trans 36 1317–21
[21] Dudev T and Lim C 2014 Chem Rev 114 538-56
[22] Gutteridge J M 1986 Biochim Biophys Acta 869 119-27
[23] Roche M, Rondeau P, Singh N R, Tarnus E and Bourdon E 2008 FEBS Letters 582 1783–87
[24] Anraku M, Chuang V T G, Maruyama T and Otagiri M 2013 Biochim Biophys Acta 1830 5465-
[25] Bourdom E, Loreau N, Lagrostl and Blache D 2005 Free Radic Res 39 15-20
[26] Nagumo K, Tanaka M, Chuang V T G, Setoyama H, Watanabe H, Yamada N, Kubota K, Tanaka M, Matsushita K, Yoshida A, Jinnouchi H, Anraku M, Kadowaki D, Ishima Y, Sasaki Y, Otagiri M and Maruyama T 2014 *PLOS-ONE* 9 e85216