

Assessment of Interaction of Human OCT 1-3 Proteins and Metformin using Silico Analyses

Running Title: Investigation of Human OCT1-3 and Metformin

Faruk Berat AKÇEŞME PhD^{1*}, Nail BEŞLİ MSc², Jorge Peña-García³, Horacio PÉREZ-SÁNCHEZ³

¹ Department of Biostatistics and Medical Informatics, Faculty of Medicine, University of Health Sciences, Istanbul, Turkey

² Department of Medical Biology, Faculty of Medicine, University of Health Sciences, Istanbul, Turkey.

³ Bioinformatics and High Performance Computing Research Group (BIO-HPC), Computer Engineering Department, Universidad Católica de Murcia (UCAM), Guadalupe, 30107, Spain.

Nail BEŞLİ: <http://orcid.org/0000-0002-6174-915X>

e-mail: beslinail@gmail.com

Horacio Pérez Sánchez: <https://orcid.org/0000-0003-4468-7898>

e-mail: hperez@ucam.edu

Jorge Peña-García

e-mail: jpena@ucam.edu

*Corresponding author: Faruk Berat AKÇEŞME

Address: University of Health Sciences, Medical Faculty, Department of Biostatistics and Medical Informatics, P.O. Box:34668, Istanbul, Turkey

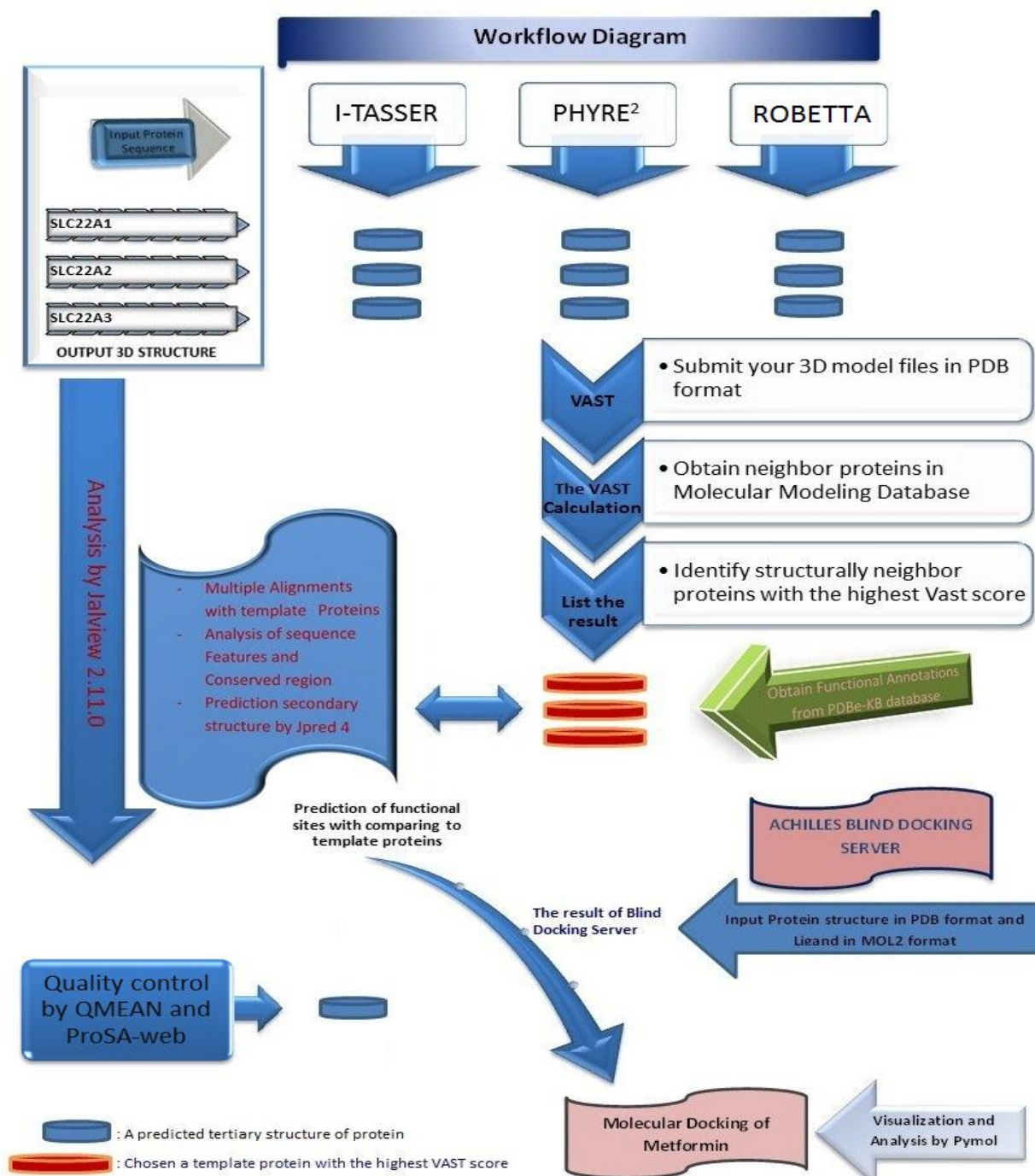
Tel: +090-537- 502 2634

E-mail: farukberat.akcesme@sbu.edu.tr

ABSTRACT

Metformin, a drug frequently used diabetic patients the first-line treatment worldwide, is positively charged and is transported into the cell through human organic cation transporters (hOCT 1-3) proteins. We aimed to mimic the cellular uptake of metformin by hOCT1-3 with various bioinformatics methods and tools. 3D structure of OCT1-3 proteins was predicted by considering the structures and function of these proteins. We predicted functional regions (active and ligand binding site) of OCT1-3 and performed comparative bioinformatics analysis. Predicted structure of hOCT1-3 was then analyzed in Blind Docking server and the results were confirmed with predicted binding site residues and conserved domain regions. We simulated the OCT1-3 and metformin docking and also validated the docking procedure with other substrates of HOCT1-3 proteins. We selected the best poses of metformin docking simulations as per binding energy (-5.27 to -4.60 kcal/mol). Lastly, we validated static description of protein-ligand (OCT-Metformin) interactions by performing MD simulation. Eventually, we obtained stable stimulation of OCT-metformin interaction.

KEYWORDS: Organic cation transporters, Metformin, Protein structure prediction, Model Validation, Docking



50

51 **Figure 1.** Identification of molecular modeling of OCT1-3 and metformin. There are four main
 52 steps in the workflow. The first step is the prediction of the 3D structure of OCT1-3 proteins and
 53 quality control of the model protein structures. The second one is the identification of template
 54 proteins with the VAST. The third one is the analysis of sequence data by Jalview 2.11.0 while
 55 the fourth is Molecular docking by Achilles Blind Docking Server.

1. Introduction

Membrane proteins are associated with prominent functions in the cell, approximately responsible for 30% genes in the human genome¹ and currently possessing 50% of pharmaceutical drug discovery.² Membrane proteins perform a broad variety of particular roles during cellular events.³ Due to its structural and physicochemical properties, plasma membrane has a selective permeability for organic and inorganic substances including cation and anion compounds. Hence, it assists to sustain the unique content both inside and outside of the cell.

One of the protein families that provide translocation of cationic organic and inorganic compounds that localized in the cell membrane is SLC (Solute carrier) family from the MFS superfamily. The SLC family is a 22-membered cell membrane transporter. A subfamily of the SLC family is SLC22A1,2 and 3 (cd17379: MFS_SLC22A1_2_3).⁴ Beside many essential cation molecules for the cell, the SLC transporters are the target of drugs with high pharmacological value. Human body constitutes more than 400 important SLC transporters for a broad range of tasks including drug metabolism as absorption, distribution, and excretion. Hence, there is a growing interest in the effects of the drug on the development and progression of interacting with these transporters.⁵

Metformin, categorized as antidiabetic medication, is uptaken by cell via SLC transporter proteins encoded by SLC22A1-3 (also named OCT1-3) genes.⁶ OCT1-3 membrane proteins are expressed at different levels in several tissues including OCT-2 in the level of renal expression is high, OCT-3 is most commonly expressed in skeletal muscle,⁷ whereas OCT-1 is the primary characterized expressed in hepatocytes.⁸ Metformin hydrophilic (logD -6.13 pH 6.0) and pKa (physiological pH) is 12.4.⁹ Functional elimination of OCT-1 in primary mouse hepatocyte culture and OCT-1 has been demonstrated to play an important role in metformin response in in vivo mouse model.¹⁰ Following metformin enters into the cell via HOCT1-3, it exhibits its anti-diabetic properties in several ways. It is broadly believed that the blood glucose-lowering impact of metformin is mediated chiefly through the repression of hepatic glucose production by decreasing gluconeogenesis and blocking glucagon-mediated signaling in the liver.^{11,12} Furthermore, some mechanisms have been suggested that metformin can activate AMPK (AMP-activated protein kinase) by the upstream liver kinase B1,¹¹ enhanced AMP/ATP rate hereby the restraint of mitochondrial respiratory chain complex I.¹³ Metformin improves the activity of the

insulin receptor and of IRS-2 (insulin receptor substrate 2) and boosts glucose uptake via enhanced translocation of glucose transporters, such as GLUT-1, to the plasma membrane.¹⁴

In the current study, we aimed to predict the three-dimensional structure of human OCT1-3 by using various computational approaches by considering the current authenticated/trusted bioinformatics tools. The molecular docking was also performed under the inspiration of the in vitro and in vivo founding. We also performed Molecular dynamics (MD) simulation of docking of Metformin and hOCT1-3 at the atomic level for validation. The computational approaches to OCT1-3 proteins, particularly structural prediction of proteins and simulations by MD, have been broadly implemented for investigating their dynamic actions.

Given the pharmacological importance of SLC22A1,2 and 3 proteins in humans, determination of the structure of these proteins, estimation of their active sites and definition of how the transport mechanism works have aroused great interest. In the present study, we have reported and defined ligand-dependent interactions of hOCT1-3 with metformin and the other ligands utilizing the computational some approaches and explored the found interactions through comparative analysis, homology modeling, and molecular dynamic studies. This study is first attempt to demonstrate OCT1-3 interaction with metformin. This interaction has characterized by docking analysis and results were validated with MD simulations. This is important study that use predicted structure of OCT proteins to stimulate the interaction which stays highly stable throughout the MD analysis.

Materials and Methods

2.1. Computational structural modeling of OCT1-3

2.1.1. Prediction of Secondary and Tertiary Structure of OCT1-3 Proteins

We retrieved OCT1,2,3 (Accession no:AAI26365.1, NP_003049.2, NP_068812.1 respectively) from GenBank in FASTA format, predicted the secondary structure of the proteins by using JPred4,¹⁵ which is the latest version of the JPred online prediction server supplying by the JNet algorithm.

Each of OCT1-3 protein structures was predicted on PHYRE^{2,16} Robetta^{17,18} AND I-TASSER^{19,21} (protein structure prediction servers). In these prediction tools, homology modeling (or comparative modeling) were used by comparing the experimentally determined proteins as templates. To quality control of model proteins, we performed to the tool of the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane)²² and ProSAweb.²³

We compared each obtained model to all PDB proteins in MMDB (Molecular Modeling Database) in order to find 3D similar structures in VAST (Vector Alignment Search Tool).²⁴ Vast found the following proteins with the highest scores, PDB Id: 4zw9_A, 4zwc_A, 5c65_A. Each of OCT1-3 protein structures obtained from Robetta server were then selected as a model since its neighbor proteins have the highest VAST score and %Id, compared to the other prediction tools (Fig1).

The Computed Atlas of Surface Topography of proteins (CASTp)²⁵ 3.0 was utilized to predict the surface of the binding pocket of the model proteins to interacting with their substrates.

2.1.2. Sequence Analysis

We performed pairwise in BLASTP²⁶ and multiple sequence alignment in Clustal OMEGA²⁷ for each of the OCT1,2,3 proteins with the selected template proteins (4zw9, 4zwc, 5c65) obtained from VAST. Parameters for alignment with Clustal OMEGA were set as –GAPEXT :0.1, ENDGAPS: 0.5, GAPDIST: 1, GAPOPEN:10 and MATRIX: BLOSUM62. We analyzed and interpreted the results in Jalview 2.11.²⁸

For analyzing sequence features, functional annotations of template proteins were retrieved from PDBe-KB database.²⁹ Then, we compared these proteins with the model proteins to identify conserved regions and to predict sequence features. In this way, we assigned Predicted Functional sites, Predicted PTM sites, Predicted Ligand binding sites, Ligand binding sites, and Interaction interfaces for our model proteins.

2.1.3. Visualization

Visualization of primary and secondary structures of the proteins was performed in the Jalview 2.11. The PyMOL³⁰ software was utilized for representing and analyzing the atomic structure of proteins.

2.1.4. Molecular docking simulations

One of the most essential steps in this study is the molecular simulation as given in the workflow in Fig1. For the preparation of docking process, hOct1-3 proteins were downloaded from Robetta server in PDB format. All ligands (Metformin, Phenformin, Norepinephrine) of hOct1-3 were retrieved as in SDF format from PubChem.³¹ We removed the water, added the polar hydrogen to the model proteins. Then, we charged the model proteins and the ligands by the computation of Gasteiger before converting to pdbqt format using AutoDock Vina.³² Before performed the docking process, for the energy minimization of 3D model protein structure was subjected to minimization method in chimera 1.14³³ as default the steepest descent:100 with 0.02 step sizes, without fixing any atoms, after that 10 steps of conjugate gradient steps with 0.02 step size (Å) minimization.

The docking study was carried out under ACHILLES BLIND DOCKING SERVER³⁴ protocol (<https://bio-hpc.ucam.edu/achilles/>). We prepared figures by using PyMOL.

2.1.5. Molecular dynamics(MD) simulations

To evaluate the structural constancy and validate static description of protein-ligand (hOCT-Metformin) interactions, we ran MD simulation using the Desmond Software³⁵. Dynamic nature of protein-ligand interactions has been studied and atomic-level interactions were investigated.

Results and Discussion

3.1. Alignments

We aligned a range of 146-445 aa of OCT-1 with of 85-397 aa of 4ZW9 and 4ZWC as explained previously. The OCT-1 sequence demonstrated the 22.77% sequence identity with 4zw9 and 4zwc. Alignment of a range of 146-540 aa of OCT-1 with of 63-477aa of 5c65 is as 22.51% identity.

167 Alignment of a range of 24-546 aa of OCT-2 with of 93-510 aa of 4ZW9, 4ZWC, and 71-438 aa
168 of 5c65 shows the same identity as 26.57% with 4zw9, 4zwc, and 5c65.

169 As for OCT-3, Alignment of a range 60-353 aa of OCT-3 with of 86-513 aa of 4zw9 and 4zwc is
170 a 24.53% Identity. Alignment of a range 84-353 aa of OCT-3 with of 64-473 aa of 5c65 is the
171 24.53% Identity.

172

173 **3.2. Analysis of Conserved Domain and Sequence Features of OCT1-3**

174 After subjecting OCT1-3 and template proteins sequences to multiple alignment, we detected the
175 conserved regions via comparative analysis in Jalview 2.11.0(Fig2). We summarized the results
176 of comparative analysis as a list in Table1.

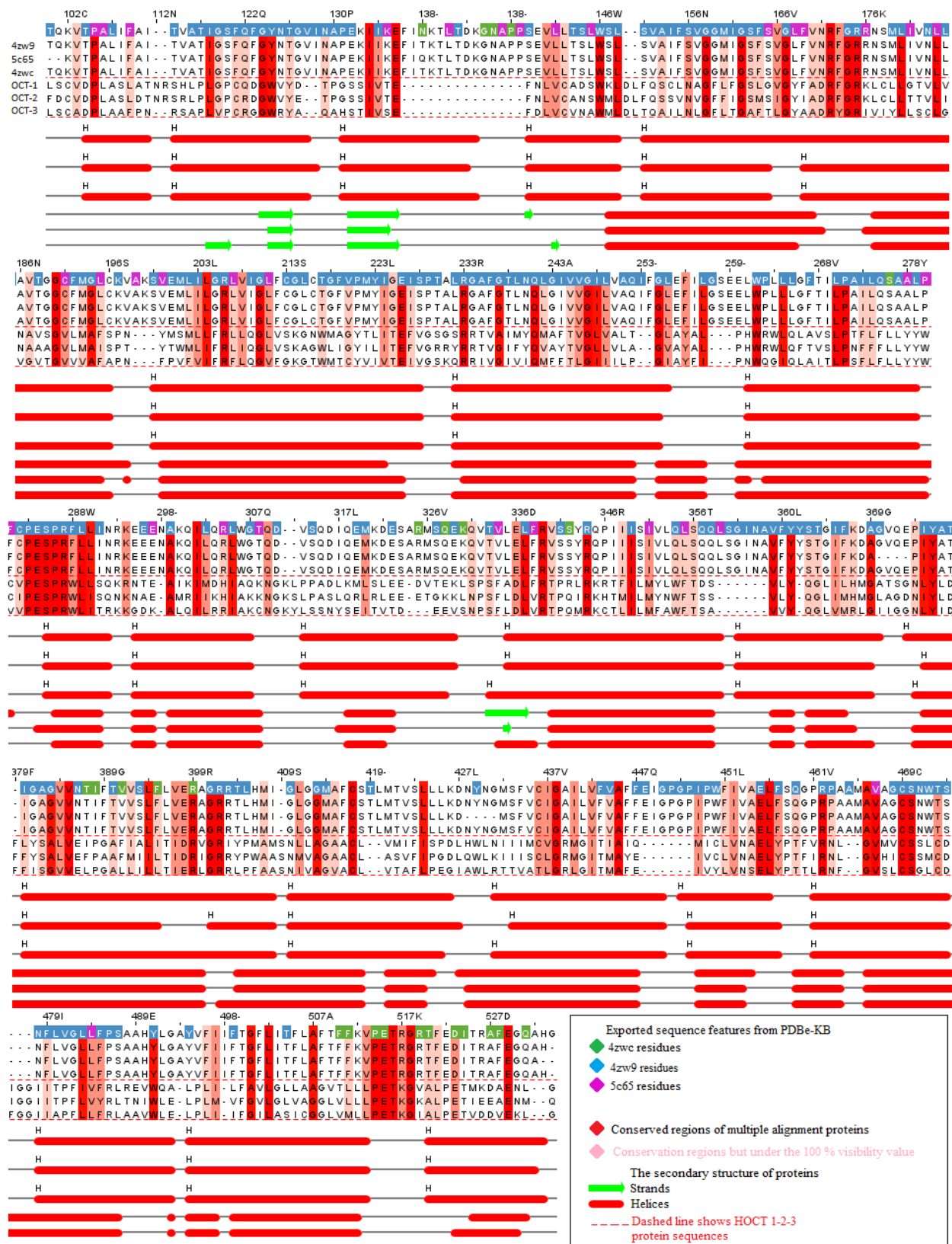


Figure 2. Representation of primary and secondary structure of 4zw9, 5c65, 4zwc, and OCT1-3 proteins. Visualization of sequence features and Conserved Domain of the protein residues are colored by analysis in Jalview 2.11.0. The probability of conserved regions decreases from dark red to pink. After multiple alignments, OCT-1 protein was set as a reference for sequence numbering. As a result of multiple sequencing, solely the overlapping regions of the proteins were exhibited.

Fig 6.Exported Functional Annotations of the templates proteins from PDBe-KB

OCT-1		
Metformin	GLU ¹³⁷	Conserved Domain
	PRO ⁴⁸¹	4zw9;Predicted Ligand binding sites, 5c65; Ligand binding sites
	ARG ⁴⁸⁸	
Phenformin	GLN ¹⁵²	4zw9;Predicted Ligand binding and functional sites, Ligand binding sites
	ASN ¹⁵⁶	4zw9;Ligand binding sites, Conserved Domain
	LYS ²¹⁴	4zw9;Predicted Ligand binding and functional sites
	TRP ³⁵⁴	4zw9;Predicted Ligand binding sites, Conserved Domain
	ASP ³⁵⁷	4zw9;Predicted Ligand binding and functional sites, Ligand binding sites
	GLN ³⁶²	4zw9;Predicted Ligand binding and functional sites, Conserved Domain
	ILE ⁴⁴⁶	
OCT-2		
Metformin	ASN ¹⁵⁷	4zw9;Predicted Ligand binding sites, Ligand binding sites, Conserved Domain
	CYS ⁴⁷⁴	4zw9;Predicted Ligand binding sites, 4zwc; Predicted PTM sites, 5c65; Ligand binding sites, Conserved Domain
	ASP ⁴⁷⁵	4zw9;Predicted Ligand binding sites
Phenformin	TYR ³⁷	
	ASN ¹⁵⁷	4zw9;Predicted Ligand binding sites, Ligand binding sites, Conserved Domain
	LYS ²¹⁵	4zw9;Predicted Ligand binding and functional sites
	TYR ²⁴⁵	4zw9;Predicted Ligand binding and functional sites, Ligand binding sites
	TYR ³⁶²	4zw9;Predicted Ligand binding sites, Ligand binding sites, Predicted funtional sites, 5c65; Predicted PTM sites, Conserved Domain
	CYS ⁴⁷⁴	4zw9;Predicted Ligand binding sites, 4zwc; Predicted PTM sites, 5c65; Ligand binding sites, Conserved Domain
	ASP ⁴⁷⁵	4zw9;Predicted Ligand binding sites
OCT-3		
Metformin	VAL ³⁷	
	ASN ¹⁶²	4zw9;Predicted Ligand binding sites, Ligand binding sites, Conserved Domain
	ARG ²¹²	4zw9;Predicted Ligand binding sites, Ligand binding sites, Conserved Domain
	GLN ³⁶⁶	4zw9;Predicted Ligand binding sites, Ligand binding sites, Conserved Domain
Norepinephrine	PHE ³⁶	
	VAL ³⁹	
	GLN ¹⁵⁸	4zw9;Predicted Ligand binding and functional sites, Ligand binding sites
	ASN ¹⁶²	4zw9;Predicted Ligand binding sites, Ligand binding sites, Conserved Domain
	ARG ²¹²	4zw9;Predicted Ligand binding sites, Ligand binding sites, Conserved Domain

201

202 **Table 1.** A list of a result of sequence features by analysis multiple alignments. The residues of

203 OCT 1-3 proteins that interact with Metformin, Phenformin and Norepinephrine and functional

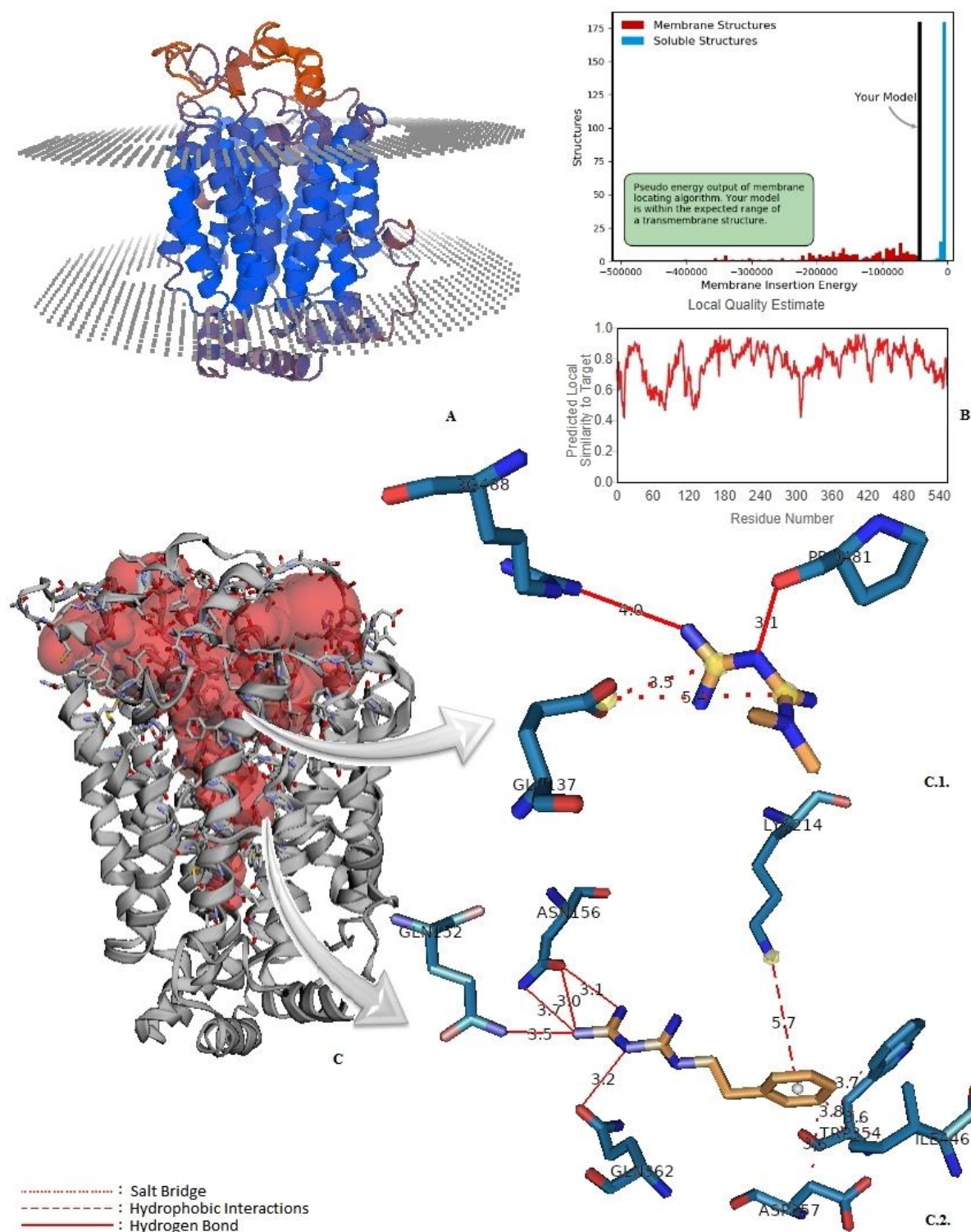
204 annotations of template proteins from PDBe-KB.

205

Cluster Populations		
The highest binding energy (kcal/mol)		
	Metformin	Phenformin
OCT-1	-4.60	-7.00
OCT-2	-5.20	-8.60
OCT-3	Metformin	Norepinephrine
	-5.27	-5.93

206

207 **Table 2.** A list of the Autodock Vina binding energy of best pose of metformin, Phenformin,
 208 Norepinephrine and hOCT1-3 proteins.



209

210 **Figure 3.** Representation of molecular modeling of OCT-1 and metformin, phenformin. A:
 211 OCT-1 transmembrane protein embedded in the plasma membrane model was predicted by

212 QMEANBrane. B: Structure validation of modeled OCT-1 with regard to membrane insertion
213 energy and local quality estimate of the residues of the model OCT-1. C: The surface of the
214 binding pocket of the model OCT-1 as computed using CASTp 3.0. Molecular simulation of the
215 best pose of the interaction of OCT1 and metformin(C.1), phenformin(C.2) with the highest
216 docking score.

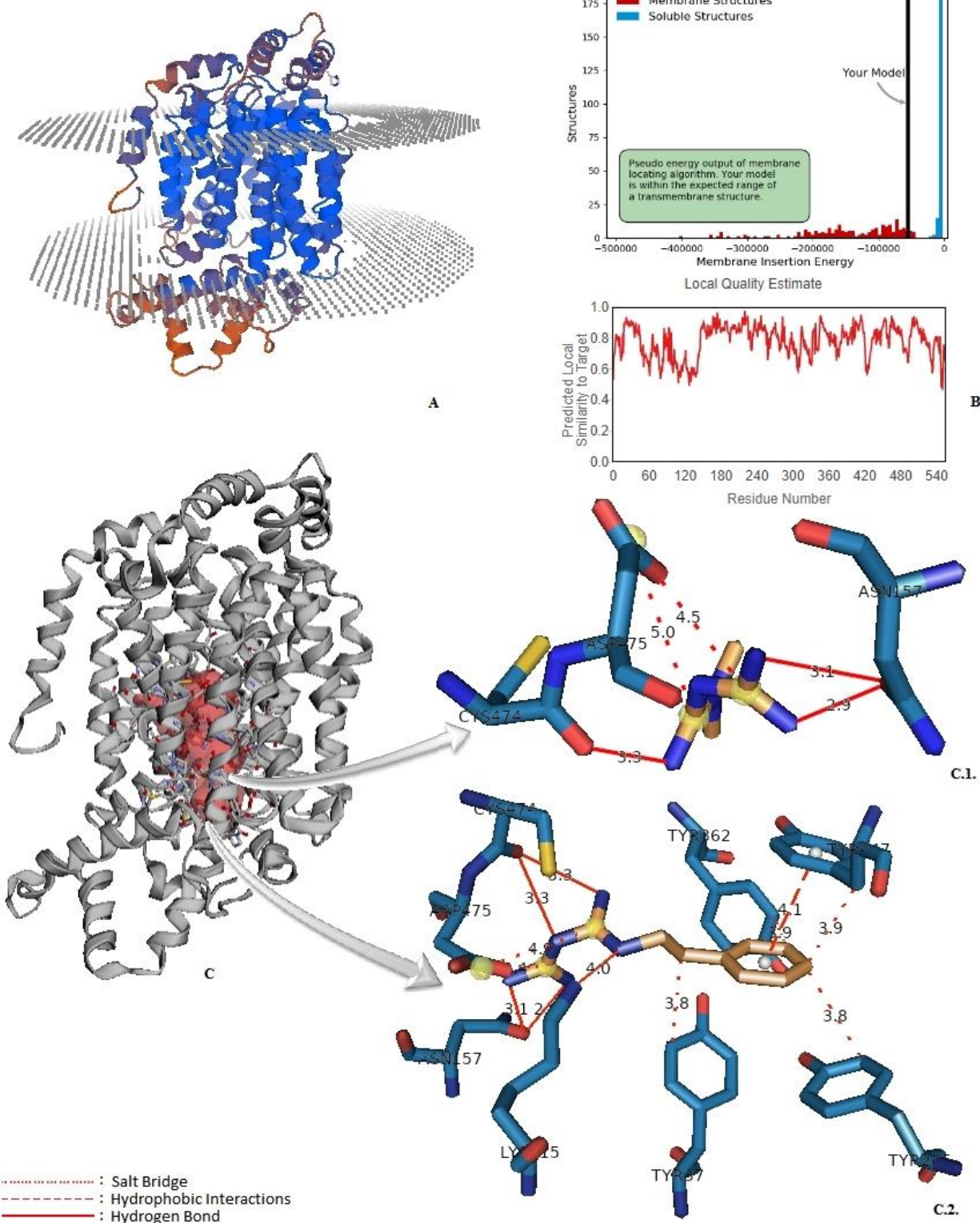
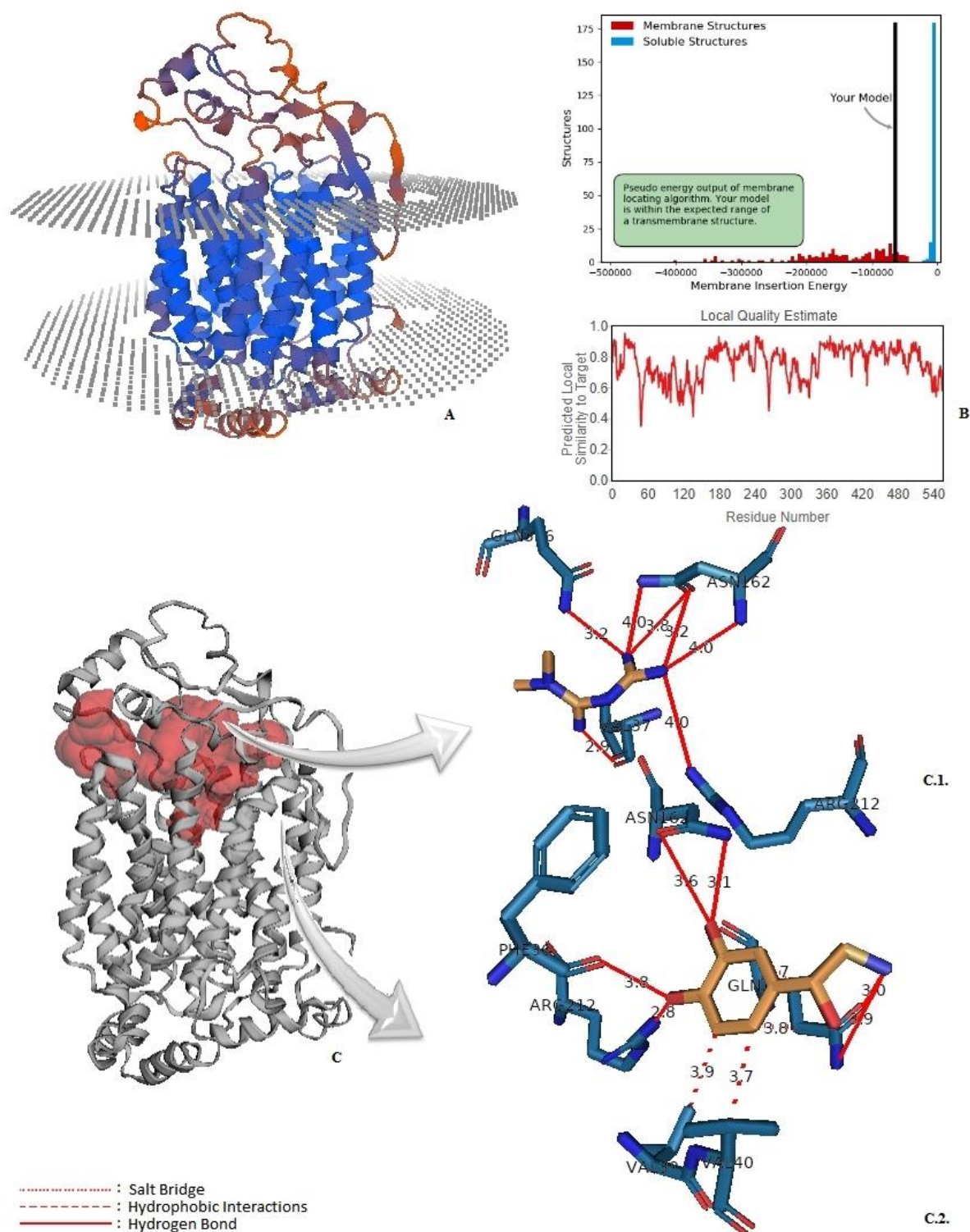


Figure 4. Representation of molecular modeling of OCT-2 and metformin, phenformin. A: OCT-2 transmembrane protein embedded in the plasma membrane model was predicted by

220 QMEANBrane. B: Structure validation of modeled OCT-1 with regard to membrane insertion
221 energy and local quality estimate of the residues of the model OCT-2. C: The surface of the
222 binding pocket of the model OCT-2 as computed using CASTp 3.0. Molecular simulation of the
223 best pose of the interaction of OCT-2 and metformin(C.1), phenformin(C.2) with the highest
224 docking score.



225

226 **Figure 5.** Representation of molecular modeling of OCT-3 and metformin, norepinephrine. A:

227 OCT-3 transmembrane protein embedded in the plasma membrane model was predicted by

QMEANBrane. B: Structure validation of modeled OCT-1 with regard to membrane insertion energy and local quality estimate of the residues of the model OCT-3. C: The surface of the binding pocket of the model OCT-3 as computed using CASTp 3.0. Molecular simulation of the best pose of the interaction of OCT-3 and metformin(C.1), norepinephrine (C.2) with the highest docking score.

The cellular and biological function of a protein is highly related to its 3D structure. The pharmacodynamics of the drug on the cell decreases or has no effect when the functional parts of these proteins are mutated in the genome. On the other hand, defining protein-ligand binding sites and explaining functional parts of the protein is a critical approach for drug discovery.³⁶ Regarding the pharmacodynamics of metformin, the 3D prediction of OCT1-3 proteins and the determination of ligand binding sites in the functional sites are critical for investigating their effects on the cell.

Recent studies and meta-analyses have shown that patients with T2DM have a lower incidence of tumor development than healthy controls and cancer patient that use metformin has a lower risk of mortality.³⁷ Metformin takes more attention after its role in cancer prevention and treatment has revealed. Improving or managing cellular uptake of therapeutic entities is mostly related to an understanding of the molecular mechanism of interaction with components of the cell membrane and therapeutic entities. This paper aimed to predict the 3D structure of OCT1-3 Proteins and identify their role in the uptake of metformin into the cells that have been studied by in vitro and in vivo studies before.^{38,39}

Sequence and structure analysis of proteins of unknown function with those of proteins of known function enables us to discover and deduce the function of unknown proteins. Characterization of protein function by in vivo and in vitro studies is time and labor-consuming. Furthermore, for some proteins, especially membrane proteins are exceedingly difficult to crystallize by experimental tools. In the modern genomic and proteomic era, a protein is mostly identified before its function is determined therefore the role of in silico studies in structural analyses of proteins become more important in recent years.

The structure of OCT1-3 proteins has not been solved yet by any experimental tools although some of the protein's structures have already known in the same protein family. This paper is important as a first attempt to study and predict the 3D structure of OCTs to reveal the information about how these proteins facilitate the uptake of metformin into the cells. Even though our analysis indicates no significant similarity between OCTs and the proteins of the database at a sequence level, the predicted OCTs have been found to be similar with its conservative regions to some carrier proteins that share a similar function.

It is known that 30 percent of all sequences are membrane proteins. Unlike globular proteins, a 3D model for membrane proteins can hardly be computed. Another important aspect of this paper is presenting a new pipeline to stimulate the docking of protein molecules in the absence of a similar sequence in the database. The recent algorithms in 3D structure prediction of proteins enable us to predict the structure of proteins in high accuracy even in the absence of sequence similarity and this paper is using the benefits of these tools. In silico analyses helped us to stimulate this biological process and propose the uptake of metformin by OCTs as it is shown in Fig 3-5.

Dakal et al. modeled the 3D structures of hOCTs by only one tool using I-TASSER in 2017.⁴⁰ In Fig 1, four key steps of this pipeline have shown as a workflow. One of the very critical points, the prediction of the 3D structure of the protein, was performed by three different tools; Iterative Threading ASSEmbly Refinement, Phyre2 that uses protein homology and Robetta. The output model proteins were then exposed to all proteins in the PDB by calculation in the VAST. This approach is reflected in our results in increasing accuracy in protein structure prediction. It is aimed to increase the accuracy of the prediction by validation of these structures to use experimentally determined proteins as templates. After obtaining a structure of OCTs, the orientation of these molecules in the plasma membrane was predicted by using QMEANBrane scoring function.

Transmembrane proteins play vital roles in a diverse range of essentially biological processes. To know about the protein position within the lipid bilayer is important and requires a computational approach because identifying the correct orientation is possible by defining the relationship between sequence, structure and the lipid environment. One of the commonly used tools to localize the structure of proteins within the lipid player by knowledge-based statistical potential,

286 QMEANBrane was used and the predicted position as exhibited in Fig 3-5. As a result, all model
287 proteins are within the expected range of transmembrane structures.

288 Models obtained from the other tools were determined to be inapplicable for the docking
289 process. Robetta is continuously evaluated with CAMEO(Continuous Automated Model
290 EvaluatiOn), which constantly assesses the accuracy and reliability of the prediction. Among
291 other prediction tools at CAMEO, Robetta and QMEANBrane are the first-line by time-based
292 statistical confidence and show reliable performance. We also used the ProSAweb to verify
293 structurally the quality of the model proteins. The Z-score designates the entire model-quality for
294 OCT1-3 respectively (Z-score:-8.59, -7.04, -5.95) in figure 6.

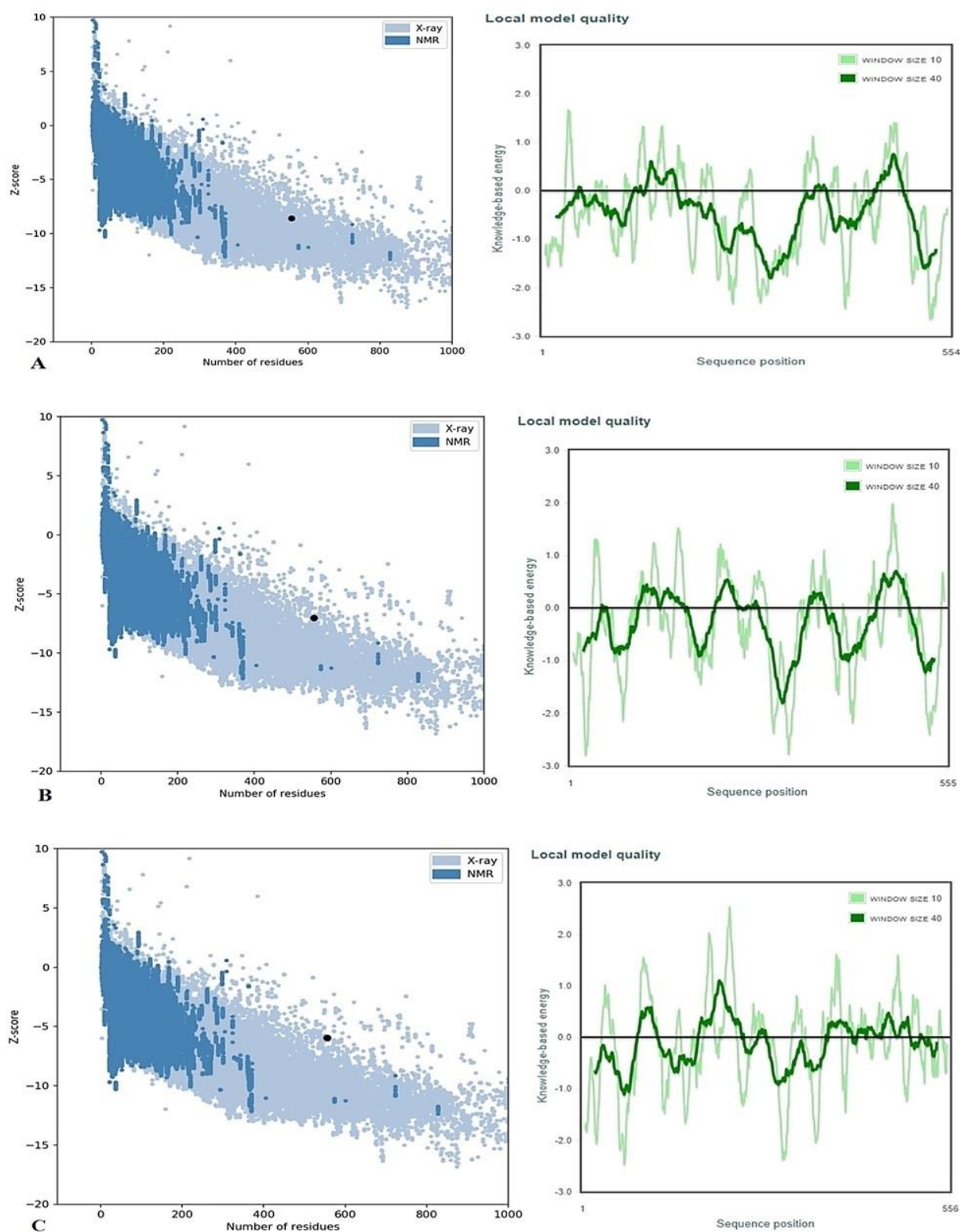


Figure 6. ProSA-web service analysis of human OCT1-3 proteins. The black points represent that model hOCT1-3 proteins are in the range of Z-score values of experimental structures

according to a number of residues. The other graph shows the local quality with regard to a number of sequence position (A;OCT-1, B;OCT-2, C;OCT-3). respectively (Z-score:-8.59, -7.04, -5.95).

In order to analyze sequence features, functional annotations of template proteins were retrieved from PDBe-KB database. Recently released database, PDBe-KB, give us a great opportunity to analyze and visualize sequence features of similar proteins that are used as a template to assign a novel function to our sequence in interest. Even though the sequence similarity is low, as the results indicate, there are significantly conserved regions. In this way, we assigned Predicted Functional sites, Predicted PTM sites, Predicted Ligand binding sites, Ligand binding sites, and interaction interfaces to OCTs.

Representation of molecular modeling of OCTs and metformin was performed by using Blind Docking server. The server mainly utilizes a customized version of Autodock Vina³² for the blind docking calculations. We obtained binding energy plots, and, in this way, the most energetically favorable docking has been selected the first best pose according to binding energy frequency (see supplementary Fig1). Due to obtaining by predicting the proteins, taking into account the model protein uncertainty as well as the small size of the metformin molecule, it is not surprising that many different ligand poses with similar scores were obtained. To cope with this, we used the CASTp bioinformatics tool and compared the prediction active sites of model proteins with the first best poses as docking results. Interestingly, both output results from two servers are similar. In addition, for the obtained results to be more meaningful, the pharmacologically important phenformin from metformin analogs was validated by the docking study of OCT-1 and OCT-2, while OCT-3 by norepinephrine compound. We also combined these outputs with outcomes from exported functional annotations of the templates proteins from PDBe-KB. We visualized the interaction of metformin, phenformin, norepinephrine and OCTs in PyMOL to better examine the poses and extract our images.

As listed in Table 1, OCT-1 forms hydrogen bonds with docked ligand molecules with the residue number of PRO481, ARG488, GLN152, ASN156, GLN362. The other four residues in the predicted site (LYS214, TRP354, ASP357, ILE446) interacted by hydrophobic and salt-bridge bonds. Chen et al.⁴¹ have reported that OCT-1 interacts with its ligands by hydrogen binding and non-covalent interaction from the ASP357, TRP354, ASN156, ILE446 residue among their predicted residues. OCT-2 interacted with both metformin and phenformin from ASN157, CYS474, ASP475 residues with noncovalent interactions such as hydrogen bond, salt bridge and hydrophobic interaction see in Figure 4. OCT-3 protein contacts with norepinephrine and metformin in the same residue (ASN162, ARG212) by hydrogen bonds. Given the extensive hydrogen bonding motif of metformin it is possible that water may be involved. This may significantly impact and alter the results and conclusions. Thus, we have considered performing the classic MD simulation for the docked complexes.

One of the residues that OCT-1 interacts with phenformin is GLN152 but OCT-3 interacts with norepinephrine in GLN158 as a same residue. The difference in number residue is because of setting as a sequenced reference. Our results suggest that, just as human OCT proteins are predominantly expressed in different tissues of the human body, the active binding sites of the proteins also vary.

Although the methodology, including template definition, comparative protein modeling, structure analysis, and molecular docking, seems pretty standard and employed in hundreds of research projects in our workflow, there is validation such as the quality control of the model proteins using Web services at almost every stage to increase reliability in achieving and evaluating meaningful results. Thus, the described pipeline is highly useful due to its ability to integrate the ligand-binding site and interaction interfaces information that is obtained from PDBe-KB database to the information that is derived from similarity analysis and prediction tools. This pipeline is promising to assign a function to the predicted 3D structure even in the absence of sequence similarity.

3.3. MD Simulations

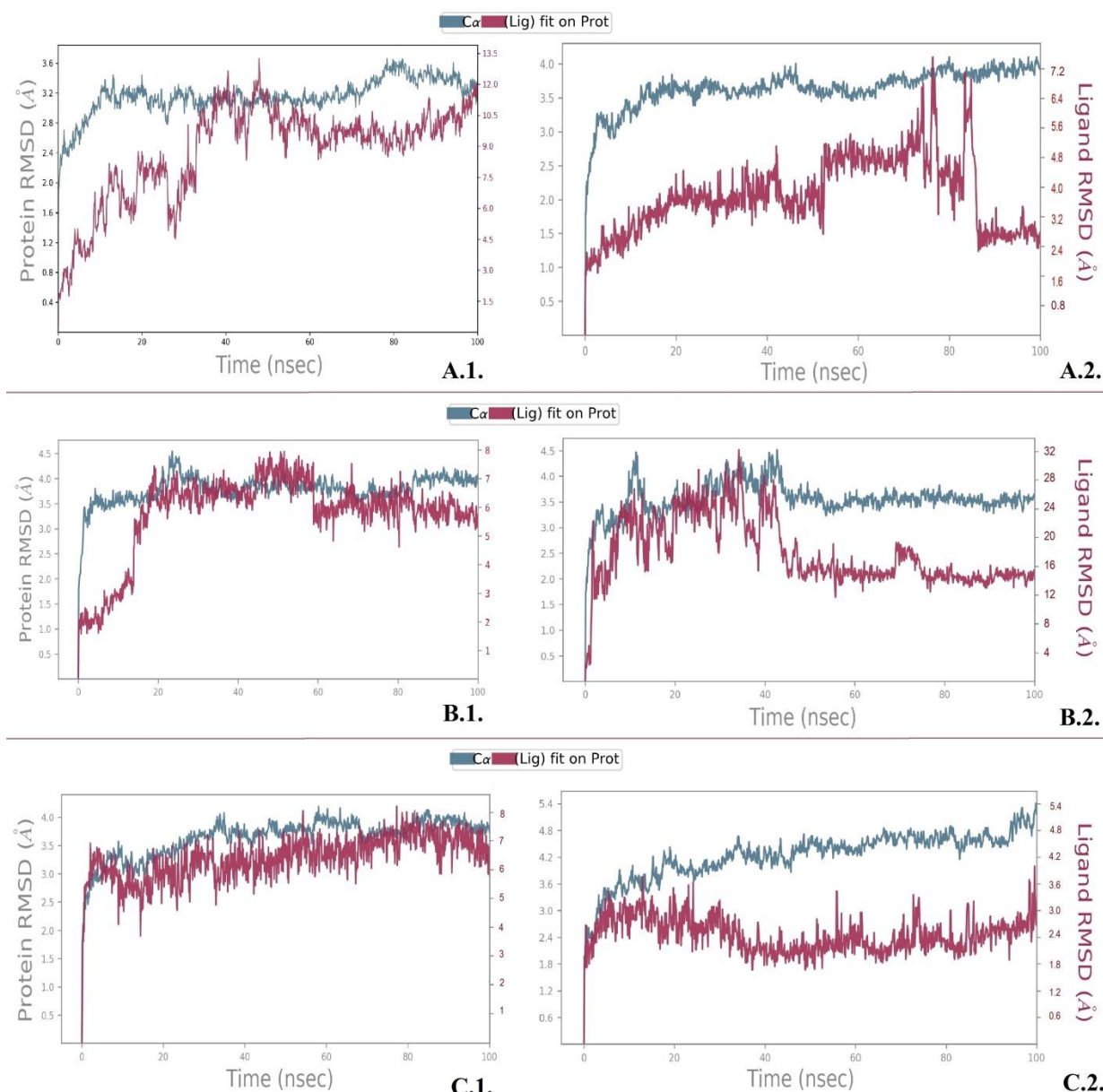
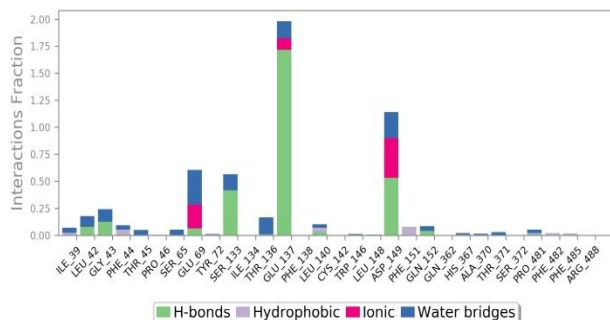


Figure 7: Desmond MD calculated Protein and Ligand RMSD: A.1.: OCT-1 and Metformin, A.2.: OCT-1 and Phenformin, B.1.: OCT-2 and Metformin, B.2.: OCT-2 and Phenformin, C.1.: OCT-3 and Metformin and C.2.: OCT-3 and Norepinephrine

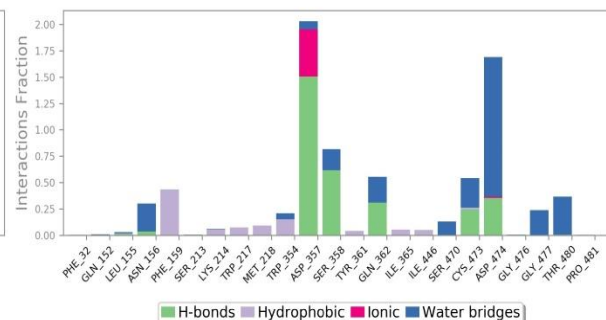
Root mean square deviation (RMD) of protein and ligand was calculated during the MDS with reference to their initial structure. RMSD of the OCTs shows its stable conformation throughout the simulation which providing an indication of the stability of the interaction with metformin and phenformin. Beside OCT3 was stable with norepinephrine throughout the simulation.

Each OCT proteins attained equilibrium in few nsec and remain stable throughout the simulation time up to 100 ncs. Initially, the RMSD plot for metformin attained equilibrium in a few nsec as well and remain stable throughout to stimulation. Some deviations observed but no bigger changes of the order of 1-3 Å are seen in our analysis. Similar RMSD scores were recorded with Phenformin interactions as well.

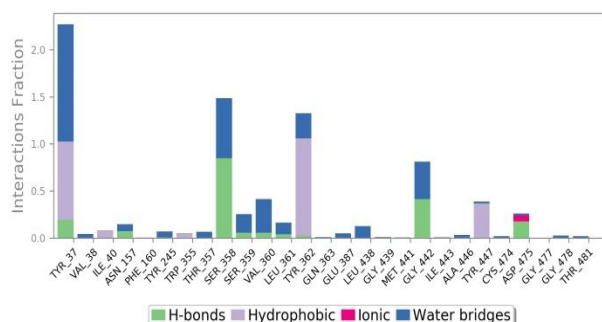
Protein-Ligand Contacts



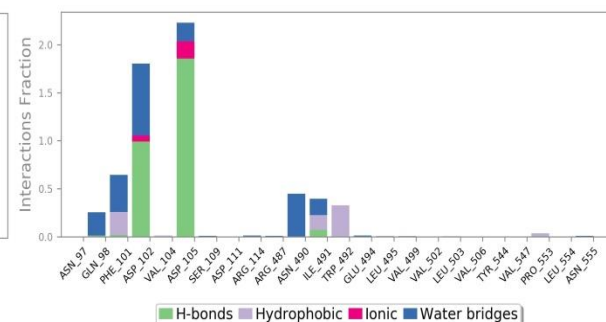
A.1.



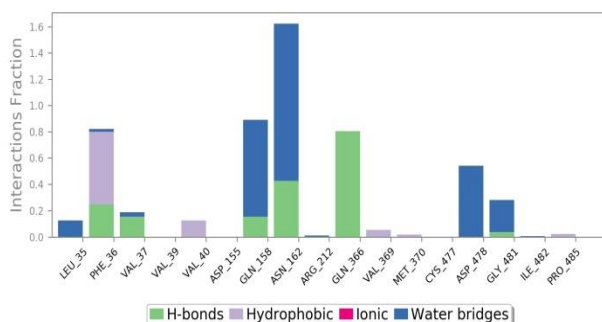
A.2.



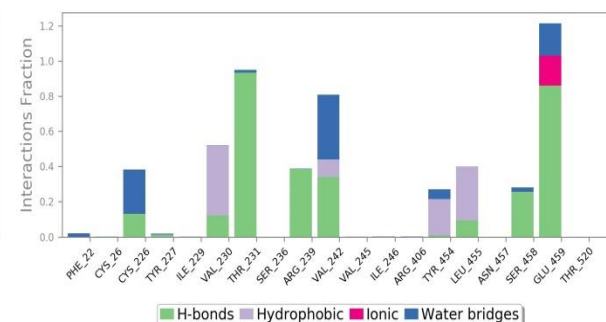
B.1.



B.2.



C.1.



C.2.

Figure 8: Protein-ligand contact interaction profile analyzed for the A.1.: OCT-1 and Metformin, A.2.: OCT-1 and Phenformin, B.1.: OCT-2 and Metformin, B.2.: OCT-2 and Phenformin, C.1.: OCT-3 and Metformin and C.2: OCT-3 and Norepinephrine

OCT1-3 interactions with the ligands were monitored throughout the simulation. Interactions are rich with H bonds that play significant role in ligand binding.

Even though the involvement of human OCT in uptake/transport of metformin is already mentioned in the literature, this study is demonstrating the hOCTs-metformin interaction at atomic level as the first time. This study describes how and where binding occurs. Mimicking this binding with the absence of the structural information of the protein was possible with the unique approach that was described in the pipeline in figure I.

CONCLUSION

The three-dimensional structure of a protein is direct to associate with its comprehensive cellular and biological function. To investigate of identifying the tertiary atomic structure of OCT1-3 proteins and their localization in the cell membrane, it is significant to evaluate the pharmacodynamics of metformin, frequently preferred in Type 2 Diabetes Mellitus medication, by determination of the residues that interact with metformin to the cell translocation of these proteins. One of the important limitations of mimicking protein-ligand interactions is the absence of protein 3D structure. The presented new pipeline is promising especially for the interaction simulation studies that are conducted with proteins with unknown structure. To determine therapeutic effect of Metformin or other life-saving drugs into the cells, further studies are needed to examine genetic variants of human OCTs in specific patient populations. Analyzing insertions, deletions, and other genetic variants effects on hOCTs in structure level is important to explore the role of these proteins in metformin pharmacokinetics and response. Our study could be a front preparation and inspiration for future positively charged drug discoveries and development by examining the atomic level of OCT proteins.

CONFLICT OF INTEREST

The authors declare that they do not have any conflict of interest.

ACKNOWLEDGEMENTS

We are really thankful to Mungo Carstairs and his coworkers from the Barton group for their support in our study in the Jalview 2.11. This work has been funded by grants from the Spanish Ministry of Economy and Competitiveness (CTQ2017-87974), and by Fundación Séneca de la Región de Murcia under Project 20988/PI/18. This research was partially supported by the supercomputing infrastructure of Poznan Supercomputing Center, the e-infrastructure program of the Research Council of Norway via the supercomputer center of UiT—the Arctic University of Norway, and by the and by the super- computing infrastructure of the NLHPC (ECM-02), Powered@NLHPC. This research project has been cofinanced by the European Union (European Regional Development Fund— ERDF)

REFERENCE

1. T. M. Bakheet, A. J. Doig, *Bioinformatics*. **2009**, 25, 451-457. DOI 10.1093/bioinformatics/btp002.
2. L. Fagerberg, K. Jonasson, G. Von Heijne, M. Uhlén, L. Berglund, *Proteomics*. **2010**, 10, 1141-1149. DOI 10.1002/pmic.200900258.
3. S. Tan, T. T. Hwee, M. C. M. Chung, *Proteomics*. **2008**, 8, 3924-3932. DOI 10.1002/pmic.200800597.
4. A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, et al., *Nucleic Acids Res*. **2016**, 45, D200-D203.
5. C. Colas, P. M. U. Ung, A. Schlessinger, *Medchemcomm*. **2016**, 7, 1069-1081. DOI 10.1039/c6md00005c.
6. I. Pernicova, M. Korbonits, *Nat. Rev. Endocrinol*. **2014**, 10, 143.
7. L. Chen, B. Pawlikowski, A. Schlessinger, S. S. More, D. Stryke, S. J. Johns, M. A. Portman, E. Chen, T. E. Ferrin, A. Sali, et al., *Pharmacogenet. Genomics*. **2010**, 20, 687. DOI 10.1097/FPC.0b013e32833fe789.
8. A. T. Nies, H. Koepsell, K. Damme, M. Schwab, *Handb. Exp. Pharmacol*. **2011**, 105-167.

- 422 DOI 10.1007/978-3-642-14541-4_3.
- 423 9. R. Saitoh, K. Sugano, N. Takata, T. Tachibana, A. Higashida, Y. Nabuchi, Y. Aso,
424 *Pharm. Res.* **2004**, *21*, 749–755.
- 425 10. D.-S. Wang, J. W. Jonker, Y. Kato, H. Kusuhara, A. H. Schinkel, Y. Sugiyama, *J.*
426 *Pharmacol. Exp. Ther.* **2002**, *302*, 510–515.
- 427 11. R. J. Shaw, K. A. Lamia, D. Vasquez, S.-H. Koo, N. Bardeesy, R. A. DePinho, M.
428 Montminy, L. C. Cantley, *Science (80-.)*. **2005**, *310*, 1642–1646.
- 429 12. H. D. McIntyre, C. A. Paterson, A. Ma, P. J. Ravenscroft, D. M. Bird, D. P. Cameron,
430 *Aust. N. Z. J. Med.* **1991**, *21*, 714–719.
- 431 13. M. R. Owen, E. Doran, A. P. Halestrap, *Biochem. J.* **2000**, *348*, 607–614.
- 432 14. J. E. Gunton, P. J. D. Delhanty, S.-I. Takahashi, R. C. Baxter, *J. Clin. Endocrinol. Metab.*
433 **2003**, *88*, 1323–1332.
- 434 15. A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, *Nucleic Acids Res.* **2015**, *43*, W389–
435 W394.
- 436 16. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **2015**,
437 *10*, 845.
- 438 17. Y. Song, F. DiMaio, R. Y.-R. Wang, D. Kim, C. Miles, T. J. Brunette, J. Thompson, D.
439 Baker, *Structure* **2013**, *21*, 1735–1742.
- 440 18. S. Raman, R. Vernon, J. Thompson, M. Tyka, R. Sadreyev, J. Pei, D. Kim, E. Kellogg, F.
441 DiMaio, O. Lange, et al., *Proteins Struct. Funct. Bioinforma.* **2009**, *77*, 89–99.
- 442 19. A. Roy, A. Kucukural, Y. Zhang, *Nat. Protoc.* **2010**, *5*, 725.
- 443 20. J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, *Nat. Methods* **2015**, *12*, 7.
- 444 21. J. Yang, Y. Zhang, *Nucleic Acids Res.* **2015**, *43*, W174-W181.
- 445 22. G. Studer, M. Biasini, T. Schwede, *Bioinformatics* **2014**, *30*, i505-i511.

- 446 23. M. Wiederstein, M. J. Sippl, *Nucleic Acids Res.* **2007**, *35*, W407-W410.
- 447 24. J.-F. Gibrat, T. Madej, S. H. Bryant, *Curr. Opin. Struct. Biol.* **1996**, *6*, 377–385.
- 448 25. W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, *Nucleic Acids Res.* **2018**, *46*, W363–W367.
- 449 26. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*,
450 403–410.
- 451 27. H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, R.
452 Lopez, *Nucleic Acids Res.* **2013**, *41*, W597-W600.
- 453 28. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, *Bioinformatics*
454 **2009**, *25*, 1189–1191.
- 455 29. Pdb.-K. consortium, *Nucleic Acids Res.* **2019**. DOI 10.1093/nar/gkz853.
- 456 30. Schrödinger, LLC, *The {PyMOL} Molecular Graphics System, Version~1.8*, **2015**.
- 457 31. S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He,
458 B. A. Shoemaker, et al., *Nucleic Acids Res.* **2016**, *44*, D1202--D1213.
- 459 32. O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455–461.
- 460 33. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng,
461 T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- 462 34. I. Sánchez-Linares, H. Pérez-Sánchez, J. M. Cecilia, J. M. García, *BMC Bioinformatics*
463 **2012**, *13*, S13.
- 464 35. S. Release, others, *Maest. Interoperability Tools*, Schrödinger, New York, NY **2017**.
- 465 36. Q. Wu, Z. Peng, Y. Zhang, J. Yang, *Nucleic Acids Res.* **2018**, *46*, W438-W442.
- 466 37. F. Zi, H. Zi, Y. Li, J. He, Q. Shi, Z. Cai, *Oncol. Lett.* **2018**, *15*, 683–690.
- 467 38. A. T. Nies, U. Hofmann, C. Resch, E. Schaeffeler, M. Rius, M. Schwab, *PLoS One* **2011**,
468 *6*, e22163.

- 469 39. M. Foretz, B. Guigas, L. Bertrand, M. Pollak, B. Viollet, *Cell Metab.* **2014**, 20, 953–966.
- 470 40. T. C. Dakal, R. Kumar, D. Ramotar, *Comput. Biol. Chem.* **2017**, 68, 153–163.
- 471 41. E. C. Chen, N. Khuri, X. Liang, A. Stecula, H.-C. Chien, S. W. Yee, Y. Huang, A. Sali, K.
- 472 M. Giacomini, *J. Med. Chem.* **2017**, 60, 2685–2696.
- 473