JURNAL ILMU KEFARMASIAN INDONESIA, April 2017, hlm. 7-16 Vol. 15, No. 1
ISSN 1693-1831

Molecular Modeling of Human 3β-Hydroxysteroid Dehydrogenase Type 2: Combined Homology Modeling, Docking and QSAR Approach

(Pemodelan Molekular Enzim *3β-Hydroxysteroid Dehydrogenase* Tipe 2: Pemodelan Kombinasi Homologi, *Docking* dan Pendekatan QSAR)

BAMBANG SULISTYO ARI SUDARMANTO* , AGUSTINUS YUSWANTO, RATNA ASMAH SUSIDARTI, SRI NOEGROHATI

Faculty of Pharmacy, Gadjah Mada University, Yogyakarta

Diterima 1 Desember 2016, Disetujui 17 Februari 2017

Abstract: A homology model of human 3β-HSD type 2 has been developed from homology modeling techniques using Phyre2 server and refined by ModRefiner. The PROCHECK, QMEAN and ProSA-web online tools were carried out to evaluate the stereochemical quality of the model. The Ramachandran plot resulted from PROCHECK showed that 84.5% residues are in the most favored region, 13.7% are in the additional allowed region, 1.5% are in the generously allowed region and 0.3% are in the disallowed region. The QMEAN (Z-score) are 0.509 (-3.006) and Z-score of ProSA-web is -7.10. The negative values of protein fold energies also found in almost all sequences. Furthermore, molecular docking was also applied to validate the model using MOE. The hydrogen bonding interactions with Tyr^{154}, Ser^{124}, and Ser^{218} are found in all docked substrates as well as known inhibitors (trilostane and epostane). A dataset of azasteroid inhibitors were also docked into the substrate active site of human 3β-HSD2. These docked structures were utilized to construct corresponding docking-based QSAR equation by employing genetic algorithm (GA) statistical analysis. The contructed best QSAR equation has a robust predictive power according to its statistical parameters, hence may be applied to supersede the default scoring function provided by docking software. These results indicate that the human 3β-HSD2 model was successfully evaluated as a good model.

Keywords: human 3β-HSD2, homology modeling, docking, QSAR.

Abstrak: Model homologi dari enzim 3β-HSD2 telah dikonstruksi menggunakan *server* Phyre2 dan dilanjutkan dengan ModRefiner. Piranti lunak daring PROCHECK, QMEAN dan ProSA-web digunakan untuk mengevaluasi kualitas model stereokimia. *Plot* Ramachandran yang dihasilkan dari PROCHECK menunjukkan bahwa 84,5% residu berada di *most favored region*, 13,7% di *additional allowed region*, 1,5% di *generously allowed region* dan 0,3% di *dissallowed region*. Nilai QMEAN (Z-*score*) adalah 0,509 (-3,006) dan Z-*score* dari ProSA-web adalah -7,10. Nilai negatif pada energi *folding* protein juga ditemukan di hampir seluruh sekuens. Selanjutnya, penambatan molekuler juga diterapkan untuk memvalidasi model menggunakan program MOE. Interaksi ikatan hidrogen dengan Tyr^{154}, Ser^{124} dan Ser^{218} ditemukan disemua substrat yang ditambatkan, seperti halnya di senyawa-senyawa inhibitor yang telah dikenal (trilostane dan epostane). Dataset *inhibitor* azasteroid juga ditambatkan ke situs aktif substrat pada enzim 3β-HSD2. Struktur yang tertambatkan digunakan untuk membangun persamaan QSAR berbasis penambatan molekuler dengan menerapkan analisis statistik *genetic algorithm* (GA). Persamaan QSAR terbaik yang terkonstruksi memiliki daya prediksi yang kuat sesuai dengan parameter statistiknya, sehingga dapat diaplikasikan untuk menggantikan fungsi *scoring default* yang disediakan oleh program MOE. Hasil ini menunjukkan bahwa model enzim 3β-HSD2 manusia berhasil dievaluasi sebagai model yang baik.

Kata kunci: enzim 3β-HSD2, pemodelan homologi, penambatan molekuler, QSAR.

* Correspondent author, Phone.
e-mail: arie_sudarmanto@ugm.ac.id

INTRODUCTION

THE 3B-HYDROXYSTEROID dehydrogenase/Δ5-4-isomerase (3β-HSD) is a bifunctional dimeric enzyme required in steroid biosynthesis pathway which catalyzes the conversion of 3β-hydroxy-Δ5-ene-steroid into the corresponding 3-oxo-Δ4-ene-steroid by sequential oxidation and isomerization, by means, oxidation of the hydroxyl group at C3 into the keto group (EC 1.1.1.145) subsequently isomerization of Δ5→Δ4 (EC 5.3.3.1). This enzyme plays a crucial role in the biosynthesis of all classes of hormonal steroids, including sex hormones (androgens, estrogens, progestins) and corticosteroids (aldosterone and cortisol)[1,2,3].

The substrates include dehydroandrosterone (which is converted into androst-5-ene-3,17-dione), pregnenolone (converted into progesterone) and 17α-hydroxypregnenolone (converted into 17α-hydroxyprogesterone). The human 3β-HSD is a single-pass membrane protein located in endoplasmic reticulum and mitochondria, consist of two isoenzymes, designed as type 1 and type 2, encoded by the HSD3B1 and HSD3B2 genes, respectively. The type 1 isoenzyme (consisting 373 amino acids, 42.252 kDa) is expressed in placenta and peripheral tissues, including prostate, mammary glands and skin. Whereas the type 2 isoenzyme (consisting 372 aminoacids, 42.052 kDa) is predominantly expressed in the adrenal glands and gonads (ovary and testis)[4].

Understanding the relationship between protein structure and functional properties provides new opportunities to improve protein functionality at the molecular level. Consequently, in order to understand the protein functions in biological system, knowing how the amino acids are distributed in three-dimensional (3D) space is also necessary. The crystal structure of 3β-HSD has never been available yet, due to the difficulty of membrane-bound enzyme isolation, purification, crystallization and structure determination. In the absence of the experimental structures, sequence homology methods are employed to build a protein 3D structure based on the understanding that proteins which share sequence similarity would also have homologous structure and function, so called homology modeling.

From the structural standpoint, the only information available mostly concern about the protein main structural motifs, which include two transmembrane-spanning domain (residues 72-89 and 283-310)[5]. Importantly, the 3β-HSD2 is a member of short-chain dehydrogenase/reductase (SDR) family of NAD- or NADP-dependent oxidoreductase, annotated as SDR11E. Alternative procedures such as fold recognition for proteins that share low sequence homology are compared to similar 3D structure of known proteins and ab-initio modelling methods may also be employed. The computational approaches to predicting protein structures are a valuable tools, so that the structure-based research become rational. Therefore, a number of three dimensional structure models of 3β-HSD have been proposed for type 1[6] and type 2[5,7] using various approaches.

In addition, molecular modeling with a wide range of computational techniques have become an important tool to learn the interaction between a small molecule and its macromolecule target in molecular state. Nowadays, common molecular modeling methods such as homology modeling, docking and QSAR become the popular computational technique and has been widely applied as a very useful tools in drug design and target identification, so that the mechanism of action can be studied. In this work, a series of computational methods including homology modeling, docking and QSAR were integrated to explore human 3β-HSD2 enzyme.

MATERIAL AND METHODS

MATERIAL. The amino acid sequence of human 3β-HSD2 enzyme was retrieved from Uniprot database (www.uniprot.org)[8], in FASTA format with accession code P26439, contains 372 amino acids. The crystal structure coordinate of protein templates were downloaded from Protein Data Bank (PDB, www.rcsb.org/pdb/)[9]. The 3β-HSD2 inhibitors were extracted from ChEMBL (www.ebi.ac.uk/chembl/)[10], accession code CHEMBL615557 and CHEMBL615558 (verified by corresponding journals), so that the inhibitors dataset contain 98 ligands (Table 1)[11]. The structure of substrates, i.e. pregnenolone (CAS Number 145-13-1), 17α-hydroxypregnenolone (CAS Number 387-79-1) and dehydroepiandrosterone (CAS Number 53-43-0) and known inhibitors, i.e. trilostane (CAS Number 13647-35-3) and epostane (CAS Number 80471-63-2) were obtained from public databases.

METHODS. Homology model of human 3β-HSD2 was built in Phyre2 (Protein Homology/analogY Recognition Engine V2.0) server (www.sbg.bio.ic.ac.uk/phyre2/)[12]. Molecular modeling, including ligand molecules preparation, docking and QSAR studies, was performed by Molecular Operating Environment (MOE)[13] software suite.

Protein Modeling of 3β-HSD2. The conserved domains with functional significance were predicted using the server, InterPro (http://www.ebi.ac.uk/interpro/)[14], where it uses predictive models, known

Table 1. The best suited templates detected from Phyre2 server, that produce models of 100% confidence.

#	Template Information					Alignment coverage		%i.d.
	PDB ID	Chain	Res.(Å)	Protein Name	Clasification	Sequence	%-tage	
1	1Z45	A	1.85	UDP-galactose 4-epimerase (*Saccharomyces cerevisiae*)	Isomerase	2-358	95%	22%
2	1Z7E	C	3.00	Protein ArnA	Hydrolase	2-356	95%	16%
3	1OC2	A	1.50	Enzyme DTDP-D-Glucose 4,6-Dehydratase (Rmlb)	Lyase	1-372	99%	15%
4	2HUN	B	2.07	dTDP-glucose 4,6-dehydratase (*Pyrococcus horikoshii*)	Lyase	1-370	99%	19%
5	1N7G	B	2.20	GDP-D-mannose-4,6-dehydratase (*Arabidopsis thaliana*)	Lyase	2-368	98%	14%
6	1N7H	A	1.80	GDP-D-mannose-4,6-dehydratase (*Arabidopsis thaliana*)	Lyase	2-368	98%	14%
7	2V6G	A	2.30	Progesterone 5{Beta}-Reductase (*Digitalis lanata*)	Oxidoreductase	2-359	95%	10%
8	4LIS	A	2.80	UDP-galactose-4-epimerase (*Aspergillus nidulans*)	Isomerase	2-365	97%	16%
9	1RKX	A	1.80	CDP-glucose-4,6-dehydratase (*Yersinia pseudotuberculosis*)	Lyase	2-372	99%	16%
10	2B69	A	1.21	UDP-glucuronate decarboxylase 1 (human)	Lyase	2-363	97%	13%
11	4EGB	C	3.00	dTDP-glucose 4,6-dehydratase (*Bacillus anthracis*)	Lyase	2-367	98%	16%
12	2Z1M	C	2.00	GDP-D-mannose dehydratase (*Aquifex aeolicus*)	Lyase	2-364	97%	14%
13	1T2A	A	1.84	GDP-D-mannose 4,6-dehydratase (human)	Lyase	2-362	96%	12%
14	1I24	A	1.20	UDP-Sulfoquinovose Synthase (*Arabidopsis thaliana*)	-	2-371	99%	13%
15	1DB3	A	2.30	GDP-mannose 4,6 dehydratase (*Escherichia coli*)	Lyase	2-363	97%	13%
16	1R6D	A	1.35	TDP-glucose-4,6-dehydratase (*Streptomyces venezuelae*)	Lyase	1-365	97%	20%
17	1WVG	A	1.80	CDP-glucose 4,6-dehydratase (*Salmonella enterica*)	Lyase	2-372	99%	15%
18	1RPN	A	2.15	GDP-D-mannose 4,6-dehydratase (*Pseudomonas aeruginosa*)	Lyase	3-363	96%	17%

as signatures, provided by different databases such as Prosite[15], ProDom[16], Gene3D[17], Pfam[18], Prints[19] and SuperFamily[20] that make up the InterPro consortium.

The amino acid sequence of human 3β-HSD2 retrieved from Uniprot database was subjected to Phyre2 server, thereof the three-dimensional (3D) structure of the protein would be constructed, which employ the high accuracy of prediction based on the Hidden Markov Models (HMM)[21] and also construct the regions with poor homology by ab initio modeling method (called Poing)[22] and afterwards joined to generate the final model. To optimize the binding site, cofactor NAD^+ and ligand were inserted to the refined model of human 3β-HSD2 by template-based superposition. The final optimized model retrieved from Phyre2 server was subjected to ModRefiner (http://zhanglab.ccmb.med.umich.edu/ModRefiner)[23] to refine the structure. Furthermore, the human 3β-HSD2 complexed with NAD and ligand was further refined in MOE program, in which Protonate3D[24], energy minimized and atomic charge assignment by means of AMBER10 forcefield, and further LowMD[25] conformational searching modules were employed.

Finally, the quality of the protein models were validated using two servers: PROCHECK (http://swissmodel.expasy.org/workspace/index.php?func=tools_structureassessment1)[26], including QMEAN and ProSA-web server (https://prosa.services.came.sbg.ac.at/prosa.php)[27]. PROCHECK determines the residue-wise stereochemical quality of the model and generates a Ramachandran plot to explain its overall quality, whereas ProSA-web calculates the overall quality of the model as Z-score based on the atomic coordinates and indicates the values in a Z-score plot of experimentally determined structures from PDB database.

Ligand Preparation. Preparation of ligands for docking studies was performed by MOE program. The extracted dataset from ChEMBL contains structural information of ligands in SMILES code, then the 3D structures of each ligands were constructed by MOE-Builder module and then subjected to a conformational searching. The lowest energy conformation of each ligands were minimized using AM1 semiempirical method until a convergence value of 0.01 kcal/Å.mol. Furthermore, all optimized structures were aligned, stored in a database file and saved in *.mdb format for molecular docking studies. The substrates and known inhibitors (trilostane and epostane) also prepared in similar way.

Validation of Protein Model by Molecular Docking. Molecular docking studies were performed using MOE-Dock module of MOE program to evaluate the protein model. The final modeled protein structure of human 3β-HSD2 was prepared using LigX module, includes hydrogen atoms addition via Protonate3D and atomic charges assignment using AMBER10 forcefield. First, cofactor NAD^+ (Nicotinamide Adenine Diphosphate) as well as the substrates (viz. pregnenolone, 17α-hydroxypregnenolone and androstenedione) were docked into their binding site (pocket), respectively. The binding pocket in the preliminary human 3β-HSD2 model was determined by superposition against the crystal structure of templates containing NAD^+ and ligand, and using this ligands position as the center of the binding site. Thus, the binding pockets were defined as a space within 5 Å region from ligand (cofactor/substrate) lined by amino acid residues. Furthermore, all ligands contained in the inhibitors dataset were docked into the substrate binding site. For each ligands in the inhibitors dataset, the best docked conformation were chosen from the

correct pose, then were scored using scoring functions, London dG, Affinity dG, Alpha-HB, ASE, GBVI/WSA dG and pKi_Andrews.

Docking-based QSAR Study. The best docked pose of 60 compounds of 3β-HSD2 inhibitors were used in QSAR study as dataset. To derive the QSAR models, 2D-, 3D- and COMBINE descriptors were calculated in MOE, and used as independent variables, while the pKi values as dependent variable. Multiple linear regression and variable selection were performed by QuaSAR-Evolution SVL script that implemented in MOE program, which employ a Genetic Algorithm based QSAR analysis. The predictive ability of the models expressed by r2, RSME, F, LOF and cross-validation Q2 values as the statistical criterions.

RESULTS AND DISCUSSIONS

Homology Modeling of Human 3β-HSD2. The human 3β-HSD2 protein is a member of SDR family. Due to the experimentally determined structure for 3β-HSD2 are not yet available in the Protein Data Bank, it is important to construct a 3D structure model of this protein. The 3D structure of protein give important insights about the molecular basis of protein function and thereby allow an effective design of experiments. That is why, in the understanding and manipulation of biochemical and cellular functions of proteins, the high-resolution 3D structure of a protein is the main key.

In this work, the 3D structure of human 3β-HSD2 has been modeled using the available protein sequence query for homology based modeling. Web-based server Phyre2 was used for homology modeling of human 3β-HSD2 through sequential steps, such as profile construction, similarity analysis and structural properties, Phyre2 selects the best suited templates and generates the protein models. Phyre2, the most popular online protein fold identification server, utilizes several datasets of known proteins taken from different reliable databases, such as Structural Classification of Proteins (SCOP) database and Protein Data Bank (PDB) and uses the Hidden Markov Method (HMM)[21] to generate alignments of a submitted protein sequence against proteins with known 3D structures. The resulting alignments against each suited identified templates are then used to produce homology-based models of the query sequence to predict its 3D structure. In addition, Phyre2 uses the following component software: Psi-pred[28] for secondary structure prediction, Disopred [29] for disorder prediction, Memsat SVM[30] for transmembrane prediction and an ab initio folding simulation called "Poing"[22] to model regions of the query with no detectable similarities to known structures, thereafter combine multiple templates of known structures to produce a final 3D structure model of the query sequence.

Phyre2 successfully detects 120 templates and build models based on 18 top rank templates (Table 1), in which the highest similar structure of PDB ID 1Z45.A (which corresponds to crystal structure of UDP-galactose 4-epimerase from *Saccharomyces cerevisiae* complexed with NAD, UDP-galactose and galactose, family of isomerase) with coverage of 95% wherein 22% identity of the sequences. Among them, six templates (PDB ID 1Z45, 1Z7E, 1OC2, 2HUN, 1N7H and 1N7G) were selected to model human 3β-HSD2 protein based on heuristics to maximize confidence, percentage identity and alignment coverage. According to the multiple sequence alignment (Fig. 1A), it has been successfully detected several conserved residues, mainly a Rossmann-fold domain ((β-α-β-α-β-α-β) domain signatured by glycine-rich 8TG-xx-G-xx-G15 motif located near N-terminal and the catalytic triad composed of Ser^{124}, Tyr^{154} and Lys^{158} at the 154Y-xx-xK158 motif. Both motifs are in state of conserved among the SDR family[31]. The 3β-HSD type 1 and 2 also have an additional 269Y-xx-S-K273. Moreover, conserved residues also found in Asp^{35}, Asp^{61}, Pro^{153}, Leu^{236}, Gly^{276}, Leu^{337} and Trp^{355} in most of templates.

Based on this alignments, about 358 (96%) residues were homology modeled based on six aforementioned templates, whereas 14 (4%) residues were modeled by ab initio, viz. the sequences near the C-terminal that fail to find the homology (belonging to membrane-spanning domain). The generated final 3D structure model had 100% confidence level. Superimposition of homology model of human 3β-HSD2 against the selected used templates are shown in Fig. 1B. Their topologies were quite similar, and most of secondary structural motifs, characteristically the Rossmann-fold motif. The 3D structural model of human 3β-HSD2 is overall similar to those of other SDR family's proteins taken as templates.

After model building, it is necessary to further refinement in quest of the best model generation to gain the more accurate model. Refinement for the final model obtained from Phyre2 server was generated by ModRefiner. It construct and refine full-atom model based on Cα traces through two steps refinement procedure. First, it constructs a main-chain model from the Cα trace with an acceptable backbone topology and main-chain hydrogen-bonding network. In the second step, side-chain atoms are added onto the backbone conformation and optimized with the

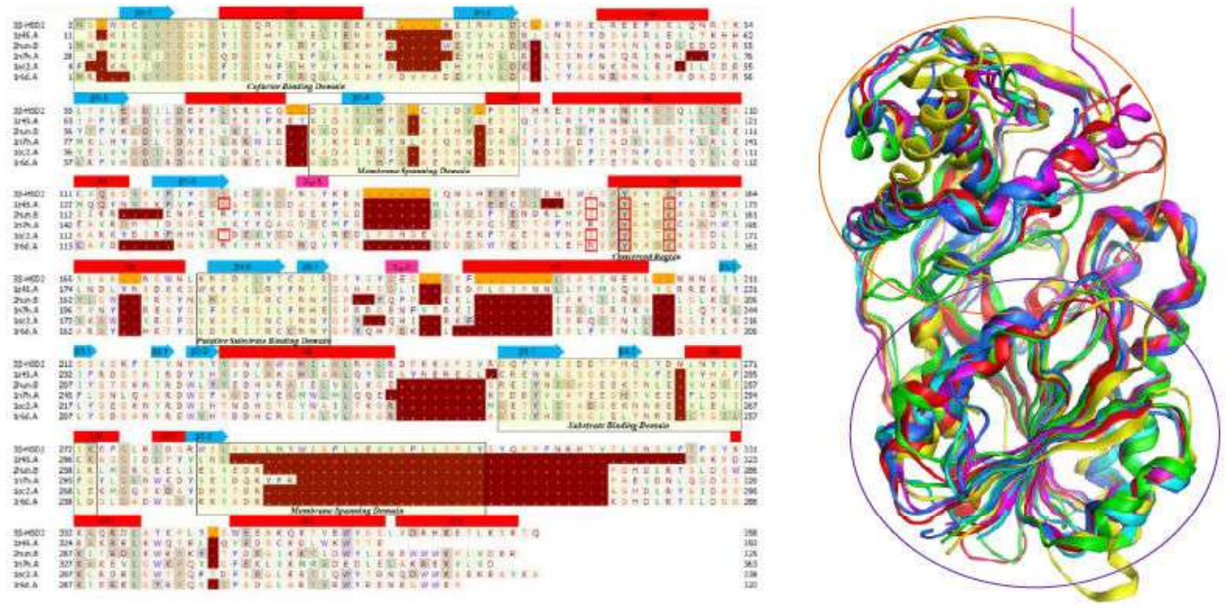

Fig. 1. (A) Multiple sequence alignment and secondary structure assignment of human 3β-HSD2 with selected templates (PDB code 1z45.A, 2hun.B, 1n7h.A, 1oc2.A, 1r6d.A). Secondary structure (PROCHECK) is displayed above the sequences: α-helices (red bars), 310-helices (pink bars), strands (cyan arrows). Domains are signed by black box. Red box point out the detected catalytic residues in crystal structures template. Residue numbering is indicated on the left of each arrow. The more identical residue are darker highlighted. (B) Superimposition of homology model of human 3β-HSD2 (red) with 1z45.A (green), 2hun.B (blue), 1n7h.A (yellow), 1oc2.A (magenta), and 1r6d.A (cyan). Cofactor NAD+ binding Rossmann-fold domain and substrate binding domain are indicated inside the purple and orange ellipses, respectively.

use of a composite physics-and knowledge-based forcefield[23].

Quality Assessment of the Human 3β-HSD2 Model. Protein model evaluation holds an important role in protein structural prediction since ultimately it is used to design further computational approach and understand the protein's biological function. Quality assessment of human 3β-HSD2 model protein was performed using PROCHECK and ProSA-web.

PROCHECK was used to check the stereochemical quality of protein structure, provide a Ramachandran plot for generated pdb structure of human 3β-HSD2 (Fig. 2A). Total number of non-glycine and non-proline residues in human 3β-HSD2 model is 329. Protein backbone conformations were evaluated by inspection of Ramachandran plot, which is an X-Y plot of phi/psi dihedral angle between N-Cα and Cα-C planar peptide bonds at the protein backbone. Both these angles are able to rotate freely in proteins (−180 to +180). Any combination of these angles is theoretically possible, but in actual, biological conditions many combinations are rarely or never seen due to steric clashes in the protein's backbone structure. The Ramachandran plot for human 3β-HSD2 shows around 278 (84.5%) residues placed in favored regions, 45 (13.7%) residues in additional allowed regions, 5 (1.5%) residues in generously allowed regions and 1 (0.3%) residues in disallowed regions.

The QMEAN Z-Score, also calculated from PROCHECK, provides an estimate of the absolute quality of a model by comparing it to same sized reference structures present in the PDB and solved by experimental techniques. It was used to estimate the "degree of nativeness" of predicted structure[32]. The QMEAN Z-score for human 3β-HSD2 model was 0.509 (Z-Score: -3.006). It was observed that C_beta interaction energy (−45.02; Z-score:−2.01), all-atom pairwise energy (−4471.35; Z score: −2.57), solvation energy (−16.47; Z-score: −1.95), torsion angle energy -8.79; Z score: −4.26), secondary structure agreement (75.8%; Z-score: -1.06) and solvent accessibility agreement (73.1%; Z-score: -1.50).

The quality of the protein fold was inspected with ProSA-web server. This calculates the energy need for the architecture of protein folds as an function of amino acid sequence. Folding energies of protein have generally negative values, and since these values correspond with the stability and nativity of the molecule. It can be seen form Fig. 2B that the negative values found in almost all parts of sequence. In addition, ProSA-web also calculate the Z-score value, that indicates of overall model quality. Its value is displayed in a plot that contains the X-score of all experimentally determined protein chains in current PDB. The final Z-score of -7.10 was obtained (Fig. 2C). In general, the more negative the Z-score is, the more accurate the model is likely to be.

According to aforementioned values, these

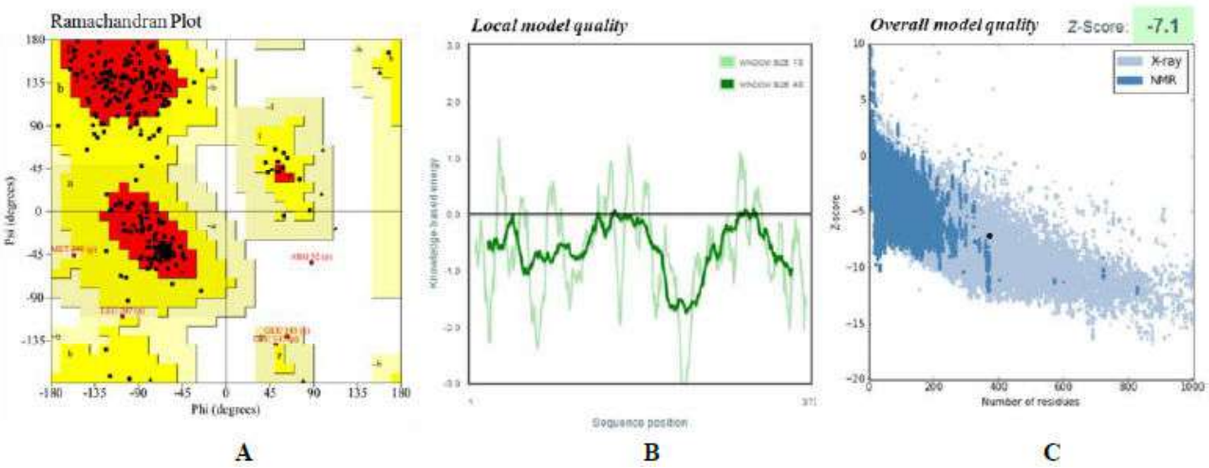

Fig. 2. Stereochemical quality of protein model. (A) Ramachandran plot of the predicted model of human 3β-HSD2, generated in PROCHECK. The most favored regions are colored red, additional allowed, generously allowed, and disallowed regions are indicated as yellow, pale yellow, and white area, respectively. (B) The ProSA-web Z-score value of overall human 3β-HSD2 model quality, displayed in plot that contain Z-scores of all experimentally determined protein chain in current PDB. (C) The ProSA-web plot of residue scores of human 3β-HSD2 model, shows local model quality by plotting energies as function of amino acid sequence position.

suggest that Phyre2 server was capable to generate a reasonably good model of human 3β-HSD2 based on its stereochemical quality. The refined final model of human 3β-HSD2 complexed with NAD^+ and substrates are depicted in Fig. 3A, while its two-dimensional topology is depicted in Fig. 3B. The modeled 3D structure of 3β-HSD2 had 13 (15.6%) strands, 12 (40.3%) α-helices and 2 (1.6%) 3_{10}-helices. The strands that builds the β-sheets are classified into 4 types, wherein β1 (parallel), β2 (parallel), β3 (parallel) and β4 (anti-parallel).

The human 3β-HSD2 comprises of two domain, the first Rossmann-fold domain binding the cofactor, encompassing a cofactor NAD^+-binding site and the second catalytic domain binding the substrate, encompassing a SBS. The Rossmann-fold domain binds NAD^+ that reversibly accepts a hydride ion, which is lost or gained by substrate in the redox reaction. The Rossmann-fold domain have an α/β folding pattern, which has a central β-sheet, composed of six α-helices found surrounding the seven mostly parallel β-sheet, like a sandwich. The strands forming the β-sheet are found in the following characteristic order 654123. The inter sheet crossover of the strands in the sheet form the cofactor NAD^+ binding pocket[32]. The cofactor-related motif of Gly-xx-Gly-xx-Gly was also conserved in the human 3β-HSD2 model. The secondary structure of the human 3β-HSD2 model exhibited a central 7-stranded all-parallel β-sheet sandwich-like structure, flanked on both sides by 4-helices (α A, α B, α H, α K) and 3-helices (αC, αE, αF), respectively (Fig. 3A).

The catalytic domain confers substrate specificity and contains amino acids involved in the precise catalysis reaction of the enzyme. The substrate binding site was characterized as non-metallo-oxidoreductase site and contained the conserved Ser^{124}, Tyr^{154} and Lys^{158} catalytic triad. It has been proposed that Ser^{124} is associated with catalysis by stabilizing the reaction intermediates, that the Tyr^{154} hydroxyl group is the proton donor involved in the electrophilic attack on the substrate carboxyl group in a reduction reaction, and that Lys^{158} facilitates the proton transfer from the hydroxyl oxygen of Tyr^{154} to the substrate.

Validation of the Human 3β-HSD2 Model using Molecular Docking. To address the predictive quality of the generated homology model, it should be evaluated in docking procedure. Molecular docking is a computer simulation procedure to calculate the conformation of a receptor-ligand complex. It is used to identify the binding affinity and interaction energy of the molecules with the target protein. Docking analysis was performed by MOE, which is an automated procedure for predicting the interaction of ligands with bio-macromolecular targets. Docking studies were performed to explore the binding modes of the native substrates as well as known inhibitors in their binding site of the human 3β-HSD2 model.

The binding modes of cofactor NAD^+, substrates and known inhibitors can be seen in Fig. 4. NAD^+ is tightly bind in its binding pocket (Fig. 4A). Adenosyl part is clamped between the H1-helix, and the ends of the β1-1, β1-2, β1-3 and β1-4 strands, hence this orientation allows the formation of hydrogen bond between the adenosyl N-amine atom and the backbone's carbonyl of Asn^{100} (2.9 Å). The diol moiety of the adenosyl A-ribose makes a hydrogen bond with Tyr^{190} (2.6 Å). The diphosphate moiety of NAD^+ is

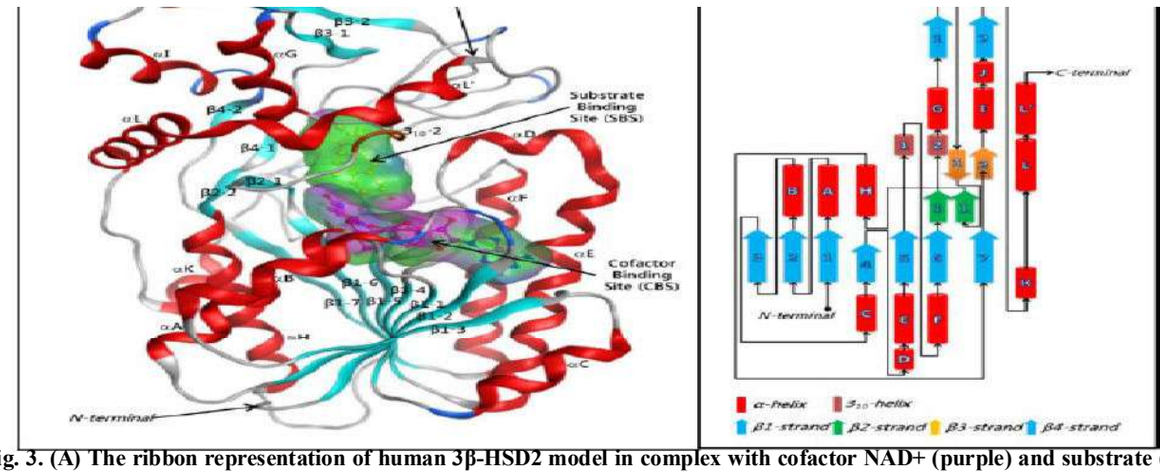

Fig. 3. (A) The ribbon representation of human 3β-HSD2 model in complex with cofactor NAD+ (purple) and substrate (yellow). The secondary structures are α-helix (red) and β-sheet (cyan) assigned by numbered H and β, respectively, whereas coil (gray), and turn (blue). Colored ligand binding site surfaces are indicated hydrophobic (green), H-bond (magenta). (B) The 2D-topology of human 3β-HSD2 model.

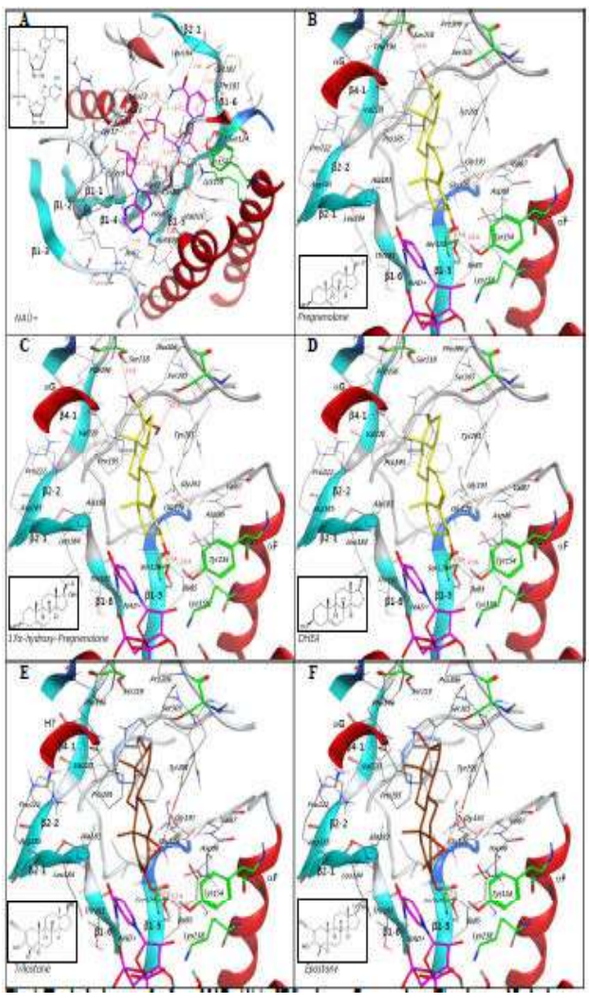

Fig. 4. The docked poses of cofactor NAD+ (A) in CBS, substrate Pregnenolone (B), substrate 17α-hydroxypregnenolone (C), substrate Dehydroepiandrosterone (D), inhibitor Trilostane (E), and inhibitor Epostane (F), in SBS. The substrates are colored yellow, NAD+ is colored magenta, residues Tyr154, Lys158, Ser124, Ser218, and Ser305 that form hydrogen bonding are colored green, while other residues colored grey.

buried completely inside in the pocket and make three strong hydrogen bonds with the backbone amide atoms of Leu^{13} (2.4 Å), Leu^{14} (2.4 Å) and Gly^{15} (2.9 Å) in the glycine-rich motif. The nicotinic carbonyl makes a hydrogen bond with CO-backbone of Leu^{184} (3.2 Å). The strong interactions occur on the diol moiety of the nicotinic ribose, where O_2 forms three strong hydrogen bonds with the sidechains of catalytic residues Tyr^{154} (2.7 Å), Lys^{158} (2.5 Å) and Ser^{124} (2.9 Å), while O_3 forms also three hydrogen bonds with backbone of Thr^{81} (2.6 Å), backbone of Thr^{122} (3.6 Å) and sidechain of Lys^{158} (3.2 Å). In addition, NAD^+ is tied up by the hydrophobic residues, particularly for adenosine and pyridyl rings (involving residues Phe^{38}, Ala^{82}, Ile^{85}, Thr^{181} and Leu^{184}). The important interactions are summarized in Table 2.

The β-face of all docked substrates (pregnenolone,

Table 2. Hydrogen bond interaction of NAD^+ complexed in cofactor binding pocket of human 3β-HSD2.

H-bond interactions				
Parts of NAD^+	**Residue**	**Distance (Å)**	**%-tage**	**Interaction with**
Adenosiyl-NH_2	Asn^{100}	2.9	43%	CO-backbone
Adenosyl-Ribose-OH	Tyr^{190}	2.6	72%	OH-sidechain
Phosphate1-O2	Leu^{13}	2.4	10%	NH-backbone
Phosphate2-O2	Leu^{14}	2.4	86%	NH-backbone
	Gly^{15}	2.9	78%	NH-backbone
Nicotinamide-Ribose-O3	Thr^{81}	2.6	20%	CO-backbone
	Thr^{122}	3.6	6%	CO-backbone
	Lys^{158}	3.2	33%	NH3-sidechain
	Tyr^{154}	2.7	59%	OH-sidechain
Nicotinamide-Ribose-O2	Lys^{158}	2.5	42%	NH3+-sidechain
	Ser^{124}	2.9	17%	OH-sidechain
Nicotinamide-CO	Leu^{184}	3.2	53%	NH-backbone

Hydrophobic interaction	
NAD^+ part	**Residues**
Nicotinamide-Pyridyl-ring	Ile^{85}, Thr^{181}, Leu^{184}
Adenosine-core	Phe^{38}, Ala^{82}

17α-hydroxypregnenolone, and androstenedione) come near to the pyridyl ring of NAD^+ (Fig. 4B,C,D). It is understandable since the catalytic center occurs at this part. The O-H•••O hydrogen bond between the β-hydroxyl group at C3-ring A of substrate and the sidechain of Tyr^{154} (~2.9 Å) was observed in all docked substrate system, which has been reported to be very important for binding and catalytic reaction by other researcher[4,6]. The residue Ser^{124} as part of the catalytic residues also constitute on substrate binding through a hydrogen bond (~2.5 Å) with this hydroxyl group. Moreover, it was found that the residues Ser^{218} and Ser^{305} could also form hydrogen bond interactions with carbonyl and hydroxyl groups at ring E substrate. The hydrophobic interactions also contribute in substrate binding, which Leu^{184}, Cys^{182}, Ala^{183}, Pro^{195} and Phe^{196} on the β-face, whilst Val^{87}, Ile^{125}, Glu^{126} and Tyr^{291} on α-face of steroids (Table 3).

The competitive mode of inhibition of human 3β-HSD2 by trilostane and epostane may be due to the overlapping of substrate steroids and these as shown by docking results in Fig. 4E,F. The similar binding mode also observed in the docked pose of the known inhibitors, trilostane and epostane. According to the docking results, it was confirmed that the modeled 3D structure of human 3β-HSD2 was reasonable.

Table 3. H-bond interaction of substrates (pregnenolone, 17α-hydroxypregnenolone, androstenedione) complexed in substrate binding pocket of human 3β-HSD2.

H-bond interactions				
Steroid part	**Residue**	**Distance (Å)**	**%-tage**	**Interaction with**
β-OH at C3-ring A (all substrate)	Tyr^{154}	2.9	48%	OH-sidechain
	Ser^{124}	2.5	58%	OH-sidechain
β-CO at C17-ring E (Preg, 17OH-Preg)	Ser^{218}	3.3	24%	OH-sidechain
α-OH at C17-ring E (17OH-Preg)	Ser^{305}	3.2	15%	OH-sidechain

Hydrophobic interaction	
Steroid part	**Residues**
β-face	Cys^{182}, Ala^{183}, Leu^{184}, Pro^{195}, Phe^{196}
α-face	Val^{87}, Ile^{125}, Glu^{126}, Tyr^{291}

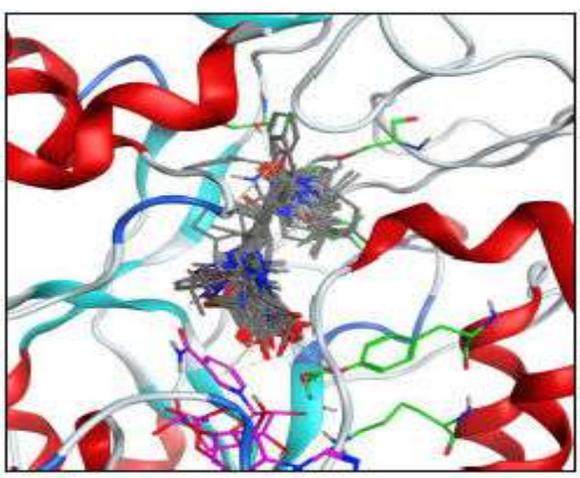

Fig. 5. The aligned inhibitors docked in the substrate binding pocket. The colored rules are similar in Fig. 4.

Docking of the Inhibitors Dataset. Among 98 ligands containing in dataset, only 50 ligands successfully docked in the substrate binding site. It occurs because the binding site is closed by a sheet formed β3-1 and β3-2 conformations and also restricted by membrane-spanning domain as well as coils connected it, so it can not accommodate the ligands containing a bulky moiety. However, all docked inhibitors have similar poses to the substrates. All docked inhibitors containing hydroxyl or carbonyl group at C3 ring A also make strong hydrogen bond interactions toward the residues Tyr^{154} and Ser^{124}. Additional interaction via hydrogen bonding also happens toward the residue Ser^{218}.

The various scoring function were utilized for the docked inhibitors, involving London dG, Affinity dG, GBVI/WSA dG (energy-based descriptors), Alpha-HB, ASE and pKi Andrews (knowledge-based descriptors). Unfortunately, all scoring function does not show a strong relationship with inhibitory activity pKi, where $R2 < 0.5$ (Fig. 6). Therefore, we need a new scoring function which have more powerful prediction ability. This could be achieved by deriving a QSAR equation, since mostly scoring function are created by regression-based.

Docking-based QSAR Analysis. The regression summary are reported in Table 4. Each models use different descriptor set. The model that employ combination of all descriptor types was successfully generate the best equation, as indicated from its statistical criterions, i.e. determination coefficient $R^2 = 0.8682$, root mean square error RMSE = 0.3433, Fisher criterion F = 28.7441, lack of fit LOF = 0.2939 and cross-validated leave-one-out Q2 = 0.7834. The best QSAR equation involves the following descriptors, pKi Andrews score, Asp^{86}_V, Gly^{194}_E, Gly^{194}_V, Ser^{218}_V, Thr^{327}_V, Tyr^{291}_E, Val^{127}_V, h_ema, h_emd, vsurf_IW3. The E and V suffix at residues denote electrostatic and hydrophobic interactions, respectively. The h_ema and h_emd denote strength of acceptor and donor, respectively, while vsurf_IW3 denote hydrophilic integy moment at -0.6.

CONCLUSION

Homology modeling was used to predict the 3D structure of human 3β-HSD2 using crystal structure of UDP-galactose 4-epimerase from *Saccharomyces cerevisiae* as the best suited template in Phyre2 server.

Table 4. Summary of QSAR result.

#	QSAR Models	Descriptor Type			Desc.	Statistical Parameters				
		SF	COMBINE	2D/3D	Total	R^2	RMSE	F	LOF	Q^2
1	COMBINE-SF_{AdG}	1	5		6	0.5917	0.6043	12.8024	0.5706	0.5343
2	COMBINE-SF_{LdG}	1	8		9	0.7244	0.4965	14.6015	0.5031	0.6336
3	COMBINE-SF_{GBdG}	1	7		8	0.6570	0.5540	12.2057	0.5706	0.5781
4	COMBINE-SF_{ASE}	1				0.6062	0.5935	11.4354	0.5992	0.4728
5	COMBINE-SF_{AHB}	1	10		11	0.8411	0.3770	23.0982	0.3543	0.7634
6	COMBINE-$SF_{Andrews}$	1	11		12	0.8485	0.3681	21.9335	0.3764	0.7739
7	COMBINE_only		5		5	0.5666	0.6226	14.1173	0.5583	0.4907
8	2D/3D_only			7	7	0.7300	0.4914	20.0832	0.4109	0.6561
9	**Mixed_Desc**	**1**	**7**	**3**	**11**	**0.8682**	**0.3433**	**28.7441**	**0.2939**	**0.7834**

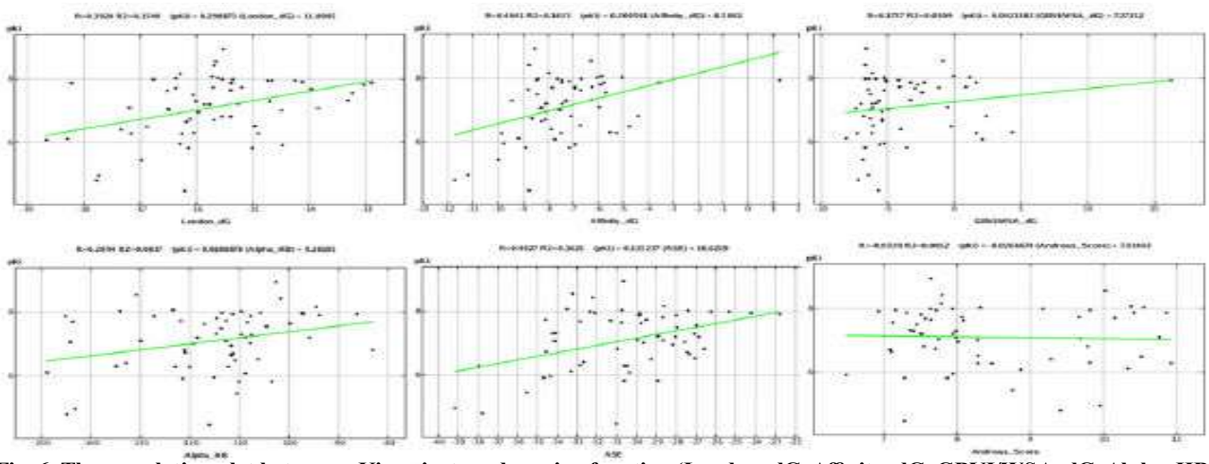

Fig. 6. The correlation plot between pKi against used scoring function (London_dG, Affinity_dG, GBVI/WSA_dG, Alpha_HB, ASE, and Andrews_Score) on the best pose docking of inhibitors.

The ModRefiner refined model was successfully evaluated in terms of its folding, stereochemical quality, ligand-receptor interaction and docking-based QSAR. Hence, this model will be suitable for further in silico structure-based designing.

ACKNOWLEDGMENT

B.S.A.S. is the recipient of BPPS scholarship from DIKTI Indonesia. This project is part of the dissertation research. Many thanks due to Prof. Dr. Narsito (alm.) as former promotor for entire supports. The authors also wish to thank the Dean of Faculty Pharmacy, UGM for providing funds to purchase the MOE package.

REFERENCES

1. Labrie F, Simard J, Luu-The V, Bélanger A., Pelletier G. Structure, function and tissue-specific gene expression of 3β-hydroxysteroid dehydrogenase/5-ene-4-ene isomerase enzymes in classical and peripheral intracrine steroidogenic tissues. J Steroid Biochem Mol Biol. 1992.43:805–26. doi:10.1016/0960-0760(92)90308-6.
2. Lachance Y, Luu-The V, Verreault H, Dumont M, Rhéaume E, Leblanc G, Labrie F. Structure of the human type II 3β-hydroxysteroid dehydrogenase/Δ5-Δ4 isomerase (3β-HSD) gene: adrenal and gonadal specificity. DNA Cell Biol. 1991.10:701–11. doi:10.1089/dna.1991.10.701.
3. Penning TM. Molecular endocrinology of hydroxysteroid dehydrogenases. Endocr Rev. 1997.18:281–305. doi:10.1210/edrv.18.3.0302.
4. Simard J, Ricketts ML, Gingras S, Soucy P, Feltus FA, Melner MH. Molecular biology of the 3β-hydroxysteroid dehydrogenase/Δ5-Δ4 isomerase gene family. Endocr Rev. 2005.26:525–82. doi:10.1210/er.2002-0050.
5. Rabbani B, Mahdieh N, Haghi Ashtiani MT, Setoodeh A, Rabbani A. In silico structural, functional and pathogenicity evaluation of a novel mutation: An overview of HSD3B2 gene mutations. Gene. 2012.503:215–21. doi:10.1016/j.gene.2012.04.080.
6. Thomas JL, Duax WL, Addlagatta A, Kacsoh B, Brandt SE, Norris WB. Structure/function aspects of human 3β-hydroxysteroid dehydrogenase. Mol Cell Endocrinol. 2004.215:73–82. doi:10.1016/j.mce.2003.11.018.
7. Goswami AM. Structural modeling and in silico analysis of non-synonymous single nucleotide polymorphisms of human 3β-hydroxysteroid dehydrogenase type 2. Meta Gene. 2015.5:162-72.
8. Apweiler R, Bairoch A, Wu CH, Barker WC,

Boeckmann B, Ferro S, *et al*. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 2004.32:D115–9. doi:10.1093/nar/gkh131.
9. Berman, H M., Westbrook J, Feng Z, Gilliland G, Bhat, T N, Weissig H, *et al*. The protein data bank. Nucleic Acids Res. 2000.28:235–42. doi:10.1093/nar/28.1.235.
10. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012.40:D1100–7. doi:10.1093/nar/gkr777.
11. Frye SV, Haffner CD, Maloney PR, Mook RA Jr, Dorsey GF Jr, Hiner RN, *et al*. 6-Azasteroids: structure-activity relationships for inhibition of type 1 and 2 human 5 alpha-reductase and human adrenal 3 beta-hydroxy-delta 5-steroid dehydrogenase/3-keto-delta 5-steroid isomerase. J Med Chem. 1994.37(15):2352-60.
12. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015.10:845–58. doi:10.1038/nprot.2015.053.
13. Molecular Operating Environment (MOE). Canada: Chemical Computing Group Inc; 2015.10.
14. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, *et al*. InterProScan 5: genome-scale protein function classification. Bioinformatics btu031. 2014. doi:10.1093/bioinformatics/btu031.
15. de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, *et al*. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 2006.34:W362–5. doi:10.1093/nar/gkl124.
16. Corpet F, Servant F, Gouzy J, Kahn D. ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. Nucleic Acids Res. 2000.28:267–9. doi:10.1093/nar/28.1.267.
17. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, *et al*. CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res. 2015.43:D376–81. doi:10.1093/nar/gku947.
18. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, *et al*. Pfam: the protein families database. Nucleic Acids Res. 2014.42:D222-30.doi:10.1093/nar/gkt1223.
19. Attwood TK. The PRINTS database: A resource for identification of protein families. Brief Bioinform. 2002.3:252–63.doi:10.1093/bib/3.3.252.
20. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol. 2001.313:903–19. doi:10.1006/jmbi.2001.5080.
21. Söding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005.21(9):2144.
22. Jefferys BR, Kelley LA, Sternberg MJE. Protein folding requires crowd control in a simulated cell. J Mol Biol. 2010.397:1329–38.doi:10.1016/j.jmb.2010.01.074.
23. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J. 2011.101:2525-34.
24. Labute P. Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. Proteins. 2009.75:187–205.doi:10.1002/prot.22234.
25. Labute P. LowModeMD: Implicit low mode velocity filtering applied to conformational search of macrocycles and protein loops. J Chem Inf Model. 2010.50:792–800.
26. Laskowski RA, MacArthur MW, Moss D, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst. 1993.26:283-91.
27. Wiederstein M & Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research. 2007.35:W407-10.
28. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999.292(2):195-202.
29. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol. 2004.337(3):635-45.
30. Jones DT. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics. 2007.23(5):538-44.
31. Kallberg Y, Oppermann U, Persson B. Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. FEBS J. 2010.277:2375-86.doi:10.1111/j.1742-4658.2010.07656.x.
32. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2011.27(3):343-50.
33. Bashton M, Chothia C. The geometry of domain combination in proteins. J Mol Biol. 2002.315:927-39. PUBMED:11812158EPMC:11812158.