Peer-reviewed version available at *Int. J. Mol. Sci.* **2020**, *21*, 893; doi:10.3390/ijms21030893

Elucidating Binding Sites and Affinities of ERα Agonist and Antagonist to Human Alpha-Fetoprotein by *in silico* Modeling and Point Mutagenesis

Nurbubu T. Moldogazieva[1*], Daria S. Ostroverkhova[1,2], Nikolai N. Kuzmich[1,3], Vladimir V. Kadochnikov[4], Alexander A. Terentiev[5], Yuri B. Porozov[1,4]

[1] Laboratory of Bioinformatics, I.M. Sechenov First Moscow State Medical University (Sechenov University); Moscow, Russia;

[2]Department of Bioengineering, M.V. Lomonosov Moscow State University; Moscow, Russia;

[3]Department of Drug Safety, I.M. Smorodintsev Research Institute of Influenza, WHO National Influenza Centre of Russia; Saint Petersburg, Russia

[4]Department of Information and Communication Technologies, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics; Saint Petersburg, Russia

[5]Deparment of Biochemistry and Molecular Biology, N.I. Pirogov Russian National Research Medical University; Moscow Russia

*Corresponding author: Nurbubu T. Moldogazieva, e-mail: nmoldogazieva@mail.ru (NM)

¶ These authors contributed equally to this work

Short title: Binding of ERα ligands to human alpha-fetoprotein

Key words: alpha-fetoprotein, estrogens, selective estrogen receptor modulators, homology-based modeling, molecular docking, protein-ligand interaction, amino acid substitutions.

Abbreviations: HAFP, human alpha-fetoprotein; HSA, human serum albumin; E2, 17β-estradiol; DES, diethylstilbestrol.

Author contribution:

Conceptualization, Nurbubu T. Moldogazieva, Alexander A. Terentiev and Yuri B. Porozov; Data curation, Nurbubu T. Moldogazieva; Formal analysis, Nurbubu T. Moldogazieva and Nikolai N. Kuzmich; Investigation, Daria S. Ostroverkhova and Vladimir V. Kadochnikov; Methodology, Nikolai N. Kuzmich; Supervision, Alexander A. Terentiev and Yuri B. Porozov; Validation, Daria S. Ostroverkhova; Visualization, Daria S. Ostroverkhova; Writing – original draft, Nurbubu T. Moldogazieva and Daria S. Ostroverkhova.

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

Abstract

Alpha-fetoprotein (AFP) is a major embryo- and tumor-associated protein capable of binding and transporting variety of hydrophobic ligands including estrogens. AFP has been shown to inhibit estrogen receptor (ER)-positive tumor growth and this can be attributed to its estrogen-binding ability. Despite AFP has long been investigated, its three-dimensional (3D) structure has not been experimentally resolved and molecular mechanisms underlying AFP-ligand interaction remain obscure. In our study we constructed homology-based 3D model of human AFP (HAFP) with the purpose to perform docking of ERα ligands, three agonists (17β-estradiol, estrone and diethylstilbestrol) and three antagonists (tamoxifen, afimoxifene and endoxifen) into the obtained structure. Based on ligand docked scoring function, we identified three putative estrogen- and antiestrogen-binding sites with different ligand binding affinities. Two high-affinity sites were located in (i) a tunnel formed within HAFP subdomains IB and IIA and (ii) opposite side of the molecule in a groove originating from cavity formed between domains I and III, while (iii) the third low-affinity site was found at the bottom of the cavity. 100 ns MD simulation allowed studying their geometries and showed that HAFP-estrogen interactions occur due to van der Waals forces, while both hydrophobic and electrostatic interactions were almost equally involved in HAFP-antiestrogen binding. MM/GBSA rescoring method estimated binding free energies (ΔG_{bind}) and showed that antiestrogens have higher affinities to HAFP as compared to estrogens. We performed *in silico* point substitutions of amino acid residues to confirm their roles in HAFP-ligand interactions and showed that Thr132, Leu138, His170, Phe172, Ser217, Gln221, His266, His316, Lys453, and Asp478 residues along two disulfide bonds, Cys224-Cys270 and Cys269-Cys277 have key roles in both HAFP-estrogen and HAFP-antiestrogen binding. Data obtained in our study contribute to understanding mechanisms underlying protein-ligand interactions and anti-cancer therapy strategies based on ER-binding ligands.

Peer-reviewed version available at *Int. J. Mol. Sci.* **2020**, *21*, 893; doi:10.3390/ijms21030893

Author summary

Investigation of protein surface topology and elucidating channels, cavities, grooves and pockets in its three-dimensional (3D) structure play important roles in predicting protein-ligand interactions and explaining molecular mechanisms underlying protein functioning. Current computational approaches are becoming powerful tools to create 3D model of a protein structure if it has not been obtained by experimental methods and to assess at atomic level mechanisms of protein-ligand binding capabilities. Mostly, this is based on ligand docking procedure followed by calculation of scoring functions with subsequent molecular dynamics (MD) simulations and binding free energy (ΔG) estimations. In this study we used the *in silico* approaches to predict three major binding sites in human alpha-fetoprotein (HAFP) and to compare binding affinities of natural and synthetic estrogens and antiestrogens. We identified key amino acid residues involved in HAFP-ligand interactions and showed that point mutation in the protein ligand-binding sites can cause preservation of the same type of interactions despite ligand moves to another residue, but with the same physicochemical properties. This phenomenon can have significance in the maintenance of protein activity and requires further investigations. Further, the data obtained in this study can be used for developing antitumor therapies based on AFP ligands.

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

Introduction

Alpha-fetoprotein (AFP) is a major mammalian embryo-specific and tumor-associated protein recognized as a "golden standard" among cancer biomarkers used in clinical practice [1]. *AFP* gene is expressed during embryonic development by liver and yolk sac and is down-regulated soon after the birth to be re-expressed in adult patients with hepatocellular carcinoma (HCC) [2-4]. Biological roles of AFP during embryonic development and tumor growth have long been investigated however remain not fully understood [5]. Experimental data evidence capability of both native and recombinant AFP of regulating cell proliferation and immune response, and binding and transporting variety of hydrophobic ligands such as estrogens, fatty acids and drugs suggesting that these functions have a role during embryo- and carcinogenesis [6-10].

High maternal serum AFP concentration during second and third trimesters of pregnancy has been shown to correlate with reduced breast cancer risk [11]. Estrogen-binding capability of AFP can contribute to decrease in concentration of free active forms of estrogens below the levels necessary to activate estrogen receptors (ERs), ERα and ERβ, in target cells. Indeed, AFP has been reported to inhibit estrogen receptor (ER)-positive human MCF-7 and MTW9A rat mammary cancer growth, and this associated with interaction of AFP with 17β-estradiol [12,13]. Moreover, synthetic human AFP-derived peptides can cause both inhibition of uterine estrogen-dependent cell proliferation and simultaneous increase in anti-tumor effect of tamoxifen [14].

Canonical estrogen-signaling pathway starts from ligand binding to ERs, which causes conformational changes, dimerization and activation of the receptor. Further, the activated receptor-hormone complex is translocated into the nucleus, where it binds to specific sequences of estrogen responsive element (ERE) in the promoter region of target genes [15]. ERs recruit variety of co-regulatory proteins, co-activators and co-repressors, which can cause alterations in hormone-receptor complex formation, DNA-binding abilities [16].

Natural estrogen, 17β-estradiol (E2) is the most potent female sex steroid hormone as compared to its metabolites, estrone (E1) and estriol (E3), in regulating cell proliferation, differentiation and homeostasis in reproductive, skeletal, cardiovascular and neural tissues [17-20]. However, various endocrine disruptors can exert structural similarity to steroid hormones and can directly interfere with estrogens for their ER-binding sites [21]. Estimation of binding affinities of 125 ER ligands, both natural and chemically synthesized, performed using competitive ligand-binding assay demonstrated, that certain structural features of estrogens such as the presence of aromatic ring may be important for their binding efficacy to ERs [22].

While estrogens are considered as ER agonists, numerous selective estrogen receptor modulators (SERMs) can serve both as ER agonists and antagonists, in a tissue- or promoter-specific and dose-dependent manner [23,24]. Among SERMs, tamoxifen is the most commonly used anti-tumor non-steroidal drug for breast cancer therapy. Tamoxifen, and its derivatives, afimoxifene (4-hydroxytamoxifen, 4OHT) and endoxifen (4-hydroxy-N-desmethyl-tamoxifen), belong to type 1 antiestrogens, which can act as both agonists and antagonists [25]. Structurally, tamoxifen is similar to diethylstilbestrol (DES), synthetic non-steroidal estrogen and pure agonist recognized as teratogen and carcinogen. However, unlike DES, tamoxifen has additional aromatic ring and a side chain (trans-isomer) with *N*-linked methyl groups [26]. Afimoxifene and endoxifen are key active metabolites of tamoxifen and experimentally have been shown to possess higher affinity and specificity to ERα as compared to tamoxifen [27].

The ability of hormone-binding transport proteins to interfere with ER-ligand interactions dictates the importance of investigation of binding affinities of both estrogens and ER disruptors not only to ERs, but also to estrogen-binding transport proteins such as AFP. Rodent AFPs have been experimentally evidenced to bind free estrogens, while human AFP (HAFP) has been shown to bind only immobilized estrogens [28,29]. Using chimeric human-rat AFP the existence of two types of estrogen-binding sites, high-affinity and low-affinity, in rat AFP has been proposed [30,31]. However, there is no experimentally obtained 3D structure of AFP for any biological species in PDB database, both with and without ligands. This dictates the necessity to create 3D model of AFP followed by molecular docking of ligands into the obtained structure to investigate AFP-ligand binding affinities and mechanisms underlying protein-ligand interactions at atomic level. This might contribute to understanding a role of estrogen and antiestrogen binding to transport proteins in functioning of endocrine system both in normal and pathologies.

HAFP contains 609 amino acid residues and is composed of three homologous domains: I (aa 19-210), II (aa 211-402) and III (aa 402-601). It belongs to the family of serum albumin (SA), which includes except SA and AFP two more proteins: alpha-albumin (afamin) and vitamin D-binding protein (VTDB). Members of this family are localized in tandem arrangement in the q11-q13 region of chromosome 4 and share high degree of sequence similarity [32]. Sufficient degree of sequence identity (up to 40%) between AFP and SA makes it possible to construct 3D model of AFP based on homology with SA, for which experimentally obtained crystal structures can be found in PDB database. Earlier, our group used this approach to build 3D structure of HAFP by homology-based modeling and to investigate HAFP-binding modes of DES, validated estrogen disruptor [33].

Later, Shen and co-workers built 3D structure of rat AFP using homology-based modeling to elucidate protein-ligand binding modes for representative AFP binders and to estimate binding

affinities of variety estrogens and estrogen disruptors [34]. The ability of rat AFP to undergo conformational changes during ligand binding and to accommodate numerous structurally and physicochemically diverse compounds has been observed. Additionally, testing rat AFP binding of a large set of structurally diverse chemical compounds revealed 53 binders from 13 structural categories and 72 non-binders [35].

In the present study, we performed homology-based modeling of 3D structure of HAFP with the aim to perform molecular docking of three ER agonists and three antagonists to the obtained model with the aim to elucidate binding modes and structures of estrogen- and antiestrogen-binding sites. MM/GBSA rescoring method based on molecular dynamics (MD) simulation was exploited to assess affinities of estrogens to HAFP as compared to affinities of antiestrogens. Further we performed *in silico* point substitutions of amino acid residues, presumably, involved in estrogen binding to confirm their roles in protein-ligand interactions. We found that there are three principal estrogen- and antiestrogen-binding sites in HAFP. Two of them are high-affinity sites located in a hollow tunnel formed within subdomains IB and IIA and in opposite side of the molecule in a groove originating from a cavity formed by U-shaped AFP structure in accordance with experimental data [36]. The third is low-affinity binding site located at the bottom of the cavity. Investigation of geometries of the binding sites allowed elucidating amino acid residues involved in HAFP-ligand binding and types of HAFP-estrogen and HAFP-antiestrogen interactions. We showed that antiestrogens have higher affinity to HAFP as compared to estrogens, both natural and synthetic, and this may be dictated by differences in their structures and amounts of rotameric states.

Results

The overall architecture and quality of constructed HAFP 3D model

Figure 1 depicts plots, diagrams and graphs for HAFP 3D model validation data. Alignment (Figure 1A) between sequences of target protein, HAFP retrieved from Uniprot knowledge base (ID: P02771), and HSA identified as the best template for homology-based modeling and retrieved from PDB database (ID: 1E78) showed identity degree to be equal to 39%, and secondary structure elements (SSE) to be represented by α-helices and random coils with no β-strands in accordance with experimental data [36].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at *Int. J. Mol. Sci.* **2020**, *21*, 893; doi:10.3390/ijms21030893

Figure 1. Validation of HAFP 3D model structure obtanied based on homology with HSA. (A) Target-template sequence alignment. Secondary structure elements represented, predominantly, by α-helices (H) are indicated. Identical residues are similarly colored according to their physicochemical properties. (B) Ramachandran map for the main chain conformation shows that 99.5% residues (black dots) are located in favoured (red) and allowed (yellow) regions. (C) RMSD plot obtained for Cα-atoms during 200 ns MD simulation shows stability of the overall HAFP 3D structure. (D) RMSF plot and (E) SSE diagrams demobstrate stability of α-helical regions. (F) Graphs and diagrams of stereochemical properties of individual residues (shown for part of residues involved in ligad binding) indicate high quality of the model structure.

Validation of constructed in our study 3D model of HAFP structure showed good quality and reliability of the obtained model. Ramachandran maps generated by PROCHECK program verified proper stereo-chemical and conformational properties of amino acid residues in the modeled HAFP structure (Figure 1B). Indeed, 92.7% of all residues were located in favored regions, while 6.8% of all residues were in additional allowed regions, and two residues, Ser84 and Glu565, were in generously allowed regions; only one residue, Ala39, occurred in disallowed region.

Relaxation and optimization of constructed HAFP model with the use of 200 ns MD simulation showed its sufficient stability. Indeed, RMSD values calculated for Cα-atoms for the optimized HAFP structure were stabilized around 3.8-4.0 Å (Fig 1C). RMSF plots obtained for residues in HAFP sequence allowed judging about local fluctuations, which were not high and correspond to regions with irregular structures and loops (Figure 1D). Additionally, high quality of the model was confirmed by analysis of SSE composition for each trajectory frame, which showed its stability during the all 200 ns MD simulation (Figure 1E). Furthermore, structural alignment score of the obtained structure against the used template, 1E78, was quite low, 0.002, along with RMSD equal to 0.196 Å. Superimposition of HAFP models before and after 200 ns MD simulation showed that the optimized structure becomes more relaxed and expanded.

Analysis of stereo-chemical parameters of the obtained structure (Figure 1F) showed that all residues have correct Cα-chirality, while only 1.2% residues have incorrect absolute deviation from mean χ1 value and 5% residues from mean ω torsion. Estimation of water accessibility areas showed that α-helical regions are, mostly, buried indicating their enrichment in hydrophobic residues necessary for interaction with variety of hydrophobic ligands. Additionally, G-factor values for the overall structure and the main chain bonds and angles along with combination of φ-ψ, χ1-χ2, χ3-χ4 correspond to good quality of the model obtained in this study and its reliability for further usage.

Analysis of the HAFP model architecture showed its U-shaped structure, in which three domains and secondary structure elements represented by α-helices can be identified (Figure 2A).

Peer-reviewed version available at *Int. J. Mol. Sci.* **2020**, *21*, 893; doi:10.3390/ijms21030893

Additionally, visualization of the model surface by Maestro suite in Schrödinger software revealed inter-domain cavity formed by HAFP domains I and III and a tunnel (Figure 2B) located between its subdomains IB and IIA. This tunnel has a length of about 29.3 Å and diameters from about 9.25 Å in its narrow part to 15.24 Å in its wide segment. Furthermore, a groove originated from the cavity (Figure 2C) formed between domains I and III may be observed at the back side of HAFP surface.

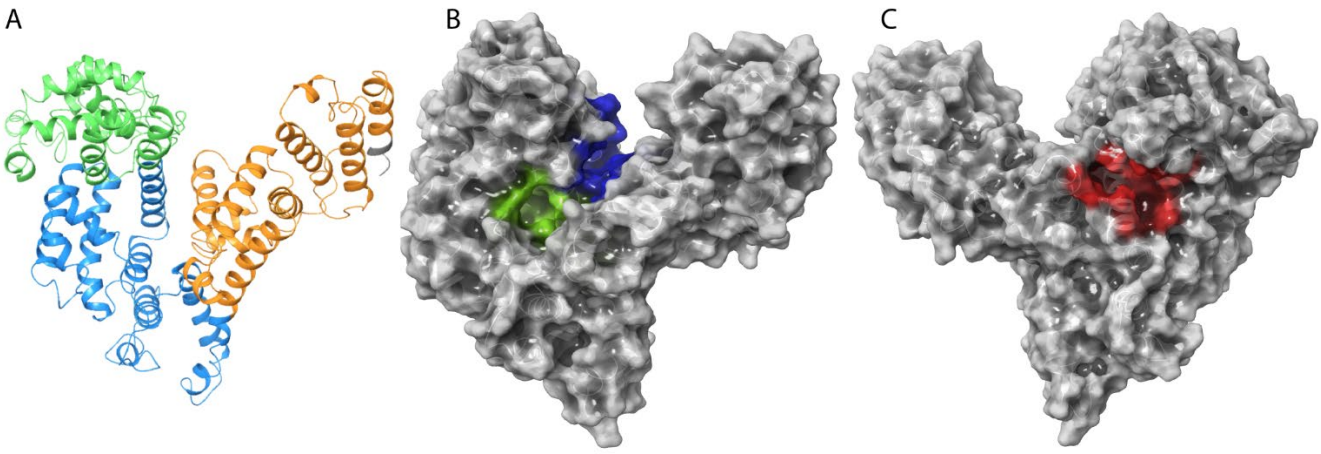

Figure 2. The overall architecture and domain organization of HAFP 3D model structure. (A) three-domain organization, where domains I, II, and III are shown in green, blue and brown, respectively; (B) topology of the entire HAFP surface in the vicinity of a tunnel (green) formed within subdomains IB and IIA and interdomain cavity (blue) formed between domains I and III; (C) a groove (red) originated from cavity.

HAFP-ligand docked poses

Docking procedure was performed to search for possible positions of 6 ligands, 3 estrogens and 3 antiestrogens (Figure 3), in HAFP potential binding sites and to estimate the ligand binding affinities. Optimized ligand molecules extracted from Pubchem database were docked into the optimized HAFP modeled structure using flexible docking protocol to enable search for all probable conformations generated due to changes in residue torsion angles. In total, 84 ligand-docked complexes were obtained based on 10 predicted by SiteMap potential binding sites.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

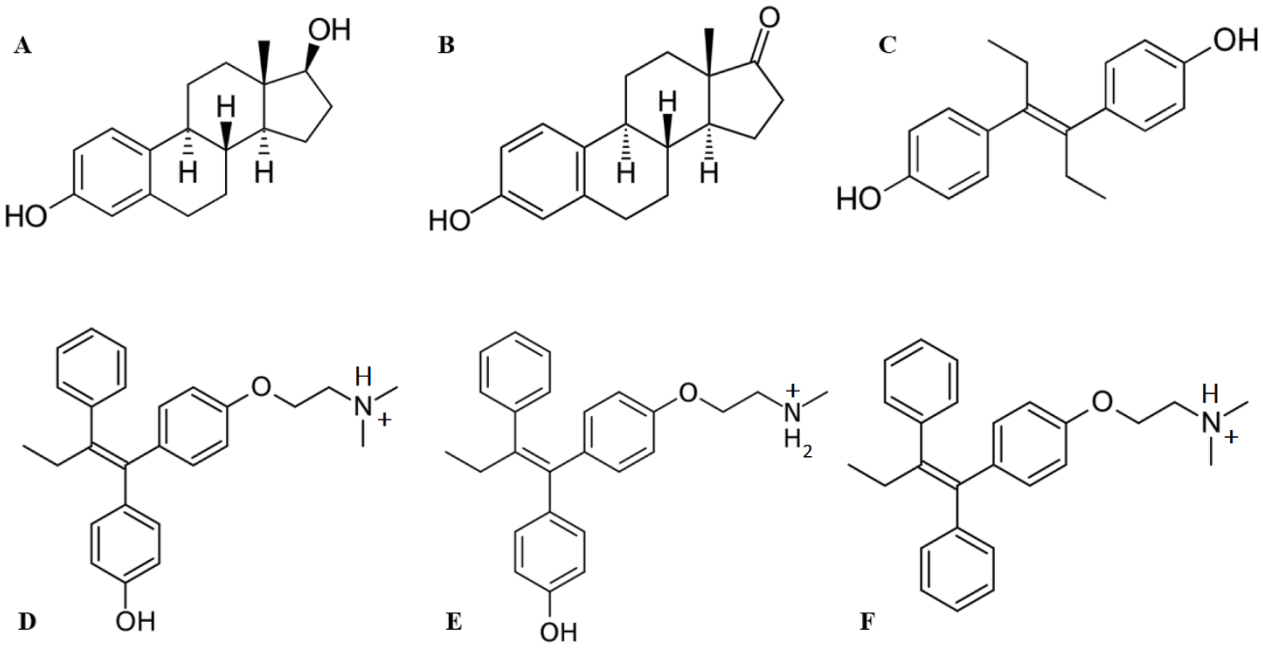

Figure 3. Structures of ligands used in this study. (A) 17β-estradiol, (B) estrone, (C) diethylstylbestrol, (D) afimoxifene, (E) endoxifen, (F) tamoxifen.

Based on calculation of ligand docked scoring functions, we elucidated sites that best accommodate ligands and selected ligands with best binding affinities (Table1). Visualization of these complexes allowed identifying three most potent estrogen- and antiestrogen-binding sites, which we designated as A and B, and C (Figure 4A-C). Site A was located in a tunnel formed between subdomains IB and IIA, while site B was at the bottom of cavity in U-shaped HAFP structure, and site C was in a groove originated from the cavity and located on the back side of HAFP molecule.

Table 1. Scoring Functions for the Best Docked HAFP-estrogen and HAFP-antiestrogen Complexes

Docked ligand	Amount of rotational bonds	Binding site	XP GScore	Glide gscore	Glide emodel
17β-estradiol	2	A	-5.420	-5.420	-49.732
		B	-4.807	-4.807	-46.737
		C	-5.384	-5.384	-35.722
Endoxifen	9	A	-6.322	-6.322	-61.786
		B	-4.247	-4.247	-50.983
		C	-5.768	-5.768	-59.989
Afimoxifene	9	A	-5.932	-5.932	-57.375
		B	-3.840	-3.840	-55.881
		C	-5.308	-5.308	-54.272
Tamoxifen	8	A	-4.805	-4.805	-53.899
		B	-2.976	-2.976	-43.853

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

		C	-4.169	-4.169	-51.536
DES	6	A	-4.934	-4.934	-37.538
		B	-3.126	-3.126	-33.573
		C	-3.440	-3.440	-37.538
Estrone	1	A	-4.553	-4.553	-45.389
		B	-4.054	-4.054	-45.389
		C	-4.408	-4.408	-38.067

Ligand docked scoring functions indicated sites A and C to have higher affinity in relation to all ligands, both estrogens and antiestrogens, as compared to site B. To confirm this and to identify dynamic HAFP-ligand binding modes and ligand binding affinities, we further carried out 100 ns MD simulation study of the 12 best protein-ligand complexes selected on the basis of the ligand docked scoring functions.

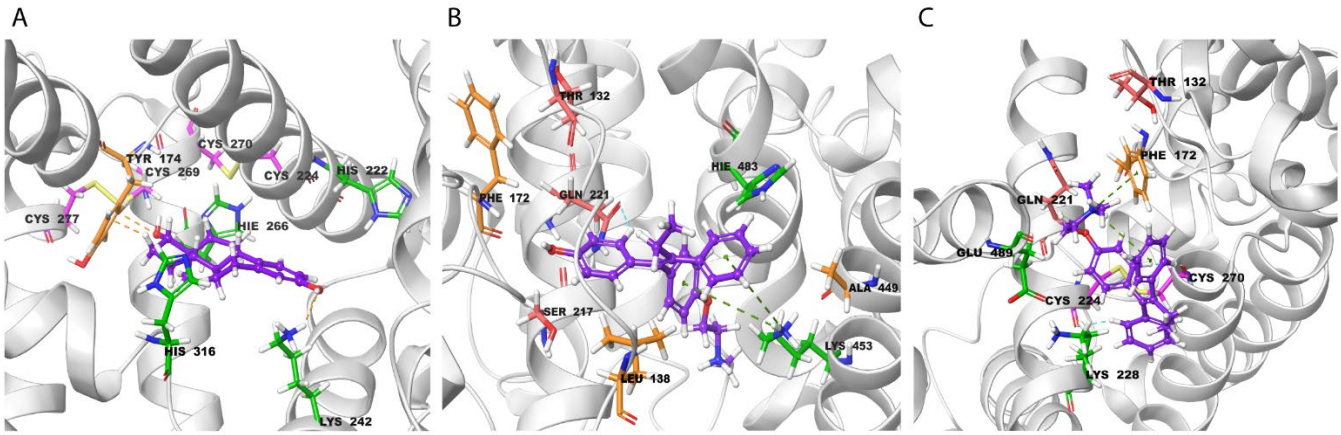

Figure 4. Amino acid residues involved in HAFP-ligand binding based on ligand docked poses. (A) 17β-estradiol in site A formed by the following amino acid residues: Tyr174, His222, Lys242, His266, His316, along with disulfides, Cys224-Cys270 and Cys269-Cys277; B) Afimoxifene in site B composed of Thr132, Leu138, Phe172, Ser217, Gln221, His483, Ala449 and Lys453; and (C) Tamoxifen in site C, which contains Lys228, Glu489 and can involve some residues from sites A and B. Ligands are shown in violet, charged amino acids in green, polar residues in red and hydrophobic residues in brown color.

Binding of ligands to HAFP studied by MD simulation

Stability of HAFP-ligand complexes

Estimation of stability of HAFP-ligand complexes was made based on protein-ligand RMSD calculations. Figure 5 depicts representative RMSD plots for optimized by 100 ns MD simulation of HAFP-ligand complexes. As shown on Figure 5A, protein conformation in HAFP-E2 complex formed in site B remains stable during 100 ns MD simulation being stabilized around 2.4-2.8 Å after 20 ns simulation. Ligand RMSD remains stable on about 2.0 Å to be

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

lower than protein RMSD until 65 ns MD simulation indicating that E2 remains kept in the binding site. As shown on Figure 5B, in HAFP-E1 complex ligand RMSD sharply increased after 15 ns MD simulation indicating that the ligand moves away from the binding site. As for HAFP-DES complex, the protein-ligand RMSD plots show the most stable interactions to be in site A (Figure 5C). Affinity of endoxifen to HAFP was high in all three sites, being the best in site A (Table 2), in which protein-ligand RMSD remains stable during 100 ns MD simulations (Figure 5D). The best HAFP-afimoxifene and HAFP-tamoxifen complexes were in site C showing enough stability during 100 ns MD simulation (Figures 5E and 5F).

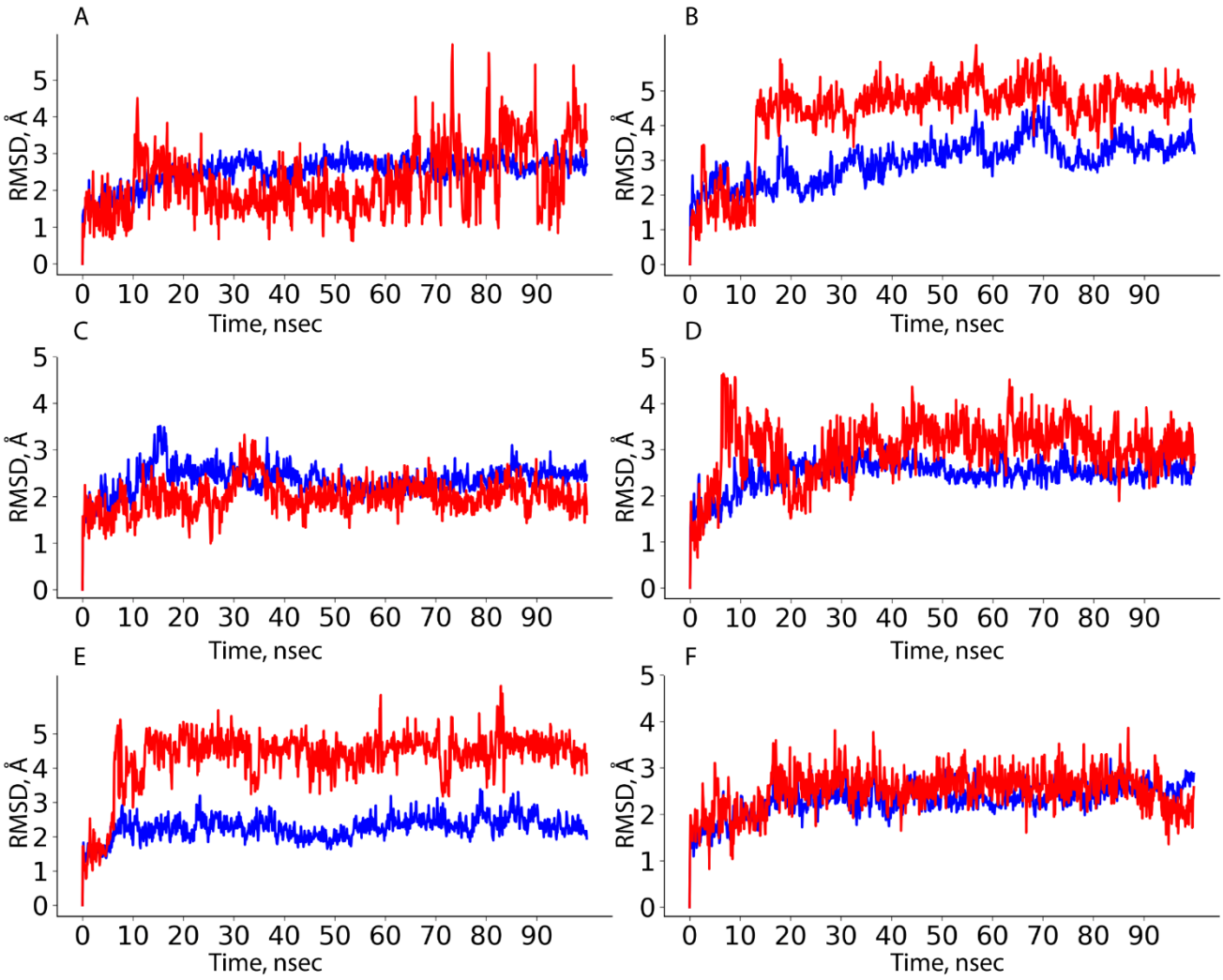

Figure 5. HAFP-ligand RMSD plots drawn during 100 ns molecular dynamics simulation in sites with most stable protein-ligand complexes. (A) HAFP-E2 complex in site B, (B) HAFP-E1 complex in site B, (C) HAFP-DES complex in site A, (D) HAFP-endoxifen complex in site A, (E) HAFP-afimoxifene complex in site C, (F) HAFP-tamoxifen complex in site C.

Ligand binding affinities

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

To assess ligand binding affinities MM/GBSA rescoring method based on 100 ns MD simulation was implemented. For this purpose, we calculated MM/GBSA values for each of protein-ligand RMSD snapshots obtained during 100 ns MD simulation and then calculated average MM/GBSA values for each trajectory. Table 2 shows that MM/GBSA ΔG_{bind} values, mainly, decrease and, consequently, binding affinities increase after optimization of HAFP-ligand complexes. Exceptions are HAFP-afimoxifene complex in site A and HAFP-endoxifen and HAFP-DES complexes in site C, in which MM/GBSA values do not decrease after optimization. This may be explained by both ligand motions during MD simulations and local arrangements of amino acid residues in ligand binding sites. This can depend on both structural features of a ligand (the presence of extra OH-group in afimoxifene, endoxifene and DES) and geometry of a binding site such as narrow tunnel (site A) or groove (site C), which creates restrictions for conformational motions. To verify this, we further estimated dynamic changes in protein-ligand contacts and interaction modes followed by *in silico* substitutions of key amino acid residues presumably involved in protein-ligand interactions.

Table 2. Values of Calculated Binding Energies and Energy Terms for HAFP-ligand Docked Poses and HAFP-ligand Complexes During 100 ns MD Simulation

Complex	MM/GBSA ΔG_{bind} docked poses	MM/GBSA ΔG_{bind} optimized complexes	$\Delta G_{Coulomb}$ Optimized	ΔG_{vdW} Optimized	ΔG_{GB} Optimized	ΔG_{lipo} Optimized	Ligand efficiency
Site A							
HAFP–17β-estradiol	-35.03	-38.29	-9.55	-33.20	20.50	-15.99	-9.58
HAFP–DES	-49.14	-50.99	-10.68	-39.69	20.06	-21.11	-12.76
HAFP–endoxifen	-64.50	-64.18	-42.01	-47.54	56.66	-28.06	-14.81
HAFP–afimoxifene	-69.34	-57.23	-36.93	-47.05	49.76	-25.18	-13.10
HAFP–tamoxifen	-30.53	-48.09	-32.86	-42.11	45.60	-19.74	-11.10
Site B							
HAFP–17β-estradiol	-42.85	-44.22	-15.23	-29.95	15.18	-13.00	-11.05
HAFP–estrone	-42.51	-50.31	-10.81	-36.08	13.86	-16.32	-12.59
HAFP–endoxifen	-55.15	-59.61	-41.92	-43.93	43.95	-18.22	-13.76
Site C							
HAFP–endoxifen	-59.13	-45.38	-39.41	-39.13	45.95	-15.91	-10.47

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

HAFP–afimoxifene	-53.80	-61.32	-39.94	-47.06	46.92	-22.23	-14.04
HAFP–tamoxifen	-52.68	-58.74	-33.62	-45.91	41.10	-23.04	-13.56
HAFP–DES	-44.47	-36.33	-9.55	-22.88	14.52	-17.71	-9.09

HAFP-ligand interaction modes

Specific protein-ligand contacts arise due to multiple weak, low-energy (1-5kcal/mol), non-covalent interactions such as H-bonds, ionic, and hydrophobic forces at short distance ranges enough for bonding, 2.6-3.4 Å [37]. Protein-ligand interaction diagrams generated by Schrödinger software throughout 100 ns MD simulation trajectory and normalized on types of non-covalent bonds were used to estimate HAFP-ligand dynamic interaction modes. Here, we estimated (i) amino acid residues involved in specific types of interactions during more than 50% of MD simulation time; (ii) amount of contacts, which the protein formed with a ligand at each frame of MD trajectory. Figure 6 depicts amino acid residues involved in HAFP-ligand interactions with indication how long a specific contact and types of non-covalent forces were maintained. In total, Tyr174, Ser216, His222, Ile238, Thr239, His266, Asp478, Ile479, and Cys269-Cys277 disulfide bond were involved in HAFP-ligand interactions in site A. For example, HAFP-E2 interactions were maintained, by Tyr174, Leu219, Lys242, Val262, and His 266, where His266 was involved, mainly, in H-bonds, while Lys242 can form any of four types of non-covalent interactions with the ligand (Figure 6A). While, His222, Thr239, Asp478, and Ile479 were in contact with DES in this site during more than 50% of MD simulation time (Figure 6B)

Site B involves Leu138, His170, Phe172, Ser217, Gln221, Lys453 residues. Here, HAFP-E1 interaction diagram shows that after 15 ns MD simulation estrone moves from His170, Ser217 and His483 to Lys138, Phe172 and Lys453 (Figure 6C) keeping, however, the same types of interaction. HAFP-endoxifen contacts in the same site were maintained by Leu138, Arg169, Ser217, and Lys543 residues (Figure 6D). In site C, HAFP-afimoxifene binding was provided, mostly, by Thr132, Phe172, Gln221, Lys228 residues, which keep contacts with the ligand during half and more of MD simulation (Figure 6E). As for HAFP-tamoxifen interaction in this site, in addition to Phe172, Lys129 instead of Lys228 and disulfide bond Cys224-Cys270 were involved (Figure 6F).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

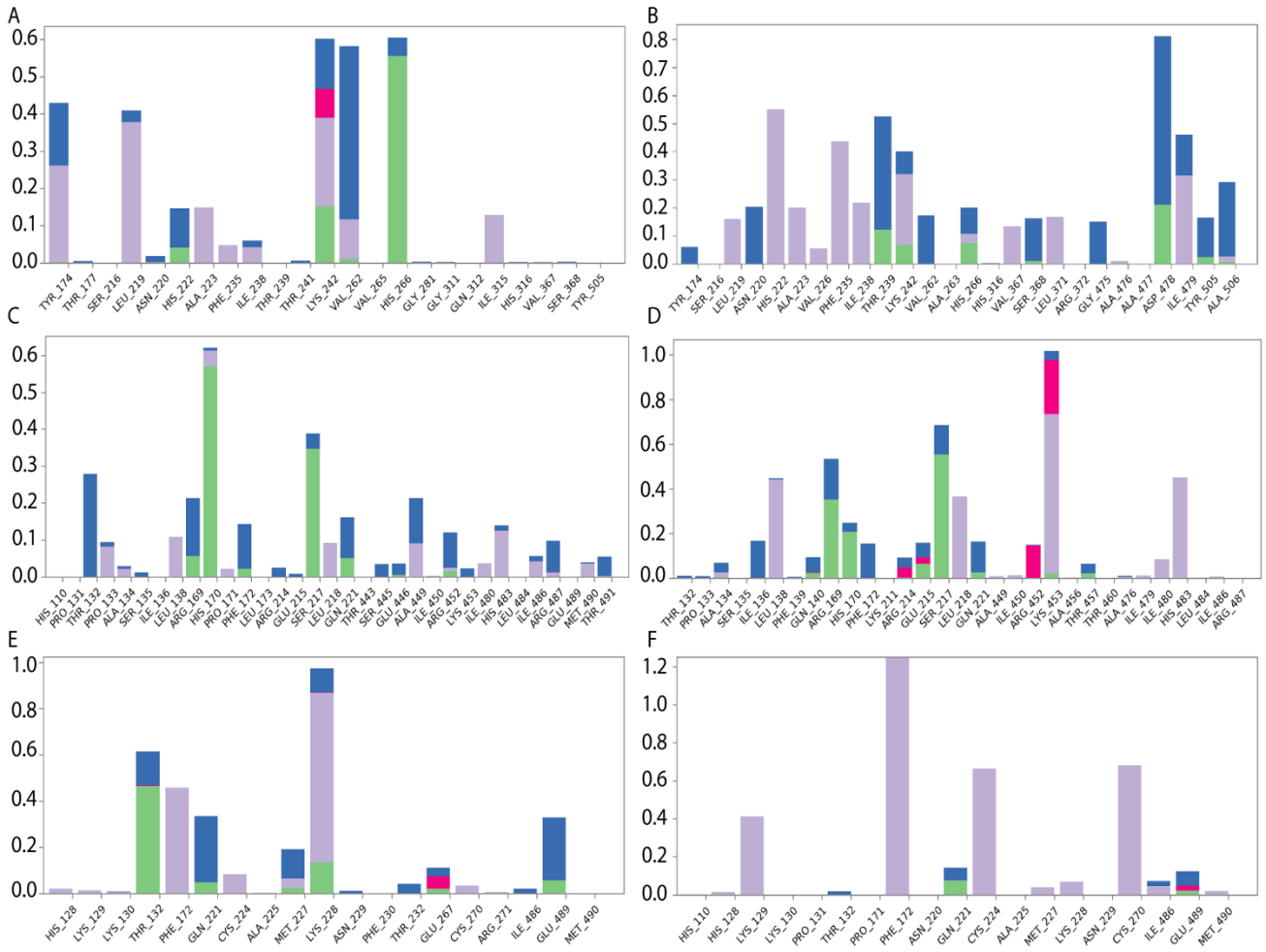

Figure 6. Diagrams of HAFP contacts with a ligand in most stable complexes with indication of amino acid residues and forces involved throughout 100 MD simulations. Protein-ligand interactions were categorized in four types: H-bonds (green), ionic (red), hydrophobic (purple) and water bridges (blue). (A) HAFP-E2 contacts in site A, (B) HAFP-DES contacts in site A, (C) HAFP-E1 contacts in site B, (D) HAFP-endoxifen contacts in site B, (E) HAFP-afimoxifene contacts in site C, (F) HAFP-tamoxifen contacts in site C.

Additionally, ligand-centered interaction maps were generated to indentify types of non-covalent interactions between functional groups of a ligand and a definite amino acid residue in HAFP. Figure 7A illustrates that HAFP-E2 interactions in site B are provided, mostly, by H-bonds between Gln221 and Ser217 to E2 OH-group at C17-atom along with π-cation interaction between aromatic A-ring of E2 and positively charged ε-amino group of Lys453. Water-bridge between Lys453 and OH-group at C3-atom of estrone and hydrophobic pocket of Leu138 and Phe172 are important in HAFP-E1 interaction in the same site (Figure 7B). In site A, DES interacts by its two OH-groups with Asp478 and Thr239 residues through water-bridges, and by its aromatic ring forms π,π-stacking to His222 supported by hydrophobic pocket consisted of Phe235 and Ile479 (Figure 7C). Endoxifen involves its side chain $^{+}NH_2$-group for H-bonding

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

with Ser216 and His316. Additionally, it interacts with Cys269-Cys277 disulfide bond and hydrophobic pocket of Val262, Leu278 residues in the binding site (Figure 7D).

Afimoxifene interacts with HAFP, mostly, through its aromatic ring, which is involved in π-stacking interactions with Phe172 and Cys224-Cys270 disulfide bond in site C. Additionally, the ligand OH-group is involved in water-mediated H-bonding and Gln221 and Glu489, while its side chain NH^+-group can form salt bridges, H-bonds and water bridges with Glu267 (Figure 7E). Tamoxifen interactions with site C were provided, mostly, by π-cation and π,π-stacking bonds to Phe172 or Lys129 as well as hydrophobic interactions with Cys224-Cys270 disulfide (Figure 7F).

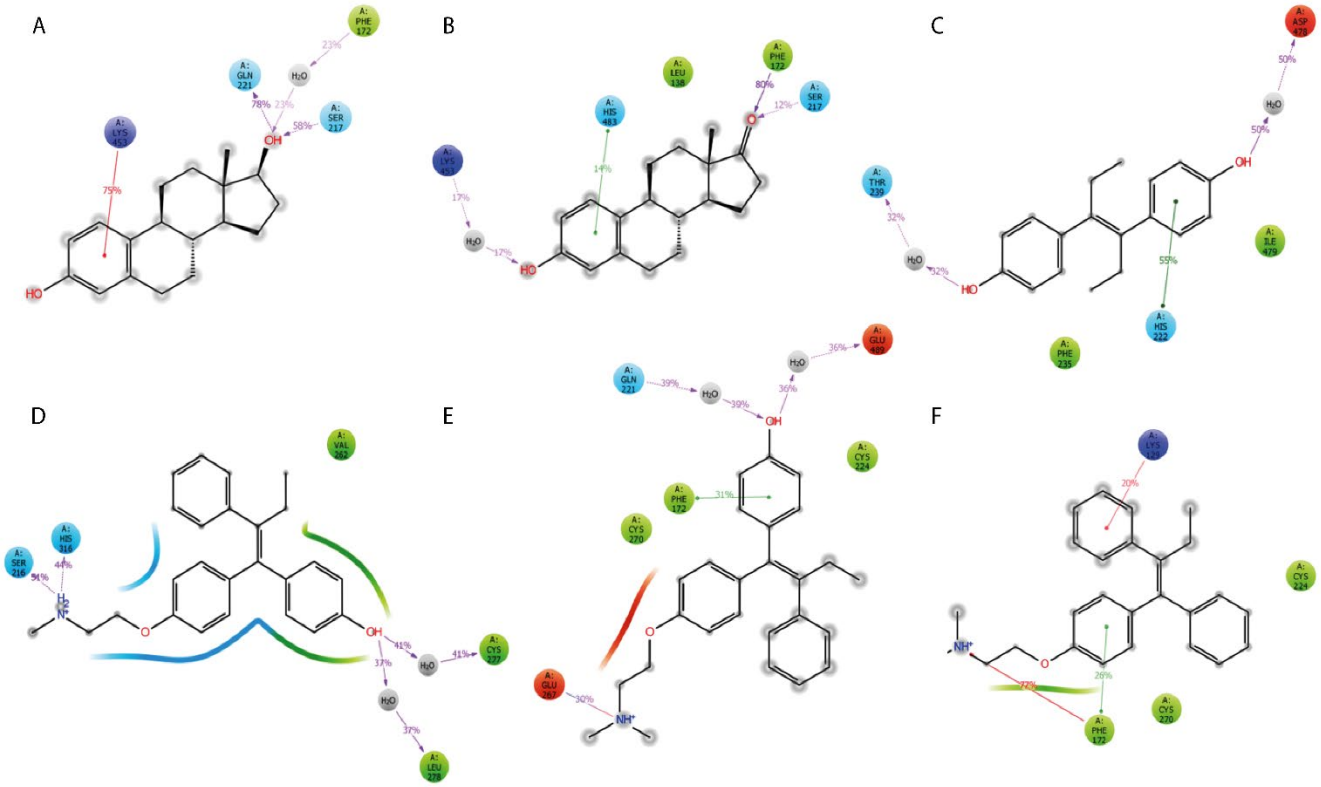

Figure 7. Types of non-covalent forces between functional groups of a ligand and HAFP in the most stable complexes throughout 100 MD simulation. (A) E2 in site B, (B) E1 in site B, (C) DES in site A, (D) Endoxifen in site A, (E) Afimoxifene in site C, (F) Tamoxifen in site C. Residues are shown as follows: green – hydrophobic, blue – polar, majenta – negatively charged, violet – positively charged, grey – water (solvent exposure). Interactions are shown as follows: red – π,cation; green – π,π-stacking, purple – hydrogen bonding.

Effects of *in silico* point amino acid substitutions

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

Table 3 demonstrates impacts of point substitutions of key amino acid residues involved in HAFP-ligand interactions performed *in silico* in all three binding sites. Two of 6 substitutions performed in HAFP-E2 complexes, His266Leu and Lys453Leu remarkably improve MM/GBSA ΔG_{bind} value, while one substitution, Gln221Val, increased binding energy. Indeed, the substitutions His266Leu performed in site A caused improvement of binding pose of 17β-estradiol in this site (Figure 8A). However, diffusion of the ligand from its binding pose in site B resulted from Gln221Val substitution was observed (Figure 8B) indicating that Gln221 has a key role in HAFP-E2 interaction.

Table 3. MM/GBSA Free Energy Changes and Energy Terms Resulted from Point Residue Substitutions

Complex	Substitution	MM/GBSA ΔG_{bind} [a]	$\Delta G_{Coulomb}$	ΔG_{vdW}	ΔG_{GB}	ΔG_{lipo}	Ligand efficiency
HAFP-17β-estradiol	Lys242Leu	-39.3628 (-38.2924)	-7.2468	-34.8330	22.0165	-19.3103	-9.8512
	His266 Leu	-42.5281 (-38.2924)	-7.8309	-35.1273	19.5618	-19.4308	-10.6434
	Ser217Ala	-45.6250 (-44.2185)	-16.5316	-30.8540	16.4249	-13.7538	-11.3184
	Gln221Val	-41.9187 (-44.2185)	-11.0667	-31.4626	16.5076	-14.9278	-10.4909
	Lys453Leu	-51.1980 (-44.2185)	-20.9175	-30.2452	17.0518	-16.4198	-12.8132
HAFP-estrone	Leu138 Ser	-47.0881 (-50.31)	-11.0476	-37.6467	16.8381	-15.1892	-11.7846
	Phe172Ala	-32.3050 (-50.31)	-7.8493	-30.9881	14.3777	-8.2677	-8.0849
	His170 Leu	-40.3051 (-43.88)	-8.1035	-33.2855	15.8254	-14.8531	-10.0870
HAFP-DES	Asp478Ala	-36.5685 (-51.00)	-11.7190	-30.5551	20.7434	-15.9298	-9.1519
	His222Leu	-48.0213 (-51.00)	-11.2453	-34.2183	16.6168	-20.4277	-12.0181
HAFP-endoxifen	Cys277Val	-54.0787 (-64.18)	-36.8418	-45.6549	55.1793	-25.6210	-12.4830
	His316Leu	-54.1489 (-64.18)	-38.4019	-45.9868	54.6261	-23.2189	-12.4991
	Lys453 Glu	-50.4113 (-59.61)	-36.0847	-40.3290	45.4848	-19.6389	-11.6364
	Ser217Ala	-50.5624 (-59.61)	-43.0608	-41.9079	51.3839	-17.3874	-11.6713
	Lys228Glu	-57.2781 (-59.61)	-42.2583	-41.4879	43.1563	-17.1490	-13.2215
HAFP-afimoxifene	Lys228Leu	-66.8559 (-50.07)	-42.6339	-48.0607	51.2632	-27.7603	-15.3083
	Thr132 Val	-44.5842	-34.4384	-41.7174	45.6890	-16.7658	-10.2087

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

		(-50.07)					
	Cys224Ala	-49.2394 (-61.32)	-40.4615	-39.3824	44.3091	-16.9701	-11.2746
	Phe172Ala	-51.4201 (-61.32)	-41.2409	-45.8675	49.0546	-16.9381	-11.7739
HAFP-tamoxifen	Phe172Leu	-51.2034 (-58.74)	-32.8152	-45.6450	43.1294	-17.2336	-11.8192
	Cys224Val	-41.0369 (-58.74)	-27.3946	-36.2815	36.6834	-16.0279	-9.4725

[a]Values of MM/GBSA/ΔG_{bind} before substitutions are given in brackets

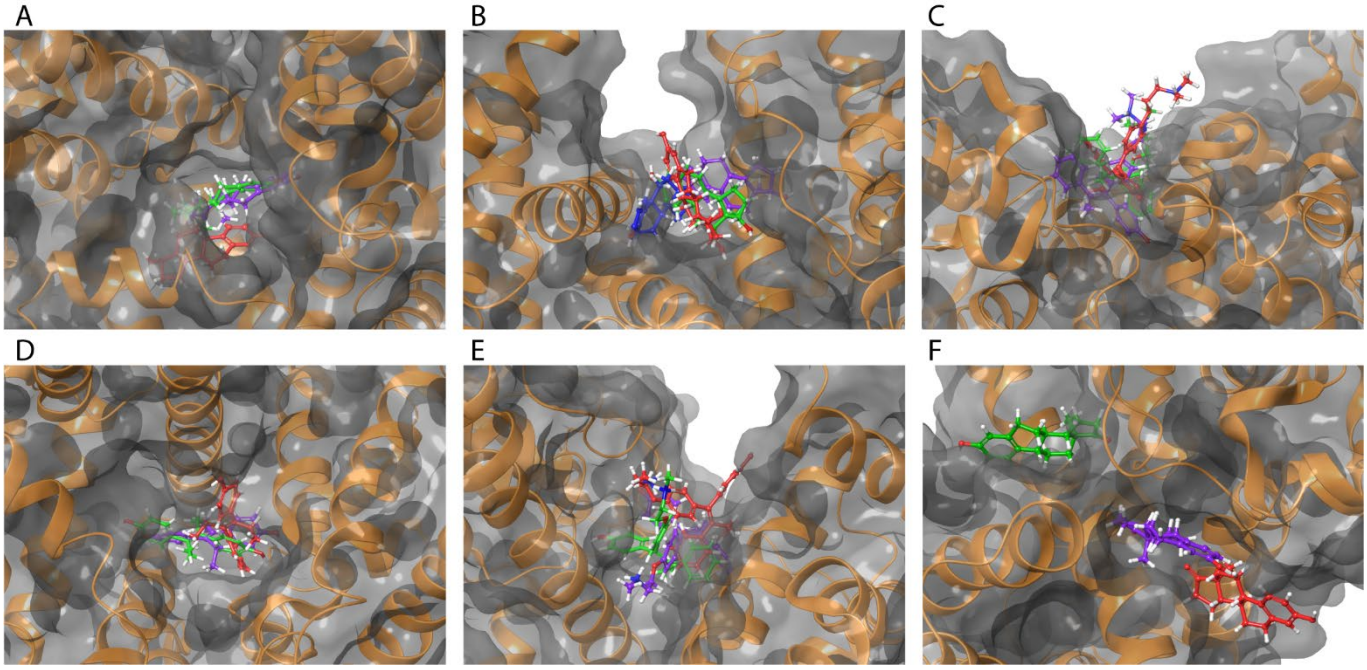

Figure 8. Superimposition of ligands in optimized HAFP-ligand complexes with mutated binding sites. Ligands in all complexes with no mutated HAFP binding sites are shown in green color. 17β-estradiol (A) in site A with substitutions Lys242Leu (violet) and His266Leu (red); (B) in site B with substitutions: Ser217Ala (violet), Lys453Leu (red) and Gln221Val (blue). Estrone in site B (C) with substitutions: Leu138Ser (violet) and Phe172Ala (red). DES in site A (D) with substitutions: His222Leu (violet) and Asp478Ala (red). Endoxifen in site B (E) with substitutions: Ser217Ala (violet) and Lys453Glu (red). Afimoxifene in site C (F) with substitutions: Phe172Ala (violet) and Cys224Ala (red).

To confirm the roles of amino acid residues in HAFP-estrone binding we performed the following substitutions: Leu138Ser, His170Leu and Phe172Ala in site B. Values of ΔG_{bind} decreased and, consequently, affinity of estrone to HAFP decreased after all substitutions. Indeed, these substitutions caused the ligand diffusion from its binding site (Figure 8C) suggesting the importance of Leu138, His170 and Phe172 in HAFP-estrone binding. Both substitutions performed for HAFP-DES complexes in site A caused decrease in ligand-binding affinities. This

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

was confirmed by visualization of the ligand binding, especially in case of Asp478Ala replacement, which caused re-arrangement of DES in the binding site (Figure 8D).

All substitutions in HAFP-endoxifen complex resulted in increase in ΔG_{bind} values (Table 3) indicating key roles of all replaced residues as well as Cys269-Cys277 disulfide bond in protein-ligand interactions. However, visualization of the complexes showed that (Figure 8E) only Lys453Glu substitution in site B caused ligand diffusion from its binding site suggesting a key role π-cation interaction between Lys453 side chain and two aromatic rings of endoxifen. All substitutions performed for HAFP-afimoxifene and HAFP-tamoxifen complexes in site C also caused decrease in binding affinities, except Lys228Leu, which improved affinity of afimoxifene binding to HAFP. Upon Lys228Leu replacement π-cation interaction was replaced by π,π-stacking between aromatic rings of Phe172 and afimoxifene, and this indicates a key role of Phe172. Visualization of afimoxifene binding in site C showed that both Cys224Ala and Phe172Ala substitutions caused escape of the ligand from its binding site confirming key roles of both Phe172 and Cys224-Cys270 disulfide in the interaction (Figure 8F).

Discussion

There are sufficient identity degrees between HAFP and HSA being up to 40% between their full length sequences and about 31, 41 and 47% between their domains I, II and III, respectively. This makes it possible to construct 3D model of HAFP based on homology with HSA, for which numerous experimentally resolved structures can be found in PDB data base. Validation of HAFP 3D model structure created in our study on the basis of homology to HSA showed its high quality and reliability enabling its usage for molecular docking of 6 ligands, three estrogens and three antiestrogens, to identify their binding modes and affinities to the protein. Importantly, identity degree between HAFP and HSA domains II and III, where principal ligand binding sites are located, are sufficiently higher than that for their domains I.

Earlier we reported construction HAFP 3D model structure using Modeller software and based on homology with the same template, HSA, PDB ID 1E78 [33]. In our previous work we studied binding of only one ligand to HAFP model, synthetic estrogen diethylstilbestrol. This was dictated by experimental data obtained by our group, which demonstrated high efficiency binding of immobilized DES to HAFP [28]. Molecular docking of DES to HAFP 3D model and subsequent MD simulation of the obtained complexes allowed identifying amino acid residues involved in HAFP-DES interaction. They included Leu138 and residues of the main α-helix located in domain III of HAFP and encompassing residues from 446 to 490, which corresponded to binding site B located at the bottom of cavity formed by domains I and III.

The purpose of the present work was to comparatively investigate binding affinities of ER agonists and antagonists to HAFP, estrogen binding and transporting protein. As a result, we found the existence of three binding sites, which differ by their ligand binding affinities. HAFP ligand-binding efficiencies were assessed using scoring functions, Glide gscore and Glide emodel. Here, Glide emodel is a scoring function primarily defined by protein-ligand Coulomb-vdW energies, and it is best suitable to select between protein-ligand complexes formed by a given ligand in different sites. Glide gscore is used to rank different by their affinities to binding site and allows discriminating SiteMap predicted sites by their ligand-binding efficiencies. In the all 3 sites (A, B and C) indentified on the basis of the scoring functions, values of gscore were better for HAFP-antiestrogens complexes than those for HAFP-estrogen complexes. Among antiestrogens, the best scoring functions were observed for docked endoxifen, while 17β-estradiol had the best docked scoring functions among estrogens. Ligand docked scoring functions were better in sites A and C as compared to site B indicating that the two first sites can be high-affinity ones, while the latter is a low-affinity site.

AFP from different mammalian species has been shown to demonstrate similar structural features, activities and functions. For example, both rodent and human AFPs can bind estrogens, retinoids, bilirubin and fatty acids [38-42]. Additionally, there are sufficiently high sequence identity degrees, 65 and 66% between HAFP-MAFP and HAFP-RAFP, respectively. Identity degrees between domains I, II and III of HAFP and rodent AFP are about 58.5, 67.0 and 71.3%, respectively. This makes it possible to compare ligand-binding affinities of rodent and human AFPs. Under *in vitro* conditions, binding of free estrogens to HAFP was not observed. However, ΔG_{exp} calculated based on experimentally obtained rodent AFP-estrogen binding K_a values (S1 and S2 Tables) evidence that 17β-estradiol has the highest binding affinity, and DES has the lowest affinity to both rat and mouse AFPs.

We calculated binding free energy values with the use of MM/GBSA rescoring method based on 100 ns MD simulation, which showed that HAFP-binding affinities of all estrogens were lower than those of antiestrogens, being highest for HAFP-endoxifen complex in site A and HAFP-afimoxifene complex in site C. Ligand efficiencies were also highest for the same complexes, being lower for estrogens as compared to antiestrogens. Analysis of MM/GBSA ΔG_{bind} energy terms showed that binding of estrogens to HAFP is provided, mostly, by van der Waals forces, while both hydrophobic and electrostatic interactions were almost equally involved in HAFP-antiestrogen binding. This can evidence that both aromatic A-ring of estrogens is a key player in HAFP-estrogen interactions, while various factors such as aromatic rings, OH-group and positive charge on N-atom in antiestrogen side chains play key roles in their interactions with the protein.

Peer-reviewed version available at *Int. J. Mol. Sci.* **2020**, *21*, 893; doi:10.3390/ijms21030893

Differences in the affinities of estrogens and antiestrogens can be explained by some distinctions in their structures along with amounts of their rotational bonds (Figure 1 and Table1). Rigid structure of estrogens that contain four condensed rings determines worse scoring functions as compared to antiestrogens, which are composed of three aromatic rings linked by relatively flexible hydrocarbon linker. The best docked pose among antiestrogens observed for endoxifen can be explained by its distinct structural features such as the presence of OH-group at C4-position, which is lacked in tamoxifen, and by the absence of second *N*-linked CH_3-group that exists in afimoxifene.

To confirm the roles of definite amino acid residues in HAFP-ligand interactions revealed by ligand docking and subsequent 100 ns MD simulation we performed *in silico* point amino acid substitutions. This allowed proposing amino acid residues, which can have key roles in binding of estrogens and antiestrogens to HAFP. Generally, point amino acid substitutions do not trigger changes in most dihedral angles in a protein and, subsequently, the entire protein conformation [43,44].

Accuracy in prediction of a role of residue substitutions in protein functions depends on residue properties such as size, hydrophobicity, charges, i.e. its type and microenvironment. In our study, we revealed that replacement of amino acid residue involved in protein-ligand interaction allows conserving the same type of interaction by ligand movement to another residue with the same physicochemical properties. This can be achieved through the ligand movement to another residue with the same physicochemical properties.

Conclusions

In our study, we showed that three major binding sites differed by values of their binding free energies can be distinguished in HAFP. Homology-based modeling, molecular docking and MD simulation studies allowed elucidating topology of the protein surface and geometries of protein ligand-binding sites along with amino acid residues involved in HAFP-ligand interactions. Point substitutions of the residues performed *in silico* allowed distinguishing those residues that can have key roles in HAFP-ligand binding. Our results are in agreement with previously obtained experimental data; however, further investigations using novel experimental approaches are required to confirm our findings.

Materials and methods

HAFP homology-based modeling

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

To perform computational HAFP homology-based 3D model construction we used a highly accurate fully-integrated Prime protein structure prediction suite of programs in Schrödinger software [45] with application of Maestro graphical interface suite [46].

Template identification

PDB [47] database was searched for experimentally obtained 3D structures of SA family proteins to be used as potential templates for HAFP 3D model building. HAFP sequence (ID: P02771) was retrieved in FASTA format from Uniprot knowledge base [48] and aligned to primary structures of candidate templates using BLAST algorithm [49]. Crystal structure of human SA (HSA) obtained at resolution of 2.6 Å (PDB ID: 1E78) was selected as the best template with sequence identity of 39% to HAFP. The edition of the target-template alignment was performed using Prime STA method and secondary structure elements were predicted using SSpro utility applied in Prime package, version 5.2.

Model generation and validation

3D model of HAFP structure was generated with the use of Protein Preparation Wizard [50] in Schrödinger software, version 2018-2. Quality and accuracy of the obtained model was validated used several methods including residue-by-residue geometry and stereochemistry assessment based on Ramachandran analysis with the implementation of PROCHECK algorithm [51].

Further, the generated model underwent refinement procedure in which any inconsistencies in the obtained structure such as incorrect bond orders, missing hydrogen atoms, and orientation of different functional groups in amino acids were checked and corrected. The model was optimized and minimized to reach the converged root mean square deviation (RMSD) of 0.30 Å with application of OPLS3e force field [52]. Additionally, to evaluate stability of the constructed model we performed RMSD and RMSF calculations for the protein backbone and Cα atoms along 200 ns MD simulation trajectory.

Model relaxation and optimization

The constructed model was subjected to optimization process during 200 ns MD simulation with application of OPLS3e force field in explicit water environment generated using TIP3P water model. For MD simulation Desmond program package, version 5.4 [53], using Maestro suite, version 11.6, in Schrödinger software, version 2018-2, was implemented. MD simulation was carried out in periodic boundary conditions created using orthorhombic box, sizes of which

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

were calculated by buffer method at distances of 10 Å along each dimension. The simulation system was electrostatically neutral in 0.15 M NaCl solution to provide physiological iso-osmotic environment. Normal pressure temperature (NPT) ensemble with implication of Nose-Hoover thermostat [54] to maintain constant temperature equal to 310 K and Martyna-Tobias-Klein barostat [55] at 1.01325 bar pressure were implemented. Cut-off radii of 9.0 Å were set for short-range Coulomb interactions. The recording interval for each frame was equal to 10 ps resulting in total of 20,000 recorded frames.

Molecular docking and scoring

Ligand preparation

Structures of ligands used in this work (Figure 1) including 17β-estradiol (ID: 5757), estrone (ID: 5870), DES (ID: 448537), tamoxifen (ID: 2733526), endoxifen (ID: 10090750) and afimoxifene (ID: 449459) were retrieved from PubChem database [56]. Comprehensive prediction of all possible tautomeric and ionization states, stereoisomers and ring conformations is very important step in ligand docking. For accurate ligand 3D model generation and conformational sampling, all ligands were subjected to structure refinement and optimization procedure using LigPrep suite [57] in Schrödinger software with application of OPLS3e force field. Protonated states were generated at pH 7.0±1.0 with implementation of Epik module, version 4.4 [58]. For each ligand, 32 possible stereo-isomeric forms were set.

Identification of protein ligand-binding sites and grid generation

SiteMap suite [59] in Schrödinger software, version 2018-2, was exploited for accurate identification and evaluation of probable binding sites in HAFP. Receptor Grid Generation suite in Schrodinger software, version 2018-2, was implemented for generation of docking grid that was centered to cover all predicted by SiteMap protein-ligand binding sites. Van der Waals radii were scaled to 1.0 Å with a partial cut-off distance of 0.25 Å to soften energy potential for non-polar parts of the receptor, while other atoms were free of scaling. Then, counter maps were generated to analyze hydrophobic and hydrophilic surfaces of the active sites.

Docking protocol

Ligand docked poses were generated by searching the most populated conformations for each protein-ligand complex. Schrödinger Glide suite, version 7.9, [60] with extra precision ligand docking protocol was employed to implement HAFP-ligand docking followed by

calculation of scoring functions. Here, ligands were sampled as flexible, while the conformer generation procedure included (non-ring) nitrogen inversions and ring conformations.

Scoring functions

To evaluate ligand docked poses, scoring functions calculated with many approximations were used. We assessed docked protein-ligand complexes using extra precision Glide score (XPG Score) tool [61], which enables calculating gscore and emodel scoring functions. Protein-ligand complexes with the best scoring functions were selected for further MD simulation.

MD simulation of protein-ligand complexes

The final ligand docked poses with the best scoring functions were subjected to 100 ns MD simulation in explicit water environment generated using TIP3P water model. Schrödinger Desmond suite, version 5.4, with application of writing intervals equal to 20 ps for each frame and OPLS3e force field was implemented. Periodic boundary conditions with box sizes calculated by buffer method at distances of 20Å along each dimension were employed. Other attributes of MD simulation protocol were the same as described earlier for HAFP model omtimization. Quality of MD simulation was assessed by Simulation Event Analysis module, and necessary data from MD simulation trajectory were extracted through various ways. Simulation Interactions Diagram Reports generated by Schrödinger software were implemented for mapping and analyzing geometry of ligand-binding sites and elucidating amino acid residues involved in HAFP-ligand interactions. The Reports include (i) protein information including secondary structure elements; (ii) ligand information including torsion angle profile and surface properties; (iii) RMSD and RMSF plots; (iv) information on protein-ligand contacts.

Protein and ligand RMSD and RMSF calculation

To assess conformational stability of the modeled protein structure during MD simulation, RMSD calculations for protein backbone and Cα atoms were implemented. RMSD analysis was also used to get an insight into conformational changes during protein-ligand interactions with and without amino acid substitutions. Analysis of ligand RMSD when it is in protein-ligand complex was performed to judge about stability of the ligand in its binding site and its ability to diffuse from the binding site during MD simulation or as a result of an amino acid substitution.

To characterize local changes along the protein polypeptide chain during MD simulation and residue-by-residue fluctuations in HAFP molecule during its interaction with a ligand,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 December 2019 doi:10.20944/preprints201912.0374.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 893; doi:10.3390/ijms21030893

calculations of root mean square fluctuation (RMSF) were made [62]. Protein RMSF plots were superimposed on secondary structure elements to evaluate their stability.

In silico point amino acid substitutions

To confirm the roles of amino acid residues involved in HAFP-ligand interactions we computationally performed point substitutions of amino acid residues in putative ligand-binding sites revealed by ligand docking and subsequent MD simulation study. Mutate Residues tool of Protein Preparation and Refinement suite in Schrödinger software, version 2018-2, was implemented to perform residue substitutions in each protein-ligand complex. Further, the mutated structures were subjected to 100 ns MD simulation using the same protocol as described above. Assessment of effects of point residue substitutions on HAFP-ligand interactions was carried out by several ways including (i) generation of Simulation Interactions Diagram Reports after each substitution; (ii) superimposition of binding poses of the same ligand in the same site with mutated residue; (iii) calculation of changes in binding free energy (ΔG_{bind}).

Binding free energy calculation by MM/GBSA rescoring

Calculation of binding free energy (ΔG_{bind}) values was exploited to estimate *in silico* ligand binding affinities. For accurate calculation of binding free energies, Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) rescoring method was used. For this purpose, we implemented Prime MM/GBSA module in Schrödinger software, version 2018-2, to generate various energy properties and contributions from different energy terms [63-65].

MM/GBSA rescoring was performed for initial ligand-docked poses with best scoring functions and for each protein-ligand complex during 100 ns MD simulation with and without amino acid substitutions in ligand-binding sites. In total, 1000 snapshots were generated from each MD simulation trajectory and average ΔG_{bind} values were calculated [66]. The free energy changes of during protein-ligand interactions were calculated with the use of OPLS3e force field and the VSGB solvent model. Before calculations all counter ions and water molecules were deleted from each system.

Binding free energy values were calculated according to the following equation (1):

$$MM/GBSA\ \Delta G_{bind} = G_{complex(optimized)} - \left(G_{protein(optimized)} + G_{ligand(optimized)}\right) \quad (1)$$

Free energy of each state, i.e. of complex, protein and ligand, was estimated by accounting molecular mechanics energies, solvation energies, and entropic terms as follows (2):

$$G = G_{int} + G_{Coulomb} + G_{vdW} + G_{GB} + G_{lipo} - TS \quad (2)$$

where G_{int}, $G_{Coulomb}$, G_{vdW} are standard MM energy terms for bond (covalent, angle and dihedral), Coulomb (electrostatic) and van der Waals interactions, G_{GB} and G_{lipo} are polar and non-polar (lipophilic) contributions to the solvation free energies, while T is an absolute temperature and S is an entropy value. Polar contribution (G_{GB}) was calculated using Generalized Born model, while non-polar contribution (G_{lipo}) was estimated based on the solvent accessible surface area (SASA).

Ligand efficiency (LE) values were calculated as follows (3):

$$LE = {}^{ln\Delta G_{bind(optimized)}}/_{n} \quad (3)$$

where n is amount of ligand heavy atoms.

To calculate experimental binding free energies (ΔG_{exp}) dissociation constant (K_d) or IC_{50} values obtained from literature data were used. ΔG_{exp} was assessed as follows (4):

$$\Delta G_{exp} = RTlnK_d \quad \text{OR} \quad \Delta G_{exp} = RT\ lnIC_{50} \quad (4)$$

Acknowledgements

This work was supported by I.M. Sechenov First Moscow State Medical University Strategic Development Program under the Russian Academic Excellence 5-100 Project.

References

1. Liu C, Xiao GQ, Yan LN, Li B, Jiang L, Wen TF, Wang WT, Xu MQ, Yang JY. Value of α-fetoprotein in association with clinicopathological features of hepatocellular carcinoma. *World J Gastroenterol* **2013**, *19(11)*, 1811-1819.
2. Sauzay C, Petit A, Bourgeois AM, Barbare JC, Chauffert B, Galmiche A, et al. Alpha-foetoprotein (AFP): A multi-purpose marker in hepatocellular carcinoma. *Clin Chim Acta*. **2016**, *463,* 39-44.
3. Zhang J, Liang R, Wei J, Ye J, He Q, Yuan C, et al. Identification of candidate biomarkers in malignant ascites from patients with hepatocellular carcinoma by iTRAQ-based quantitative proteomic analysis. *Biomed Res Int* **2018**, *2018*, 5484976.
4. Mizejewski GJ. Biological role of alpha-fetoprotein in cancer: prospects for anticancer therapy. *Expert Rev Anticancer Ther* **2002**, *2,* 709-735.
5. Terentiev AA, Moldogazieva NT. Aplha-fetoprotein: a renaissance. *Tumor Biol,* **2013**, *34,* 2075-2091.

6. Li MS, Li P.F., Yang F.Y., He S.P., Du G.G., Li G. The intracellular mechanism of AFP promoting the proliferation of NIH 393 cells. *Cell Res* **2002**, **12**, 151-156.
7. Atemezem A, Mbemba E, Marfaing R, Vaysse J, Pontet M, Saffar L, Charnaux N, Gattegno L, Human alpha-fetoprotein binds to primary macrophages. *Biochem Biophys Res Commun* **2002,** *296(3)*, 507-514.
8. Matsuura E., Kang Y., Katikawa H., Ogata P., Kotani T., Ohtaki S., Nishi S. Modulation of T-cell function by alpha-fetoprotein. An in vivo study on porcine thyroid peroxidase induced experimental autoimmune thyroiditis in transgenic mice producing human alpha-fetoprotein. *Tumor Biol* **1999**, *20*, 162-171.
9. Milligan S.R., Khan O., Nash M. Competitive binding of xenobiotic oestrogens to rat alpha-fetoprotein and sex steroid binding proteins in human and rainbow trout (Oncorhynchus mykiss) plasma. *J Comp Endocrinol* **1998**, *112*, 89-95.
10. Shahbazzadeh D. Estrogen binding activities of recombinant alpha-fetoproteins expressed in yeast. *Hokkaido Igaku Zasshi* **1995**, *70*, 473-483.
11. Aussel C, Masseyeff R. Human alpha-fetoprotein-fatty acid interaction. *Biochem Biophys Res Commun* **1983**, *115*, 38-45.
12. Mizejewski GJ. Biological roles of alpha-fetoprotein during pregnancy and perinatal development. *Exp Biol Med (Maywood)* **2004**, *229*, 439-463.
13. Sierralta W.D., Epunan M.J., Reyes J.M., Valladares L.E., Andersen T.T., Bennett J.A., Jacobson H.I., Pino A.M. A peptide derived from alpha-fetoprotein inhibits the proliferation induced by estradiol in mammary tumor cells in culture. *Oncol Rep* **2008**, *19*, 229-235.
14. Torres, C., Antileo, E., Epuñán, M.J., Pino, A.M., Valladares, L.E., & Sierralta, W.D. A cyclic peptide derived from α-fetoprotein inhibits the proliferative effects of the epidermal growth factor and estradiol in MCF7 cells. *Oncology Reports*, **2008**, *19*, 1597-1603.
15. Grober OM, Mutarelli M, Giurato G, Ravo M, Cicatiello L, De Filippo MR, et al. Global analysis of estrogen receptor beta binding to breast cancer cell genome reveals an extensive interplay with estrogen receptor alpha for target gene regulation. *BMC Genomics* **2011**, *12*, 36.
16. Zhao L, Zhou S, Gustafsson JA. Nuclear receptors: recent drug discovery for cancer therapies. *Endocr Rev* **2019**, pii:er.2018-00222.
17. Blair RM, Fang H, Branham WS, Hass BS, Dial SL, Moland CL, et al. The estrogen receptor relative binding affinities of 188 natural and xenochemicals: structural duversity of ligands. *Toxicol Sci* **2000**, *54*, 138-153.
18. Nilsson S, Makela S, Trenter E, Tujague M, Thomsen J, Andersson G, et al. Mechanisms of estrigen action. *Physiol Rev* **2001**, *81*, 1535-1565.
19. Mauvais-Jarvis F, Clegg DJ, Henever AL. The role of estrogens in control of energy balance and glucose homeostasis. *Endocr Rev* **2013**, *34*, 309-338.
20. Iorga A, Cunningham CM, Moazeni S, Ruffenach G, Umar S, Eghbali M. The protective role of estrogen and estrogen receptors in cardiovascular disease and the controversial use of estrogen therapy. *Biol Sex Differ* **2017** *8*, 33.
21. Marino M, Galluzzo P, Ascenzi P. Estrogen signaling multiple pathways to impact gene transcription. Curr Genomics **2006**, *7*, 497-508.
22. Hong H, Branham WS, Ng HW, Moland CL, Dial SL, Fang H, et al. Human sex hormone-binding globulin binding affinities of 125 structurally diverse chemocals and comparison with their binding to androgen receptor, estrogen receptor, and alpha-fetoprotein. *Toxicol Sci* **2015**, *143(2)*, 333-348.
23. Clarke BL, Khosla S. Modulators of androgen and estrogen receptor activity. *Crit Rev Eukaryot Gene Expr* **2010**, *20*, 275-294.

24. Cornil CA, Ball GF, Balthazart J. The dual action of estrogen hypothesis. *Trends Neusci* **2015**, *38*, 408-416.
25. Kojetin DJ, Burris TP, Jensen EV, Khan SA. Implications of the binding of tamoxifen to the coactivator recognition site of the estrogen receptor. Endocr Relat Cancer. 2008; 15: 851-870.
26. Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, Greene GL. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95*, 927-937.
27. Dai S, Chalmers MJ, Bruning J, Bramlett KS, Osborne HE, Montrose-Rafizadeh C, Barr RJ, Wang Y, Wang M, Burris TP, Dodge J, Prediction of the tissue-specificity of selective estrogen receptor modulators by using a single biochemical method. *Proc Natl Acad Sci USA* **2008**, *105*, 7171-7176.
28. Tatarinov YuS, Terentiev AA, Moldogazieva NT, Tagirova AK, Human alpha-fetoprotein and its purification by chromatography on immobilized estrogens. *Tumor Biol* **1991**, *12*, 125-130.
29. Terentiev AA, Moldogazieva NT. Structural and functional mapping of alpha-fetoprotein. *Biochemistry (Mosc)* **2006**, *71*, 120-132.
30. Nishi S, Matsue H, Yoshida H, Yamamoto R, Sakai M. Location of the estrogen-binding site of alpha-fetoprotein in the chimeric human-rat proteins. *Proc Natl Acad Sci USA* **1991**, *88(8)*, 3102-3105.
31. Nishi S, Shahbazzadeh D, Azuma M, Sakai M, Estrogen binding site of rat AFP. *Tumour Biol* **1993**, *14(4)*, 234-237.
32. Nishio H, Dugaiszyk A. Complete structure of the human alpha-albumin gene, a new member of the serum albumin multigene family. *Proc Natl Acad Sci USA* **1996**, *93*, 7557-7561.
33. Terentiev AA, Moldogazieva NT, Levtsova OV, Maximenko DM, Borozdenko DA, Shaitan KV. Modeling of three dimensional structure of human aplha-fetoprotein complexed with diethylstilbestrol: docking and molecular dynamics simulation study. *J Bioinform Comput Biol* **2012**, *10*, 1241012.
34. Shen J, Zhang W, Fang H, Perkins R, Tong W, Hong H. Homology modeling, molecular docking, and molecular dynamics simulations elucidated α-fetoprotein binding modes. *BMC Biology* **2013**, *14 (Suppl 14)*, 56.
35. Hong H, Branham WS, Dial SL, Moland CL, Fang H, Shen J, et al. Rat α-fetoprotein binding affinities to a large set of structurally diverse chemicals elucidated the relationships betweeen structures and binding affinities. *Chem. Res. Toxicol* **2012**, *25(11)*, 2553-2566.
36. Luft AJ, Lorscheider FL. Structural analysis of human and bovine α-fetoprotein by electron microscopy, image processing and circular dichroism. *Biochemistry* **1983**, *22*, 5971-5978.
37. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. *Proteins*, **2004**, *55*, 351-367.
38. Uriel J, Bouillon D, Aussel C, Dupiers M. Alpha-fetoprotein: the major high-affinity estrogen binder in rat uterine cytosols. *Proc Natl Acad Sci USA* **1976**, *73(5)*, 1452-1456.
39. Versee V, Barel AO. Binding specificityof estrogens and norandrogens to rat alpha-fetoprotein (AFP). *FEBS Letters* **1978**, *96(1)*, 155-158.
40. Savu L, Benassayag C, Vallette G, Christeff N, Nunez EA. Mouse α-fetoproptein and albumin. A comparison of their binding properties with estrogen and fatty acid ligands. *J Biol Chem* **1981**, *256(18)*, 9414-9418.
41. Savu L, Benassayag C, Vallette G, Nunez EA. Ligand properties of diethylstylbestrol: studies with purified native and fatty acid-free rat alpha-fetoprotein and albumin. *Steroids* **1979**, *34(7)*, 737-748.
42. Keel BA, Abney TO. The kinetics of estrogen binding to rat alpha-fetoprotein. *Experientia* **1984**, *40*, 503-504.

43. Moldogazieva NT, Terentiev AA, Shaitan KV. Relationship between structure and function of alpha-fetoprotein: conformational changes and biological activity. *Biomed Khim* **2005**, *51(2)*, 127-151.
44. Feyfant E, Sali A, Fiser A. Modeling mutations in protein structures. *Protein Sci* **2007**, *16*, 2030-2041.
45. Schrodinger Release 2018-2. Prime. Schrodinger, LLC; New-York, NY, USA, 2018.
46. Schrödinger, Release.2018-2. Maestro. Schrödinger, LLC; New York, NY, USA, 2018.
47. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28*, 235-242.
48. Chen C, Huang H, Wu CH. Protein bionformatics databases and resources. *Methods Mol Biol* **2017**, *1558*, 3-39.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*, **1990**, *215*, 403-410.
50. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screenong enrichments. *J Comput Aided Mol Des*, **2013**, *27*, 221-234.
51. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to heck the stereochemical quality of protein structures. *J Appl Cryst*, **1993**, *26*, 283-291.
52. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J Chem Theory Comput* **2016**, *12*, 281-296.
53. Schrödinger Release 2018-2. Desmond Molecular Dynamics System. Schrodinger, LLC; New York, NY, USA, 2018.
54. Hünenberger H. Thermostat algorythms for molecular dynamics simulations. In: Holm C, Kremer K, editors. *Advanced computer simulation*. Springer: New York, USA, 2005. pp. 105-149.
55. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *J Chem Phys* **1994**, *101*, 4177-4189.
56. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A. PubChem substance and compound databases. *Nucleic Acid Res* **2016**, *44*, D1202-D1213.
57. Schrödinger Release 2018-2. LigPrep. Schrödinger, LLC; New York, NY, USA, 2018.
58. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* **2007**, *21*, 681-691.
59. Halgren TA, Chem J. Identifying and characterizing binding sites and assessing druggability. *Inf Model* **2009**, *49*, 377-389.
60. Cho AE, Guallar V, Berne BJ, Friesner R. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J Comput Chem* **2005**, *26*, 915-931.
61. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004**, 47, 1739-1749.
62. Kufareva I, Abagyan R. Methods of protein structure comparison. *Methods Mol Biol* **2012**, *857*, 231-257.
63. Aqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng* **1994**, 7, 385-391.
64. Genheden S, Ryde U. The MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* **2015**, *10*, 449-461.

65. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* **2011**, *51*, 69-82.
66. Zhang X, Perez-Sanchez H, Lightstone FC. A comprehensive docking and MM/GBSA rescoring study of ligand recognition upon binding antithrombin. *Curr Top Med Chem* **2017**, *17*, 1631-1639.
67. Wilkinson MD, Dumontier M, Aalsbersberg IJJ, Appleton G, Axton M, Baark A. et al. The FAIR guiding principles for scientific data management and sweardship. *Sci Data* **2016**, *3*, 160080.
68. Elofsson A, Hess B, Lindahl E, Onufriev A, van der Spoel D, Wallqvist A. Ten simple rules on how to create open access and reproducible molecular simulations of biological systems. *PLoS Comput Biol* **2019**, *15(1)*, e1006649.

Supplementary materials

S1 Table. Experimental binding affinities of estrogens to rat AFP

S2 Table. Experimental binding affinities of estrogens to mouse AFP

MD simulation input, output and trajectories files. They are deposited in Harvard Dataverse Network according to FAIR data principles [67, 68]. They can be found through the following link: http://doi.org/10/7910/DVN/ATG6NH

S1 Table. Experimental binding affinities of estrogens to rat AFP

Ligand	K_a RAFP (M^{-1})	K_d RAFP (M)	ΔG_{RAFP} (kcal/mol)	References
17β-estradiol	From 9.3 x 10^8 To 11.4 x 10^8	From 0.107 x 10^{-8} To 0.088 x 10^{-8}	From -12.232 To -12.347	[30]
17β-estradiol	2.83±0.78 x10^8	0.353 x 10^{-8}	-11.525	[38]
17β-estradiol	5.0±1.0 x 10^7	0.200 x 10^{-7}	-9.135	[39]
17β-estradiol	From 0.6 x10^8 To 1.4 x10^8	From 1.667 x 10^{-8} To 0.714 x 10^{-8}	From -10.606 To -11.106	[38]
Estrone	5.5±1.01 x 10^8	0.182 x 10^{-8}	-11.917	[39]
Estrone	9.0 x 10^7	0.111 x 10^{-7}	-10.847	[38]
DES	1.5 x 10^6	0.667 x 10^{-6}	-8.452	[41]

S2 Table. Experimental binding affinities of estrogens to mouse AFP

Ligand	K_a MAFP (M^{-1})	K_d MAFP (M)	ΔG_{MAFP} (kcal/mol)	References
17β-estradiol	0.8 x10^8	1.25 x 10^{-8}	-10.777	[40]
DES	0.2 x 10^7	5.00 x 10^{-7}	-8.592	[40]