Heliyon 10 (2024) e37297

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Exploring pathogenic SNPs and estrogen receptor alpha interactions in breast cancer: An *in silico* approach

Ahmad M. Alamri [a,b,**], Faris A. Alkhilaiwi [c,d], Najeeb Ullah Khan [e,***], Reham Mahmoud Mashat [f], Munazzah Tasleem [g,*]

[a] Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Khalid University, Abha, 61413, Saudi Arabia
[b] Cancer Research Unit, King Khalid University, Abha, 61413, Saudi Arabia
[c] Department of Natural Products and Alternative Medicine, Faculty of Pharmacy, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[d] Regenerative Medicine Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[e] Institute of Biotechnology and Genetic Engineering (Health Division), The University of Agriculture Peshawar, 25130, Pakistan
[f] College of Science, Department of Biology, King Khalid University, Abha, 61413, Saudi Arabia
[g] Center for Global Health and Research, Saveetha Medical College and Hospital, Chennai, 602105, India

ARTICLE INFO

Keywords:
Single nucleotide polymorphism
Mutation
Molecular docking
Breast cancer
Intra-molecular interaction

ABSTRACT

The estrogen receptor 1 gene (*ESR1*) plays a crucial role in breast and mammary development in humans. Alterations such as gene amplification, genomic rearrangements, and missense mutations in the *ESR1* gene are reported to increase the risk of breast cancer in humans. The purpose of this study is to analyze the missense mutations and molecular modeling of *ESR1*, focusing on the pathogenic SNP H516N, for a better understanding of disease risk and future benefits for therapeutic benefits. This SNP was selected based on its location in the binding pocket of ESR1 and its predicted impact on drug binding. The *in silico* analysis was performed by applying various computational approaches to identify highly pathogenic SNPs in the binding pocket of ESR1. The effect of the SNP was explored through docking and intra-molecular interaction studies. All SNPs in *ESR1* were identified followed by the identification of the highly pathogenic variant located in the binding pocket of ESR1. The mutant model of the pathogenic SNP H516N was generated, and hydroxytamoxifen was docked with the wild-type and the mutant model. The mutant model lost the formation of stable hydrogen bonds with the active site residues and hydroxytamoxifen, which may result in reduced binding affinity and therefore, will predict the patient's response to estrogenic inhibitors.

1. Introduction

Estrogen is a steroid hormone produced by ovaries in the reproductive cycle stages and plays a crucial role in the growth and development of the breast [1]. Estrogen binds to estrogen receptors (ERs) of the nucleus forming a receptor-ligand complex, which binds to the promoter region of estrogen response elements in estrogen receptor genes via protein-protein interaction [2]. This event

* Corresponding author.
** Corresponding author. Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Khalid University, Abha 61413, Saudi Arabia.
*** Corresponding author.
E-mail addresses: aalamri@kku.edu.sa (A.M. Alamri), najeebkhan@aup.edu.pk (N.U. Khan), munazzah.t@gmail.com (M. Tasleem).

https://doi.org/10.1016/j.heliyon.2024.e37297
Received 11 February 2023; Received in revised form 28 August 2024; Accepted 30 August 2024
Available online 31 August 2024

recruits other co-regulatory proteins towards the promoter sequence for increasing and decreasing the expression levels.

ER functions as a ligand-dependent transcription factor consisting of two estrogen receptor isoforms i.e. ERα and ERβ, sharing sequence homology [3]. ER plays a vital role in mammary development, whereas the function of ER is unknown; nonetheless, investigations have shown that the absence of this gene does not result in reduced function [4,5]. *ESR1* encodes the Estrogen Receptor alpha (ERα), which is composed of 595 amino acids, that encode the ERα gene, a molecular weight of (300 Kb), consisting of seven introns and nine exons with the first eight exons encoding the full-length isoform of ERα [6]. ERα consists of six structural domains, which include two binding domains i.e. DNA binding domain named DBD/C-domain and the ligand binding domain named LBD/E-domain [7]. The ligand binding domain of ERα consists of a C-terminal helix named H12 which controls the antagonist state of the receptor [8]. The receptor is activated when the ligand binds to LBD causing the stabilization of H12 in active conformation and allowing the subsequent binding of co-activators [9]. On the other hand, receptor activity is inhibited after an antagonist binds to LBD and causes stabilization of H12 inactive conformation which results in the binding of co-repressors [4].

The prevalence of different cancers is significantly influenced by genetic variation. The most frequent genome variation in humans associated with diseases is Single Nucleotide Polymorphism (SNP). It accounts for 90 % of all genetic variations and exists after every 100–300 nucleotide bases in the genome sequence [10]. Although SNPs exist in both coding and non-coding regions of the genome with variable densities, they occur frequently in non-coding regions of the human genome such as introns, regulatory regions, and untranslated regions [11]. SNPs in UTRs result in the alteration of transcriptional activity, while SNPs in regulatory regions affect the gene expression and binding ability of transcriptional factors [12]. On the other hand, SNPs in the coding regions cause changes in amino acids resulting in variable protein product formation and leading to several different diseases [13].

Polymorphism in the *ESR1* results in breast cancer including gene amplification, genomic rearrangements, and missense mutations [14]. A multitude of SNPs in *ESR1* associated with breast cancer incidence have been identified in recent years and there is much research under process for comprehending the role of these SNPs in breast cancer [15–17]. Recent reports displayed that about 70 % of breast cancers are estrogen receptor-positive [18]. Numerous cellular activities, including cellular development, differentiation, and reproductive system operation, are regulated by *ESR1*. The development and upkeep of the skeleton, as well as the healthy operation of the neurological and cardiovascular systems, are all regulated by estrogen. The development or changed expression of mutant and/or variant versions of the estrogen receptor is one mechanism that has been proposed to contribute to the transition of human breast cancer from hormone reliance to independence [19].

Genetic epidemiology requires identifying disease-causing gene variations. SNPs can be used to study many genetic traits and their problems [20]. Many of the SNPs identified in *ESR1* have not yet been thoroughly studied for their potential to cause breast cancer [18]. The ligand-binding domain of ESR1 is critical for its function, as it directly interacts with ligands and co-regulatory proteins. Mutations within this domain can significantly impact the receptor's function and its interaction with therapeutic agents. It has been demonstrated that hydroxytamoxifen, a selective estrogen receptor modulator (SERM) and active metabolite of tamoxifen, is a drug of choice for treating estrogen receptor-positive breast cancer due to its ability to bind to the LBD and inhibit receptor activity [21,22]. Therefore, the structure of ESR1 bound with hydroxytamoxifen (PDB ID: 7UJ8) provides a valuable framework for studying these interactions [23].

In the present study, highly pathogenic SNPs that lie in the binding pocket of ESR1 are identified. We aim to explore the structural and functional consequences of the mutation on drug binding, particularly with hydroxytamoxifen. This study provides insights into the potential impact of this mutation on the effectiveness of estrogenic inhibitors in breast cancer treatment. The effect of the SNP on drug binding is assessed by applying a molecular docking approach. To create more effective and strong inhibitors to target these mutant receptors in the therapy, it is essential to have a thorough understanding of how these and other gain-of-function mutations affect the structure and function of ESR1.

2. Materials and methods

2.1. Dataset retrieval

Data for *ESR1* variant analysis was obtained from the online gnomAD database (V2.1) released on October 17, 2018 (https://gnomad.broadinstitute.org). During data extraction from gnomAD, a variety of filters were implemented, namely loss of function [24], 3' & 5′ UTRs, splice region, synonymous, and missense (non-synonymous). The 75 missense mutations that probably result in breast cancer were identified using Variation Viewer. The UniProt ID: P03372, "NM_000125.4" was used for the analysis of *ESR1* Variants [25].

2.2. Structural and motif analysis of ESR1

To identify the motifs within the ESR1 protein sequence, we used the MEME (Multiple EM for Motif Elicitation) tool version 5.5 [26]. The protein sequence of ESR1 was retrieved from the UniProt database (UniProt ID: P03372). The tool was set to search for motifs with a minimum width of 6 and a maximum width of 50. The Jpred 4 server [27] was used for secondary structure prediction, predicting secondary structure elements like alpha-helices, beta-strands, and coils using multiple sequence alignments and neural networks. Additionally, the ESR1 sequence's topology was analyzed using InterProScan, a comprehensive tool that integrates protein signature databases, to identify domains, repeats, and key sites [28].

2.3. Deleterious variants

A web-based tool called CADD (Combined annotation dependent depletion) (https://cadd.gs.washington.edu/) has over 60 annotations [29]. This program provides a C-score and a thorough analysis of variations. The C-score, generally referred to as the "Phred score," reliably predicts whether a variation would have a high or low deleterious effect on the protein. The results were generated after a variant list was uploaded to CADD. The remaining variants were then investigated for further analysis after a C-score filter ($\geq$15) was applied to the CADD file.

2.4. Pathogenicity of SNPs

The pathogenicity of amino acid substitution and its molecular processes are predicted using the MutPred analysis tool (http://mutpred.mutdb.org/) [30]. It is used to classify amino acid changes in humans as either benign or pathogenic/deleterious. MutPred has a cut-off value of >0.5, indicating that this substitution is pathogenic or would have negative effects. The prediction for disruptive features either functional or structural can be characterized for a given mutation.

2.5. Stability analysis of SNPs

The impact of amino acid substitutions on protein stability was assessed using the computational tools I-Mutant 2.0 [31] and MUpro [32]. I-Mutant 2.0 was used to predict the change in free energy ($\Delta\Delta G$) based on the protein sequence, while MUpro utilized neural networks and support vector machines to estimate $\Delta\Delta G$ without requiring structural information. For each variant of interest, the DDG score was obtained from both I-Mutant 2.0 and MUpro. Positive DDG values were interpreted as indicating increased protein stability, while negative values suggested the mutation would destabilize the protein.

2.6. Ligand preparation

The 3D structure of Hydroxytamoxifen was retrieved through PubChem [33]. The compound was prepared for the docking process by standardizing charges and adding hydrogens in Swiss-PdbViewer (http://www.expasy.org/spdbv/) [34].

2.7. Target protein preparation

The three-dimensional structure of the target protein ESR1 PDB ID 7UJ8_A was obtained via the RCSB Protein Data Bank [35]. The X-ray crystal structure of ESR1 with inhibitor 4- hydroxytamoxifen attached was obtained from the RCSB Protein Data Bank, crystallized at 2.38 Å with a molecular weight of 60.84 kDa and 262 amino acid residues in a single unique chain [23]. The mutant model of ESR1 H516N was built by the "Mutagenesis" module from PyMOL [36]. The target protein was prepared by adding missing atoms, correcting connectivity, correcting names, and inserting missing loops using Swiss-PdbViewer. Energy minimization of the prepared protein was carried out by Swiss-Pdb Viewer. The binding pocket of 7UJ8 was selected by creating a sphere of radius 13.12 Å around the inhibitor OHT with x, y, and z coordinates 21.191095, −26.165193, and 12.794570, respectively.

2.8. Molecular docking and intramolecular interaction assessment

The iGEMDOCK was utilized to complete the docking of the hydroxytamoxifen. Lib Dock takes care of everything from ligand and protein preparation through molecular docking. While docking, scoring functions are employed to differentiate between active and

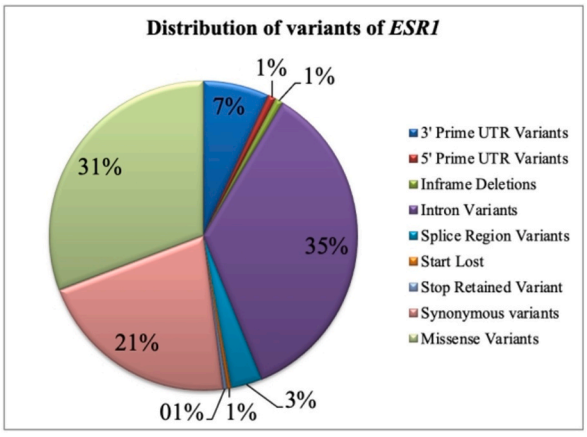

Fig. 1. The distribution of SNPs across the ESR1 gene revealed by gnomAD database.

random compounds and to predict binding free energies in ligand-protein docking [37]. iGEMDOCK provides the post-analysis tools by using k-means and hierarchical clustering methods based on the docked poses (protein-ligand interactions) and compound properties (atomic compositions) [38]. Intramolecular interactions were analyzed by the "View Interaction" module from Discovery Studio Visualizer [BIOVIA, Dassault Systèmes, Discovery Studio Visualizer, v21.1, San Diego: Dassault Systèmes, 2021].

3. Results

3.1. SNP datasets

The gnomAD database (V2.1) (http://www.gnomad-sg.org/) released on October 17, 2018; was used to retrieve the pathogenic nsSNPs using filters [39]. A total of 244 SNPs were reported in the Human *ESR1* gene. On further selection, 6.9 % were found to be 3′ UTRs, 0.8 % as 5' UTRs, 35.2 % as intron variants, 0.8 % in-frame deletion/insertions, 0.4 % as start lost/gain, 3.2 % as splice regions, 21.31 % were synonymous and 30.7 % were nonsynonymous as depicted (Fig. 1). After filtering large data, 75 nsSNPs were further analyzed for the investigation of most pathogenic variants.

3.2. Structural and motif analysis of ESR1

The ESR1 protein sequence was analyzed using the MEME tool, revealing two key motifs with high statistical significance. These motifs, sequence "**NWAKRVPGFV"** and sequence "**NQGKSVEGMV,"** indicate critical regions within the protein that may play significant roles in its structural stability and function. These motifs also indicate potential binding sites or interaction regions essential for the protein's activity in estrogen receptor signaling pathways. The Jpred 4 analysis predicted the secondary structure elements of ESR1, including alpha-helices and beta-strands, which are crucial for receptor stability and function, Figure S 1. InterProScan analysis of the ESR1 protein sequence revealed the presence of crucial domains essential for its function. The nuclear receptor (NR) ligand-binding domain (LBD) (residues start: 325- end: 546), along with the ligand-binding domain of the nuclear hormone receptor and HOLI domain, were identified. Furthermore, the ESR1 was annotated for functions, which highlighted its involvement in key biological processes such as the intracellular estrogen receptor signaling pathway, regulation of transcription by RNA polymerase II, and cellular response to estrogen stimulus. At the molecular level, ESR1 exhibits functions including estrogen response element binding and nuclear receptor activity. In terms of cellular localization, ESR1 is predominantly located in the chromatin and nucleus, highlighting its pivotal role in transcriptional regulation and chromatin dynamics.

3.3. Prediction of pathogenic nsSNPs of the ESR1 using the CADD database

A total of 75 variants were selected for the analysis of most pathogenic variants using the CADD database. After passing them from different predictive algorithms (SIFT, Polyphen2, MutationTaster, PROVEAN, CADD) nsSNPs showed hazardous impacts on ESR1 proteins. This prediction was done by observing the scores of variants as mentioned (Table S1). SIFT scores range from 0 to 1. Scores less than 0.05 are predicted to be deleterious, while scores greater than or equal to 0.05 are predicted to be tolerated. PolyPhen-2 scores range from 0.0 (tolerated) to 1.0 (deleterious). Scores from 0.0 to 0.15 are predicted to be benign. Scores from 0.15 to 1.0 are possibly damaging. Scores from 0.85 to 1.0 are more confidently predicted to be damaging. PROVEAN scores range from −14 to 14. A default score threshold of −2.5 is used to classify variants as deleterious (score ≤ -2.5) or neutral (score > -2.5). MutationTaster provides a probability value (ranging from 0 to 1) that the alteration is either pathogenic or benign. Values closer to 1 indicate a high 'security' of the prediction. The CADD Phred score greater than 15 is considered deleterious, as shown in Fig. 2.

3.4. Pathogenic SNPs

Out of the 50 variants that were retrieved from the CADD Database, MutPred analyses the nine most highly pathogenic variants.

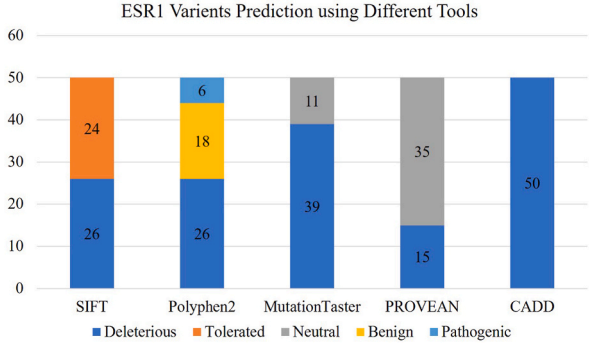

Fig. 2. The distribution of deleterious and tolerated nsSNPs by numbers identified by SIFT, Polyphen 2, Mutation Taster, PROVEAN, and CADD.

Based on their scores, disruptive structural and functional features of the pathogenic variants, such as loss of loop, altered disordered interface, and altered transmembrane protein, were also analyzed (Table 1). All the highly pathogenic variants predicted by MutPred have a greater than 0.5 score.

3.5. Stability analysis

The predicted impacts of the tested amino acid substitutions on protein stability are summarized in Table 2. The change in free energy (ΔΔG) values were obtained.

3.6. Molecular docking and intramolecular interaction assessment

The structure of ESR1 was retrieved through the RCSB Protein Data Bank. The structure lacks the atomic coordinates of some residues, which were built using Swiss-PdbViewer (SPDBV). The energy of the ESR1 (PDB ID: 7UJ8_A) was −6717.289 kJ/mol as calculated by SPDBV. The structure was then energy minimized to −6869.719 kJ/mol. The wild-type structure of 7UJ8_A starts from residue A307 and ends at A546, as shown in Fig. 3. The mutant model of ESR1 H516N was generated using the "Mutagenesis" module from PyMOL. The energy of the mutant model was estimated to be −10133.405 kJ/mol, which was further minimized to −10272.704 kJ/mol. Docking is a structural biology tool for studying how two molecules interact with one another. In the current study, we have examined how the ESR1 protein and its specific mutant H516N docks with hydroxytamoxifen. The drug hydroxytamoxifen was observed to bind in the binding pocket of ESR1 and form stable hydrogen bond interactions with Gly420 and Asp351, as shown in Fig. 4. However, the covalent hydrogen bonds are lost in the docking of hydroxytamoxifen with the ESR1 mutant. Other close intramolecular interactions formed between the hydroxytamoxifen and the ESR1 and mutant ESR1 are shown in Tables 3 and S2.

4. Discussion

Pathogenic SNPs were discovered in the Human ESR1 gene after analysis of a total of 244 SNPs. If the SNPs are found in the binding pocket, they may influence how the drug binds. As a result, research was done on the pathogenic SNP H516N that is present in the binding pocket.

Structural analysis revealed the identification of two significant motifs, "NWAKRVPGFV" and "NQGKSVEGMV," using MEME analysis highlights important areas that are likely responsible for maintaining structural integrity [40]. In addition, the secondary structure predictions made by Jpred 4 indicate a clearly defined arrangement of alpha-helices and beta-strands that are crucial for maintaining the stability and functional integrity of ESR1 [41]. In addition, the InterProScan analysis provides evidence of the existence of important domains, including the nuclear receptor ligand-binding domain (LBD) [42] and the HOLI domain. This is supported by X-ray crystallography data (Fig. 3). The presence of these domains highlights the significant role of ESR1 as a key participant in the regulation of transcription mediated by estrogen. Moreover, functional annotation provides additional clarification on its key role in biological processes, such as intracellular estrogen signaling, transcriptional regulation by RNA polymerase II, and cellular responses to estrogen stimuli. ESR1 demonstrates distinct functions at the molecular level, including binding to estrogen response elements and acting as a nuclear receptor. These functions indicate its role in regulating gene expression. Cellular localization studies demonstrate that ESR1 is primarily located in the chromatin and nucleus, suggesting its essential function in regulating chromatin dynamics and transcriptional activities related to estrogen signaling [43].

Estrogen receptor alpha activation by the hormone promotes breast cancer in many ways. When estrogen attaches, the receptor alters shape, exposing a surface for coactivators. After a coactivator binds to the estrogen receptor, the cancer cell grows, divides, infiltrates surrounding tissues, and spreads. Antiestrogen medications change the receptor's structure to impede coactivator binding

Table 1
Highly pathogenic variants of *ESR1* predicted by MutPred algorithm.

Variation	Predicted Score	Structural and Functional Features of ESR Receptor
A65P	0.562	Gain of Loop (Pr = 0.27 |P = 0.04); Loss of Sulfation at Y60 (Pr = 0.03 |P = 0.02); Loss of GPI-anchor amidation at N63 (Pr = 0.01 |P = 0.02)
Y73C	0.672	Loss of Sulfation at Y73 (Pr = 0.01 |P = 0.04)
S118P	0.641	Loss of Loop (Pr = 0.28 |P = 0.02); Gain of B-factor (Pr = 0.25 |P = 0.03)
E181A	0.578	Altered Transmembrane protein (Pr = 0.11 |P = 0.03)
Q226E	0.714	Altered Disordered interface (Pr = 0.29 |P = 0.03); Gain of Disulfide linkage at C221 (Pr = 0.26 |P = 3.4e-03); Altered Transmembrane protein (Pr = 0.20 |P = 5.9e-03); Loss of GPI-anchor amidation at N225 (Pr = 0.01 |P = 0.03)
R256Q	0.618	Altered Disordered interface (Pr = 0.51 |P = 2.4e-03); Loss of Helix (Pr = 0.29 |P = 0.02); Gain of Acetylation at K257 (Pr = 0.25 |P = 0.01); Altered DNA binding (Pr = 0.15 |P = 0.04)
K268N	0.535	Altered Disordered interface (Pr = 0.37 |P = 8.3e-03); Gain of Acetylation at K266 (Pr = 0.21 |P = 0.03); Loss of SUMOylation at K268 (Pr = 0.19 |P = 0.04); Gain of Proteolytic cleavage at R263 (Pr = 0.13 |P = 0.02)
R269C	0.696	Altered Disordered interface (Pr = 0.41 |P = 6.2e-03); Loss of B-factor (Pr = 0.26 |P = 0.04); Gain of Acetylation at K266 (Pr = 0.23 | P = 0.02)
R269H	0.503	Altered Disordered interface (Pr = 0.40 |P = 6.6e-03); Loss of Acetylation at K266 (Pr = 0.21 |P = 0.03)
H516N	0.565	Altered Coiled coil (Pr = 0.15 |P = 0.03); Gain of GPI-anchor amidation at N519 (Pr = 0.01 |P = 0.03)

Table 2
Predicted protein stability changes.

Variation	I-Mutant 2.0 (ΔΔG)	MUPRO (ΔΔG)
A65P	−0.98	−1.28209
Y73C	1.75	−0.50018
S118P	0.17	−0.86991
E181A	−0.29	−1.1997
Q226E	0.02	−0.44146
R256Q	−1.1	−0.64152
K268N	0.28	−0.33979
R269C	−0.61	−0.83622
R269H	−0.21	−1.06756
H516N	−1.66	−1.04068

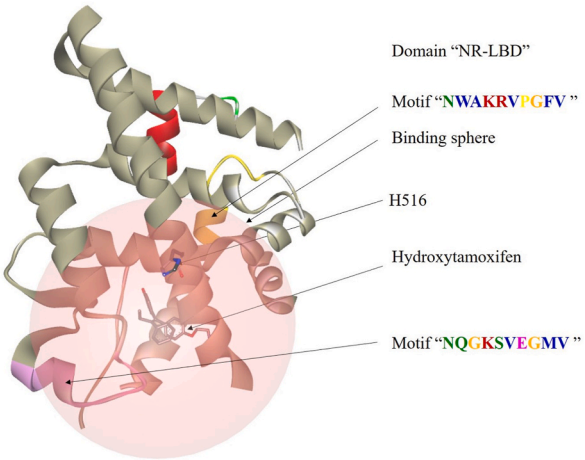

Fig. 3. Cartoon representation of ESR1 structure (PDB: 7UJ8_A), highlighting the SNP H516N (Shown in sticks), NR-LBD domain highlighted in olive color, motifs, and binding sphere around the bound ligand hydroxytamoxifen.

and compete with estrogen for receptor binding [44]. However, recent studies show that many metastatic breast cancer patients have mutations in the estrogen receptor alpha gene. The free binding energy of the docked complexes shows insignificant changes; however, loss of a stable hydrogen bond interaction was noted in the mutant-hydroxytamoxifen complex, which may result in reduced binding affinity leading to decreased effectiveness in the inhibitory effect of the anti-estrogen drug. The directionality of drug-receptor hydrogen bonds is crucial in regulating the specificity of drug-receptor binding. Hydrogen bonds between drugs and receptors are regarded as key factors in binding. However, hydrophobic interactions increase the binding of the drug and contribute significantly to the affinity of most drugs to the receptors [45]. The common hydrophobic interactions formed between the wild-type ESR1, mutant ESR1 H516N and hydroxytamoxifen Ala350, Leu346, Ala350, Leu525, Met421, and Leu387. However, the hydrophobic interactions with Trp383 and Met481 are lost in the mutant ESR1 H516N. The knowledge acquired from these frequent alterations can be applied in the development of new antiestrogens to successfully treat breast cancer.

5. Conclusion

The current study aimed to fill the gap between genetic analysis and therapeutic methodologies for treating breast cancer. This study is based on *in silico* studies of the *ESR1* and its SNPs. The revealed pathogenic SNPs, one of these is located in the binding pocket was further studied for its effect on drug binding. The mutant model of the SNP H516N was generated, where hydroxytamoxifen was docked and the interactions were compared with the wild-type ESR1. The SNP H516N was observed to loose some specific hydrogen bond and hydrophobic interactions that may result in reduced binding affinity of the drug. Therefore, as an approach of precision medicine this study will be helpful in predicting the patient's response to estrogenic inhibitors. Moreover, a novel drug can be designed to cure breast cancer patients with H516N mutation that can effectively bind to the mutant ESR1.

6. Limitations and Future Perspective

The study provides valuable insights into the effect of ESR1 SNP H516N, however, future experimental validation is required. Current limitations of the study include a single SNP and a single ligand, therefore, future studies are required to explore broader SNP profiles and additional therapeutic agents for better understanding and treatment strategies. Future avenues for research could greatly

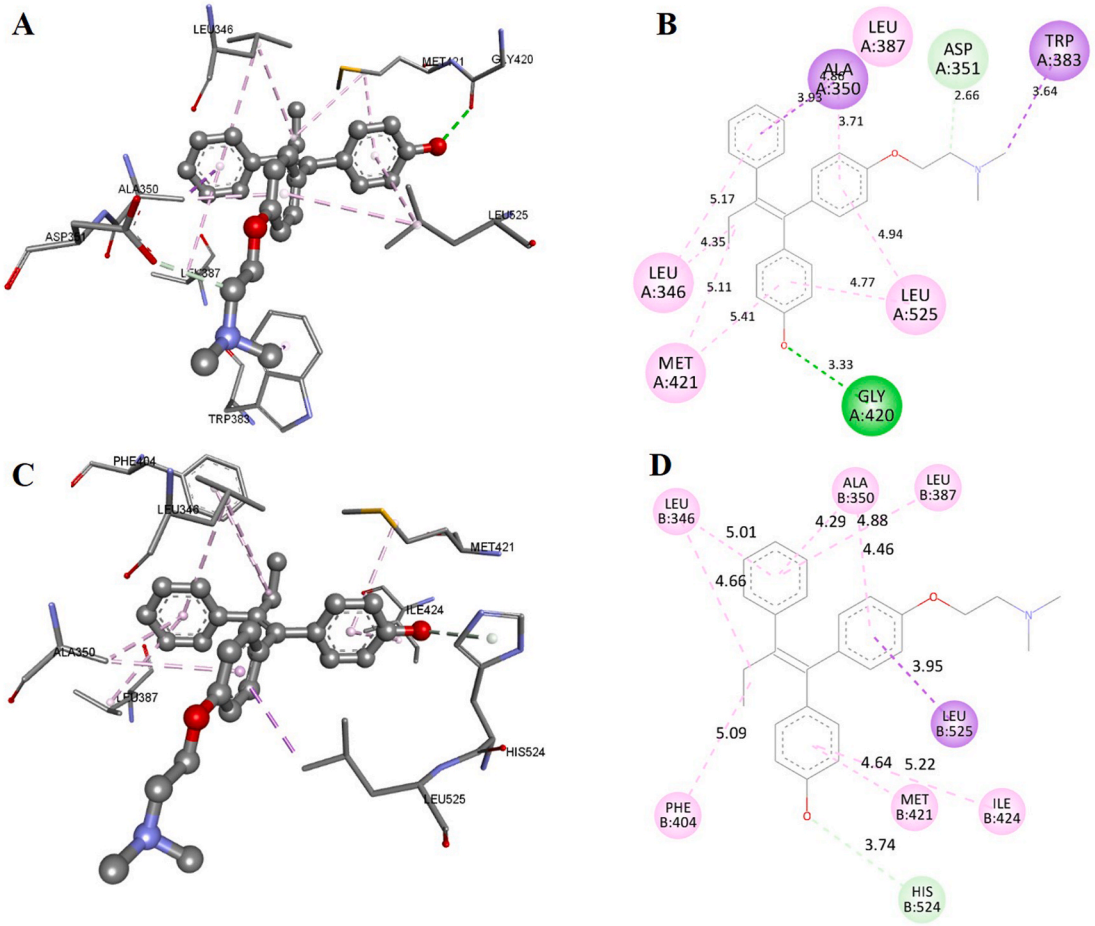

Fig. 4. Docked complexes and their interaction in the binding pocket of wild-type and mutant ESR1, the ligand-hydroxytamoxifen is shown in ball and sticks. The interactions are shown in dashed lines and labeled with the distance in Å. (A) and (B) 3D and 2D interaction of hydroxytamoxifen in the binding pocket of wild-type ESR1, respectively.(C) and (D) 3D and 2D interaction of hydroxytamoxifen in the binding pocket of mutant H516N, ESR1, respectively.

Table 3
Intramolecular Interactions formed between hydroxytamoxifen and the ESR1 wild-type and mutant protein.

Target Protein	Fitness	Intramolecular Interaction
ESR1	−100.25	Conventional Hydrogen Bond: Gly420. Carbon hydrogen bond: Asp351. Pi-sigma: Ala350, Trp383. Alkyl: Leu346, Met481. Pi-alkyl: Ala350, Leu525, Met421, Leu346, Leu387.
ESR1 mutant H516N	−100.24	Pi-Donor Hydrogen Bond: His524. Pi-Sigma: Leu525. Alkyl: Leu346. Pi-Alkyl: Ala350, Met421, Ile424, Leu346, Leu387, Phe404.

improve our knowledge and treatment approaches for breast cancer. Our future studies include, integration of molecular dynamics (MD) simulations to investigate the dynamic behavior of ESR1 and its variants that will provide a more profound understanding of their structural dynamics and intra-molecular interactions. For validating the predicted effects of SNP H516N on ESR1 function and drug binding, *in vitro* and *in vivo* approaches will be employed to improve the reliability and applicability of the computational findings. Additionally, expanding the focus to encompass thorough SNP profiling throughout ESR1 and associated genes may reveal some novel findings impacting treatment outcomes and breast cancer risk. These initiatives have the potential to direct the development of customized treatment plans, such as personalized medicines.

Funding

Supported by the Deanship of Scientific Research, King Khalid University, project number (RGP2/426/44).

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

Data availability statement

All the necessary data are included in the main manuscript, further questions may directly contact with corresponding author/s.

CRediT authorship contribution statement

Ahmad M. Alamri: Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition. **Faris A. Alkhilaiwi:** Writing – review & editing, Supervision, Formal analysis, Data curation. **Najeeb Ullah Khan:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Project administration, Investigation, Formal analysis, Data curation, Conceptualization. **Reham Mahmoud Mashat:** Resources, Formal analysis. **Munazzah Tasleem:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Ahmad M. Alamri reports financial support was provided by King Khalid University. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author appreciates the support of the Deanship of Scientific Research at King Khalid University Abha, Saudi Arabia through a project number (RGP2/426/44).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e37297.

References

[1] A.C. Wolff, et al., American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer, Arch. Pathol. Lab Med. 131 (1) (2007) 18–43.
[2] C.K. Osborne, R. Schiff, Mechanisms of endocrine resistance in breast cancer, Annu. Rev. Med. 62 (2011) 233–247.
[3] S. Mosselman, J. Polman, R. Dijkema, ERβ: identification and characterization of a novel human estrogen receptor, FEBS Lett. 392 (1) (1996) 49–53.
[4] M. Carausu, et al., ESR1 mutations: a new biomarker in breast cancer, Expert Rev. Mol. Diagn. 19 (7) (2019) 599–611.
[5] K.J. Hamilton, et al., Estrogen hormone biology, Curr. Top. Dev. Biol. 125 (2017) 109–146.
[6] L.-Y. Tang, et al., Effects of passive smoking on breast cancer risk in pre/post-menopausal women as modified by polymorphisms of PARP1 and ESR1, Gene 524 (2) (2013) 84–89.
[7] J.-F. Arnal, et al., Membrane and nuclear estrogen receptor alpha actions: from tissue specificity to medical implications, Physiol. Rev. 97 (3) (2017) 1045–1087.
[8] Y. Arao, K.S. Korach, The physiological role of estrogen receptor functional domains, Essays Biochem. 65 (6) (2021) 867–875.
[9] K.J. Hamilton, et al., Estrogen hormone biology, Curr. Top. Dev. Biol. 125 (2017) 109–146.
[10] P. D'Haeseleer, How does gene expression clustering work? Nat. Biotechnol. 23 (12) (2005) 1499–1501.
[11] F. Zhang, J.R. Lupski, Non-coding genetic variants in human disease, Hum. Mol. Genet. 24 (R1) (2015) R102–R110.
[12] M. Rebai, A. Rebaï*, In silico characterization of functional SNP within the oestrogen receptor gene, J. Genet. 95 (2016) 865–874.
[13] J. Hampe, et al., A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1, Nat. Genet. 39 (2) (2007) 207–211.
[14] R. Jeselsohn, et al., ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer, Nat. Rev. Clin. Oncol. 12 (10) (2015) 573–583.
[15] J.-C. Yu, et al., Genetic variation in the genome-wide predicted estrogen response element-related sequences is associated with breast cancer development, Breast Cancer Res. 13 (1) (2011) 1–8.
[16] A.C. Antoniou, et al., Breast-cancer risk in families with mutations in PALB2, N. Engl. J. Med. 371 (6) (2014) 497–506.

[17] D.D. Paskulin, et al., ESR1 rs9340799 is associated with endometriosis-related infertility and in vitro fertilization failure, Dis. Markers 35 (6) (2013) 907–913.
[18] D. Dustin, G. Gu, S.A. Fuqua, ESR1 mutations in breast cancer, Cancer 125 (21) (2019) 3714–3728.
[19] D. Dustin, G. Gu, S.A.W. Fuqua, ESR1 mutations in breast cancer, Cancer 125 (21) (2019) 3714–3728.
[20] A.G. Heidema, et al., The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases, BMC Genet. 7 (2006) 1–15.
[21] W. Toy, et al., ESR1 ligand-binding domain mutations in hormone-resistant breast cancer, Nat. Genet. 45 (12) (2013) 1439–1445.
[22] J.T. Mah, E.S. Low, E. Lee, In silico SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery, Drug Discov. Today 16 (17–18) (2011) 800–809.
[23] D.J. Hosfield, et al., Stereospecific lasofoxifene derivatives reveal the interplay between estrogen receptor alpha stability and antagonistic activity in ESR1 mutant breast cancer cells, Elife 11 (2022).
[24] B. Wallner, A. Elofsson, Can correct protein models be identified? Protein Sci. 12 (5) (2003) 1073–1086.
[25] G. Flouriot, et al., Identification of a new isoform of the human estrogen receptor-alpha (hER-alpha) that is encoded by distinct transcripts and that is able to repress hER-alpha activation function 1, EMBO J. 19 (17) (2000) 4688–4700.
[26] T.L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in bipolymers, 1994, pp. 28–36.
[27] A. Drozdetskiy, et al., JPred4: a protein secondary structure prediction server, Nucleic Acids Res. 43 (W1) (2015) W389–W394.
[28] T. Paysan-Lafosse, et al., InterPro in 2022, Nucleic Acids Res. 51 (D1) (2023) D418–D427.
[29] M. Kircher, et al., A general framework for estimating the relative pathogenicity of human genetic variants, Nat. Genet. 46 (3) (2014) 310–315.
[30] V. Pejaver, et al., Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, Nat. Commun. 11 (1) (2020) 5918.
[31] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (suppl_2) (2005) W306–W310.
[32] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, Proteins: Struct., Funct., Bioinf. 62 (4) (2006) 1125–1132.
[33] N.C.f.B. Information, PubChem compound summary for CID 449459, 4-hydroxytamoxifen [cited 2023 25January]; Available from: https://pubchem.ncbi.nlm.nih.gov/compound/4-Hydroxytamoxifen, 2023.
[34] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (15) (1997) 2714–2723.
[35] S.K. Burley, et al., RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences, Nucleic Acids Res. 49 (D1) (2021) D437–d451.
[36] W.L. DeLano, Pymol: An open-source molecular graphics tool, CCP4 Newsl. Protein Crystallogr. 40 (1) (2002) 82–92.
[37] R. Zrieq, et al., Tomatidine and patchouli alcohol as inhibitors of SARS-CoV-2 enzymes (3CLpro, PLpro and NSP15) by molecular docking and molecular dynamics simulations, Int. J. Mol. Sci. 22 (19) (2021).
[38] J.M. Yang, C.C. Chen, GEMDOCK: a generic evolutionary method for molecular docking, Proteins 55 (2) (2004) 288–304.
[39] K.J. Karczewski, et al., The mutational constraint spectrum quantified from variation in 141,456 humans, Nature 581 (7809) (2020) 434–443.
[40] C.O. Mackenzie, G. Grigoryan, Protein structural motifs in prediction and design, Curr. Opin. Struct. Biol. 44 (2017) 161–167.
[41] E.D. Chao, Critical analysis of secondary structure prediction algorithms. Molecular Biology Journal, Kluwer Academic Publishers, 2002.
[42] E.R. Weikum, X. Liu, E.A. Ortlund, The nuclear receptor superfamily: a structural perspective, Protein Sci. 27 (11) (2018) 1876–1892.
[43] N. Fuentes, P. Silveyra, Estrogen receptor signaling mechanisms, Adv Protein Chem Struct Biol 116 (2019) 135–170.
[44] S.W. Fanning, et al., Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation, Elife 5 (2016).
[45] A.M. Davis, S.J. Teague, Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis, Angew Chem. Int. Ed. Engl. 38 (6) (1999) 736–749.