From Genome to Personalised Medicine: Cancer Treatment and Discovery of Novel Variants in Qatar

Shunzhou Wan[1], Deepak Kumar[2], Valentin Ilyin[2], Ussama Al Homsi[3], Gulab Sher[4], Peter V. Coveney[1,5]

[1]Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom
[2]Computational Biology, Carnegie Mellon University in Qatar (CMU-Q), Doha, Qatar.
[3]Hematology and Oncology Department, National Center for Cancer Care & Research, Hamad Medical Corporation, Doha, Qatar.
[4]Interim Translational Research Institute, Hamad Medical Corporation, Doha, Qatar
[5]Faculty of Science, Informatics Institute, University of Amsterdam (UvA), 1090 GH Amsterdam, Netherlands.

1. Introduction

Breast cancer is the most commonly diagnosed cancer in females and the leading cause of death in women. Given the ever increasing cases of breast cancer, it is pertinent that we devise high-throughput experimental and computational methods that provide a comprehensive and holistic understanding of the cause of cancer. In the 'one size does not fit all' era, personalised medicine is the way forward, considering the improved ability provided by the methodology to predict treatments that would work effectively for specific patients. Advances in genomic profiling of breast cancer have led to the identification of several key mutations in the disease. An in-depth understanding of the mechanisms of the disease requires not only knowledge of the genome and variants but the correct tools to fully interpret the knowledge. The pathways for the disease are routed via proteins, and it is their interactions that are amenable to treatment. This leads in turn to clinical decision support for personalised drug treatment.

An optimal selection of sequencing techniques is crucial to generate genomic libraries for specific patients, depending on sample size and the genomic targets. When interrogating a small region of DNA on a limited number of samples or genomic targets Sanger sequencing is a good choice. The estrogen receptor (ER) protein encoded by the ESR1 gene is expressed in about 70% of breast cancers. ER also plays a vital role in classifying breast cancer subtypes and assigning therapeutic strategies; moreover, clinical research has established the central role of ER in the initiation and progression of breast cancers. For this reason, we focus here on the ESR1 gene and the ER protein.

In our study we identify genetic aberrations in 50 breast cancer patients from a population cohort in the state of Qatar using Sanger sequencing targeted on the specific gene ESR1, and perform ESMACS (enhanced sampling of molecular dynamics with approximation of continuum solvent) (1) and TIES (thermodynamic integration with enhanced sampling) (2) binding free energy studies to understand the effect of these mutations in a manner that could be used in the development of novel therapeutic strategies to inhibit these ER mutants and

substantially improve treatment outcomes.

2. Materials and Methods

Genome sequencing. Fifty breast cancer samples were collected from a population cohort of breast cancer patients in the state of Qatar and were subjected to Sanger sequencing. The Sanger sequencing method was applied to the ten coding exons of the ESR1 gene of the collected samples to detect aberrant mutants. A mutation report was generated for each patient. Chromatogram analysis was performed on the sequenced data to detect artefacts such as mis-called-nucleotides and aberrations. A list of SNPs (synonymous and nonsynonymous) was thus generated, consisting of the patient number, mutation and its novelty or known status based on variant databases such as dbSNP.

Molecular dynamics based investigation of protein-drug interactions. Mutations obtained from sequencing analysis were subjected to a modelling and simulation study in order to understand the effect of these variants on the binding affinity of drugs to ER. Binding affinities were obtained for 5 ER drugs or drug metabolites: toremifene (TOR), endoxifen (EDO), raloxifene (RAL), afimoxifene (AFI) and tamoxifen (TAM), and the natural ligand estrogen (EST) for ER. Two x-ray structures of the estrogen receptor, PDB codes 3ERT and 1QKU, were used for this study, which represent the open and closed forms with the respect to the position of the H12 helix (Figure 1).

ESMACS (1) studies employed an ensemble molecular dynamics approach which consists of 25 replicas simulations. For each replica, the same initial coordinates were used for a given ligand-receptor complex, with different initial velocities randomly assigned to the atoms according to a Maxwell-Boltzmann distribution at constant temperature. Free energy was evaluated approximately on the basis of the MMPBSA (molecular mechanics Poisson−Boltzmann surface area) method.

We have recently extended our TIES approach (2) to study the free energy changes caused by protein mutations, a TIES variant we call TIES-PM (3). We have established a standard protocol for TIES-PM, in which thirteen windows, consisting of the two endpoints representing the two physical states (WT and mutant ERs) and 11 intermediate states, are simulated for the alchemical process of protein mutation. Simulations were performed for both ligand-protein complexes and apo-proteins. Five replicas were used for each window, from which the energy deviations and the statistical errors were calculated (3). The binding free energy differences were then calculated as the difference of the alchemical free energy changes in the apo-proteins and ligand bound complexes. Four residue mutations identified in the current sequencing study were selected for the TIES-PM study: L384V, L387R, K529N and R548P.

Simulations. The binding affinity calculator (BAC) (4) software tool was used to perform ESMACS and TIES-PM studies. BAC constitutes a computational pipeline built from preparation and setup of the simulations. Standard protocols for ESMACS (1) and TIES-PM (3) have been applied, in which energy minimisation and 2 ns equilibration were conducted before 4 ns production runs were performed for each replica of the ESMACS and TIES-PM studies. All simulations were run on the BlueWaters supercomputer at the National Center for

Supercomputing Applications of the University of Illinois at Urbana−Champaign (https://bluewaters.ncsa.illinois.edu).

3. Result and Discussion

From our sequencing study 23 mutations were identified, of which seven were in the ligand binding domain and nonsynonymous, shown in Figure 1. Two of these mutations are at the binding site: L384V and L387R; while two are located at or near helix 11 or 12: K529N and R548P, which are important for the orientation of helix H12 (Figure 1).

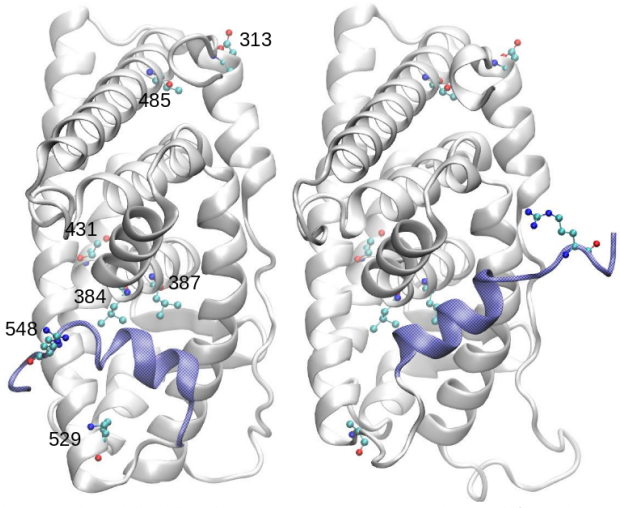

Figure 1. *Positions of the mutations identified from 50 breast cancer patients of Qatari nationals, in both the open and closed conformations. The PDB code of the open conformation crystal structure is 3ERT, and the closed conformation is 1QKU. The H12 helix is highlighted in blue.*

These four mutations, L384V, L387R, K529N and R548P, are directly involved in ligand binding or protein activation, and are further investigated by our ESMACS and TIES approaches. The other three mutations, D313N, T431A and T485I, occur away from the ligand binding site or the helices H11/H12; these are not expected to affect the ligand binding or protein activation directly, and no further investigation was performed for these mutations using molecular modelling approaches.

The predicted binding affinities of L384V and L387R from ESMACS were compared with the wild-type results. The results show that L384V and L387R induce resistance in all of the studied ligands, evidenced by the less negative binding free energies for the mutated ERs compared with the wild-type ER with the corresponding ligands. L384V and L387R occur in the binding pocket and directly interact with the ligands. The L387R mutation, especially, introduces not only steric bulk but a net charge change. It induces significantly larger free energy changes for the ligands Tam and Tor than the mutation L384V does. Large changes in the size of the residues and the charge distributions can confer resistance to and even completely block access to the ligands.

The TIES-PM calculations also show that the two mutations induce resistance, which is in line with the ESMACS predictions. The L387R mutation introduces a large change in the binding affinity, making it likely to block the binding of all these ligands, including the native estradiol.

The estrogen receptor exists in at least two conformational states: active and inactive

(Figures 1). The TIES-PM approach can deliver accurate and precise predictions, and is used here to investigate the relative binding free energy changes in the two states caused by protein mutations. The results show that the L384V and K529N mutations make a moderate change with regard to the stability of these two protein states, rendering the active state slightly more favourable for the mutant protein than the wild-type. By contrast, the L387R and R548P mutations have a large impact on the preference of the two states, making the active state significantly more favourable for the mutant proteins than for the wild-type.

4. Conclusions

From our study, performed on 50 breast cancer patients in the population cohort of Qatar, we are able to bring a holistic understanding of the effect of deleterious mutations on the binding affinity of prevalent breast cancer drugs available in the market. In addition, we have performed a rigorous and in depth molecular modelling study of estrogen receptor with sequential variations obtained from the gene sequencing study in this project. The sequencing study identified a few mutations from breast cancer patients in Qatar. Some of these mutations are of considerable interest, and have not been previously reported.

The molecular modelling approaches were used for the newly identified mutations in the ligand binding domain of the receptor. The predicted binding free energies provide a clear explanation for the effects of these mutations. The mutations at the binding site, L384V and L387R, induce resistance for the drugs studied in the project; the mutations L387R and R548P play an important role in the activation of the estrogen receptor. Our findings provide the basis for a clinical decision support tool for patient specific drug treatment, in which the combination of genomics and rapid BAC based calculations provides a powerful and reliable way to identify a subgroup of patients for which an effective treatment regimen based on their genetic profile may be selected.

Acknowledgements

The authors would like to acknowledge the support of the Qatar National Research Fund (7-1083-1-191) and the EU H2020 project CompBioMed (Grant No. 675451).

Reference

1. Wan S, Knapp B, Wright DW, Deane CM, Coveney PV. *J Chem Theory Comput.* **2015**;11(7):3346-56.
2. Bhati AP, Wan S, Wright DW, Coveney PV. *J Chem Theory Comput.* **2017**;13(1):210-22.
3. Bhati AP, Wan S, Coveney PV. *J Chem Theory Comput.* **2019**;15(2):1265-77.
4. Sadiq SK, Wright D, Watson SJ, Zasada SJ, Stoica I, Coveney PV. *J Chem Inf Model.* **2008**;48(9):1909-19.