INTRODUCTION
Pharmacogenomics is the study of how inherited and acquired variations in the human genome affect the response to medication. The presence of genetic polymorphisms and/or mutations have been documented to alter drug absorption, metabolism and excretion. This is particularly relevant to the administration of selective estrogen receptor modulators (SERM) and their use in the treatment in patients with estrogen receptor alpha (also named ESR1) positive breast cancer. The most widely used estrogen antagonist, tamoxifen, requires activation by the cytochrome P450 2D6 (CYP2D6) enzyme to form its functional metabolites 4-hydroxy tamoxifen (OH-Tam) and endoxifen. The latter two are thought to have a much greater affinity for the ESR1 protein-product and the ability to inhibit cell proliferation than can tamoxifen alone. Polymorphisms in the CYP2D6 gene result in alterations in the metabolism of tamoxifen, leading to the identification of specific phenotypes with poor, intermediate, extensive and ultra-fast capabilities in the metabolism of this SERM. Patients with polymorphism(s) resulting in poor metabolism of this SERM may represent only a subset of those breast cancer patients that are eventually found to be estrogen receptor positive but tamoxifen-unresponsive. Since approximately 30-40% of estrogen receptor(+) (ER+) breast cancer patients do not derive any benefit from being treated with this SERM, other mechanisms need to be identified in order to provide more cost-effective, precision-based, medical care.1 Loss of ER expression, endocrine adaption, pharmacologic tolerance, alterations in co-regulator activity or changes in cellular signal transduction pathways are just a few of the possible venues that have been investigated. In contrast, an alternative plausible mechanism in ER(+) tumor indifference to the administration of a SERM is genetic variation. Genetic polymorphisms and/or mutations in critical regions of specific genes have been demonstrated to result in the alteration of drug efficacy used to treat chronic myelogenous leukemia and adenocarcinoma of the lung.2,4 Recent studies have documented that single nucleotide variations in the ESR1 gene can be found in the malignant cells of breast cancer patients who have demonstrated tumor progression despite being previously treated with a SERM.5-8 In the ESR1 gene, the binding sites of OH-Tam to that gene’s protein product have previously been identified through
crystallography and correspond to a highly conserved region. This highly conserved region is also near the site where polymorphisms/mutations have been recently reported to be present in samples taken from patients with tumor progression after treatment with a SERM. It is postulated that the presence of these nucleotide variants may confer an advantage and hence clonal selection to those tumor cells harboring them. Although their presence in previously treated patients has now been established, the frequency of these single nucleotide alterations in treatment naïve breast cancer patients has not received as much attention. Also, these reports of mutations, although in the ligand binding domain, did not query the sequences responsible for the amino acids enumerated by Shiau to be responsible for actual ligand binding. Clinically, if polymorphisms do exist in these sequences of the ESR1 gene, identifying them could theoretically stratify breast cancer patients between responders and non-responders, the latter being those with de novo resistance. This would increase the level of precision in the oncologic care of breast cancer patients by decreasing medical costs and reducing the risk of side effects in potentially unresponsive patients. To this end, a small cohort of ER(+), treatment naïve breast cancer patients were identified and the DNA from their tumors recovered for sequencing of the corresponding regions, previously identified by crystallography studies, to be responsible for ligand binding.

Figure 1. Static diagram of the functional domains of the ESR1 gene with a ribbon diagram of the ligand binding domain demonstrating the hydrophobic cavity formed by this domain’s 12 helices. The critical amino acids that play a role in ligand binding belong to parts of these 12 helices.
METHODS
Identification of Critical Ligand Binding Sites in the ESR1 Gene
The critical regions of the 595 amino acid residue long ESR1 protein responsible for the binding of OH-Tam has been previously reported by Shiau.\(^9\) The critical region is located in the ligand binding domain, one of five functionally demarcated regions of this protein.\(^{10}\) The ligand binding domain is created by the 3-dimensional positioning of 12 helices that form a central hydrophobic pocket (Figure 1).\(^{11}\) The intermolecular attractions between the ESR1 gene product and OH-Tam that promote ligand binding include both van der Waals forces and Hydrogen bonding. Van der Waal forces are the weak, temporary, fluctuating dipole interactions between different molecules that cause attractions between different regions of different molecules. According to Shiau’s previous crystallography study, there are seventeen amino acids involved with van der Waals binding between the ER gene product and OH-Tam. The ESR1 gene product consists of 595 amino acids. The critical amino acids (and their amino acid positions) are Alanine (350); Aspartic acid (351); Glutamic acid (419); Glycine (420 and 521); Histidine (524); Isoleucine (424); Leucine (346, 349, 391, 428 and 525); Methionine (343 and 421); Phenylalanine (401) and Tryphtophan (383) (Figure 2). The stronger hydrogen bonding between the protein product and ligand involves the amino acids Glutamic acid (353) and Arginine (394). The corresponding nucleotide sequences for these amino acids were identified by manual curation of the estrogen receptor alpha reference gene (GeneBank GI:201023303).

Figure 2. Partial listing of the amino acid sequence for the ESR1 gene, with the seventeen residues (and their positions) implicated to interact with OH-Tam via van der Waals forces highlighted in yellow and the two residues involved with hydrogen bonding highlighted in orange. Numbers in blue rectangles indicate amino acid residue position.

Sequence Specific Primer Design
The amino acids involved in both van der Waal’s and hydrogen binding were culled into four separate groups that encompassed amino acids 343 (Methionine) to 353 (Arginine), 383 (Tryphtophane) to 401 (Phenylalanine), 419 (Glutamic acid) to 428 (Leucine) and 521 (Glycine) to 538 (Aspartic acid). The last amino acid implicated in Shiau’s crystallography study with attractive forces on OH-Tam was...
Histidine at position 524. Aspartic acid, at position 538, was not mentioned as participating with either van der Waals forces or hydrogen binding, but was included because of reported variations in DNA sequences in previous studies. Primer pairs were designed to be large enough to include each region of interest using Invitrogen’s OligoPerfectTM Designer software and yet small enough considering that the DNA would be coming from formalin fixed, paraffin embedded material, which is typically degraded and limited to intact sequences of less than 500 base pairs in length. For the first sequence, the primer pair was (forward) 5'-ACATGAGAGCTGCCAACCTT-3'; (reverse) 5'-CTCTAAAAAGCTGCGCTTCG-3'; the second sequence primer pair was (forward) 5'-TGGATTTGAGCCTCCATGAT-3'; (reverse) 5'-CCCCAATGCACTCTTTTGTT-3'; the third sequence primer pair was (forward) 5'-TCCATGAAGACAATGGCTGA-3'; and the fourth sequence primer pair was (forward) 5'-CCCCTTCTAGGGATTTCAGC-3'; (reverse) 5'-ATGAAGTAGAGCCCGCAGTG-3'. Although not considered to be part of the critical amino acids responsible for ligand binding, primer pair four was designed to be inclusive of additional sequences corresponding to amino acids previously reported to participate in the binding of OH-Tam. The four separate primer pairs created amplified sequences of 271, 271, 240 and 249 base pairs in length, respectively.

Patient and Archived Tissue Selection
Fifteen cases of invasive ductal carcinoma of the breast, in the embodiment of formalin fixed, paraffin-embedded tissue, were identified from the departmental database and retrieved from the archival storage files. This project received approval by the Institutional Review Board of the University at Buffalo. Selected cases were from patients with primary tumors that had not been subjected to any prior therapeutic intervention prior to surgery. The fifteen cases came from 14 women and 1 man, all with a diagnosis of infiltrating ductal carcinoma, not otherwise specified. The breakdown of the Nottingham scores for the fifteen cases were as follows: grade I = 5 cases; grade II = 7 cases; grade III = 3 cases. The immunoprofile of the fifteen cases were as follows: estrogen and progesterone receptors both positive with negative Her-2 staining = 13 cases; estrogen and progesterone receptors (+) with positive Her-2 staining = 1 case; estrogen receptor (+), progesterone receptor (-) and negative Her-2 staining = 1 case. Corresponding archived immunohistochemistry slides for the estrogen receptor were reviewed in each case to confirm the expression of this protein in the tumor cell nuclei (Figure 3). Each case was microscopically examined to select for a representative block that possessed greater than 80% of the tissue specimen as tumor. Twenty 5 micron thick sections were then cut from the selected tissue blocks from each case and collected in separate 1.5 ml microcentrifuge tubes.

Figure 3. Representative photomicrographs of the breast cancer tumors (3A) and the expression of the estrogen receptor by immunohistochemistry (3B). Original magnification 10X.

DNA Isolation and Recovery
DNA was isolated using the QIAamp DNA FFPE Tissue kit (catalog #56404; Qiagen, Valencia, CA) and by following the stated protocol. Briefly, twenty 5 micron thick sections of FFPE material from blocks of tissue with > 80% tumor were sectioned and placed into 2 ml microcentrifuge tubes. One milliliter of xylene was then added to each tube and the contents vortexed. The tube was then centrifuged at 13,300 rpm using a Spectrafuge 24D benchtop microcentrifuge (United Lab Plastics, St, Louis, MO) for 2 minutes at room temperature. The supernatant was then pipetted off and the remaining pellet resuspended with 1 ml of 100% ethanol with mixing. The tubes were then centrifuged at full speed again for 2 minutes at room temperature and the supernatant removed. The pellet was then allowed to air dry at room
temperature for 10 minutes, followed by resuspension with 180 ul of commercial kit buffer “ATL”. Twenty microliters of proteinase K was then added, the mixture vortexed and incubated at 56°C for one hour, then 90°C for another hour. The mixture was then briefly spun down and 200 ul of commercial kit buffer “AL” was added and the resulting solution mixed by vortexing. A final volume of 200 ul of 100% ethanol was added and the solution again mixed by vortexing. The resulting lysate that was recovered and then transferred to the kit’s QIAamp MinElute column. The column was then centrifuged at 6000 x g for 1 minute. The MinElute column was then fitted with another collection tube and 500 ul of commercial kit wash buffer added, followed by centrifugation at 6000 x g for 1 minute. After an additional wash, the column was spun down at full speed for 3 minutes. The column was then transferred to another receiver microcentrifuge tube, loaded with commercial kit buffer “ATE”, allowed to sit at room temperature for 1 minute and then spun down at maximum speed for 1 minute. Recovered DNA was quantitated using the NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific, Wilmington, DE).

Figure 4. Representative chromatograms of the four separate primer pairs demonstrating the absence of variants in the sequence corresponding to the critical amino acids involved in ligand binding. Amino acid residues and their corresponding codon are indicated in the blue boxes and yellow triangles. In primer pair four, the primer pair was designed to include the sequences for amino acids in position 536-8 (gray boxes).
Polymerase Chain Reaction (Pcr)
PCR conditions were optimized on a Stratagene MX3006P thermocycler using the master mix from the commercially available FastStart Universal SYBR Green Master (Roxy) mix (Roche, Indianapolis, IN) kit. Forty cycles of PCR were performed using the following conditions for each primer pair: 95°C for denaturation; 60°C for annealing; and 72°C for extension. The amplified products were purified for sequencing using the QIAquick Gel Extraction kit (Qiagen, Valencia, CA). Briefly, 10 ul of PCR sample from each case was added to commercial kit buffer “QG”, vortexed and isopropanol added. The mixture was then placed into a QIAquick spin column with a collection tube and spun at maximum speed for 1 minute. The flow-through was discarded and several washes performed using the provided kit’s buffers and centrifugation. A final volume of 20 ul of nuclease free water was added to elute of the bound DNA. The recovered DNA was quantitated using spectrophotometry, and, along with their accompanying primer pairs, brought to the Roswell Park Cancer Institute’s Biopolymer facility for sequencing using the Big Dye Terminator approach.

Single Nucleotide Polymorphism/ Mutation Detection
Prior to tissue collection, the National Center for Biotechnology Information’s (NCBI) Single Nucleotide Polymorphism Database (dbSNP) was queried to determine if polymorphisms in the critical ligand binding regions of the ER gene have been previously documented. The dbSNP is a public online resource built to archive all genetic variation that has been found and reported in homo sapiens. The resulting sequences were then queried against the ESR1 reference gene using NCBI’s Basic Local Alignment Search Tool (BLAST).

RESULTS
The ESR1 gene in the NCBI database exists as a 419779 base pair sequence and consists of 10 exons. The sequences of interest corresponded to nucleotide position 258941 to 25873; 326211 to 326270, 375515 to 375544 and 413244 to 413297 in the ESR1 gene. These sequences reside in exons 6, 7, 8 and 10, respectively. The query of the dbSNP database found no reported polymorphisms in the critical amino acids involved in the binding of OH-Tam to the E1R gene.

Total DNA recovered from the FFPE tissue ranged from 560 to 39,650 ng. The A260/280 ratios ranged from 1.05 to 2.12. Products were detectable by real-time PCR in all fifteen cases as demonstrated by the production of good amplification curves. Sequences that initially did not produce a product were repeated and resulted in an amplicon. In sequences for primer pairs 1, 2 and 3, no variations were identified (Figure 4). In primer pair 4, three separate cases possessed a single base pair variation (g.413259_413530insA), resulting in the possible insertion of an Adenine in front of a Thymidine after amino acid 525 (Leucine). This would have changed the reading frame from the “Tyrosine, Serine, Methionine, Lysine” sequence to an “Isoleucine, Glutamine, Histidine, Glutamic acid” sequence. However, this variant was not substantiated when the samples were re-sequenced in the reverse direction. Another possible insertion altered the codon sequence from amino acids AGST to AGIY. As this alteration was towards the end of the amplified sequence, it was not included in the reverse sequence analysis.

DISCUSSION
Targeted therapies to certain malignancies have improved survival times for cancer patients. For breast cancer, the most widely used targeted therapy is tamoxifen, a SERM that acts in an antagonistic manner when bound to expressed estrogen receptors in breast cancer cells. Tamoxifen has been in clinical use for the past few decades and has resulted in reducing the risk of death and tumor recurrence due to breast cancer. However, the recent increasingly wide use of additional targeted agents to other types of cancers has highlighted a previously well-known axiom, that alterations in the genetic sequence can result in changes in the translated protein. These alterations in the genetic sequence may result in changes of the molecular dynamics of a protein, causing changes to the philicity/phobicity of certain regions of a protein, or modifications in protein structure or function. These changes may unduly influence the binding kinetics of ligands. An example of how genetic sequence alterations can negatively impact ligand binding and hence drug efficacy can be seen with the drug imatinib and chronic myelogenous leukemia. Imatinib is a tyrosine kinase inhibitor that is thought to bind to the ATP binding site of the target protein resulting in the inactivation of that protein’s enzymatic activity. However, point mutations in the bcr-abl fusion protein product have been shown to interfere with the binding of imatinib and lead to tumor relapse.

In breast cancer, a number of different mechanisms have been attributed to the ineffectiveness of tamoxifen in some cases of ESR1(+) breast cancer. However, sequence polymorphism had previously not been one of them. This, despite a number of separate functional studies that have shown the negative implications of ESR1 molecular alterations with ligand binding using cell line. To the best of our knowledge, no directed studies regarding the frequency of nucleotide variations in the critical ligand binding domain region in actual patient material has been reported.

However, recent reports have been published documenting the presence of mutations in helix 12 and involving amino acids at positions 536, 537 or 538. These amino acids were not predicted in Shiau’s original crystallography study to be involved with ligand binding. Instead, it has been hypothesized that alterations in this region of the ESR1 gene may alter the coupling between ligand binding, changes in receptor conformation and receptor activation. A study by Wang et al using droplet digital PCR directed for the region encompassing amino acids 537 and 538 recently identified mutations in primary (un-treated) breast cancers at a
frequency of 7%. With the addition of our current study, alterations involving amino acids 536-8 in primary tumor material reduced this number to 5% (n = 61) of the cases of treatment naïve, ER(+) breast cancer. In contrast, the vast majority of mutations involving these particular amino acids occurred in patients experiencing tumor progression (recurrence or metastasis) and had prior treatment with estrogen. In recently published studies, the percentage of cases found, collectively, to harbor alterations in the ESR1 gene sequence corresponding to amino acids 536-8 in patients with progressive disease, amounted to 54%.5,6,8 These findings have led to a growing sense that patients who harbor these changes are in fact mutations that have transformed these individuals into non-responders to any further treatment benefit with estrogen.

In the current study, no polymorphisms were identified in the regions purported by crystallography to be critical for ligand binding in ER(+) breast cancer patients who not yet been treated. There are no other studies in the scientific literature that specifically examined the ESR1 gene sequences corresponding to these critical ligand binding amino acids. There are, however, a growing number of studies that have examined other, non-critical, sequences in the ESR1 ligand binding domain. It should be noted that in Shiau’s crystallography study, the region in the ESR1 gene corresponding to amino acids 536-8 were not implicated as being critical to ligand binding. One weakness of this study is the small cohort. Although the cohort in the current study was small, it is in keeping with the sequencing data available from The Cancer Genome Atlas whole exome study on ER(+) breast cancers. In that study, no polymorphisms were identified in the ESR1 gene in 390 cases of treatment naïve breast cancer patients.6 The latter study provided more data in terms of cohort number and genetic sequence coverage because it was whole exome in nature, unlike the current study which was more focused on specific regions of the ESR1 gene.

CONCLUSION

Alterations in the DNA sequence of the ESR1 gene for the nineteen amino acids considered to be critical for the binding of OH-Tam were not identified in this study. These findings are in keeping with a larger, sequencing based study that was not specifically directed to these binding sites, but included them.6 Instead, mutations involving amino acids not considered to be critical by Shiau’s crystallography study, specifically amino acids 536-38 but still part of the ligand binding domain, have been associated with tumor progression in recent reports and a growing opinion that these mutations render the tumor refractory to efficacious treatment with estrogen. It does not appear at this point in time that sequencing for critical ligand binding site polymorphisms is warranted in patients with ER(+) breast cancer, whether by PCR or next generation sequencing methods. It does appear however, based on other studies and ours, that assaying for the presence of polymorphisms/mutations in the ESR1 gene corresponding to amino acids 536-38 are variable and clinically important. These amino acids, which were not part of Shiau’s critical ligand binding site amino acids, may affect therapeutic efficacy. Sequencing polymorphisms at these amino acids therefore, may be warranted for guiding appropriate medical treatment decisions and care.

CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

REFERENCES

