Chapter 5

VIRTUAL SCREENING, DOCKING, CRYSTALLIZATION AND STRUCTURE DETERMINATION OF TETRAZOLYL INDOLE DERIVATIVES
Chapter 5

5.1 Introduction

There has been considerable progress in the synthesis of structurally modified selective estrogen mimetics that can act as antagonists for breast/uterine cancer or which can interfere with the endocrine mediated events leading to the implantation of the fertilized ovum to the endometrium (anti-implantation agents). These mimetics can also act as progesterone antagonists that could affect early post-implantation events. As the molecular structure of the hormones and their receptors associated with reproductive processes are being unraveled, the desire for better and safer contraceptives has driven efforts to increase the chemical diversity in non-steroidal molecules to act as estrogen antagonists for fertility control and in diseases like uterine or breast cancer. Despite the work already carried out, researchers in this field face considerable challenges in the search for an ‘ideal’ molecule with superior pharmacology and reduced side effects.

Specific molecular recognition of heterocyclic compounds, in particular of nitrogen heterocycles, has been of considerable importance in biological chemistry (Hajela et al, 2001). Among the nitrogen heterocycles, the indole nucleus is a fundamental constituent of a large number of natural and synthetic products with biological activities (Kikuchi et al, 1999). The 2-phenyl indole nucleus caught the attention of chemists and biologists as a possible estrogen mimetic due to its structural similarity to estradiol. Several estrogen antagonists have been developed incorporating the 2-phenyl indole moiety for example ZK119010 (Angerer et al, 1990) (1), pipendoxifene (ERA-923, 2), bazedoxifene (TSE-424, 3) (Miller et al, 2001) and have suggested to be promising antiestrogens or selective estrogen receptor modulators (SERMs). Recently, some tetracyclic indoles like indenoindoles (Angerer et al, 1990) and benzocarbazoles (Miller et al, 2001) have also been shown to possess mixed estrogen agonist/antagonist activity (Pickar et al, 2001 & Miller et al, 2000). These tetracyclic compounds are more rigid than the 2-phenyl indoles and resemble the natural estrogens.

Structural studies on the estrogen receptors (ER) have suggested that there is an ample unoccupied space within the ligand-binding domain (LBD) of ER and it can accommodate a wide variety of structurally distinct ligands in its binding pocket (Anstead et al, 1997). The antagonistic activity of non-steroidal antiestrogens is due to the presence of a
long carbon chain with a terminal amino functionality at an appropriate position on the ligand is now well established (Lednicer et al, 1966 & Grese et al, 1997). This side chain, an integral part of almost all non-steroidal antiestrogens establishes hydrophobic interactions with amino acids of the estrogen receptor in the ligand binding domain (Watanabe et al, 2003). Nature of side chain is also an important determinant of selective action and it has been demonstrated that even small changes in the nature or structure of side chain result in large changes in the tissue selectivity profile of the molecule (Robertson et al, 1982). This realization of the important role of side chain in modulating activity of ER modulators in various physiological processes has resulted in extensive research in this area. In medicinal chemistry, tetrazole group is often used as metabolism resistant isosteric replacement of acidic groups and also the delocalization of the negative charge around the tetrazole ring is considered favorable for better receptor substrate interaction (Her et al, 2002).

To extend the scope of ER ligands through the modification of basic side chain, it was thought (by virtual screening) to synthesize new molecules with altered carbon chain through insertion of 5-substituted N-ethyl amino tetrazolyl moiety in place of conventional phenoxy alkyl amino group at the nitrogen of tetracyclic indole scaffold used as core structure for mimicking 17β-estradiol. The purpose is to facilitate favorable receptor ligand recognition due to greater distribution of charge around the tetrazole group for enhanced binding affinity/activity and to find new estrogen antagonists with promising anti-implantation (post-coital contraceptive) activity (Pandey et al, 1999).

The present chapter initially describes the virtual screening and docking studies involving novel indole derivatives to explore the possibility of new types of antagonists for the ERα Ligand Binding Domain (ERα LBD). The chapter also reports the crystallization, data collection, structure solution of two of the best scoring compounds in the study.
Chapter 5

5.2 Experimental section

The flowchart showing the systematic methodology used in the virtual screening studies is shown below. The details of the compound synthesis and anti-implantation assays are not reported here, as these are not part of the thesis. These have been reported by the collaborating group elsewhere (Singh et al, 2008).

5.2.1 Molecular docking studies protocol

Docking, molecular dynamics, energy minimization and molecular visualisation studies were performed using a Silicon graphics OCTANE workstation (M/s Silicon Graphics Inc., USA). The genetic algorithm implemented in the AutoDock 3.0 program has been employed for the docking studies involving the tetrazoyl indole derivatives into the active sites of the estrogen receptor. The crystal structure of the estrogen receptor ligand binding domain (LBD) in complex with the endogenous estrogen, 17β-estradiol (II) and Raloxifene (Brzozowski et al, 1997) and reference protein coordinates used for docking were recovered from the Brookhaven Protein Data Bank (http://www.rcsb.org/pdb/home) (entry codes 1ERE
and 1ERR, respectively). Ligand structures of 4a and 5a were obtained in pdb-format from the respective crystal structures. The whole docking operation can be summarized as follows. Firstly, the ligand molecules were checked for polar hydrogens and the Gasteiger-Hückel (Purcell et al, 1967) partial atomic charges were assigned. Flexible torsions were defined with the help of AutoTors. This allowed for the conformational search of ligands during the process of docking. The PDBQ file was created for ligands. For the ER structure, polar hydrogens were added and Kollman-all atom (Weiner et al, 1984) atomic charges were taken and the atomic salvation parameters were also assigned using the ADDSOL utility of AutoDock 3.0.5 program. Secondly, the 3-D grid of 0.375 Å resolutions was centered on the active site using the AutoGrid algorithm (Goodsell et al, 1990) to evaluate the interaction energies between the ligands and the ER LBD. At this stage, the ER was embedded in the 3-D grid and a probe atom was placed at each grid point. The affinity and electrostatic potential grid were calculated for each type of atom in the ligands. Thirdly, a series of the docking parameters were set on. Not only the atom types but also the generations and the number of runs for GA algorithm were edited and properly assigned. The number of generations, energy evaluations, and GA runs were sets of 27,000, 200,000 and 15, respectively. Finally, the docked complexes of ligand-receptor are selected according to the criteria of interacting energy combined with geometrical matching quality. These complexes were used for comparative study and correlation between the activity and its structural conformations. The total binding free energy was empirically calibrated based on the above-stated terms and set of coefficient factors. The same rational was applied to the system of 4a and 5a compounds and estrogen receptor. The total binding free energy and corresponding inhibitory constant between compound and ER was calculated according to the algorithm in the AutoDock 3.0 program.

5.2.2 Affinity Docking

The Affinity module of InsightII (M/s. Accelrys inc.) can carry out docking of a single ligand with the protein at a time. The docking method employs a combination of Monte Carlo (MC) and Simulated Annealing (SA) approaches. In the initial phase, an MC search generated 20
unique substrate-binding orientations, the 10 most favorable of which were minimized and subjected to the SA docking phase. The SA phase consisted of 50 cycles of 100 fs molecular dynamics (MD) simulations. The temperature is re-scaled at each cycle, with the first cycle running at 500 K and the final cycle at 300 K. Following the SA phase, the structure is minimized by 1000 steps of conjugate gradients. With Affinity, a distance dependent dielectric function was used to simulate charge screening by water molecules. This docking method models flexibility of both the ligand and receptor molecule but is highly computational intensive and therefore not suitable for a general virtual screening purpose, hence the procedure was used only for two structures.

5.2.3 Crystallization, data collection and structure solution of 4a and 5a
The compounds that were suggested by the virtual screening and docking studies were synthesized in Dr. Kanchan Hajela’s lab, Medicinal & Process Chemistry division, Central Drug Research Institute, Lucknow, India. The synthesized compounds were crystallized by the slow evaporation method after being dissolved in methanol. The resultant crystals were mounted on a glass capillary. The data collection was done on Bruker AXS using P4 diffractometer. ShelXTL was used for structure solution.

5.3 Results and Discussion
5.3.1 Molecular docking studies
By looking at the spatial conformations of the two regioisomers 4a and 5a shown in the crystal structure (Fig 5.4 & 5.5), the display of low order of antagonism shown by N-2 regioisomers was surprising. It can be clearly seen that the carbon chain in the N-2 regioisomer 5a orients itself perpendicular to the indole core structure giving the favorable conformation shown by most of the estrogen antagonists. This molecular similarity was further substantiated by superimposition of N-2 regioisomer 5a over 17b-estradiol and raloxifene as shown in Fig. 5.1. The relative disposition of the tetracyclic rings of 5a superimposed closely over both the endogenous hormone and raloxifene and also the side chain is well overlayed on the raloxifene side chain. In contrast, the carbon chain in N-1
regioisomer 4a bends inwards over the tetracyclic indole scaffold giving a folded molecule that may give a better fit into the receptor cavity as shown by pure agonists.

To gain insights to the binding mode of each regioisomer through the interaction of salient amino acids known to anchor both the estrogenic and antiestrogenic ligands in the binding pocket of LBD of ER and to explain the low potency of *in vivo* biological effects, molecular docking studies were performed with both regioisomers 4a and 5a. The interaction mechanism currently suggested is that both estrogenic (estradiol) and antiestrogenic (raloxifene) ligands anchor to the binding pocket (aminoacids Glu353 and Arg394) by means of the OH group located at their phenolic site. The second OH group interacts with His524 in both the ligands, whereas the side chain in raloxifene is projected towards Asp351 (Fig. 5.2 and 5.3b).

Though it is suggested that a molecule may not adopt similar conformation as shown by the crystal structure in the LBD domain, it was observed that on docking of crystal structure of 4a, the molecule adopts a very unfavorable orientation as shown in Fig. 4.3a. The tetracyclic core fails to penetrate the ER binding pocket and occupies a space in the cavity by displacing all the salient amino acids responsible for binding of the ligand from the binding pockets. The side chain is placed far away from Asp351 and since the molecule shows no interaction with any of the amino acids, it justifies the weak binding affinity and agonist activity shown by N-1 isomers. However, on docking of N-2 regioisomer 5a (Fig. 5.3b), the molecule assumes an orientation somewhat comparable to raloxifene. The tetracyclic core is analogously aligned like estradiol or raloxifene. The phenyl ring of tetracyclic core shares binding pocket with His524 lying in close proximity, but the binding pocket of Glu353 and Arg394 is shifted a little away because of the puckering of seven membered ring giving an upward twist to the benzothiepine core. The side chain projects towards Asp351 making contact with the nitrogen of the piperidine ring, also the tetrazole ring shows interaction with another amino acid Thr347. The binding pockets comprising the salient amino acids are conserved and this could explain the marginally better receptor binding and moderate antagonism of N-2 regioisomers than N-1 regioisomers.
Fig 5.1 Structural superposition of the test compound 5a (magenta) with Raloxifene (pink) and Estradiol (cyan).
Fig 5.2. Control docking

(A) Structural superposition of docked Estradiol (magenta) with Estradiol (green) of Estrogen Receptor Ligand Binding Domain (PDB: 1ere). (B) Structural superposition of docked Raloxifene (magenta) with Raloxifene (green) of Estrogen Receptor Ligand Binding Domain (PDB: 1ERR).
Table I Predicted docking energies of selected compounds. Raloxifene and Estradiol refers to the control docking.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Compounds</th>
<th>Structure</th>
<th>Binding energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>4a</td>
<td><img src="image" alt="Structure of 4a" /></td>
<td>-7.23</td>
</tr>
<tr>
<td>2.</td>
<td>5a</td>
<td><img src="image" alt="Structure of 5a" /></td>
<td>-14.68</td>
</tr>
<tr>
<td>3.</td>
<td>Raloxifene</td>
<td><img src="image" alt="Structure of Raloxifene" /></td>
<td>-6.48</td>
</tr>
<tr>
<td>4.</td>
<td>Estradiol</td>
<td><img src="image" alt="Structure of Estradiol" /></td>
<td>-7.53</td>
</tr>
</tbody>
</table>
Fig 5.3 Amino acid residues involved in the interactions with compounds 4a (a) and 5a (b) in the binding site of ER resulting from docking experiments. The binding cavity of ER shows interaction with neighboring amino acids through hydrogen bonding (dotted lines). The active site residues and those forming H-bonds with the test molecule are colored blue.

5.3.2 Crystallographic details

Table II Data collection and refinement statistics for the 4a and 5a compounds.

Test Compound 4a

Crystal data
C26H30N6S
Mr = 458.68
Monoclinic, P21/c
a = 11.462 (2) Å
b = 23.115 (1) Å
c = 21.518 (2) Å
β = 94.57 (1)°

Data collection
Bruker AXS P4 Diffractometer
Absorption correction: multi-scan
(SADABS; Bruker, 1997; Blessing, 1995)

Refinement
R [F² > 2σ(F²)] = 0.0428
wR(F²) = 0.0948
S = 1.016
2853 reflections
299 parameters

V = 2352.2 (5) Å³
Z = 4
Mo- Kα radiation
μ = 0.16 mm⁻¹
T = 294 (2) K
0.25 x0.20 x 0.20 mm
F (000) = 976
5441 measured reflections
4741 independent reflections
2853 reflections with I > 2σ (I)
Rint = 0.020
2 restraints
H-atom parameters constrained
Δρ max = 0.36 e Å⁻³
Δρ min = 0.25 e Å⁻³
Test Compound 5a

Crystal data
C26H30N6S
Mr = 458.68
Triclinic, P-1
a = 12.09 (2) Å
b = 12.239 (2) Å
c = 16.180 (2) Å
α = 90.49(1)
γ = 98.18(1)
V = 2366.3 (6) Å³
Z = 2
Mo- Kα radiation
μ = 0.16 mm⁻¹
T = 294 (2) K
0.25 x0.20 x 0.15 mm
β = 93.14 (1)°

Data collection
Bruker AXS P4
Diffractometer
Absorption correction: multi-scan
(SADABS; Bruker, 1997; Blessing, 1995)
9647 measured reflections
8345 independent reflections
2589 reflections with I > 2σ (I)
Rint = 0.0202

Refinement
R [F² > 2σ (F²)] = 0.0871
wR(F²) = 0.1855
S = 0.934
2589 reflections
596 parameters
2 restraints
H-atom parameters constrained
Δρmax = 0.36 e Å⁻³
Δρmin = 0.25 e Å⁻³
(Δ/σ)max = 000

Refinement
The weighted R-factor wR and goodness of fit S are based on F², conventional R-factors R are based on F, with F set to zero for negative F². The threshold expression of F² > 2σ (F²) is used only for calculating R-factors etc. and is not relevant to the choice of reflections for
refinement. $R$-factors based on $F^2$ are statistically about twice as large as those based on $F$, and $R$-factors based on ALL data will be even larger.

Table 1. Hydrogen-bond geometry (Å, °) and short contacts participating in packing of the molecule in crystal structure

<table>
<thead>
<tr>
<th>D-H...A</th>
<th>D-H</th>
<th>H...A</th>
<th>D...A</th>
<th>D-H...A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test compound 4a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6-H6B...N</td>
<td>0.97</td>
<td>2.68</td>
<td>3.601 (3)</td>
<td>158.9</td>
</tr>
<tr>
<td>C20 – H20A ... S1</td>
<td>0.97</td>
<td>2.85</td>
<td>3.635 (3)</td>
<td>138.2</td>
</tr>
<tr>
<td>Test Compound 5a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9-H9A...N8</td>
<td>0.97</td>
<td>2.69</td>
<td>3.604</td>
<td>156.87</td>
</tr>
<tr>
<td>C35-H35B...N2</td>
<td>0.96</td>
<td>2.72</td>
<td>3.632</td>
<td>156.97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X-H</th>
<th>Cg</th>
<th>ARU</th>
<th>H...Cg</th>
<th>X... Cg</th>
<th>X-H-Cg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound 4a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C20 – H20B</td>
<td>Cg (I)</td>
<td>[1555.01]</td>
<td>2.73(10)</td>
<td>3.521</td>
<td>139</td>
</tr>
<tr>
<td>C22-H22B</td>
<td>Cg (4)</td>
<td>[3565.01]</td>
<td>2.94(20)</td>
<td>3.721</td>
<td>138</td>
</tr>
<tr>
<td>Compound 5a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C25-H25A</td>
<td>Cg (2)</td>
<td>[1456.01]</td>
<td>2.96</td>
<td>3.793</td>
<td>145</td>
</tr>
<tr>
<td>C40-H40</td>
<td>Cg (1)</td>
<td>[2675.01]</td>
<td>2.96</td>
<td>3.739</td>
<td>142</td>
</tr>
<tr>
<td>C49-H49B</td>
<td>Cg (14)</td>
<td>[1655.02]</td>
<td>2.97</td>
<td>3.799</td>
<td>144</td>
</tr>
</tbody>
</table>

Symmetry code: (1456), -1+x, y, 1+z; 2675, 1-x, 2-y, z; 1655, 1+x, y, z; 3565, -x, 1-y, -z
Fig 5.4 The molecular structure and atomic numbering scheme of the title compound 4a, showing 30% probability ellipsoids. Hydrogen atoms are not shown for clarity.
Fig 5.5 The molecular structure and atomic numbering scheme of the title compound 5a, showing 30% probability ellipsoids.
5.3.3 Summary
In summary, we have explored some new indole derivatives by introducing for the first time a 5-substituted N-ethyl amino tetrazolyl moiety in place of the conventional phenoxy alkyl amino basic side chain as potential estrogen antagonists for contraceptive activity. We envisaged promising activity for N-2 regioisomers showing favorable crystal structure spatial conformation comparable to known antagonists and supported by molecular docking studies. However, display of weak antagonism by these isomers may be explained due to the larger size of the side chain resulting in its twisting and thereby reducing the binding affinity to the active sites within the receptor cavity, the absence of hydroxyl groups on the tetracyclic core and steric hindering of an active conformational change of the receptor ligand complex due to puckering of 7-membered of benzothiepine ring. These studies give credence to the view that formation of desired active conformation at a "given position in space" is essential requisite for ligands to show the desired biological effect.