Understanding the Selectivity of Genistein for Human Estrogen Receptor-β Using X-Ray Crystallography and Computational Methods

Eric S. Manas,1,* Zhang B. Xu,2 Rayomand J. Unwalla,1 and William S. Somers2
1Department of Chemical and Screening Sciences
Wyeth Research
500 Arcola Road
Collegeville, Pennsylvania 19426
2Department of Chemical and Screening Sciences
Wyeth Research
200 Cambridge Park Drive
Cambridge, Massachusetts 02140

Summary
We present X-ray crystallographic and molecular modeling studies of estrogen receptors-α and -β complexed with the estrogen receptor-β-selective phytoestrogen genistein, and coactivator-derived NR box peptides containing an LXXLL motif. We demonstrate that the ligand binding mode is essentially identical when genistein is bound to both isoforms, despite the considerably weaker affinity of this ligand for estrogen receptor-α. In addition, we examine subtle differences between binding site residues, providing an explanation for why genistein is modestly selective for the β isoform. To this end, we also present the results of quantum chemical studies and thermodynamic arguments that yield insight to the nature of the interactions leading to estrogen receptor-β selectivity. The importance of our analysis to structure-based drug design is discussed.

Introduction
Since its discovery in 1996 (Kuiper et al., 1996), the estrogen receptor (ER) isoform known as ERβ has become the focus of intense investigation as a potential drug target. The existence of clear-cut differences in ERβ and ERα expression suggests that tissues could be differentially targeted with ligands selective for either isoform (Couse et al., 1997; Enmark et al., 1997; Lau et al., 1998; Shughrue et al., 1996). In particular, the fact that ERβ is widely expressed but not the primary estrogen receptor in, for example, the uterus (where estrogenic effects are mediated via ERα) (Harris et al., 2002b) opens up the possibility of targeting other tissues while avoiding certain classical estrogenic effects. Until very recently (Harris et al., 2003), the specific therapeutic utility of ERβ ligand has been unclear. As described previously (Harris et al., 2003), one way to validate ERβ as a target is to design highly ERβ-selective ligands and then use them as a tool to probe the physiological function of ERβ. However, without a fundamental understanding of how ligands can differentiate between ERβ and ERα at the molecular level, this design process can prove to be very difficult.

A major advance toward understanding how some phytoestrogens achieve modest ERβ selectivity was the X-ray structure determination of the ERβ ligand binding domain (LBD) complexed with genistein (GEN, see Figure 1) (Pike et al., 1999), a 40-fold ERβ-selective ligand (Harris et al., 2002a). This study clearly showed that there are only two residue substitutions in close proximity to GEN: ERα Leu354 is replaced by ERβ Met354; and ERα Met421 is replaced by ERβ Ile423. However, despite this observation, the authors found it difficult to reconcile the selectivity of GEN with the protein-ligand interactions. In fact, although it was suggested that the ERα Met354 → ERβ Met354 substitution may allow ERβ to accommodate more polar substituents at the distal end of the binding cavity (Pike et al., 1999), it remains unclear exactly why this might be the case.

In the current paper, we present the X-ray structure determination of the ERα LBD and ERβ LBD, each complexed with GEN. In both cases, the complexes are also cocrystallized with a coactivator fragment containing an LXXLL motif. Using ab initio quantum chemical calculations, we examine the contribution of both residue substitutions, ERα Leu354 → ERβ Met354 and ERα Met421 → ERβ Ile423, to the observed selectivity of GEN. This analysis provides insight into possible mechanisms for enhancing the ERβ selectivity of other scaffolds.

Results
General Observations
The overall ERβ and ERα LBD structures are similar to those previously reported (Brozowska et al., 1997; Pike et al., 1999; Shiau et al., 1998; Warnmark et al., 2002), and thus they will not be described in detail here. We do point out that the helix-12 conformation described for both the ERα and ERβ LBD structures is the same as that observed for other bound agonists (Brozowska et al., 1997; Pike et al., 2000; Shiau et al., 1998; Warnmark et al., 2002), in contrast to the antagonist-like conformation of helix-12 previously reported for ERβ/GEN (Pike et al., 1999). This result is not extremely surprising, but it does deserve a brief explanation. Presumably, the relative free energies of the two helix-12 conformations are similar when GEN is bound and the complex is crystallized, with the “agonist-like” state being slightly more stable. It is likely that the binding of the coactivator fragment to the ERβ/GEN complex provides the additional stabilization required to observe an “agonist-like” state in the structures presented here.

Unbiased electron density difference maps unambiguously define the ligand binding mode for both complexes, as shown in Figure 2. The interactions made by GEN are similar to what was reported previously (Pike et al., 1999). Briefly, the phenol group mimics the EZ “A ring” (Anstead et al., 1997), with the phenolic hydroxyl group involved in a hydrogen bonding network between ERα residues Glu305 and Arg146 (ERβ residues Glu305 and Arg146), and a highly ordered water molecule. An-
other hydrogen bond is formed between the isoflavone 7-OH and N$_\text{H}^\alpha$ of ER$\alpha$ His$_{324}$ (ER$\beta$ His$_{325}$). The core scaffold fills the remainder of the primarily hydrophobic pocket. Interestingly, despite the 40-fold ER$\beta$ selectivity of GEN, the ligand binding modes are nearly identical in both complexes. However, since 40-fold selectivity corresponds to only a 2.2 kcal/mol difference in the free energy of binding to ER$\alpha$ versus ER$\beta$ at room temperature (using $\Delta G = -RT \ln \left[ K_{i} / K_{j} \right]$), one might expect to observe subtle rather than dramatic differences in the way GEN binds to the two isoforms. For the remainder of this section, we will shift our attention to these subtle interactions.

Interactions with ER$\alpha$ Leu$_{384}$/ER$\beta$ Met$_{336}$

The GEN $\gamma$-pyrone B ring is in close proximity to the ER$\alpha$ Leu$_{384}$ $\rightarrow$ ER$\beta$ Met$_{336}$ residue substitution, with the B ring centroid approximately 4.0–4.5 Å (measuring distances in both monomer units) from the ER$\beta$ Met$_{336}$ C$_\text{C}$ atom, versus 6.2 Å from the ER$\alpha$ Leu$_{384}$ C$_\text{C}$ atom. One would expect some contribution to ER$\beta$ selectivity on the basis of this differential B ring-methyl distance alone. In addition, the carbon at the 3-position of the isoflavone ring system is approximately 4.2–4.6 Å from the ER$\beta$ Met$_{336}$ S$_\text{C}$ atom, versus 4.8 Å from the ER$\alpha$ Leu$_{384}$ C$_\text{C}$ atom, although it is not immediately obvious how the methyl $\rightarrow$ sulfur difference also affects the interaction. We have observed differences similar to what is described above when comparing X-ray crystal structures of an ER$\beta$-selective benzofuran complexed with ER$\alpha$ versus ER$\beta$ (Manas et al., 2004). In addition, we have seen a similar potential contribution to core template selectivity for compounds as simple as 4-OH biphenyl, which place a phenyl ring in a region nearly identical to that of the $\gamma$-pyrone of GEN when docked to ER$\beta$ (Edsall et al., 2003; Yang et al., 2004). These observations suggest that compounds placing an aryl or heteroaryl ring at a similar position and orientation as the $\gamma$-pyrone of GEN are able to achieve better contact and stronger interactions with ER$\beta$ Met$_{336}$ than ER$\alpha$ Leu$_{384}$.

Although methionine-aromatic interactions are thought to stabilize proteins (Pal and Chakrabarti, 2001; Pranata, 1997; Viguera and Serrano, 1995), the magnitude of such an interaction relative to the leucine-aromatic interaction, particularly given the specific protein-ligand contacts we observe in ER$\alpha$/$\beta$, is unclear. Thus, in order to support our hypothesis that an aryl group in approximately the same region as the GEN B ring is able to achieve more favorable interactions with ER$\beta$ Met$_{336}$ than ER$\alpha$ Leu$_{384}$, we performed ab initio quantum chemical calculations of the interaction between a phenyl ring (representative of a “generic” aryl group) and the respective side chains of ER$\beta$ Met$_{336}$ and ER$\alpha$ Leu$_{384}$. We modeled the methionine side chain as dimethyl sulfide and the leucine side chain as propane, representing the atoms closest to the B ring. In addition, we utilized approximately the same relative orientation between each side chain and the GEN B ring observed in the X-ray structures, as shown in Figure 3A. Figure 3B shows how the interaction energy varies as a function of the distances shown in Figure 3A. The curves in Figure 3B clearly demonstrate that, given the way in which the ER$\beta$ Met$_{336}$ and ER$\alpha$ Leu$_{384}$ side chains are presented to the pocket, ER$\beta$ Met$_{336}$ clearly has a greater potential to achieve an attractive interaction with an aryl ring than ER$\alpha$ Leu$_{384}$. Furthermore, given the actual distances of closest approach determined from the crystal structures presented above, it is possible that the B ring interaction with ER$\beta$ Met$_{336}$ relative to ER$\alpha$ Leu$_{384}$ can lead to a differential interaction of roughly 1.2–2.2 kcal/mol (apart from a small zero-point energy correction; the range of values reflects the range of distances observed for both monomer units in the ER$\beta$ structure). This represents anywhere between an 8- to 41-fold contribution to selectivity.

Interactions with ER$\alpha$ Met$_{421}$/ER$\beta$ Ile$_{373}$

The GEN 5-OH group, which is most likely involved in an intramolecular hydrogen bond with the 4-keto moiety (Fang et al., 2001; Ferte et al., 1999; Kozerski et al., 2003; Mazurek et al., 1998; Michalak et al., 2001; Nishiyama et al., 2002), is in close proximity to the ER$\alpha$ Met$_{421}$ $\rightarrow$ ER$\beta$ Ile$_{373}$ residue substitution. Solvent descreening upon burial of the ligand is likely to further stabilize this intramolecular hydrogen bond, preventing the 5-OH group from forming a hydrogen bond with the S$_\text{O}$ atom of ER$\alpha$ Met$_{421}$. This is in contrast to what has been observed for certain RAR$\gamma$-selective retinoids possessing an alcohol or oxime moiety, which are able to form a hydrogen bond with the RAR$\gamma$ Met$_{272}$ S$_\text{O}$ atom (Charpentier et al., 1995; Egea et al., 2000; Klaholz and Moras, 1998, 2002; Klaholz et al., 1998, 2000). Even prior to our determination of the ER$\alpha$/GEN X-ray crystal structure, the above points suggested to us the possibility that the 5-OH oxygen and the methionine sulfur atom could be involved in a repulsive interaction, given the fact that oxygen is relatively nonpolarizable and possesses a partial negative charge. However, after determining the ER$\alpha$/GEN structure, it became clear that ER$\alpha$ Met$_{421}$ adopts a different rotamer state compared to that of ER$\alpha$ complexed with E2 or DES. In this new state, the S$_\text{O}$ atom of ER$\alpha$ Met$_{421}$ is approximately 5.2 Å from the GEN 5-OH oxygen, with the sulfur lone pairs facing away from the oxygen atom (see Figures 2A and 4). For comparison, overlaying the ER$\alpha$/GEN structure with ER$\alpha$/E2 (either 1ERE or 1GWR) and ER$\alpha$/DES yields distances of approximately 3.6–4.0 Å and 2.9–3.1 Å, respectively.
Figure 2. Electron Density Difference Maps for the ERα/GEN and ERβ/GEN Complexes

(A and B) Stereoimage of unbiased 3f o-2fc maps contoured at σ, showing the electron density for GEN when complexed with (A) ERα and (B) ERβ. As described in the text, to avoid model bias, the ligand, the loop connecting H8-H9, the C- and N-terminal helices, and the coactivator peptide were omitted from the search models during molecular replacement. Although the density for (B) shows slight disorder within the plane of the A ring, both (A) and (B) demonstrate clear overall electron density for the ligand and demonstrate that the ligand binding mode is well determined in both cases.

with the sulfur and oxygen lone pairs facing the same region of space. Thus, it is possible that a repulsive interaction between the GEN 5-OH oxygen and ERα Met421 S contributes to the change in rotamer state. In contrast, the 5-OH group is approximately 4.0–4.2 Å from the ERβ Ile373 Cα atom (distance of closest approach), and hence we do not expect a repulsive interaction with this residue.

To gain a better understanding of whether or not a repulsive interaction might occur between the GEN 5-OH group and the ERα Met421 side chain relative to that of ERβ Ile373 when the former is in a rotamer state similar to that found in ERα/E2 or ERα/DES, we performed ab initio quantum chemical calculations similar to those described above. For calculations with the aug-cc-pVTZ basis set, the OH group was capped with a methyl group for the sake of computational efficiency (the results described below were confirmed to be qualitatively similar with a phenolic OH using a 6-31G+(d,p) basis set). The relative orientation with respect to dimethyl sulfide was chosen such that the hydroxyl hydrogen is directed away from the sulfur atom. We then substituted the sulfur atom with a methylene, maintaining the same approximate relative orientation (after minimization of the propane fragment), in order to estimate the interaction with ERβ Ile373 when the two side chains occupy the same region of space. This was done to show that the space explored by the sulfur-containing side chain of ERα Met421 is more restricted than that of the aliphatic side chain of ERβ Ile373 due to the different electronic character of the two side chains. The relative orientations and interaction energies are shown in Figure 5. Clearly, interaction with the sulfur-containing chain is purely repulsive, whereas a weakly attractive interaction is possible with the approximately isosteric aliphatic group. The maximal differential interaction of 2 kcal/mol occurs at an intermolecular separation of 3.2 Å, close to the distance that would be observed if Met421 adopted a “DES-like” rotamer. Interestingly, if such a differential interaction were maintained, it would contribute roughly 30-fold to the ERβ selectivity. When combined with the ERα Leu384 → ERβ Met336 contribution to selectivity discussed above, this would then lead to an ERβ selectivity of at least several hundred-fold! However, the above results are clearly expected to be dependent on the relative orientation of the interacting moieties (the orientation we chose represents an “extreme” case). Furthermore, as mentioned above, ERα Met421
seems able to adopt an alternate rotamer to eliminate any repulsive interaction.

In order to see if there is any “residual” differential interaction between GEN and the ERα Met421/ERβ Ile373 side chains when ERα Met421 adopts the alternate rotamer described above, we calculated the interaction energy of the GEN C ring, modeled as 6-acetyl resorcinol, with dimethyl sulfide and propane. As with the B ring interaction, we utilized approximately the same relative orientation between each side chain and the GEN C ring observed in the X-ray structures, as shown in Figure 6. We found that there is no potential for a differential interaction favoring binding to ERβ when each of these residues adopts their respective observed rotamers. The interaction of the 6-acetyl resorcinol with both dimethyl sulfide and propane is predicted to be favorable at the distances observed in the ERα GEN and ERβ GEN crystal structures, with a value of approximately –1.4 kcal/mol (apart from a small zero-point energy correction) at the LMP2/aug-cc-pVTZ(-f) level of theory.

We do point out that repulsion between the GEN 5-OH group and ERα Met421 can still “contribute” to ERβ selectivity, in a sense. This can be seen by considering daidzein, the 5-deoxy analog of GEN (see Figure 1), which is only about 5-fold ERβ selective. In the absence of the 5-OH group, it is conceivable that ERα Met421 could adjust to make a more favorable interaction with the ligand relative to ERβ Ile373, thus leading to a differential interaction that actually favors binding to ERα. This would act in opposition to the differential interaction with ERα Leu384/ERβ Met336 and lower the net ERβ selectivity, consistent with what we observe experimentally. We emphasize that the above hypothesis is also consistent with the lowered ERβ and ERα binding affinities of daidzein (552 nM and 360 nM, respectively) relative to those of GEN (9 nM and 360 nM, respectively). Upon binding to either ERα or ERβ, the intramolecular hydrogen bond between the GEN 5-OH group and the adjacent keto moiety becomes highly descreened from solvent, which should act to strengthen this hydrogen bond further and offset the desolvation penalty of both moieties. This in turn would act to improve binding affinity to both ERα and ERβ. When this OH group is removed (i.e., daidzein), the descreened intramolecular interaction is no longer present, but the keto moiety still becomes desolvated upon binding, which represents a penalty. Thus, both the ERα and ERβ binding affinities are weakened. In ERα, this may be offset by slightly stronger interactions with Met421 relative to Met336 as discussed above, but not enough to improve the ERα binding affinity of daidzein to better than 3 μM. Since the corresponding ERβ binding affinity is 552 nM, daidzein is only ~5-fold ERβ selective. In summary, the GEN 5-OH group can contribute to ERβ selectivity by preventing an interaction from occurring that would otherwise improve binding to ERα relative to ERβ. At the same time, it can contribute to both ERα and ERβ potency by forming a buried intramolecular hydrogen bond with the adjacent keto moiety, which offsets the ligand desolvation penalty.

**Entropic Considerations**

We have also considered a potential contribution to ERβ selectivity due to the relative reduction in side chain...
Understanding the Selectivity of Genistein for ERβ

Figure 6. Structures Used to Estimate the Interaction of the Genistein C Ring, Modeled Here as 6-Acetyl Resorcinol, with the ERα Met421 and ERβ Ile373 Side Chains

As in Figure 3, the side chains were modeled as dimethyl sulfide and propane, respectively, and the relative orientations were adapted from the ERα/GEN and ERβ/GEN crystal structures. Potential energies calculated by using counterpoise-corrected LMP2/aug-cc-pVTZ(-f) show no evidence for a differential interaction favoring ERβ.

Figure 5. Quantum Chemical Calculation of the Interaction between Hydroxyl Oxygen and a Sulfur-Containing versus a Purely Aliphatic Side Chain

(A) Schematic representation of the structures used to determine the interaction of hydroxyl with dimethyl sulfide, when the OH is anti with respect to the sulfur atom. Methanol was oriented with respect to dimethyl sulfide by constraining both the C-O-S-C dihedral and the O-S-C-C improper dihedral to 120° and both the C-O-S and O-S-C angles to approximately 109.5°. Interaction of hydroxyl with propane was determined by substituting the sulfur atom with a methylene group (C₃) and then optimizing the propane geometry as described in the text.

(B) Counterpoise-corrected LMP2/aug-cc-pVTZ potential energy curves are shown for hydroxyl interacting with dimethyl sulfide (triangles) and propane (squares). The intermolecular distances scanned are shown in Figure 5A.

The methionine side chain has approximately 27 possible low-energy rotamer states, based on trans, gauche+, and gauche− rotations for each of three dihedral angles, χ₁, χ₂, and χ₃. Although, based on an analysis of protein crystal structures, the effective number of rotamers has been estimated to be roughly 14.9 (Pickett and Stemberg, 1993). In comparison, the number of effective rotamers for leucine and isoleucine is approximately 3.7 and 4.5, respectively (Pickett and Stemberg, 1993). Upon burial of the residue, either by protein folding or ligand binding, the number of rotamers is likely to be reduced to a number closer to unity (Smith et al., 1991). Given that the ligand binding pocket in ER is encapsulated by the rest of the protein, and that the solvent-exposed surface areas of ERα Leu421/Met421 and ERβ Met421/Ile423 are very low even when no ligand is bound, it is likely that the effective number of rotamers for these residues in the unliganded states is considerably lower than what is reported above. Therefore, we consider RT ln(14.9) = 1.6 kcal/mol to be a high estimate of the conformational entropy penalty associated with restricting the methionine side chain rotamers upon ligand binding, with RT ln(3.7) = 0.77 kcal/mol and RT ln(4.5) = 0.89 kcal/mol representing the corresponding penalties for leucine and isoleucine, respectively. Thus,
taken individually, a Met → Leu/Ile substitution should in principle be capable of contributing 3- to 4-fold to selectivity at most. However, for ERα/β, differences in the entropic penalty term of Equation 2 for the ERα Leu384 → ERβ Met384 and ERα Met384 → ERβ Ile373 residue substitutions approximately cancel (we point out that, upon ligand binding, differential desolvation effects for each of these two residue substitutions are also expected to act in opposition to one another). Thus, ligands that do not make differential interactions with either set of residues, as well as those whose differential interactions cancel, will tend to be nonselective. In contrast, GEN does appear to make a net differential interaction, which favors binding to ERβ over ERα.

Discussion

The results of the above study suggest that the interaction of the GEN B ring with ERα Leu384 relative to ERβ Met384, and to some extent the interaction of the GEN 5-OH group with ERα Met421 relative to ERβ Ile373, are capable of contributing significantly to the observed ERβ selectivity of GEN. Using ab initio quantum chemical calculations, we have investigated the ability of both groups to make a differential interaction with the nearest ERβ residue relative to its ERα counterpart. For example, we find that an aryl group in the B ring region of GEN is capable of making a more favorable interaction with ERβ Met384 than with ERα Leu384, based on the position and orientation of each side chain relative to the ligand. Furthermore, the GEN 5-OH group is capable of making a repulsive interaction with the side chain of ERα Met321 (depending on the rotamer state and relative orientation), which does not occur with the purely aliphatic side chain of ERβ Ile323. The unfavorable interaction with ERα Met321 appears to be due to a combination of two factors. First, the fact that both the oxygen and sulfur atoms possess partial negative charges leads to an unfavorable (i.e., tail-to-tail) dipole-dipole interaction. Second, since oxygen is relatively nonpolarizable, dispersive, and inductive interactions with the methionine side chain are likely to be small and thus are unable to compensate for the leading repulsive electrostatic term of the interaction. These effects are due to the electronic character of the methionine side chain relative to the isoleucine side chain, and not due to differences in residue size or shape.

Interestingly, the binding of GEN to ERα appears to force the Met321 side chain into a rotamer state in which any repulsion with the 5-OH group is eliminated (different from that observed when nonselective ligands are bound to ERα). We do not actually observe any maintained repulsive interaction between the GEN 5-OH group and ERα Met321 in the ERα/GEN crystal structure. In fact, our model calculations suggest that there is no significant contribution of the ERα Met321 → ERβ Ile323 residue substitution to the ERβ selectivity of GEN. However, based on a comparison of binding data for GEN and daidzein, the GEN 5-OH group itself does appear to affect ERβ selectivity. One possibility is that it may prevent the ERα Met321 side chain from making a more favorable interaction with the ligand compared to ERβ, which would act in opposition to the differential interaction with the B ring and thus lower the net ERβ selectivity. These results suggest that a more selective ligand could be designed by introducing functional groups to GEN or a GEN-like scaffold that penetrate more deeply into the ERα Met321/ERβ Ile373 pocket, such that either (1) a repulsive interaction is maintained with ERα Met321, while preserving optimal contact with ERβ Ile373 (i.e., the methionine side chain cannot “escape” from the unfavorable interaction), or (2) the ligand incurs an energetic strain penalty when bound to ERα in order to alleviate the repulsion with ERα Met321. So far, we have been successful at enhancing ERβ selectivity to beyond 100-fold by applying such an approach to ligands utilizing 2-phenyl benzofuran and 2-phenyl benzoxazole scaffolds (Manas et al., 2004).

We point out that the overall free energy of the protein may also be slightly increased by the ERα Met321 rotamer change. In addition, another less obvious possibility is that the GEN 5-OH group modulates the interaction of the chromone moiety with ERβ Met384 via the intramolecular hydrogen bond with the 4-keto group. Thus, the GEN 5-OH group might indirectly affect the differential interaction with ERα Leu384/ERβ Met386. We are currently investigating such substituent-induced effects on the methionine-aromatic interaction.

The conserved ligand binding mode we observe for GEN bound to ERα relative to ERβ may also have general implications for structure-based design involving enhancement of selectivity. Introduction of functional groups to modulate affinity of a ligand for one isofrom of a protein or another introduces the possibilities that both the protein and ligand can adjust in response to the perturbation. For GEN binding to ERα and ERβ, clearly adjustment of the protein, i.e., ERα Met321, represents the more favorable free energy change compared to, say, a change in binding orientation. However, the same is not necessarily true for other scaffolds. For example, if the free energy difference between binding orientations (keeping the protein conformation fixed) is less than the energetic cost of protein readjustment (keeping the ligand orientation fixed), then it is conceivable that a different binding mode could be observed for ERα relative to ERβ. In this case, it would be important for structure-based design applications to know the alternate binding mode(s), especially if the design strategy involves targeting a repulsive interaction in one isofrom (e.g., ERα) relative to the other (e.g., ERβ). It might also be important to consider modifications to the scaffold that prevent any major changes in the ERα binding mode.

In general, side chain entropy considerations tend to favor methionine → leucine or isoleucine substitutions by roughly 3- to 4-fold at most, since burial of methionine upon ligand binding is likely to lead to a greater restriction of rotamer states than for leucine or isoleucine, simply because methionine has more effective rotamer states in the unliganded protein. For the case of GEN binding to ERα/β, the small penalties associated with ERα Leu384 → ERβ Met386 and ERα Met321 → ERβ Ile323 appear to oppose one another. However, it may be possible for the degree of cancellation to vary during lead optimization. For example, an unoptimized lead mole-
cule with moderate selectivity may allow for some disorder in Erα, Met424, or Erβ Met332, even when the ligand is bound. However, after lead optimization, either methionine nine side chain could become conformationally restricted upon binding. Thus, if one attempted to optimize ER selectivity by introducing functional groups that form a more favorable interaction with Erβ Ile328 relative to Erα Met232, the differential interaction would tend to act synergistically with the differential entropic penalty associated with restricting these residue side chains. In contrast, a differential interaction with Erα Leu324 relative to Erβ Met332 favoring Erβ would act in opposition to the differential entropic penalty of restricting the two side chains.

Finally, we have demonstrated that the helix-12 conformation in both the Erα/GEN and Erβ/GEN structures is similar to that observed for other agonists complexed to Erα/β (Brzozowski et al., 1997; Pike et al., 2000; Shiau et al., 1998; Warmark et al., 2002), in contrast to the antagonist-like conformation of helix-12 previously reported for Erβ/GEN (Pike et al., 1999). This serves to highlight that the binding of the coactivator fragment itself is able to influence the conformational state of helix-12, and that care should be taken when making inferences about the functional behavior of ligands bound to nuclear receptors in the presence of coactivator fragments.

**Experimental Procedures**

**Materials**

Coactivator peptide fragments Biotin-SGSHKLVLQQLTTT-COOH and Biotin-SGSHKLRLLEQ-COOH were obtained from the Wyeth DNA and peptide synthesis lab. The two peptides used in this study were derived from the steroid receptor coactivator 1 and 3 (SRC-1 and SRC-3) NR box domains. SRC-1 and SRC-3 are both known to enhance ligand-dependent transcriptional activation of the estrogen receptor, and they have been shown to be specifically recruited by two cycles of French press (SLM Instrument) at 20,000 psi in a buffer containing 5 M ligand. Purity was estimated to be >98% by SDS-PAGE. Eluted protein was concentrated by a Millipore Ultrafree centrifugal filtration device and desalted with a BioRad disposable desalting column equilibrated with 50 mM ammonium bicarbonate (pH 7.5).

Cloning, Expression, and Purification of Human ERβ LBD

Human ERα cDNA was generated from human testis RNA by RT-PCR and cloned into mammalian expression vector pcDNA3. Amino acids 261–500 of the LBD were amplified from the cloned cDNA by PCR with the forward primer 5′-GAACCATGGACGACGCCCTGAG and the reverse primer 5′-GAAGCTGATCGAGCGCTGAGCTAGTG-3′ and the reverse primer 5′-GGACTCGAGTTAAGTGGACACGCTGAGCCGCTAGTG-3′. The PCR fragment was inserted into E. coli expression vector pET16b (Novagen) between the Ncol and Xhol restriction sites. The primers used encode three extra asp codons, one before the codon for Ile334 and two after Leu336. The expressed LBD thus has the following sequence: M[D(Dar)526]DD.

ERβ LBD was overexpressed from a high-density culture of E. coli BL21DE3 host cells (Strategene) in a Biostat C-10 bioreactor (B. Braun Biotech). Cultures were induced with 1.0 mM IPTG final for 4 hr at 37°C. Cells pellets were quick-frozen in liquid nitrogen prior to storage at -80°C.

ERβ LBD was purified as follows. Harvested cells were lysed by two cycles of French press (SLM Instrument) at 20,000 psi in a buffer of 20 mM Tris-Cl (pH 7.5), 0.5 M NaCl, 5 mM DTT, 1 mM EDTA (10 mg/ml of cells). Lysate was clarified by centrifugation at 45,000 × g for 45 min at 4°C then applied to a Q Sepharose (Pharmacia) column. The flow through was then applied to a 5 ml estradiol-Sepharose fast flow column (PTI Research, Inc.) and washed with 300 ml of 10 mM Tris-HCl (pH 7.5) containing 0.5 M NaCl, 1 mM EDTA (buffer A). The column was then reequilibrated with 50 ml of 10 mM Tris-HCl (pH 7.5), 0.2 M NaCl, 1 mM EDTA (buffer B), and the protein was carboxymethylated by using 50 ml buffer B containing 5 mM iodoacetic acid. Then, the column was washed by 500 ml buffer A, followed by elution in buffer A containing 200 μM ligand. Finally, the eluate was concentrated by ultrafiltration and size exclusion chromatography (Sephadex 200, Pharmacia) by using the elution buffer containing 5 μM ligand. Purity was estimated to be >98% by SDS-PAGE. Excess ligand was removed by passing the solution through a G-25 column (Pharmacia).

**Crystallography**

The ERα/GEN complex was concentrated to 11.0 mg/ml and then mixed with the SRC-3 peptide at a molar ratio of 1:1.5 protein-ligand peptide. Screening of crystallization conditions was performed at 18°C by using the hanging drop vapor diffusion method (McPherson, 1976). Crystals were grown from a drop containing a mixture of protein solution (50 mM ammonium bicarbonate (pH 7.5)) and reservoir solution of 15% PEG3350 (v/v), 0.2 M NaI, 0.1 M HEPES (pH 7.4). The ERα/GEN complex was concentrated to 12.5 mg/ml in 0.2M NaCl, 1 mM EDTA, 5 mM DTT, 10 mM Tris-HCl (pH 7.5). A molar excess of the SRC-3-peptide was then added to ERα/GEN at a ratio of 1:5:1. Crystals were grown by using the same technique and appeared over wells containing 25% PEG2000 (v/v), 0.15 M.
Table 1. X-ray Data Collection and Refinement Statistics for Structures Studied in This Paper

<table>
<thead>
<tr>
<th>Structure</th>
<th>ERα/GEN</th>
<th>ERβ/GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P 2_12_2</td>
<td>P 2_12_2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>60.88</td>
<td>52.23</td>
</tr>
<tr>
<td>b (Å)</td>
<td>69.83</td>
<td>87.67</td>
</tr>
<tr>
<td>c (Å)</td>
<td>57.37</td>
<td>100.11</td>
</tr>
<tr>
<td>Maximal resolution (Å)</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Unique observations</td>
<td>17,143</td>
<td>21,240</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.0</td>
<td>99.9</td>
</tr>
<tr>
<td>Rmerge&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.1%</td>
<td>6.3%</td>
</tr>
<tr>
<td>Mean I/(I)</td>
<td>26.1</td>
<td>23.4</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.75</td>
<td>5.1</td>
</tr>
<tr>
<td>Highest resolution bin (Å)</td>
<td>2.07–2.00</td>
<td>2.34–2.30</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>91.6%</td>
<td>100%</td>
</tr>
<tr>
<td>Rmerge&lt;sup&gt;a&lt;/sup&gt;</td>
<td>33.5%</td>
<td>21.4%</td>
</tr>
<tr>
<td>Mean I/(I)</td>
<td>3.8</td>
<td>9.0</td>
</tr>
<tr>
<td>Mosaicity</td>
<td>0.37</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of molecules per asymmetric unit</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Protein atoms&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1,872</td>
<td>3,554</td>
</tr>
<tr>
<td>Other atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ligand&lt;sup&gt;c&lt;/sup&gt;</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Coactivator peptide&lt;sup&gt;c&lt;/sup&gt;</td>
<td>82</td>
<td>162</td>
</tr>
<tr>
<td>Water</td>
<td>115</td>
<td>78</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>15–2.0</td>
<td>15–2.3</td>
</tr>
<tr>
<td>Reflections used</td>
<td>15,809</td>
<td>20,690</td>
</tr>
<tr>
<td>Reflections in working set</td>
<td>15,036</td>
<td>19,683</td>
</tr>
<tr>
<td>Reflections in test set</td>
<td>773</td>
<td>1,007</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.224</td>
<td>0.232</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.269</td>
<td>0.282</td>
</tr>
<tr>
<td>Rms bond length&lt;sup&gt;d&lt;/sup&gt; (Å)</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>Rms bond angles&lt;sup&gt;e&lt;/sup&gt; (°)</td>
<td>1.111</td>
<td>1.024</td>
</tr>
<tr>
<td>Mean B factor (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>28.5</td>
<td>29.9</td>
</tr>
<tr>
<td>Main chain atoms</td>
<td>27.3</td>
<td>29.6</td>
</tr>
<tr>
<td>Side chain atoms</td>
<td>29.4</td>
<td>30.4</td>
</tr>
<tr>
<td>Ligand</td>
<td>20.5</td>
<td>15.5</td>
</tr>
<tr>
<td>Water</td>
<td>35.1</td>
<td>29.8</td>
</tr>
<tr>
<td>Rms backbone ∆B (Å&lt;sup&gt;2&lt;/sup&gt;)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.875</td>
<td>0.685</td>
</tr>
<tr>
<td>% A, B, L (a, b, l, p)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>100</td>
<td>99.8</td>
</tr>
</tbody>
</table>

<sup>a</sup>R<sub>merge</sub> = \(\frac{\sum_{hkl} I_{hkl} - \langle I_{hkl} \rangle}{\sum_{hkl} I_{hkl}}\), where I<sub>n</sub> is the n<sup>th</sup> observation of reflection hkl, and \(\langle I_{hkl} \rangle\) denotes an average of reflection hkl over n observations.

<sup>b</sup>Per monomer unit.

<sup>c</sup>Crystallographic R factors were computed by using R<sub>work</sub> = \(\frac{\sum_{hkl} |F_{hkl}| - |F_{hkl}|}{\sum_{hkl} |F_{hkl}|}\) and R<sub>free</sub> = \(\frac{\sum_{hkl} |F_{hkl}| - |F_{hkl}|}{\sum_{hkl} |F_{hkl}|}\). R<sub>free</sub> values were calculated in the same manner as R<sub>work</sub>, except over approximately 4% of the data was excluded from the refinement.

<sup>d</sup>Root mean square deviation in bond length and bond angle distances from Engh and Huber ideal values.

<sup>e</sup>Rms mean square deviation between B factors for bonded main chain atoms.

<sup>f</sup>Percentage of residues located in most favored (additional) regions of the Ramachandran plot as determined by PROCHECK (Laskowski et al., 1993).

MgCl<sub>2</sub>, 20 mM hexaminecobalt trichloride, and 0.1 M MES (pH 6.0). Prior to data collection, all crystals were briefly soaked in a solution containing mother liquor and 15%–20% glycerol (v/v).

X-ray data were collected at 100 K by using the Quantum-4 CCD area detector at the Advanced Light Source (ALS, Berkeley, CA) and were processed by using DENZO and SCALEPACK (Otwinowski and Minor, 1994). Crystal structures were solved by molecular replacement with AMORE (Bailey, 1994) by using the cocrystal structure of ER<sub>α</sub> complexed with diethylstilbestrol (Shiau et al., 1998) as a search model (pdb code: 2ERD). To avoid model bias, the ligand, the loop connecting H8-H9, the C- and N-terminal helices, and the coactivator peptide were omitted from the search models. Structures were refined with the program CNS (Brunger, 1998). The resulting difference electron density maps showed clear electron density for the ligand, residues within the binding site, helix 12, and the coactivator peptide. Apart from the conformation of helix-12, which is clearly influenced by the presence of a coactivator peptide, the ERβ/GEN structure was found to be in good agreement with the published structure of ERα/GEN without a coactivator peptide (Pike et al., 1999) (pdb code: 1QKM). The final model of the ERα/GEN crystal structure contains a monomer, with the ligand and coactivator peptide bound to their respective binding sites, and 144 water molecules. The model for ERβ/GEN contains a dimer, with a ligand and coactivator peptide bound to each monomer unit, and 78 water molecules. Cysteine modifications and some flexible loop residues were not included into the models due to poor electron density. Table 1 gives the data collection and refinement details for both complexes.
Understanding the Selectivity of Genistein for ERα

Molecular Modeling

All quantum chemical calculations were performed by using the Jaguar software package (Jaguar 5.5; Schrodinger, LLC), as described previously (Manas et al., 2004). Briefly, intermolecular potential energy curves were calculated by constructing a z-matrix for both molecules, and then varying the intermolecular distance while holding the relative orientation of the two molecules fixed. Molecular geometries were optimized for the isolated molecules at the LMP2 level of theory by using the 6-31G** + i and 6-311G** + i basis sets (Clark et al., 1983; Frisch et al., 1984; Krishnan et al., 1980; McLean and Chandler, 1980). Potential energies were then calculated for the optimized structures by using the augmented correlation-consistent polarized valence triple zeta (aug-cc-pVTZ) basis set (Dunning, 1989; Kendall et al., 1992; Woon and Dunning, 1993, 1994), also at the LMP2 level of theory. For calculations involving a phenyl ring, the aug-cc-pVTZ-f basis set was used. The LMP2/local-Moller-Plesset second-order perturbation theory method (Pulay and Saeboe, 1986; Saebø and Pulay, 1988, 1993, 2001; Saebø et al., 1993) is known to minimize basis set superposition error (BSSE), and therefore counterpoise (CP) calculations (Boys and Bernardi, 1970; Gutowsky et al., 1987) were performed only to correct the Hartree-Fock (HF) term for BSSE.

We utilized the above approach to estimate the interaction of various components of GEN with methionine versus leucine or isoleucine, in order to understand the nature of the interactions leading to ERα selectivity. In particular, we were concerned about accurately capturing both the polarity and polarizability of the methionine side chain, especially when estimating interactions with the core scaffold. Although a sulfur-aromatic nonbonded interaction can be accounted for by empirical energy functions (Nemethy and Scheraga, 1981; Pranta, 1997), we felt that a quantum chemical approach would be better able to capture the balance between electrostatic, dispersive, and inductive effects. In addition, although previous quantum chemical studies have shown that canonical MP2/6-31G(d) calculations (uncorrected for BSSE) on the sulfur-aromatic interaction are qualitatively reproduced by forcefields like AMBER95 and OPLS-AA (Pranta, 1997), we felt that higher-level quantum chemical calculations were necessary, using the specific relative orientations observed in our crystal structures, to adequately capture the subtle differences between methionine and leucine or isoleucine. Finally, since the interactions studied in this paper are completely buried inside the ER pocket and are thus fully desolvated from solvent action, we are only concerned with understanding intermolecular interactions of GEN with ERα relative to ERβ, we neglect solvation in our calculations.

As discussed previously (Manas et al., 2004), the 6-311G** + i basis set yields qualitatively similar results compared to aug-cc-pVTZ when used to evaluate intermolecular potential energies in conjunction with LMP2. However, these interaction energies are typically on the order of 0.5 kcal/mol lower with aug-cc-pVTZ when Charpentier, B., Bernardon, J.M., Eustache, J., Millois, C., Martin, S., and Saebo, G. (2004). Understanding the Selectivity of Genistein for ERα

Received: May 25, 2004
Revised: August 18, 2004
Accepted: September 13, 2004
Published: December 7, 2004

References


Enmark, E., Peltohakkio, M., Grandien, K., Lagercrantz, S., La-


Accession Numbers

Atomic coordinates have been deposited in the Protein Data Bank with accession codes 1X7R and 1X7J for the ERalpha/GEN and ERbeta/GEN complexes, respectively.