

Rapid Creation of a Novel Protein Function by *in Vitro* Coevolution

Zhilei Chen¹ and Huimin Zhao^{1,2,3*}

¹Center for Biophysics and Computational Biology, University of Illinois, Urbana IL 61801, USA

²Department of Chemical and Biomolecular Engineering University of Illinois, Urbana IL 61801, USA

³Department of Chemistry University of Illinois, Urbana IL 61801, USA

We have developed a simple and efficient method for creation of novel protein functions in an existing protein scaffold. The *in vitro* coevolution method involves design of a hypothetical pathway for the target function followed by stepwise directed evolution of the corresponding protein along the pathway. As a test case, this strategy was used to engineer variants of human estrogen receptor α ligand-binding domain (hER α LBD) with novel corticosterone activity. Two steroids, testosterone and progesterone, that provide a progressive structural bridge between 17 β -estradiol and corticosterone, were chosen to assist the directed evolution of hER α LBD. A total of approximately 10⁶ variants were screened in four rounds of random mutagenesis, resulting in two hER α LBD variants that respond to corticosterone. Creation of this new ligand activity required the presence of four simultaneous mutations. In addition, several required mutations were located outside the ligand binding pocket and yet exerted important action on ligand binding. Our results demonstrate the ability of *in vitro* coevolution to create novel protein function that is difficult or impossible to achieve by existing protein engineering approaches and also shed light on the natural evolution of nuclear hormone receptors. This *in vitro* coevolution approach should provide a powerful, broadly applicable tool for engineering biological molecules and systems with novel functions.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: protein engineering; rational design; directed evolution; nuclear hormone receptor; nuclear receptor evolution

*Corresponding author

Introduction

Protein engineers now possess an unprecedented capability to design, modify and engineer naturally occurring proteins at the molecular level. Two different yet complementary approaches have been developed in the past two decades: rational design and directed evolution.^{1–3} Rational design involves the rational alterations of selected residues in a protein *via* site-directed mutagenesis, and requires detailed knowledge of protein folding, structure, function, and dynamics. In contrast, directed evolution mimics the process of natural evolution in the test-tube, involving repeated cycles of creating molecular diversity by random mutagenesis and/or

gene recombination and screening/selecting the functionally improved variants. Both approaches have been used successfully to engineer a wide variety of protein functions such as stability, activity, affinity, selectivity and pH profiles.^{1–5}

However, most of the protein engineering studies are concerned with improvement of existing protein functions, whereas creation of novel protein functions remains an overwhelming challenge in protein engineering.^{1,2,6} This challenge may come from the fact that multiple simultaneous or synergistic mutations are required for the creation of a novel protein function, which represents a huge jump in the sequence space.⁷ Currently, identifying these simultaneous residues proves to be too difficult for either rational design or direct evolution. For example, conventional directed evolution generally requires a starting protein with some detectable activity toward the desired function.⁸ Moreover, the potential library size of a typical protein of 300 amino acid residues with three simultaneous mutations is $\sim 3 \times 10^{10}$, which is often too large to be experimentally screened in a conventional

Abbreviations used: hER α , human estrogen receptor α ; LBD, ligand-binding domain; E₂, 17 β -estradiol; HEC, human endometrial cancer; ER, estrogen receptor; PR, progesterone receptor; MR, mineralocorticoid receptor; AR, androgen receptor; GR, glucocorticoid receptor.

E-mail address of the corresponding author: zhao5@uiuc.edu

directed evolution experiment. On the other hand, despite recent advances in computational protein design and structural proteomics, identifying the molecular basis for the desired protein function by rational design is not straightforward and reliable.³ In addition, recent findings that many protein functions are not confined to a small set of amino acids but are affected by residues far away from active sites^{9–12} have added another level of difficulty.

Here, we describe a novel protein engineering strategy, *in vitro* coevolution, that allows creation of novel protein functions. Coevolution is a ubiquitous process that is responsible for the parallel adaptive evolution of plant–insect interactions, host–parasite interactions, protein–protein interactions, and protein–DNA/RNA interactions.^{13–17} Our *in vitro* coevolution approach mimics the process of natural coevolution in the test-tube, and involves the design of a hypothetical evolutionary pathway for the target function followed by stepwise directed evolution of the corresponding protein along the pathway (Figure 1). Since the target function cannot be screened experimentally for the wild-type protein, the hypothetical pathway is designed such that each intermediate function can be screened experimentally. With the progression of directed evolution, these evolved intermediate functions will eventually bridge the functional gap between the wild-type protein function and the novel protein function.

To demonstrate the utility of this *in vitro* coevolution strategy, we have chosen human estrogen receptor α ligand-binding domain (hER α LBD) as a model system. Human estrogen receptor (hER) is a ligand-regulated transcription factor that mediates the actions of estrogen in different target tissues including the reproductive, pituitary, hypothalamus, bone, liver, and cardiovascular systems.¹⁸ It is a member of the nuclear receptor superfamily that comprises steroid receptors, non-steroid receptors, and orphan receptors.¹⁹ Like other members of the

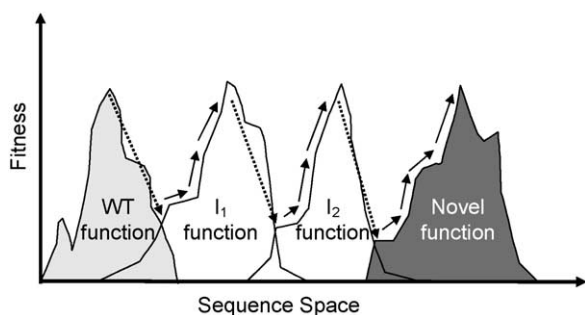


Figure 1. The strategy of *in vitro* coevolution for engineering novel protein functions. The wild-type (WT) protein function and the novel protein function are separated by an inactive region of sequence space, which may be filled by two intermediate functions (I_1 and I_2) that are amenable to conventional directed evolution. The arrows illustrate a potential evolutionary path leading to the novel protein function.

superfamily, the hER has three modular structural domains: an amino-terminal ligand-independent transactivation domain, a central DNA-binding domain (DBD), and a carboxy-terminal ligand-binding domain (LBD). The hER LBD interacts specifically with its physiological ligand 17 β -estradiol (E_2) and contains a dimerization function and a ligand-dependent activation function. The hER has been linked with several human diseases such as breast cancer and osteoporosis, and considerable effort has been directed at understanding the molecular basis of the estrogen receptor and ligand interactions.^{18–21} It should be noted that, despite the low level of sequence homology, as low as 20% between the LBDs of different nuclear receptors, all these proteins share a similar fold consisting of 11 or 12 α -helices and a small β -sheet arranged in an antiparallel sandwich structure.

Here, we have for the first time demonstrated that *in vitro* coevolution can be used to create novel corticosterone activity in the hER α LBD. Remarkably, creation of this novel ligand activity required four simultaneous mutations. In addition, several required mutations were located outside the ligand binding pocket and yet exerted important action on ligand binding. Our findings provide new insights into the natural evolution and molecular basis of ligand specificity of nuclear hormone receptors. More importantly, our *in vitro* coevolution approach allows creation of novel protein functions, which is difficult to achieve by rational design and conventional directed evolution, and may provide a powerful, broadly applicable tool for engineering novel protein functions.

Results and Discussion

Corticosterone activity is a novel function to the hER α LBD

The hER α has exquisite ligand specificity and sensitivity that enables it to discriminate between different classes of steroids with closely related structures.^{22,23} For example, although the chemical structure of testosterone (a C_{19} steroid) and E_2 (a C_{18} steroid) differ only slightly in the A-ring region, the activation of the hER α requires at least 10,000-fold higher concentration of testosterone than that of E_2 .¹⁰ Since corticosterone (a C_{21} steroid) differs from E_2 in four positions in their chemical structures, we reasoned that corticosterone may not be able to activate the hER α .

To test this hypothesis, we used a yeast two-hybrid-based cell growth assay to determine the dose-response profiles of E_2 , testosterone, progesterone, and corticosterone to the wild-type hER α LBD. As shown in Figure 2, the hER α LBD responds to sub-nanomolar concentrations of E_2 and micromolar concentrations of testosterone. It barely responds to progesterone at saturating ligand concentrations ($\sim 10^{-5}$ M), and does not respond at all to corticosterone at saturating ligand

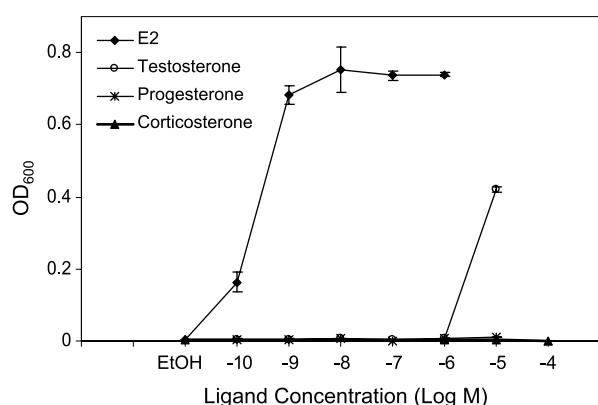


Figure 2. Dose-response profiles of E₂, testosterone, progesterone, and corticosterone to the wild-type hER α LBD determined by a yeast two-hybrid-based cell growth assay.

concentration ($\sim 10^{-4}$ M). Furthermore, we determined the dose-response profile of corticosterone to the full-length hER α in ER-negative human endometrial cancer (HEC-1) cells. Although the HEC-1 cell transactivation assay is more sensitive than the yeast cell transactivation assay,¹⁰ the hER α still failed to respond to corticosterone at 10^{-4} M (data not shown). Thus, we suggest that the corticosterone activity is a novel function to the hER α LBD in the transcriptional activation assays.

In vitro coevolution of novel corticosterone activity in the hER α LBD

To create a variant of the hER α LBD that responds to corticosterone, we first used testosterone and progesterone to construct a hypothetical evolutionary

pathway between E₂ and corticosterone (Figure 3(a)), and then we used directed evolution to evolve, sequentially, hER α LBD variants that act on these two hypothetical evolutionary intermediates. It should be noted that steroids E₂, testosterone, progesterone, and corticosterone are the physiological ligands for estrogen receptor (ER), androgen receptor (AR), progesterone receptor (PR), and glucocorticoid receptor (GR), respectively. In addition, these four steroids are important intermediates in the same biochemical pathway of cholesterol biosynthesis.

The first and second rounds of directed evolution were carried out to obtain hER α LBD variants with increased potency to testosterone, whereas the third and fourth rounds were to obtain hER α LBD variants with increased potency to progesterone. In each round, error-prone PCR was used to introduce a low frequency of random point mutations (one or two amino acid substitutions per gene on average.²⁴) into the LBD fragment consisting of hER α amino acid residues 312–595. A total of approximately 10^6 variants were screened using a yeast two-hybrid system (Figure 3(b)).¹⁰

The first two rounds of directed evolution resulted in a hER α LBD variant, T17-2, that showed >500-fold increased sensitivity toward testosterone in yeast compared to the wild-type hER α LBD, and also responded to progesterone at micromolar concentrations (Figure 4).¹⁰ The wild-type hER α LBD had an almost undetectable response to progesterone at saturating ligand concentration of 10^{-5} M in yeast (Figures 2 and 4). The subsequent two rounds of directed evolution led to two new variants, Pg10-1 and Pg10-16, that showed responses to progesterone at nanomolar concentrations, and more importantly, showed significant responses to corticosterone (10^{-4} M) within 24

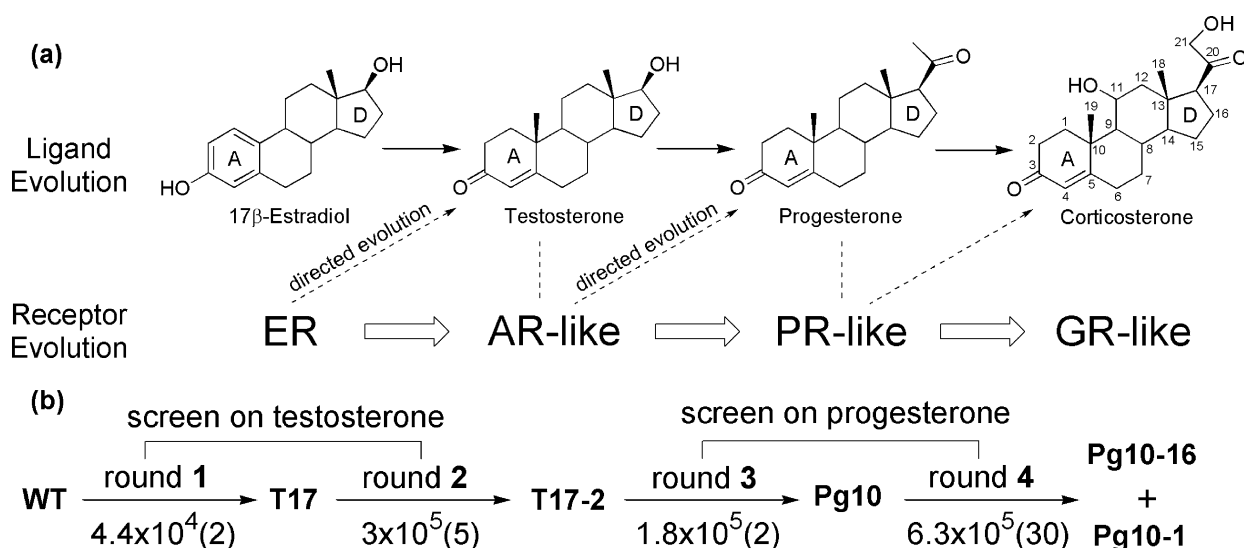


Figure 3. In vitro coevolution of novel corticosterone activity in the hER α LBD. (a) Schematic diagram of the strategy. (b) Experimental implementation of the strategy. The library size and the number of positive variants (in parentheses) in each round of directed evolution are shown below the arrow. The best mutant identified from each round was selected as the parent for the next round of directed evolution.

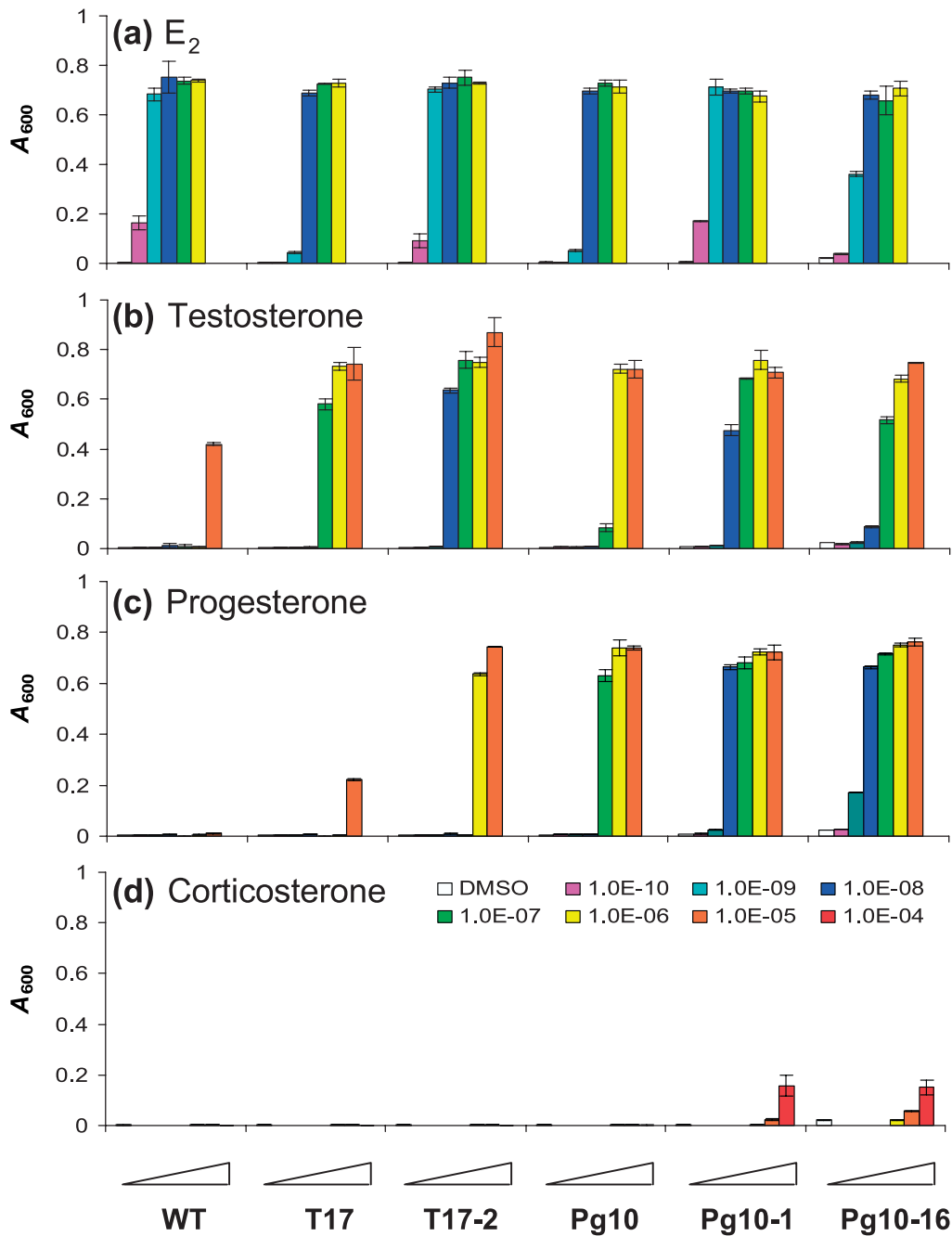


Figure 4. Comparison of dose-response profiles of the wild-type hER α LBD and various hER α LBD mutants determined by a yeast two-hybrid-based cell growth assay. (a) E₂, (b) testosterone, (c) progesterone, and (d) corticosterone. E₂, testosterone and progesterone were dissolved in 100% ethanol, whereas corticosterone was dissolved in 100% DMSO. The ligand concentrations are shown in molarity.

hours in yeast. In comparison, all the other evolved hER α LBD variants and the wild-type hER α LBD showed no corticosterone-dependent response in yeast, even after incubation at 30 °C for four days (data not shown). However, it should be noted that the level of corticosterone response is still quite low compared to that of E₂ response. Further improvement of the corticosterone response by another round of directed evolution and saturation mutagenesis at the residue positions identified from the last round failed to yield any improved variants, suggesting

that a local maximum in the sequence space may be reached.

It should be noted that although the ligand is not experimentally evolved in the above-described *in vitro* coevolution strategy, testosterone and progesterone may be considered as the intermediates of a hypothetical evolutionary pathway between E₂ and corticosterone because of the progressive changes in their chemical structures. The evolution of the corticosterone activity in the hER α LBD protein is directed by these ligands, which, to a large extent, mimics the principle of natural coevolution.

Characterization of the evolved hER α LBD variants

To confirm the evolved corticosterone activity in mammalian cells, we determined the dose-response profiles of corticosterone to the full-length wild-type hER α and the variants identified from the last round of directed evolution in HEC-1 cells. Plasmids containing the full-length wild-type hER α and variants were transiently transfected into HEC-1 cells, together with an internal control and a reporter plasmid. The wild-type hER α did not respond to corticosterone at the saturating concentration, 10^{-4} M, while both Pg10-1 and Pg10-16 responded to corticosterone with an EC₅₀ value of $\sim 3 \times 10^{-6}$ M (data not shown). The maximum transactivation potency of Pg10-1 and Pg10-16 is $\sim 80\%$ that of the wild-type hER α in response to 10^{-8} M E₂ and both variants showed slightly elevated basal level responses.

We also performed *in vitro* hormone binding assays to determine whether the observed transcriptional activities of various variants correspond to their ligand affinities. Wild-type and mutant hER α LBDs (amino acid residues 312–595) were cloned into the pET15b vector and expressed in BL21(DE3) cells. This plasmid introduced a 6 \times histidine tag to the N terminus of the expressed ER. The expressed receptors were purified to near-homogeneity using a nickel resin column and used in the ligand binding assays. The E₂ binding affinities of wild-type and mutant hER α LBDs, $K_d^{E_2}$, were determined using a direct hormone binding assay. As shown in Table 1, T17, Pg10, and Pg10-16 all showed decreased affinity toward E₂ compared to the wild-type hER α LBD. This result agreed well with the *in vivo* E₂ activity measured in yeast, in which T17 and Pg10, and Pg10-16 showed \sim tenfold and \sim fivefold decreased E₂ response, respectively, compared to that of the wild-type hER α LBD. The $K_d^{E_2}$ s of T17-2 and Pg10-1 were similar to that of the wild-type hER α LBD, so did their dose-response profiles measured in yeast. The binding affinities of wild-type and mutant hER α LBDs for testosterone,

progesterone and corticosterone were determined using a competitive hormone binding assay and characterized as the relative binding affinities (RBAs). The wild-type hER α had similar RBAs to testosterone, progesterone and corticosterone. However, it showed different dose-response profiles for testosterone, progesterone, and corticosterone in both yeast and mammalian cells. In addition, although Pg10, Pg10-1 and Pg10-16 have similar RBAs to corticosterone, only Pg10-1 and Pg10-16 showed sub-millimolar corticosterone activity in yeast cells.

The lack of full correlation between ligand binding and transactivation activity is not surprising. As a multi-domain, multifunctional protein, the transcriptional activity of hER depends not only on ligand binding, but also on dimerization, DNA binding, and more importantly, cofactor (coactivator or corepressor) recruitment. Consequently, the results from cell-based transactivation assay may likely be different from those from *in vitro* ligand binding assays. For example, although the *in vitro* competitive hormone binding assay suggested some kind of weak ligand interaction between corticosterone and the wild-type hER α LBD protein (Table 1), the cell-based transactivation assays did not show any corticosterone activity. This could be due to the extra large size of the hER α ligand binding pocket (450 \AA^3 determined from the crystal structure *versus* 245 \AA^3 of E₂).²⁵ It is possible that a compound with a scaffold similar to E₂ may fit into the ligand binding pocket and displace E₂ at sufficiently high concentrations ($\sim 10^6$ -fold higher than E₂) in the *in vitro* hormone binding assays, even though the ligand cannot activate the hER α .

Molecular basis for the creation of novel corticosterone activity

To identify the molecular basis for the creation of this novel ligand activity, we sequenced all five evolved variants and found seven non-synonymous mutations (Table 2). Mutation E353Q is located in the ligand binding pocket and alters the

Table 1. The ligand binding affinities of the wild-type and mutant hER α LBDs to E₂, testosterone (T), progesterone (Pg), and corticosterone (Cs)

Estrogen receptor	$K_d^{E_2}$ (nM) ^a	RBA ^b			K_d (nM) ^c		
		T	Pg	Cs	T	Pg	Cs
Wild-type	0.21 ± 0.12 (3)	$< 10^{-4}$ (2)	$< 10^{-4}$ (2)	$< 10^{-4}$ (2)			
T17	1.01 ± 0.44 (2)	0.52 ± 0.18 (2)	0.016 ± 0.014 (2)	$< 10^{-4}$ (2)	193	6334	
T17-2	0.31 ± 0.13 (2)	0.97 ± 0.20 (3)	0.017 ± 0.005 (3)	$< 10^{-3}$ (2)	32	1801	
Pg10	1.28 ± 0.09 (2)	0.15 ± 0.08 (2)	0.674 ± 0.063 (2)	0.017 ± 0.002 (2)	832	191	7612
Pg10-1	0.81 ± 0.23 (3)	0.21 ± 0.06 (3)	0.679 ± 0.021 (3)	0.008 ± 0.001 (3)	388	119	10,288
Pg10-16	2.95 ± 0.27 (2)	0.10 ± 0.03 (3)	2.184 ± 1.3 (3)	0.015 ± 0.009 (3)	2872	135	20,227

^a $K_d^{E_2}$ values were determined by Scatchard analysis from multiple independent experiments ($n=2$ or 3), and the error bounds represent the range ($n=2$) or S.E. ($n>2$).

^b RBA values were determined with 2 nM [³H]E₂ for wild-type and all mutants. $RBA = EC_{50}^{E_2}/EC_{50}^{ligand} \times 100$. Values represent the average of multiple independent determinations ($n=2$ or 3).

^c The binding affinity of testosterone, progesterone or corticosterone was calculated with $K_d^{ligand} = (K_d^{E_2}/RBA) \times 100$.

Table 2. DNA sequencing results of the selected evolved hER α LBD variants

hER α variants	Amino acid substitutions
T17	E353Q
T17-2	E353Q, G390D
Pg10	E353Q, G390D, H524N
Pg10-1	E353Q, G390D, H524N, L536H, T585S
Pg10-16	E353Q, G390D, H524N, M528L, L536P

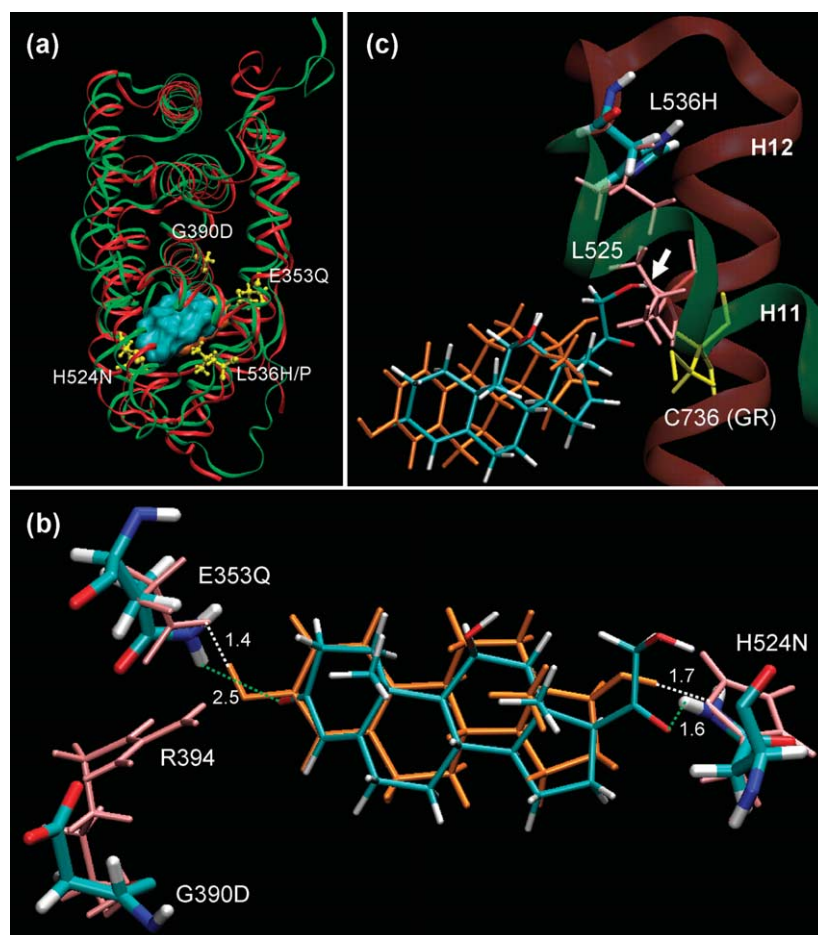
hydrogen-bonding pattern near the A-ring of the ligand between the receptor and the ligand (Figure 5). E353 (a hydrogen bond acceptor) pairs well with the 3-phenolic group of E₂ (a hydrogen bond donor), whereas Q353 (a hydrogen bond donor) pairs well with the 3-keto group of testosterone, progesterone, or corticosterone (a hydrogen bond acceptor). Thus, mutation E353Q may account for the emergence of ligand activity of the evolved hER α LBD variants toward 3-ketosteroids. Indeed, residue Q353 is conserved in AR, PR, GR and mineralocorticoid receptor (MR).

Mutation G390D is not within the ligand binding pocket. Molecular modeling suggests that this mutation may form a new electrostatic interaction with R394 to compensate the loss of the electrostatic

interaction formed between E353 and R394 in the wild-type hER α LBD (Figure 5)¹⁰ thus stabilizing the overall interactions between the receptor and the ligand.

Mutation H524N might abolish the hydrogen bond formed between the δ nitrogen atom of histidine and the 17 β -hydroxyl group of E₂, while establishing a new hydrogen bond between the 20-keto group of progesterone or corticosterone and the γ -amino group of asparagine (Figure 5). Indeed, Pg10 (E353Q,G390D,H524N) showed \sim tenfold higher sensitivity to progesterone, and \sim tenfold and \sim 50-fold lower sensitivity to E₂ and testosterone, respectively, than T17-2 (E353Q,G390D) (Figure 4) in yeast cells. *In vitro* hormone binding assay also suggested that Pg-10 showed increased $K_d^{E_2}$ and K_d^T , and decreased K_d^{Pg} (Table 1). Similarly, the single mutant H524N showed a slightly increased sensitivity to progesterone, and \sim fivefold decreased sensitivity to E₂ compared to the wild-type hER α LBD in yeast cells (data not shown).

None of the three selected variants from the first three rounds of directed evolution (i.e. T17, T17-2, and Pg10) showed any response to corticosterone (Figure 4); only the two fourth round variants, Pg10-1 and Pg10-16, responded to corticosterone at submillimolar concentrations in yeast cells. In



LBD (yellow) and mutation His536 (elemental color) are shown as sticks. The helices 11 (H11) and 12 (H12) of hER α are shown in red and green, respectively.

Figure 5. Molecule modeling studies. (a) Ribbon diagram of the superimposed three-dimensional structures of hER α LBD (red) complexed with E₂ (orange) and human GR LBD (green) complexed with docked corticosterone (cyan). The functional mutations except for T585S are shown in CPK model (yellow). T585S is not located in the LBD. (b) The structural model showing the E353Q, H524N and G390D mutations and Arg394 and the ligands (E₂, orange; corticosterone, elemental color). All mutated residues are shown in elemental colors while the original residues are in pink. The altered hydrogen bonds are highlighted in broken lines (yellow in the wild-type hER α LBD and E₂ complex, green in the mutant hER α LBD and corticosterone complex). The distances (Å) shown are from hydrogen bond acceptor atoms to hydrogen atoms. (c) The structural model showing the mutation L536H. Leu525 in the wild-type hER α LBD sterically clashes with the large substituent at the C17 α position of corticosterone (indicated by an arrow). Leu525 is 2.55 Å away from the closest atom in Leu536. Residues from hER α LBD (pink) and human GR

comparison with Pg10, both Pg10-1 and Pg10-16 contained two additional mutations, with one occurring at the same position (L536). Four quadruple mutants, Pg10+L536H, Pg10+L536P, Pg10+T585S, and Pg10+M528L, were created by site-directed mutagenesis and assayed for their trans-activation activity in yeast cells. All of them showed increased sensitivity to progesterone, but only the first three showed response to corticosterone.

None of these four mutations is within the ligand binding pocket. Residue L536 is located in the loop connecting helix 11 and helix 12, and is thought to be critical in coupling the binding of ligand to the modulation of the conformation and activity of the hER α .²⁶ Consistent with our previous findings,¹⁰ the functions of both L536H and L536P are context-dependent: quadruple mutants Pg10+L536H and Pg10+L536P showed no or negligible ligand-independent response, whereas single mutants L536H and L536P showed significantly elevated ligand-independent response in yeast cells.

Molecular modeling (Figure 5) suggests that L525 (located on helix 11) forms a van der Waals interaction with L536 and, unlike its corresponding residue in human GR (C736), L525 sterically clashes with the larger substituent at the C17 α position of corticosterone compared with the corresponding substituent in E₂, testosterone or progesterone. Thus, the substitution of L536 by a residue with a smaller side-chain (L536H or L536P) will likely shift the side-chain position of L525, resulting in a larger side pocket of hER α near the C17 atom of E₂ to accommodate the large substituent at the C17 α position of corticosterone. Mutation T585S is not located in the ligand binding domain, and its effect on ligand binding is unclear. Saturation mutagenesis at residue 536 or 585 on the Pg-10 background failed to produce variants with further increased corticosterone activity. However, a few amino acid substitutions, mostly smaller or polar residues, at the 536 position rendered quadruple mutants with low but distinct corticosterone activity.

Of note, none of the single mutants containing each of the seven mutations showed any response to corticosterone. In addition, using corticosterone as a selection ligand, we screened two hER α LBD libraries ($\sim 10^6$ variants per library) created by error-prone PCR with low and high mutagenesis rate (1.7 and 11 nucleotide substitutions per gene, respectively), and failed to identify any mutants responding to corticosterone. Furthermore, triple mutants (E353Q+G390D+L536H, E353Q+G390D+L536P, E353Q+G390D+T585S, and E353Q+G390D+M528L) did not show any corticosterone-dependent response in yeast cells. Together, we suggest that creation of the corticosterone activity in the wild-type hER α LBD using the yeast two-hybrid system-based screening method requires at least four simultaneous mutations and that, while we could not obtain these changes directly by a one-step directed evolution using corticosterone as the selection ligand, we could access them efficiently

by our progressive ligand-receptor coevolution strategy.

Implications in molecular evolution of nuclear receptors

Our results shed new light on molecular evolution of nuclear steroid receptors. Six evolutionarily related steroid receptors have been discovered, including estrogen receptors α and β (ER α and ER β), PR, AR, GR, and MR. Molecular phylogenetic analysis suggests that all steroid receptors have evolved from an ancestral estrogen receptor through a series of gene duplication and divergent evolution.²⁷ A ligand exploitation model was proposed as an evolutionary mechanism for the creation of a novel ligand-receptor pair. New hormones emerged when duplicated receptors evolved increased affinity for biochemical intermediates in a biosynthetic pathway.^{28,29}

Consistent with this model, our results indicate that a novel corticosterone activity could be readily created in the laboratory by coevolving the estrogen receptor and biochemical intermediates including testosterone and progesterone from the cholesterol biosynthetic pathway. However, it should be noted that, unlike the naturally occurring steroid receptors, the laboratory-evolved hER α variants are promiscuous receptors. The broadening of substrate/ligand specificity is a widely observed phenomenon in conventional directed evolution experiments to evolve new substrate/ligand specificity.³⁰ Such findings might not be surprising, given that only a positive selection was applied in most conventional directed evolution experiments as well as in our work. When both positive selection and negative selection are used, more specific receptor mutants might be created. For example, using yeast two-hybrid system-based positive and negative selection, we have successfully reduced the ligand sensitivity of hER α LBD towards E₂ by ~ 140 -fold while increasing its sensitivity towards 4,4'-dihydroxybenzil by ~ 50 -fold, resulting in an hER α LBD variant with 7000-fold ligand selectivity alteration compared to the wild-type hER α LBD.³¹ Thus, our laboratory evolution experiments suggest that both positive and negative selection may operate simultaneously in the natural evolution of human estrogen receptors. The ligand promiscuity may account for the creation of new steroid receptors by providing a head start towards being captured by adaptive evolution, with further mutations providing the required specificity toward the newly adopted ligand, as is the case in the natural evolution of novel enzyme activities.³²

In vitro coevolution as a novel protein engineering approach

The described *in vitro* coevolution approach has some key distinctions from rational design and conventional directed evolution approaches. Although structure-based computational design

allows a vast number of protein variants to be screened *in silico* ($>10^{14}$), the search for mutations is limited to a particular region, i.e. residues forming direct contacts with the substrate or ligand.^{33,34} As shown here and in many other studies,^{9–11} residues far away from the enzyme active site or ligand-binding pocket can exert their effects on protein functions through long-range interactions whose analysis is still beyond the capability of existing computational design approaches. Thus, rational design is not as well suited for engineering novel protein functions.

On the other hand, conventional directed evolution generally requires a screening or selection method that can detect the target function in the wild-type protein. As such, conventional directed evolution is typically involved in improving an existing function or a promiscuous function that was already present in the wild-type protein. In addition, the power of conventional directed evolution is limited by the number of sequences (library size) that can be screened experimentally (about 10^{14} for library panning and 10^7 for high throughput screening).³³ Thus, conventional directed evolution is best at hill-climbing in the sequence landscape (fine-tuning the protein function), and is not well suited for creating a completely novel function that may require multiple simultaneous mutations. However, by using *in vitro* coevolution, the target novel function may be divided into a few intermediate functions that are amenable to conventional directed evolution. Since single or double mutations will likely show beneficial effects in these intermediate functions, only a small library of protein variants (less than 10^4 – 10^5) need to be screened in each round of directed evolution. The accumulation of these beneficial mutations will eventually lead to the creation of the target novel function that requires multiple simultaneous mutations.

Take the hER α LBD for example; the number of possible hER α LBD variants containing four simultaneous mutations is 6.9×10^{13} , which is extremely difficult to be screened experimentally by any existing technology. In other words, the chance of identifying the four simultaneous mutations required for the corticosterone activity as described here is almost zero. Consequently, it is not surprising that screening of large libraries of hER α LBD variants ($>10^6$) directly for corticosterone activity failed to yield any positive mutants. However, by coevolving the receptor and the ligand, we were able to create the novel corticosterone activity in the hER α LBD. In total, we only screened approximately 10^6 variants in four rounds, and the number of identified positive variants ranged from two (rounds 1 and 3) to 30 (round 4) (Figure 3). Thus, the *in vitro* coevolution approach overcomes the library size limitation in conventional directed evolution and represents a new strategy for creating novel protein functions. It may also represent a general approach to engineering biomolecules and biosystems such as receptors, enzymes, antibodies,

ribosozymes, DNazymes, and viruses with novel functions.

Materials and Methods

Reagents and plasmids

All restriction enzymes and DNA modifying enzymes were obtained from New England BioLabs (Beverly, MA). Yeast strain YRG-2 (*Mat a ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3 112 gal4-542 gal80-538 LYS2::UAS_{GAL1}-TATA_{GAL1}-HIS3 URA3::UAS_{GAL4} 17mers(x3)-TATA_{CYCI}-lacZ*) was from Stratagene (La Jolla, CA). *Taq* DNA polymerase was from Promega (Madison, WI). QIAprep spin plasmid mini-prep kit, QIAEX II gel purification kit, and QIAquick PCR purification kit were purchased from Qiagen (Valencia, CA). Various oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). Unless otherwise specified, all chemicals were obtained from Sigma (St. Louis, MO). Plasmid pBD-Gal4 hER α containing hER α amino acid residues 312–595 fused to the Gal4 DNA-binding domain and plasmid pGAD424 SRC-1 containing the full-length coactivator SRC-1 fused to the Gal4 activation domain were constructed as described.¹⁰ Plasmid pBD-Gal4-Cam was obtained from Stratagene (San Diego, CA). The cloning of hER α LBD mutant constructs into the mammalian expression vector pCMV5 and *Escherichia coli* expression vector pET15b were as described.¹⁰

Library construction and screening

Library construction and screening have been described.¹⁰ Briefly, a two-tiered strategy consisting of an agar plate-based selection followed by a 96-well plate-based screening was used. In the selection method, the mutagenized LBD fragments containing hER α amino acid residues 312–595 were co-transformed with the EcoRI-SalI-digested pBD-Gal4-Cam hER α vector into *Saccharomyces cerevisiae* YRG2 cells harboring pGAD424 SRC-1. The transformed cells were plated on an agar plate containing minimum medium lacking tryptophan, leucine and histidine, and supplemented with an appropriate ligand concentration. The ligand concentration was chosen such that yeast cells bearing the parental hER α LBD in each round of directed evolution cannot form colonies, whereas yeast cells bearing a variant with five- to ten-fold improvement in ligand binding affinity may form colonies. To eliminate the false positives caused by mutations resulting in ligand-independent responses, the colonies that appeared on agar plates after incubation at 30 °C for two days were picked with toothpicks and grown in minimal medium lacking tryptophan and leucine in the 96-well plate. The overnight cultures were then transferred into two replica 96-well plates containing the minimum medium lacking tryptophan, leucine and histidine and supplemented with the same ligand concentration as during the selection or ethanol (EtOH). Variants with ligand-dependent growth responses were selected and assayed using the ligand dose-response assay. The first and second libraries were screened on 5×10^{-7} M and 10^{-8} M testosterone, respectively, whereas the third and fourth libraries of variants were screened on 5×10^{-8} M and 5×10^{-9} M progesterone, respectively.

Mutagenic PCR

Mutagenic PCR was performed as described.¹⁰ The average mutagenic rate was determined by DNA sequencing of nine randomly picked clones from the mutant library.

Site-directed mutagenesis

The single, triple and quadruple site-directed mutants were created using overlap extension PCR and yeast *in vivo* recombination.¹⁰ Plasmids of the different site-directed mutants were rescued from yeast cells, transferred into *E. coli*, and sequenced to confirm the presence of the introduced specific mutations and the absence of PCR-associated random mutations.

Ligand dose-response assay

A yeast two-hybrid-based cell growth assay was used to quantify the ligand activity of the wild-type and mutant hER α LBD in 96-well plates.¹⁰ Briefly, yeast cells harboring the plasmid containing the target hER α LBD and plasmid pGAD424 SRC-1 were grown to saturation (A_{600} 4–5) in 2–3 ml of minimal medium lacking tryptophan and leucine, and then diluted to A_{600} 0.002 using minimal medium lacking tryptophan, leucine and histidine. Each well contained 200 μ l of diluted yeast cells and 0.2 μ l of specified ligand dissolved in 100% ethanol (E_2 , testosterone, and progesterone) or DMSO (corticosterone). The 96-well plates were incubated at 30 °C for 24 hours and the cell density was measured at 600 nm using a SpectraMax plate reader (Molecular Devices, Sunnyvale, CA).

Molecular modeling

Ligand corticosterone, generated using the Builder function of MOE (Molecular Operating Environment, Chemical Computing Group Inc., Montreal, Quebec, Canada) and energy minimized under the MMFF94s forcefield, was docked into the ligand binding pocket of human GR LBD (PDB code 1M2Z) using the MOE Dock function. The lowest-energy docked conformation was further energy minimized. The resulting three-dimensional structure of human GR LBD complexed with corticosterone was structurally aligned with the crystal structure of hER α LBD complexed with E_2 (PDB code 1GWR) and imported into Visual Molecular Dynamics (VMD).³⁵ Residues Glu353, Gly390, His524, and Leu536 were mutated to Gln, Asp, Asn, and His, respectively, using the MOE Rotamer Explorer and the appropriate conformations were manually selected.

Mammalian cell transfection and luciferase assay

Mammalian transfection and luciferase assay were performed as described.¹⁰

Hormone binding assay

Wild-type and mutant hER α LBD (312–595) were cloned into pET15b vector and transformed into *E. coli* BL21(DE3) cells. For protein expression, cells were grown in LB medium supplemented with 10% (w/v) sucrose³⁶ at 37 °C until A_{600} =0.6 and induced with 0.5 mM isopropyl- β -D-thiogalactopyranoside at 25 °C for eight hours. Cells were harvested by centrifugation and

resuspended in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0), and lysed by a French press according to the standard protocol suggested by the manufacturer. To separate the cell debris from soluble proteins, lysed cells were centrifuged for 30 minutes at 25,000g. The resulting supernatants (10 ml) were mixed with 0.25 ml of Ni-NTA agarose (QIAGEN) and incubated at 4 °C for two hours with constant mixing, and proteins were purified according to the QIAGEN standard protocol to near homogeneity, and used in all hormone binding assays.

The competitive hormone binding assay and Scatchard analysis were performed as described.¹⁰ Briefly, for RBA measurement, purified wild-type or mutant hER α LBD was diluted to 2 nM in buffer B (50 mM Tris (pH 8.0), 10% (v/v) glycerol, 10 mM β -mercaptoethanol) and incubated with 2 nM [³H] E_2 together with various concentrations of the unlabeled competitor at 0 °C for 18–24 hours. The protein was stable under these conditions. For Scatchard analysis, the protein was diluted to 0.1–0.3 nM in buffer B and incubated with various concentrations of [³H] E_2 (Amersham, Piscataway, NJ) in the absence or presence of a 100-fold excess of the unlabeled ligand for 12–14 hours at 0 °C. Aliquots of the incubation solution were used to determine the total concentration of [³H] E_2 in the sample. The incubation solutions were then assayed by adsorption onto hydroxyapatite (BioRad, Hercules, CA).

Acknowledgements

We thank J. A. Katzenellenbogen and J. A. Gerlt for valuable comments on the manuscript. This work was supported by a National Science Foundation CAREER Award (BES-0348107).

References

1. Brannigan, J. A. & Wilkinson, A. J. (2002). Protein engineering 20 years on. *Nature Rev. Mol. Cell. Biol.* **3**, 964–970.
2. Penning, T. M. & Jez, J. M. (2001). Enzyme redesign. *Chem. Rev.* **101**, 3027–3046.
3. Arnold, F. H. (2001). Combinatorial and computational challenges for biocatalyst design. *Nature*, **409**, 253–257.
4. Arnold, F. H. (1998). Design by directed evolution. *Accts. Chem. Res.* **31**, 125–131.
5. Schmidt-Dannert, C. (2001). Directed evolution of single proteins, metabolic pathways, and viruses. *Biochemistry*, **40**, 13125–13136.
6. Lo Surdo, P., Walsh, M. A. & Sollazzo, M. (2004). A novel ADP- and zinc-binding fold from function-directed in vitro evolution. *Nature Struct. Mol. Biol.* **11**, 382–383.
7. Bolon, D. N., Voigt, C. A. & Mayo, S. L. (2002). *De novo* design of biocatalysts. *Curr. Opin. Chem. Biol.* **6**, 125–129.
8. Shao, Z. & Arnold, F. H. (1996). Engineering new functions and altering existing functions. *Curr. Opin. Struct. Biol.* **6**, 513–518.
9. Yano, T., Oue, S. & Kagamiyama, H. (1998). Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl Acad. Sci. USA*, **95**, 5511–5515.
10. Chen, Z., Katzenellenbogen, B. S., Katzenellenbogen,

- J. A. & Zhao, H. (2004). Directed evolution of human estrogen receptor variants with significantly enhanced androgen specificity and affinity. *J. Biol. Chem.* **279**, 33855–33864.
11. Nettles, K. W., Sun, J., Radek, J. T., Sheng, S., Rodriguez, A. L., Katzenellenbogen, J. A. *et al.* (2004). Allosteric control of ligand selectivity between estrogen receptors alpha and beta. Implications for other nuclear receptors. *Mol. Cell*, **13**, 317–327.
 12. Orenca, M. C., Hanson, M. A. & Stevens, R. C. (2000). Structural analysis of affinity matured antibodies and laboratory-evolved enzymes. *Advan. Protein Chem.* **55**, 227–259.
 13. Moyle, W. R., Campbell, R. K., Myers, R. V., Bernard, M. P., Han, Y. & Wang, X. (1994). Co-evolution of ligand-receptor pairs. *Nature*, **368**, 251–255.
 14. Athanikar, J. N. & Osborne, T. F. (1998). Specificity in cholesterol regulation of gene expression by coevolution of sterol regulatory DNA element and its binding protein. *Proc. Natl Acad. Sci. USA*, **95**, 4935–4940.
 15. Schmidt, T. R., Wu, W., Goodman, M. & Grossman, L. I. (2001). Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome *c* oxidase. *Mol. Biol. Evol.* **18**, 563–569.
 16. Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. (2002). Evolutionary rate in the protein interaction network. *Science*, **296**, 750–752.
 17. Salazar, J. C., Ahel, I., Orellana, O., Tumbula-Hansen, D., Krieger, R., Daniels, L. & Soll, D. (2003). Coevolution of an aminoacyl-tRNA synthetase with its tRNA substrates. *Proc. Natl Acad. Sci. USA*, **100**, 13863–13868.
 18. Katzenellenbogen, J. A. & Katzenellenbogen, B. S. (1996). Nuclear hormone receptors: ligand-activated regulators of transcription and diverse cell responses. *Chem. Biol.* **3**, 529–536.
 19. Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K. *et al.* (1995). The nuclear receptor superfamily: the second decade. *Cell*, **83**, 835–839.
 20. Tenbaum, S. & Baniahmad, A. (1997). Nuclear receptors: Structure, function and involvement in disease. *Int. J. Biochem. Cell Biol.* **29**, 1325–1341.
 21. Nilsson, S., Makela, S., Treuter, E., Tujague, M., Thomsen, J., Andersson, G. *et al.* (2001). Mechanisms of estrogen action. *Physiol. Rev.* **81**, 1535–1565.
 22. Kuiper, G. G., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S. & Gustafsson, J. A. (1997). Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology*, **138**, 863–870.
 23. Ekena, K., Katzenellenbogen, J. A. & Katzenellenbogen, B. S. (1998). Determinants of ligand specificity of estrogen receptor-alpha: estrogen *versus* androgen discrimination. *J. Biol. Chem.* **273**, 693–699.
 24. Zhao, H., Moore, J. C., Volkov, A. A. & Arnold, F. H. (1999). Methods for optimizing industrial enzymes by directed evolution. In *Manual of Industrial Microbiology and Biotechnology* (Demain, A. L. & Davies, J. E., eds), 2nd edit., pp. 597–604, ASM Press, Washington, DC.
 25. Anstead, G. M., Carlson, K. E. & Katzenellenbogen, J. A. (1997). The estradiol pharmacophore: ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site. *Steroids*, **62**, 268–303.
 26. Zhao, C., Koide, A., Abrams, J., Deighton-Collins, S., Martinez, A., Schwartz, J. A. *et al.* (2003). Mutation of Leu-536 in human estrogen receptor-alpha alters the coupling between ligand binding, transcription activation, and receptor conformation. *J. Biol. Chem.* **278**, 27278–27286.
 27. Laudet, V. (1997). Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.* **19**, 207–226.
 28. Thornton, J. W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc. Natl Acad. Sci. USA*, **98**, 5671–5676.
 29. Thornton, J. W., Need, E. & Crews, D. (2003). Resurrecting the ancestral steroid receptor: Ancient origin of estrogen signaling. *Science*, **301**, 1714–1717.
 30. Matsumura, I. & Ellington, A. D. (2001). *In vitro* evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J. Mol. Biol.* **305**, 331–339.
 31. Chockalingam, K., Chen, Z., Katzenellenbogen, J. A. & Zhao, H. (2005). Directed evolution of specific receptor-ligand pairs for use in the creation of gene switches. *Proc. Natl. Acad. Sci. USA*. In press.
 32. O'Brien, P. J. & Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105.
 33. Hayes, R. J., Bentzien, J., Ary, M. L., Hwang, M. Y., Jacinto, J. M., Vielmetter, J. *et al.* (2002). Combining computational and experimental screening for rapid optimization of protein properties. *Proc. Natl Acad. Sci. USA*, **99**, 15926–15931.
 34. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003). Computational design of receptor and sensor proteins with novel functions. *Nature*, **423**, 185–190.
 35. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.
 36. Eiler, S., Gangloff, M., Duclaud, S., Moras, D. & Ruff, M. (2001). Overexpression, purification, and crystal structure of native ER alpha LBD. *Protein Exp. Purif.* **22**, 165–173.

Edited by F. Schmid

(Received 3 October 2004; received in revised form 11 February 2005; accepted 16 February 2005)