Mass Spectrometric Characterization of the Human Androgen Receptor Ligand-Binding Domain Expressed in *Escherichia coli*†

Zixin Zhu,‡ Robert R. Becklin,§,|| Dominic M. Desiderio,§,||,⊥ and James T. Dalton*,‡,#

Department of Pharmaceutical Sciences, The Charles B. Stout Neuroscience Mass Spectrometry Laboratory, Department of Neurology, and Department of Biochemistry, University of Tennessee Health Science Center, Memphis, Tennessee 38163

Received March 13, 2001; Revised Manuscript Received June 21, 2001

ABSTRACT: The ligand-binding domain (LBD) of the human androgen receptor (hAR LBD), encompassing amino acids (AAs) 647−919, was expressed in *Escherichia coli* with an N-terminal polyhistidine tag (His_{10}-hAR LBD) from a pET-16b vector. The overexpressed protein was initially insoluble in inclusion bodies, and was subsequently solubilized in 8 M guanidine hydrochloride (GdnHCl). The solubilized His_{10}-hAR LBD was purified to apparent homogeneity by metal ion affinity chromatography in the presence of 6 M GdnHCl. The isolated protein migrated as a single band in sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) with an apparent molecular mass of 33−34 kDa, as expected from the plasmid construct. Immunoblot analysis with C-terminal antibodies raised against a peptide corresponding to the last 19 AAs (AAs 901−919) of hAR revealed that the purified protein contained an immunoreactive epitope present within the AR and was of the appropriate size. Further characterization, using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI/TOF-MS), showed a single protein species of average mass 34 580 Da, confirming the size and purity of the purified His_{10}-hAR LBD. Detailed tryptic peptide mapping analysis, using MALDI/TOF-MS, identified a total of eight peptides with a 30% coverage of the LBD, including the last tryptic peptide in the hAR sequence. These data confirm that the purified protein was the intact hAR LBD. AA sequencing of these tryptic peptides, using an HPLC-coupled electrospray ionization ion trap mass spectrometer (LC/ESI-ITMS and MS/MS), unambiguously confirmed that the peptides were from the hAR LBD. The purified His_{10}-hAR LBD in 6 M GdnHCl could be renatured as determined by ligand-binding activity, with a similar equilibrium dissociation constant (K_d) for [^{3}H]-mibolerone and a similar steroid specificity to the AR isolated from rat ventral prostate.

The androgen receptor (AR)[1] is a ligand-dependent transcriptional regulator that belongs to the superfamily of nuclear receptors (*1*, *2*), which includes steroid receptors, thyroid receptor, vitamin D receptor, and retinoic acid receptor, as well as orphan receptors for which the ligands are not known. Glucocorticoid receptor (GR), progesterone receptor (PR), mineralocorticoid receptor (MR), estrogen receptor (ER), and the AR belong to the steroid receptor subfamily. Structural analysis of the amino acid (AA) sequence of the human androgen receptor (hAR) revealed that, like other steroid receptors, the AR is functionally organized into four domains: an N-terminal transactivation domain, a central DNA-binding domain, a hinge region, and a C-terminal ligand-binding domain (LBD) (*2*−*6*). The LBD of steroid receptors is responsible for ligand binding, heat shock protein binding, and receptor dimerization (*7*−*9*). This domain also displays an inhibitory function for transcriptional activation in the absence of hormone (*6*, *10*, *11*). Deletion of the LBD constitutively activates gene transcription (*6*). The LBD of steroid receptors is large (about 250 AAs), compared to the small size of the ligands. Therefore, only a very small number of AAs within this domain form a hydrophobic binding pocket for the hydrophobic natural steroids or synthetic ligands.

Techniques for characterizing the interactions between receptors and their ligands include X-ray crystallography, affinity labeling, NMR, mass spectrometry, and site-directed mutagenesis. Crystallographic analysis is the most comprehensive method, and can provide details of the interaction of the ligand with each AA in the binding cavity. However,

† This work was supported by grants from the National Institutes of Health (1 R29 CA68096 to J.T.D. and RR 10522 to D.M.D.), the St. Francis of Assisi Foundation of Memphis (to J.T.D.), and the National Science Foundation (DBI 9604633 to D.M.D.).

* Correspondence should be addressed to this author.

‡ Department of Pharmaceutical Sciences.

§ The Charles B. Stout Neuroscience Mass Spectrometry Laboratory.

|| Department of Biochemistry.

⊥ Department of Neurology.

Present address: College of Pharmacy, Division of Pharmaceutics, The Ohio State University, 242 L.M. Parks Hall, 500 W. 12th Ave., Columbus, OH 43210. Tel: +1-614-688-3797; fax: +1-614-292-7766; e-mail: dalton.1@osu.edu.

[1] Abbreviations: AR, androgen receptor; hER, human estrogen receptor; LBD, ligand-binding domain; AA, amino acid; PCR, polymerase chain reaction; IPTG, isopropyl-β-thiogalactopyranoside; GdnHCl, guanidine hydrochloride; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; CHCA, α-cyano-4-hydroxy-cinnamic acid; SA, sinapinic acid; TFA, trifluoroacetic acid; MS, mass spectrometry; HPLC, high-performance liquid chromatography; MALDI/TOF-MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; LC/ESI-ITMS and MS/MS, HPLC-coupled electrospray ionization ion trap mass spectrometry; DTT, dithiothreitol; MIB, mibolerone; DHT, 5α-dihydrotestosterone; K_d, equilibrium dissociation constant; B_{max}, binding capacity.

10.1021/bi010525j CCC: $20.00
Published on Web 08/14/2001

crystallization studies require a large amount and a homogeneous preparation of biologically active receptor or receptor LBD. Such a preparation can be achieved by heterologous expression followed by purification. Although the LBD of many steroid receptors has been expressed efficiently in a number of systems (*12*−*14*), the hydrophobic character of the LBD has limited the solubility in aqueous systems. However, the LBDs of ER and PR overexpressed in *Escherichia coli* (*E. coli*) have been used successfully for crystallography studies, using a small amount either of soluble LBD or of LBD solubilized from inclusion bodies (*15*−*18*). Overexpression of the LBD of the hAR (hAR LBD) in *E. coli* for an X-ray crystallography study was recently reported (*19*). The 3-keto oxygen of methyltrienolone (R1881), an AR agonist, is hydrogen bonded to Arg^{752}, whereas the 17β-hydroxyl is hydrogen bonded to Asn^{705} and Thr^{877} (*19*). Correspondingly, the AR in the human prostate cancer cell line LNCaP has a $Thr^{877}Ala$ mutation, and can be activated by antagonists. However, crystallography structures of the hAR LBD complexed to other agonists and antagonists await exploration.

Microheterogeneity of the "purified" preparation of the expressed protein can interfere with structural studies by X-ray and NMR analysis (*20*−*23*). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI/TOF-MS) and HPLC-coupled electrospray ionization ion trap mass spectrometry (LC/ESI-ITMS and MS/MS) allow us to characterize proteins for their identity and purity (*20*−*23*), and to provide detailed structural information such as microheterogeneity (*20*−*23*), phosphorylation, and glycosylation.

In this report, we describe the development of a high-yield expression system for the production of an N-terminal decahistidine-tagged hAR LBD (His_{10}-hAR LBD) in *E. coli*, as well as its solubilization and purification, using metal ion affinity chromatography in the presence of 6 M guanidine hydrochloride (GdnHCl). We also characterized the purified His_{10}-hAR LBD and some of its tryptic peptides by MALDI/TOF-MS. Furthermore, LC/ESI-ITMS and MS/MS analysis unambiguously confirmed the AA sequences of the tryptic peptides, which included the extreme C-terminal tryptic peptide in the hAR sequence, with the latter data indicating that the purified His_{10}-hAR LBD was intact at the C-terminus. The renatured His_{10}-hAR LBD maintained a similar binding affinity for [^{3}H]-mibolerone ([^{3}H]-MIB) and steroid specificities as the AR isolated from rat ventral prostate. However, the renaturation efficiency was very low. In summary, the expression, purification, and renaturation system reported herein should provide a homogeneous and functional preparation of the hAR LBD for X-ray crystallography and NMR studies, provided that the renatured LBD is further separated from the unfolded form, using other techniques.

EXPERIMENTAL PROCEDURES

Materials. [17α-*methyl*-^{3}H]-Mibolerone ([^{3}H]-MIB, 83.5 Ci/mmol) and unlabeled MIB were purchased from Dupont Research NEN Products, Boston, MA. Tris base, sucrose, sodium molybdate, dithiothreitol (DTT), silver nitrate, sodium thiosulfate, sodium carbonate, and α-cyano-4-hydroxy-cinnamic acid (CHCA) were purchased from Sigma Chemical Co., St. Louis, MO. Guanidine hydrochloride and 3,5-dimethoxy-4-hydroxy-cinnamic acid (sinapinic acid, SA) were purchased from Aldrich Chemical, Milwaukee, WI. Hydroxyapatite (HAP), sodium dodecyl sulfate (SDS), ammonium persulfate, TEMED, glycine, 30% acrylamide mix, nitrocellulose membranes, Tween 20, and bromophenol blue were purchased from Bio-Rad Laboratories, Hercules, CA. EcoLite(+) scintillation cocktail was purchased from ICN Research Products Division, Costa Mesa, CA. Rabbit polyclonal anti-AR antibody AR (C-19) was purchased from Santa Cruz Biotechnology Inc., Santa Cruz, CA. Secondary donkey−anti-rabbit horseradish peroxidase-conjugated antibody and the substrates were purchased from Amersham Pharmacia Biotech Inc., Piscataway, NJ. Talon metal affinity resin was purchased from Clontech Laboratories, Inc., Palo Alto, CA. $ZipTip_{C18}$ were purchased from Millipore, Bedford, MA. The pET-16b vector and BL21 *E. coli* strain for protein expression were purchased from Novagen Inc., Madison, WI. DH5α-competent cells were purchased from GIBCO BRL, Gaithersburg, MD. Sequencing-grade modified trypsin was purchased from Promega, Madison, WI. All materials were used as received from the manufacturers.

Buffers. The renaturation binding buffer for solubilized/purified His_{10}-hAR LBD in 6 M GdnHCl was routine binding buffer (10 mM Tris, pH 7.4, 0.25 M sucrose, and 10 mM sodium molybdate) plus 5 mM DTT. The wash buffer was 50 mM Tris (pH 7.2). Use of routine binding buffer without DTT or with 1 mM DTT did not yield satisfactory renaturation, whereas the inclusion of 5 mM DTT in the renaturation binding buffer was found to be critical for renaturation, suggesting that this concentration was optimal for inhibition of disulfide bridge formation during the renaturation step.

Cell lysis buffer consisted of 50 mM Tris (pH 7.4), 50 mM NaCl, 1 mM EDTA, and 1 mM DTT (with 1, 4, or 6 M urea). The His_{10}-hAR LBD purification wash buffer consisted of 40 mM Tris (pH 7.5), 6 M GdnHCl, and 0.3 M NaCl. The elution buffer was wash buffer with 150 mM imidazole.

Construction of hAR LBD Expression Vector. The 5′-end primer (AACA*GGATCC*G^{1939}AGCACCACCAGCCCCACTGAGG^{1960}) and 3′-end primer (AGCG*GGATCC*^{2760}TCACTGGGTGTGGAAATAGATGGG^{2737}) with *Bam*HI sites were used to amplify the LBD DNA sequence (corresponding to AAs 647−919) from the mammalian hAR expression vector p9, using the polymerase chain reaction (PCR). The 0.8 kb PCR product was subcloned into the *Bam*HI site of pET-16b (a bacterial expression vector that expresses the cloned gene with an N-terminal decahistidine tag and a factor Xa cleavage site). DNA sequencing was used to confirm the in-frame ligation. The final construct was named pET-16b-hAR LBD.

Expression and Purification of His_{10}-hAR LBD. The pET-16b-hAR LBD was transformed into the *E. coli* strain BL21 for protein expression. Conditions suggested in the manual provided by the makers of the pET-16b plasmid (Novagen) for cell protein induction were followed precisely. Cells were grown in selective LB (Luria Bertani) medium, with a 50 μg/mL final concentration of ampicillin, to an absorbance of 0.6 (at 600 nm), and induced with 1 mM isopropyl-β-thiogalactopyranoside (IPTG). Initial studies were performed to determine the optimal time and temperature for induction. Cells were induced for 1, 2, 3, 4, and 6 h or overnight at 25, 30, or 37 °C. Cells were pelleted by centrifugation. SDS

sample buffer was added to the pellet, and the mixture was boiled. After centrifugation, supernatant proteins were separated by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) followed by Coomassie blue staining.

For subcellular localization studies, cells were lysed in routine binding buffer by sonication. After centrifugation, the supernatant and the pellet fraction were subjected to SDS−PAGE followed by Coomassie blue staining.

For subsequent protein purification experiments, cells were induced with IPTG for 6 h at 25 °C. After centrifugation, cell pellets were resuspended in lysis buffer that contained 6 M urea, and lysed by sonication (4 × 30 s) on ice. After centrifugation at 12000*g* for 10 min, the supernatant was discarded; the pellet was washed with distilled water, and solubilized with 8 M GdnHCl. Talon metal affinity resin (2 mL slurry) was equilibrated with 10 mL of purification wash buffer by gentle agitation, and centrifuged at 700*g* for 2 min to pellet the resin. The solubilized sample was centrifuged; the supernatant was diluted to a final concentration of 6 M GdnHCl, and incubated with the preequilibrated resin for 30 min at room temperature. The resin was centrifuged, and the pellet was washed 3 times with 10 mL of purification wash buffer. The resin was transferred to a column, and the adsorbed proteins were eluted with 20 mL of elution buffer. Fractions of 1.4 mL were collected. An aliquot (0.5 mL) of the second fraction was dialyzed overnight against lysis buffer that contained 8 M urea, using a 10k Slide-A-Lyzer dialysis cassette (Pierce, Rockford, IL). An aliquot from the dialyzed sample was analyzed by SDS−PAGE followed by Coomassie blue staining or immunoblotting. The protein content of the samples was measured by the Bradford method (*24*).

ZipTip$_{C18}$ Purification and MALDI/TOF-MS Analysis of the His$_{10}$-hAR LBD. To assay the His$_{10}$-hAR LBD protein by MALDI/TOF-MS, an aliquot (3.3 μL) of the purified His$_{10}$-hAR LBD in 6 M GdnHCl was added to 2 μL of 2.5% trifluoroacetic acid (TFA) and 4.7 μL of water (final volume 10 μL, 0.5% TFA, 2 M GdnHCl). A reverse-phase ZipTip$_{C18}$ microcolumn (a tip coated with spherical silica-based C18 resin for peptide concentration, desalting, and fractionation) was preequilibrated with 50% acetonitrile and was washed 3 times with 0.1% TFA. The diluted sample in 2 M GdnHCl was drawn into the tip to allow protein binding, and was washed 4 times with 10 μL of 0.1% TFA to remove any contaminants that might interfere with matrix−protein cocrystallization and/or ionization. The protein was eluted with 3 μL of elution solution (75% acetonitrile, 0.1% TFA, and 10 mg/mL SA as matrix). Separately, a matrix solution that contained 30 mg/mL SA in 90% acetonitrile was prepared, and was used to prepare the base-coat solution (1 part of matrix solution:5 parts of acetonitrile:2 parts of distilled water). An aliquot (0.9 μL) of the base-coat solution was spotted and dried onto the MALDI sample plate. This matrix was overlayed with 1 μL of the eluate from the ZipTip$_{C18}$ for MALDI/TOF-MS analysis.

MALDI/TOF-MS analysis of the His$_{10}$-hAR LBD protein was performed on a Voyager-DE RP MALDI/TOF instrument (PerSeptive Biosystems, Framingham, MA) equipped with a nitrogen laser (λ = 337 nm, 3 ns pulse width), a single-stage reflector, and a timed ion selector. The spectra were obtained in the positive-ion linear mode with delayed extraction, using standard conditions. The acceleration voltage was 25 kV, and the delay time (the time between the laser pulse and the acceleration voltage pulse) was 750 ns. Bovine serum albumin (BSA), $[M + nH]^{n+} = 66\,431$ (average), was used as the external mass calibration standard.

Silver Staining, In-Gel Digestion, and Peptide Extraction. Protein gel electrophoresis was performed with either 12% or 15% polyacrylamide slab gels in the presence of 0.1% SDS in a Protean Minigel apparatus (Bio-Rad Laboratories, Hercules, CA) at a current of 56 mA per slab gel. Following electrophoresis, the proteins were Coomassie blue-stained, silver-stained (see below), or transferred to a nitrocellulose membrane for immunoblotting.

A modified silver-staining method (*25*) was used for subsequent in-gel trypsin digestion to ensure compatibility with mass spectrometric analysis. After electrophoresis, the gel was fixed with 50% methanol/5% acetic acid for 20 min, washed with 50% methanol for 10 min, and washed with distilled water for 10 min. The gel was incubated with 0.02% sodium thiosulfate for 1 min, followed by two 1-min washes with distilled water. The gel was submerged in 0.1% silver nitrate for 20 min, followed by two 1-min washes with distilled water. The gel was incubated with freshly prepared 2% sodium carbonate with 0.04% formalin until the desired intensity was achieved. The development was stopped by incubating the gel in 5% acetic acid for 10 min. The gel was washed with distilled water for 5 min to make it ready for band-excision and in-gel digestion.

The 34 kDa band from one lane in a silver-stained gel (ca. 100 pmol of His$_{10}$-hAR LBD) was excised, minced into 1 mm^3 pieces, and transferred to a siliconized 0.6 mL tube. The gel pieces were washed once with distilled water, dehydrated 3 times with 200 μL of acetonitrile, and dried under vacuum (Vacufuge, Eppendorf, Hamburg, Germany). The sample was rehydrated with 50 mM NH_4HCO_3, pH 8.0, and was digested with 0.5 μg of trypsin (=13 000 units/mg of protein) overnight at 37 °C. The supernatant was collected, and the peptides were extracted from the gel pieces once with 20 μL of 20 mM NH_4HCO_3, pH 8.0, and 3 times with 20 μL of 50% acetonitrile/5% formic acid, using sonication (20 min each step), according to a protocol described by Haynes et al. (*26*). Supernatants from each extraction were combined.

ZipTip$_{C18}$ Purification and Tryptic Peptide Mapping of the His$_{10}$-hAR LBD with MALDI/TOF-MS. For MALDI/TOF-MS analyses, 100 pmol of His$_{10}$-hAR LBD was used for in-gel digestion. The combined supernatant from above was vacuum-dried (Vacufuge, Eppendorf, Hamburg, Germany). The sample was reconstituted in 10 μL of 0.1% formic acid. A ZipTip$_{C18}$ microcolumn was preequilibrated with 50% acetonitrile, and was washed with 0.1% formic acid. The reconstituted sample was drawn into the tip to allow peptide binding, and was washed 3 times with 10 μL of 0.1% formic acid. The peptides were eluted with 2 μL of 75% acetonitrile that contained 10 mg/mL CHCA as matrix. Separately, a matrix solution that contained 10 mg/mL CHCA in 50% acetonitrile was prepared, which was used to prepare a base-coat solution (1 part matrix solution:3 parts acetonitrile:4 parts water). An aliquot (0.5 μL) of base-coat solution was spotted, and was dried on the MALDI sample plate. This matrix was overlayed with 1 μL of the eluate from the ZipTip$_{C18}$ for MALDI/TOF-MS analyses.

Mass spectrometric analyses of the tryptic digest were performed on the same Voyager-DE RP MALDI/TOF instrument as mentioned above, but the spectra were obtained in the positive-ion reflector mode with delayed extraction, using standard conditions. The acceleration voltage was 20 kV, and the delay time was 150 ns. The mass scale was calibrated with a mixture of des-Arg^1-Bradykinin, $[M + H]^+ = 904.5$ (monoisotopic), and angiotensin, $[M + H]^+ = 1296.7$ (monoisotopic), or with a mixture of angiotensin, $[M + H]^+ = 1296.7$ (monoisotopic), and bovine insulin, $[M + nH]^{n+} = 5734.6$ (average) (Sequazyme peptide mass standards kit, PerSeptive Biosystems).

Partial AA Sequencing of the His$_{10}$-hAR LBD with LC/ESI-ITMS and MS/MS. For LC/ESI-ITMS and MS/MS analyses, 100−300 pmol of His$_{10}$-hAR LBD was used for in-gel digestion. Following peptide extraction from the gel pieces, the combined supernatant was filtered through a Microcon YM-30 (MW cutoff 30 000, Millipore, Bedford, MA) to remove residual gel pieces. The filtrate was dried, and was reconstituted in 27 μL of 2% TFA. For peptide separation, 23 μL was injected onto a RP-HPLC microbore column.

A Bruker-HP Esquire-LC system (HPLC-coupled electrospray ionization ion trap mass spectrometer) was used for AA sequencing of the HPLC-separated tryptic peptides. Data acquisition was controlled by a Hewlett-Packard ChemStation, and data processing was done by Esquire-LC data analysis software. Collection of all MS and MS/MS data was performed on samples coupled through an LC/ESI-ITMS interface with a low-TFA reversed-phase microbore C_{18} mass spectrometry column (Vydac, 1.0 mm i.d. × 250 mm long, 5 μm particles, and 300 Å pore size; Hesperia, CA) operating at 50 μL/min. The HPLC eluent was directed to the ESI-ITMS probe for nebulization. Detection was set for the UV amide absorbance at 210 nm.

Binary gradient elution conditions were used for all analyses. Step gradients were generated for elution from 0 to 9% acetonitrile over 20 min, from 9 to 45% acetonitrile from 20 to 60 min, and from 45 to 81% acetonitrile from 60 to 65 min in the presence of 0.1% formic acid for all steps. The solvents and reagents used for HPLC, ESI-ITMS, and MS/MS analyses were HPLC-grade, and were used without further purification. MS/MS was set at the automatic mode, and the three most abundant precursor ions, with intensities above a threshold of 20 000 arbitrary units (corresponds to the mass spectrum abundance), were fragmented. Masses from the MS (precursor ion) and MS/MS (product ions) spectra were entered into the MS-Tag subprogram in the University of California San Francisco (UCSF) ProteinProspector program (http://prospector.ucsf.edu/), and were searched against the National Center for Biotechnology Information (NCBI)'s nonredundant (nr) protein database.

Ligand-Binding Assays. For ligand-binding assays of the purified and renatured His$_{10}$-hAR LBD in 6 M GdnHCl, a method described by Xie et al. (*27*) was used. Preparations in 6 M GdnHCl were first diluted 40-fold with purification wash buffer, which contained 6 M GdnHCl. An aliquot (5 μL) of this diluted sample was diluted 10-fold into renaturation binding buffer that contained 5 mM DTT but not 6 M GdnHCl, and was kept on ice for 15 min. (i) For saturation binding assays, the sample was further diluted 5-fold (i.e., final 2000-fold dilution of sample, 50-fold dilution of GdnHCl) with renaturation binding buffer that contained the indicated concentration of [^{3}H]-MIB in the presence (nonspecific binding) or absence (total binding) of a 1000-fold excess of unlabeled MIB at 4 °C for 18 h in a final volume of 250 μL. The specifically bound radioactivity was determined by the hydroxyapatite (HAP) method, as described previously (*28*). The nonspecific binding in each case was less than 10% of the total binding. The equilibrium dissociation constant (K_d) and binding capacity (B_{max}) were determined with WinNonlin software (Pharsight Corp., Mountain View, CA). (ii) For steroid-specificity assays, the sample was further diluted 5-fold (i.e., final 2000-fold dilution of sample, 50-fold dilution of GdnHCl) with the renaturation binding buffer that contained 2 nM [^{3}H]-MIB (total binding in the absence of competitors), 2 nM [^{3}H]-MIB, and 200 nM of other competing steroids (total binding in the presence of competitors), or 2 nM [^{3}H]-MIB and 1000 nM of unlabeled MIB (nonspecific binding). Specific binding was calculated as total binding minus nonspecific binding. The specific binding in the absence of competitors was set at 100% as control, and the specific binding in the presence of competitor was compared to it. A lower percentage of control indicated a higher AR affinity.

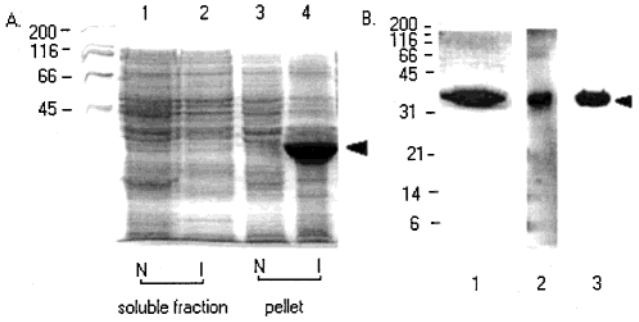

FIGURE 1: Purification of the His$_{10}$-hAR LBD from inclusion bodies. (A) Total cellular proteins from *E. coli* cells were analyzed by 12% SDS−PAGE followed by Coomassie blue staining. Lanes 1 (uninduced) and 2 (induced) were proteins from the soluble cytosol fraction. Lanes 3 (uninduced) and 4 (induced) were proteins from the insoluble pellet fraction. (B) The purified His$_{10}$-hAR LBD was analyzed by 15% SDS−PAGE followed by Coomassie blue staining (lane 1), silver staining (lane 2), or immunoblot with a C-terminal antibody, AR-C19 (lane 3). The His$_{10}$-hAR LBD band is indicated with an arrowhead.

RESULTS

Plasmid Construction, His$_{10}$-hAR LBD Expression in E. coli, and Purification. The His$_{10}$-hAR LBD expression vector pET-16b-hAR LBD was constructed and transformed into *E. coli* strain BL21 for protein expression. Upon IPTG induction for 6 h at 37 °C, a 33−34 kDa protein was overexpressed as expected from the construct. Initial studies to determine the time-course and optimal temperature for protein expression induction, using the total bacterial protein preparation, showed that IPTG induction at 25 °C for 6 h was the optimal condition for maximum protein expression (data not shown). However, under those optimal expression conditions, all of the expressed proteins were trapped in inclusion bodies and were present in the pellet upon cell lysis and fractionation; i.e., they were insoluble and inactive (Figure 1A), similar to the hER LBD expressed in *E. coli* (*29*). Initial attempts to solubilize His$_{10}$-hAR LBD from inclusion bodies with lysis buffer that contained 1 or 4 M

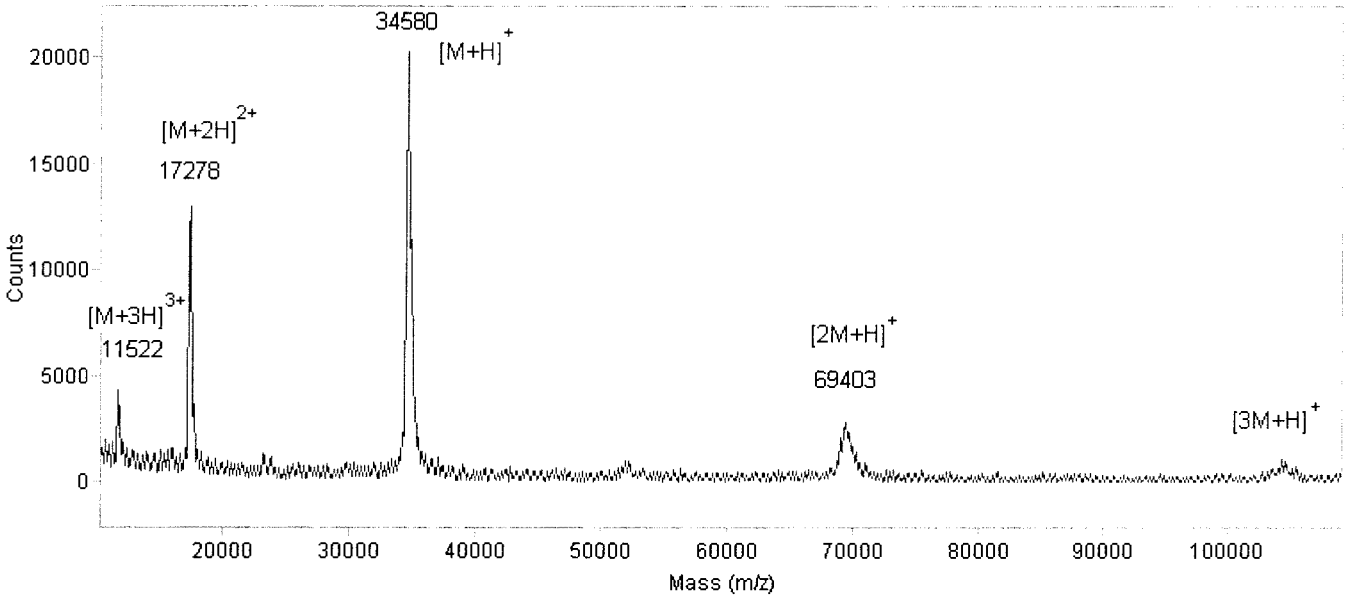

FIGURE 2: Analysis of the His_{10}-hAR LBD by MALDI/TOF-MS. His_{10}-hAR LBD purified in the presence of 6 M GdnHCl was diluted to 2 M GdnHCl with 0.5% TFA final concentration. A $ZipTip_{C18}$ microcolumn was used to adsorb the proteins, followed by several washes to remove salts and GdnHCl. The adsorbed His_{10}-hAR LBD was eluted, and was loaded onto a MALDI sample plate.

Table 1: The 12 Peptides Detected by MALDI/TOF-MS and/or Detected/Sequenced by LC/ESI-ITMS and MS/MS

peptide number	location of AAs	peptide sequence	calculated mass (Da)	$[M+H]^+$ observed in MALDI	retention time in LC/MS (min)	$[M+2H]^{2+}$ observed in LC/MS	$[M+H]^+$ observed in LC/MS
1	721−726	ALPGFR	660.38	660.5	6.6		660.4
2	753−760	SFTNVNSR	924.45	924.7	4.4	462.9	924.4
3	761−774	MLYFAPDLVFNEYR	1777.86	1778.3	62.5	889.7	
4	826−831	FFDELR	826.41	826.6	13.4	414.0	826.5
5	855−861	RFYQLTK	955.54	955.8	5.2	478.4	955.5
6	862−871	LLDSVQPIAR	1111.65	1111.9	7.9	556.5	1111.6
7	872−883	ELHQFTFDLLIK	1503.82	1504.2			
8	911−919	VKPIYFHTQ	1132.62	1132.9	5.7	567.0	1132.6
9	711−717	QLVHVVK	822.52		4.9	412.0	822.5
10	832−836	MNYIK	668.34		5.0		668.4
11	856−861	FYQLTK	799.44		5.6		799.5
12	906−910	ILSGK	517.34		4.6		517.3

urea so that the solubilized protein would be active for ligand binding without the need for renaturation were not successful; those results were different from similar studies on the hER LBD expressed in *E. coli* (*29*). Attempts with up to 8 M urea or 6 M GdnHCl still only solubilized a small fraction of the His_{10}-hAR LBD from the inclusion bodies. In addition, alternative strategies to decrease the induction temperature and/or IPTG concentration, aimed at increasing the soluble fraction of the expressed His_{10}-hAR LBD instead of solubilizing the pellet fraction, did not give satisfactory results. Therefore, we used 8 M GdnHCl to solubilize the His_{10}-hAR LBD protein from inclusion bodies, and to purify it in the presence of 6 M GdnHCl, using metal ion affinity chromatography.

Analysis of the Size and Purity of the Purified His_{10}-hAR LBD by SDS−PAGE. The purified preparation in the presence of 6 M GdnHCl was dialyzed against 8 M urea and was subjected to SDS−PAGE. Coomassie blue staining revealed that the preparation was pure with a single band at 33−34 kDa (Figure 1B, lane 1). No other protein bands were visible, even when the gel was purposely overloaded and the proteins detected with a modified silver-staining method that is compatible with in-gel trypsin digestion and MALDI/TOF-MS (Figure 1B, lane 2). Immunoblot analysis of this purified preparation revealed that the C-terminal antibody raised against a peptide that corresponds to AAs 901−919 (the last 19 AAs) of hAR recognized this 33−34 kDa protein band, indicating that the purified protein was hAR LBD and suggesting that the C-terminal part of the LBD was intact (Figure 1B, lane 3). A protein concentration assay showed a concentration of 0.9 μg/μL (or 26 pmol/μL) of His_{10}-hAR LBD in the purified preparation.

Analysis of the His_{10}-hAR LBD and Its Tryptic Peptides by MALDI/TOF-MS. MALDI/TOF-MS was used to determine the molecular weight and the purity of the His_{10}-hAR LBD purified in the presence of 6 M GdnHCl. Very strong signals for $[M + H]^+$, $[M + 2H]^{2+}$, and $[M + 3H]^{3+}$ protonated molecule ions, as well as the dimer and trimer ions, were detected; those data indicated an average molecular mass of 34 580 Da, as shown in Figure 2. The estimated mass was consistent with the SDS−PAGE data (33−34 kDa), and was only 0.5% different than the calculated mass from the construct (34 768.2 Da). No other significant masses were detected in this MS trace; therefore, the purity of the preparation is indicated.

Tryptic peptide mapping with MALDI/TOF-MS showed a total of eight peptides with a 30% coverage within the LBD (Table 1; peptide numbers are also used to describe those

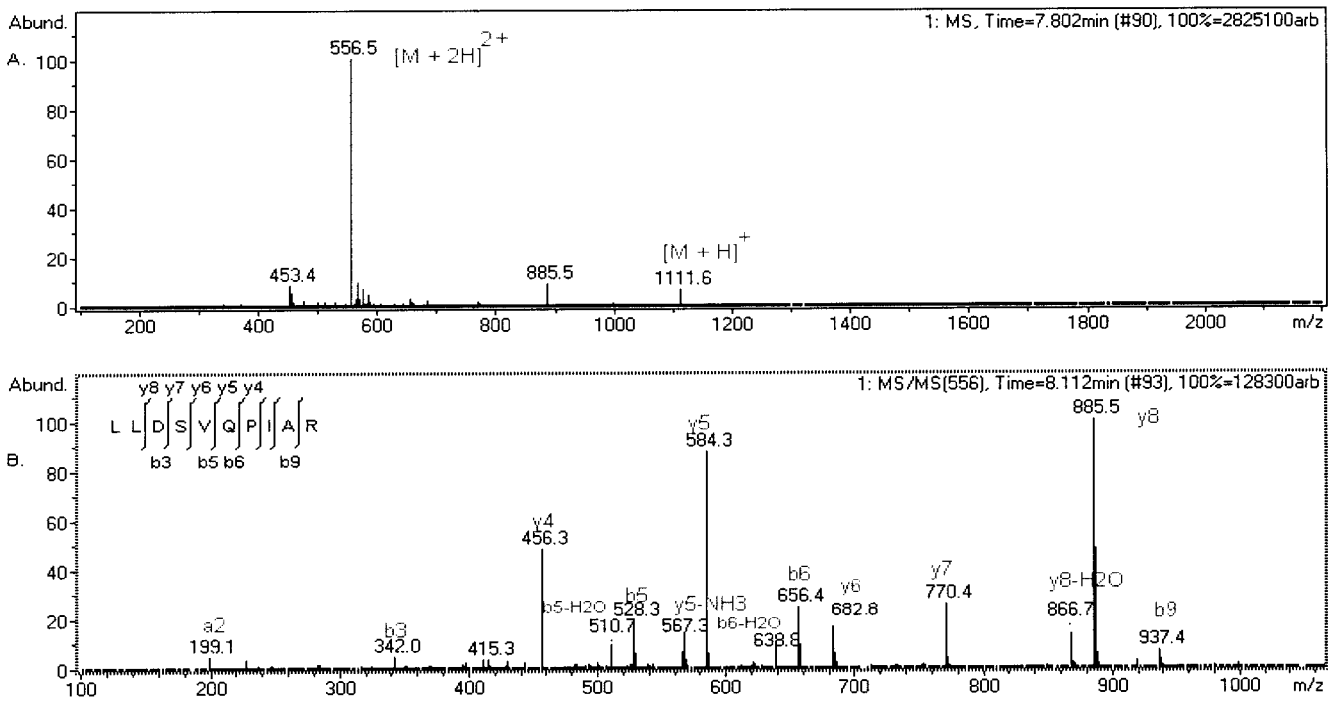

FIGURE 3: AA sequencing analysis of the hAR tryptic peptide ^{862}LLDSVQPIAR^{871}. (A) MS spectrum at a retention time of 7.8 min during an HPLC elution. *m*/*z* range: 100−2200. (B) MS/MS spectrum of precursor ion ($[M + 2H]^{2+} = 556$, corresponds to $[M + H]^{+} = 1111$) fragmentation. *m*/*z* range: 100−2200.

detected in LC/ESI-ITMS; see next section), including the C-terminal tryptic peptide (peptide 8, $[M + H]^{+} = 1132$) in the hAR sequence. These data confirmed that the His_{10}-hAR LBD thus expressed and purified was intact. The locations of these tryptic peptides are discussed below, together with the LC/ESI-ITMS data. Four unidentified contaminant peaks with *m*/*z* of 842.7 (trypsin autolysis product), 863.7, 1178.9, and 1994.5 were observed. These peaks also appeared in a control sample (trypsin digestion of a blank gel piece) and were presumably from trypsin autolysis or trypsin digestion of human skin and hair keratin.

Partial AA Sequencing of the His_{10}-hAR LBD with LC/ESI-ITMS and MS/MS. To further characterize the tryptic peptides, a microbore HPLC system coupled to an LC/ESI-ITMS interface was used to separate the tryptic peptides. Initial separation of the tryptic peptides, using a linear gradient of 1.8−81% acetonitrile over 90 min, revealed that most of the tryptic peptides eluted within the first few minutes. Therefore, a modified step gradient with 0−9% acetonitrile over 20 min was used, and the separation of the peptides was satisfactory. In contrast to the MALDI/TOF-MS analysis, most of the same peptides detected by LC/ESI-ITMS were doubly charged (Table 1). The peptides detected by MALDI/TOF-MS were all detected in LC/ESI-ITMS, except peak 7 ($[M + H]^{+} = 1504$) (Table 1). Three new peptides, and a fourth (peptide 11) that is included within peptide 5, were detected by LC/ESI-MS (Table 1), increasing the coverage to 32%. An automatic mode of MS/MS with an intensity-threshold of 20 000 arbitrary units (corresponds to the mass spectrum abundance) was used, and the three most abundant precursor ions at any given moment, with intensities above the threshold, were fragmented into AA sequence-determining ions. All 11 peptides were eluted with intensities above the threshold, and thus were fragmented and sequenced. A set of a typical MS spectrum (precursor ion) and MS/MS spectrum (product ions) is shown in Figure 3A and 3B, respectively, using peptide 6 ($[M + H]^{+} = 1111$) as an example. For AA sequencing purposes, the masses of the precursor ion and the product ions were entered into the MS-Tag subprogram in the UCSF-ProteinProspector program, and were searched against the NCBI's nonredundant (nr) protein database. The searched results showed that the top 10 matches unambiguously identified the tryptic peptide ^{862}LLDSVQPIAR^{871} of either the *Homo sapiens* AR or the rat AR. The masses in the MS/MS spectrum were compared to those from a list of theoretical product ions of this peptide provided by the MS-Product subprogram of the same program, and the matched ions are shown (Figure 3B). Most of the product ions were y-ions (C-terminus-containing fragment ions) and b-ions (N-terminus-containing fragment ions) (Figure 3B). The locations of the peptides detected by MALDI/TOF-MS and/or detected/sequenced by LC/ESI-ITMS and MS/MS are illustrated in Table 1 and Figure 4. Of particular note, the extreme C-terminal tryptic peptide ^{911}VKPIYFHTQ^{919} in the hAR sequence was also detected and sequenced, confirming that the purified His_{10}-hAR LBD was intact at the C-terminus (Figure 4).

Characterization of the Ligand-Binding Properties of the His_{10}-hAR LBD. In contrast to the hER LBD expressed in *E. coli* (*29*), our His_{10}-hAR LBD contained in inclusion bodies could not be solubilized in a low concentration (1−5 M) of urea, so that the solubilized receptor can have ligand-binding activity directly without the need for renaturation. Our expressed His_{10}-hAR LBD could only be solubilized with a high concentration of denaturant (8 M GdnHCl) and purified in the presence of 6 M GdnHCl. The purified His_{10}-hAR LBD was renatured to bind ligand by a final 2000-fold sample dilution (50-fold GdnHCl dilution) into renaturation binding buffer that contained 5 mM DTT. In saturation binding analysis, a typical saturation binding isotherm was

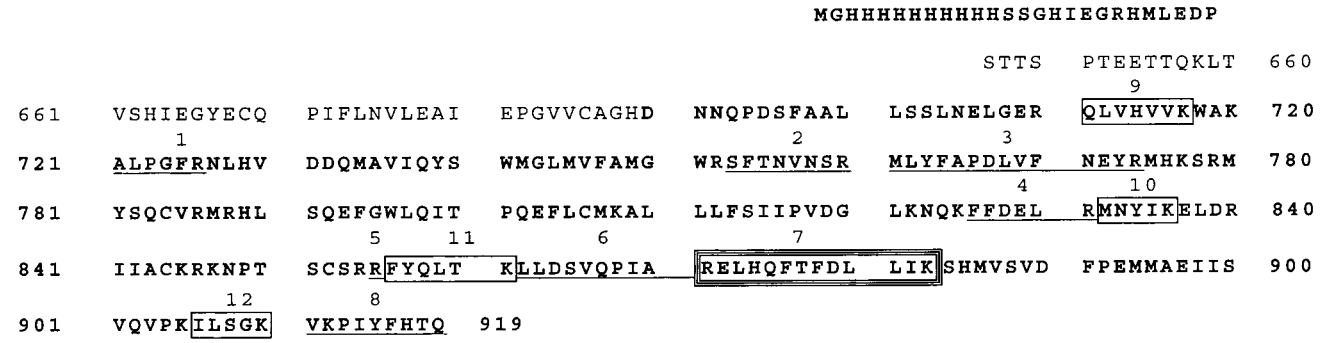

FIGURE 4: Schematic illustration of the locations of the 12 peptides detected by MALDI/TOF-MS and/or detected/sequenced by LC/ESI-ITMS and MS/MS. Boldface letters at the N-terminus represent the recombinant sequence with the decahistidine tag from the constructed vector; boldface letters from AAs 690 to 919 represent the LBD; underlined peptides were detected in both MALDI and LC/MS; the peptide in a triple-lined box was only detected in MALDI; peptides in a single-lined box were only detected in LC/MS.

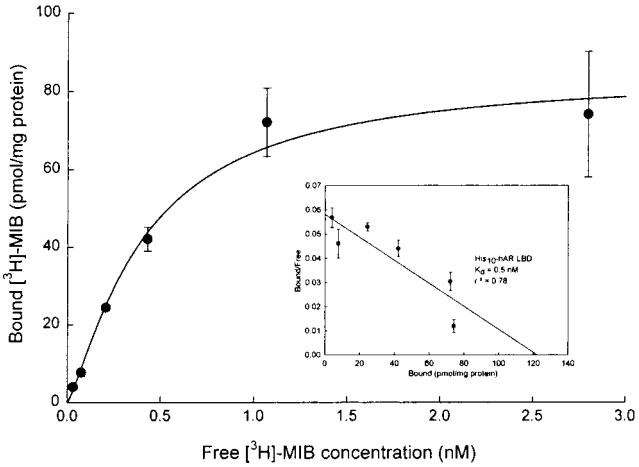

FIGURE 5: Saturation binding analysis of the renatured His_{10}-hAR LBD. Purified sample in 6 M GdnHCl was diluted into renaturation binding buffer with an increasing concentration of [^{3}H]-MIB for 18 h at 4 °C, and specifically bound radioactivity was measured. Binding data were fitted to the equation: $B = B_{max}[L]/(K_d + [L])$, using WinNonlin software ($n = 3$ for each point, mean ± standard deviation). Inset: Scatchard analysis of the same set of saturation binding data.

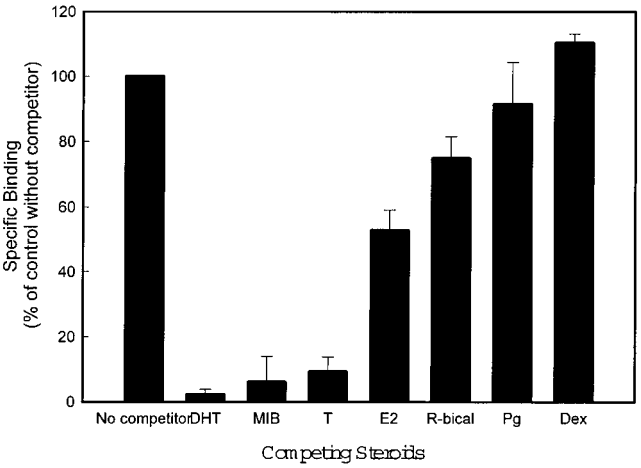

FIGURE 6: Steroid specificities of the renatured His_{10}-hAR LBD. Purified sample in 6 M GdnHCl was diluted into the renaturation binding buffer with 2 nM [^{3}H]-MIB and 200 nM unlabeled steroids for 18 h at 4 °C, and specifically bound radioactivity was measured. Unlabeled MIB (1000 nM) was included instead to determine nonspecific binding. Specific binding (total binding minus nonspecific binding) in the absence of any competing ligand was set at 100, and specific binding in the presence of the unlabeled steroid was compared to it. DHT, dihydrotestosterone; MIB, mibolerone; T, testosterone; E2, estradiol; R-bical, (*R*)-bicalutamide; Pg, progesterone; Dex, dexamethasone ($n = 3$ for each bar, mean ± standard deviation).

found (Figure 5). Specific binding versus free [^{3}H]-MIB concentration was fitted to the single-receptor-type saturation binding equation: $B = B_{max}[L]/(K_d + [L])$, where B is the number of binding sites, B_{max} is the number of total binding sites, K_d is the equilibrium dissociation constant, and [L] is the [^{3}H]-MIB concentration. A K_d of 0.5 ± 0.2 nM was determined (Figure 5), which was similar to that of the AR isolated from rat and human ventral prostate (*28, 30*). However, a binding capacity of 89 ± 22 pmol of binding sites/mg of protein indicated that the renaturation efficiency was very low (<1% by estimation). The mathematical transformation of the saturation binding data for Scatchard analyses ($B/F = B_{max}/K_d - B/K_d$, with B and F indicating the bound and free [^{3}H]-MIB, respectively) gave a straight line. Those data confirm that there was a single type of binding site for MIB in this receptor preparation ($r^2 = 0.78$, $P < 0.02$, Figure 5, inset).

To test the binding specificity for steroids, different steroids were included as competitors for [^{3}H]-MIB binding, and the specifically bound radioactivity was measured. The specific binding in the absence of competitor was set at 100% as control, and specific binding in the presence of competitor was compared to it. A lower percentage of control indicated a higher AR affinity. DHT, unlabeled MIB, and testosterone (T) had higher affinities than the other steroids (Figure 6). Lower affinities were found for (*R*)-bicalutamide, progesterone, and estradiol. Dexamethasone did not appear to have any affinity for the receptors. Therefore, the purified and renatured His_{10}-hAR LBD maintained a similar steroid specificity as other insect cell-expressed hAR (*27, 31*).

DISCUSSION

In this report, we described the development of an expression system for the hAR LBD in *E. coli*. The His_{10}-hAR LBD was present in inclusion bodies, and could only be solubilized with a high concentration of denaturant (8 M GdnHCl). The solubilized His_{10}-hAR LBD was purified in the presence of 6 M GdnHCl. MALDI/TOF-MS analysis of the purified His_{10}-hAR LBD showed an average mass of 34 580 Da, which is 188.2 Da less than the calculated mass from the construct (34 768.2 Da). This suggests the presence of mutations, posttranslational modifications, or, especially,

small deletions in the His_{10}-hAR LBD primary structure. The hER LBD expressed and purified from the bacterial system was shown to have small deletions at the C-terminus by mass spectrometric techniques (*29*). In the current study, detailed tryptic peptide mapping analysis using MALDI/TOF-MS and AA sequencing using LC/ESI-ITMS and MS/MS confirmed a total of 11 peptides with a 32% coverage of the LBD, including the last tryptic peptide in the hAR sequence. These results indicate that the purified His_{10}-hAR LBD was intact at the C-terminus, and modifications to the primary structure of His_{10}-hAR LBD, if present, could only occur in the remaining 68% of the sequence. The fact that the His_{10}-hAR LBD was purified using metal ion affinity chromatography indicated that the protein was also intact at the N-terminus. However, interestingly, the lack of 188.2 Da compared to the calculated mass from the construct coincides with a hypothetical deletion of the first two AAs (Met 131.04 Da and Gly 57.02 Da) from the expressed protein (Figure 6), leaving the following histidine residues intact for their normal function during protein purification. Further studies using LC/ESI-ITMS and MS/MS coupled with trypsin and other protease digestion of the His_{10}-hAR LBD may be able to reveal the underlying mechanism.

The His_{10}-hAR LBD purified in the presence of 6 M GdnHCl could be renatured to have ligand-binding activity. However, the microheterogeneity of the purified proteins would complicate X-ray and NMR studies (*20*−*23*). Therefore, should this system be used for those studies, the heterogeneity (renatured versus unfolded species) derived from the renaturation step should be addressed by using other purification techniques, such as testosterone−biotin resin, to separate the renatured from the unfolded form. Nevertheless, given the fact of the rapid, unlimited protein resource in a prokaryotic expression system, and the apparent homogeneous, intact nature of our purified His_{10}-hAR LBD, this system could prove to be valuable for X-ray and NMR studies.

MALDI/TOF-MS, and HPLC/ESI-ITMS and MS/MS provide us with the opportunity to study the hAR sequence and posttranslational modifications. Future studies that involve ligand−receptor interaction (such as affinity labels), ligand-induced conformational change (such as different protease digestion patterns), signal transduction upon ligand occupation (such as phosphorylation), and protein−protein interaction (such as discovery of novel coactivators/repressors) are likely to be accomplished by using mass spectrometry techniques.

REFERENCES

1. Beato, M. (1989) *Cell 56*, 335−344.
2. Evans, R. M. (1988) *Science 240*, 889−895.
3. Lubahn, D. B., Joseph, D. R., Sar, M., Tan, J., Higgs, H. N., Larson, R. E., French, F. S., and Wilson, E. M. (1988) *Mol. Endocrinol. 2*, 1265−1275.
4. Chang, C. S., Kokontis, J., and Liao, S. T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7211−7215.
5. Carson-Jurica, M. A., Schrader, W. T., and O'Malley, B. W. (1990) *Endocr. Rev. 11*, 201−220.
6. Jenster, G., van der Korput, J. A., Trapman, J., and Brinkmann, A. O. (1992) *J. Steroid Biochem. Mol. Biol. 41*, 671−675.
7. Wong, C. I., Zhou, Z. X., Sar, M., and Wilson, E. M. (1993) *J. Biol. Chem. 268*, 19004−19012.
8. Housley, P. R., Sanchez, E. R., Danielsen, M., Ringold, G. M., and Pratt, W. B. (1990) *J. Biol. Chem. 265*, 12778−12781.
9. Smith, D. F., and Toft, D. O. (1993) *Mol. Endocrinol. 7*, 4−11.
10. Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. R., and Chambon, P. (1987) *Cell 51*, 941−951.
11. Tsai, M. J., and O'Malley, B. W. (1994) *Annu. Rev. Biochem. 63*, 451−486.
12. Srinivasan, G. (1992) *Mol. Endocrinol. 6*, 857−860.
13. Ahrens, H., Schuh, T. J., Rainish, B. L., Furlow, J. D., Gorski, J., and Mueller, G. C. (1992) *Receptor 2*, 77−92.
14. Wooge, C. H., Nilsson, G. M., Heierson, A., McDonnell, D. P., and Katzenellenbogen, B. S. (1992) *Mol. Endocrinol. 6*, 861−869.
15. Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A., and Greene, G. L. (1998) *Cell 95*, 927−937.
16. Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engstrom, O., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. (1997) *Nature 389*, 753−758.
17. Williams, S. P., and Sigler, P. B. (1998) *Nature 393*, 392−396.
18. Tanenbaum, D. M., Wang, Y., Williams, S. P., and Sigler, P. B. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 5998−6003.
19. Matias, P. M., Donner, P., Coelho, R., Thomaz, M., Peixoto, C., Macedo, S., Otto, N., Joschko, S., Scholz, P., Wegg, A., Basler, S., Schafer, M., Egner, U., and Carrondo, M. A. (2000) *J. Biol. Chem. 275*, 26164−26171.
20. Bayer, E. A., Ben-Hur, H., Hiller, Y., and Wilchek, M. (1989) *Biochem. J. 259*, 369−376.
21. Chait, B. T., and Kent, S. B. (1992) *Science 257*, 1885−1894.
22. Beavis, R. C., and Chait, B. T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6873−6877.
23. Aebersold, R. (1993) *Curr. Opin. Biotechnol. 4*, 412−419.
24. Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
25. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Anal. Chem. 68*, 850−858.
26. Haynes, P. A., Fripp, N., and Aebersold, R. (1998) *Electrophoresis 19*, 939−945.
27. Xie, Y. B., Sui, Y. P., Shan, L. X., Palvimo, J. J., Phillips, D. M., and Janne, O. A. (1992) *J. Biol. Chem. 267*, 4939−4948.
28. Mukherjee, A., Kirkovsky, L., Yao, X. T., Yates, C. R., Miller, D. D., and Dalton, J. T. (1996) *Xenobiotica 26*, 117−122.
29. Seielstad, D. A., Carlson, K. E., Katzenellenbogen, J. A., Kushner, P. J., and Greene, G. L. (1995) *Mol. Endocrinol. 9*, 647−658.
30. Murthy, L. R., Johnson, M. P., Rowley, D. R., Young, C. Y., Scardino, P. T., and Tindall, D. J. (1986) *Prostate 8*, 241−253.
31. Beitel, L. K., Sabbaghian, N., Alarifi, A., Alvarado, C., Pinsky, L., and Trifiro, M. (1995) *J. Mol. Endocrinol. 15*, 117−128.

BI010525J