Analysis of the Structural Core of the Human Estrogen Receptor Ligand Binding Domain by Selective Proteolysis/Mass Spectrometric Analysis†

Donald A. Seielstad,‡ Kathryn E. Carlson,‡ Peter J. Kushner,§ Geoffrey L. Greene,|| and John A. Katzenellenbogen*,‡

Department of Chemistry, University of Illinois, 600 South Mathews Avenue, Urbana, Illinois 61801, Metabolic Research Unit, University of California, 1141 Health Sciences West, San Francisco, California 94143, and Ben May Institute, University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

Received March 30, 1995; Revised Manuscript Received July 31, 1995⊗

ABSTRACT: The structure of the ca. 250 amino acid hormone binding domain of the human estrogen receptor (hER-LBD), expressed in *E. coli* and purified as a complex with estradiol, has been probed by selective proteolysis, with analysis of the protein fragments both by classical methods (SDS-PAGE and Edman N-terminal sequencing) and by mass spectrometry (HPLC-coupled electrospray ionization mass spectrometry (LC/ESI-MS)). Rapid cleavage by several proteases (trypsin, chymotrypsin, thermolysin, and Asp-N endoproteinase) is observed within a localized region (residues 297−303) at the N-terminus. In contrast, proteolytic scission at the C-terminus is less localized and more progressive; initial cuts by trypsin, chymotrypsin, thermolysin, V8, and Asp-N proteinases are observed to occur in the region 553−571, followed by further cleavage with thermolysin (548) and trypsin (548, 531, and 529). Thus, N_{304} and K_{529} define the protease-resistant N- and C-termini of a core structure for this domain that appears to contain the elements sufficient for ligand binding. The remaining segment of this domain (530−553), which is known to embody elements essential for ligand-modulated transcription activation (AF-2), is likely a surface-exposed region that, through these studies, is shown to be accessible to proteases. Only a single region within the 26 kDa ligand-binding core (N_{304}−K_{529}) has been identified as being readily accessible to proteases; rapid proteolysis using the proteases trypsin, chymotrypsin, and thermolysin, is localized to residues 465−468, with cleavage occurring at residues K_{467}, L_{466}, and both T_{465} and S_{468}, respectively. The flexibility implied by the cuts in this internal 465−468 region suggest that the hER-LBD may actually consist of two subdomains. These proteolysis studies provide a substantially refined view of the conformational nature of the human estrogen receptor ligand binding domain.

The nuclear hormone receptors comprise a superfamily of proteins that act as ligand-modulated regulators of transcription; they are multidomain proteins in which the functions of DNA binding and ligand binding are ascribed to distinct protein regions, namely domain C and domain E, respectively (Parker, 1991; Evans, 1988; Beato, 1988; Krust et al., 1986). The junctions between these domains have been defined by two approaches, mutational deletions coupled with functional analysis and multisequence alignments. The latter approach is useful because there is high homology among all members of this superfamily in the DNA-binding domain C and considerable homology in the hormone-binding domain E, but little homology elsewhere (Parker, 1991).

An alternate approach to elucidate domain junctions involves the probing of protein topology by protease sensitivity (Carrey, 1989; Aitken et al., 1989). Proteinases can cleave peptide sequences only in regions that can adopt rather extended conformations (Hubbard et al., 1991, 1994). Thus, proteolytic clips occur readily in the less structured, more flexible regions that exist between the densely packed core domains of multidomain proteins, whereas cleavages within the domain cores themselves are rare (Carrey, 1989; Aitken et al., 1989; Hubbard et al., 1991, 1994). Thus, the pattern and the sequence of the proteolytic cuts of a protein may be sensitive probes of its domain structure and its conformation.

It is known that proteolysis of numerous nuclear hormone receptors under mild conditions results in rapid cleavage to a stable ligand binding domain (LBD)[1] protein core of mass *25−30 kDa*,[2] comprised largely of domain E, which retains the capacity for binding ligand (Sherman et al., 1978; Katzenellenbogen, B. S., et al., 1987; Elliston & Katzenellenbogen, 1988; Chakraborti et al., 1992; Pavlik & Katzenel-

† This work was supported in part by grants from the National Institutes of Health (PHS 5R37DK 15556 to J.A.K., and PHS 5RO1 CA02897 to G.L.G.), the U.S. Army Medical Research Core (DAMD 17-94-J-44228 to G.L.G.), and the American Cancer Society (BE61E to P.J.K.). Mass spectrometric analysis was performed on instruments whose acquisition was supported in part by a grant from the Division of Research Resources, National Institutes of Health (RR 07141).

* To whom correspondence should be addressed: Dr. John A. Katzenellenbogen, 461 Roger Adams Lab 37-5, Department of Chemistry, University of Illinois, 600 S. Mathews Ave., Urbana, IL 61801. 217 333 6310 (Phone), 217 333 7325 (FAX), katzenel@aries.scs.uiuc.edu.

‡ University of Illinois.

§ University of California.

|| University of Chicago.

⊗ Abstract published in *Advance ACS Abstracts,* September 15, 1995.

[1] Abbreviations used are: SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; LC/ESI-MS, liquid chromatography electrospray ionization mass spectrometry; hER-LBD, human estrogen receptor ligand binding domain; rGR, rat glucocorticoid receptor.

[2] Throughout the manuscript, molecular masses estimated by SDS-PAGE are given in italics (e.g., *12 kDa* fragment), while masses to the nearest kDa determined by ESI-MS are in standard typeset (e.g., 12 kDa fragment). Exact masses as determined by ESI-MS are given to the nearest Da.

0006-2960/95/0434-12605$09.00/0

lenbogen, 1980; Katzenellenbogen, J. A., et al., 1983; Thole, 1993; Fritsch et al., 1993). The use of selective probes (ligand affinity labels and monoclonal antibodies) and the availability of recombinant preparations of receptors have enabled the proteolysis of the *~30 kDa* ligand binding core of several of the receptors systems to be followed yet further (Sherman et al., 1978; Katzenellenbogen et al., 1987; Elliston & Katzenellenbogen, 1988; Chakraborti et al., 1992; Pavlik & Katzenellenbogen, 1980; Katzenellenbogen et al., 1983; Thole, 1993; Fritsch et al., 1993). In each case, a discrete set of protein fragments ranging from *7−26 kDa* in mass is generated, suggesting that the ligand binding domain contains core sub-structures that are relatively protease resistant (Carrey, 1989; Aitken et al., 1989; Hubbard et al., 1991, 1994). While classical protein chemistry methods have been used to identify the N-termini of some of these fragments, determination of their C-termini is more difficult. Indeed, these termini have, in most cases, simply been inferred based upon the electrophoretically-determined molecular weight and the expected amino acid specificity of the proteases used in their generation (Chakraborti et al., 1992; Thole, 1993).

Recent developments in mass spectrometry have revolutionized the analysis of peptide and protein composition and sequencing (Aebersold, 1993; Chait, 1994; Gross, 1993). In particular, electrospray ionization mass spectrometry (ESI-MS)[1] enables the analysis of picomole quantities of peptides and proteins with 0.01% mass resolution, for biopolymers of up to 300000 Da. When coupled with high pressure liquid chromatography, LC/ESI-MS[1] becomes a selective, sensitive, and comparatively high resolution method for the analysis of peptides, proteins, and protein mixtures (Aebersold, 1993; Chait, 1994; Gross, 1993).

In this report, we describe the use of both classical protein chemistry and recently introduced LC/ESI-MS methods to analyze the pattern of proteolysis of the ligand binding domain of the human estrogen receptor (hER-LBD)[1] at the molecular level. Using an hER-LBD, produced in *E. coli* and isolated as a complex with estradiol (Seielstad et al., 1995), we have studied its proteolysis by trypsin, chymotrypsin, thermolysin, *S. aureus* V8 proteinase, and Asp-N endoproteinases. By analysis of the composition and sequence of many of the protein fragments generated, we have found that the hormone binding domain of ER has a protease-resistant N-terminus, but suffers more progressive cleavages from the C-terminus; it also has an internal site that is accessible to proteolysis. Thus, structural aspects of surface accessibility and flexibility of the termini and central core region of this domain are revealed.

EXPERIMENTAL PROCEDURES

Materials. Materials were obtained from the following sources: Kaleidoscope prestained molecular weight standards, Bio-Rad; TPCK-trypsin, TLCK-chymotrypsin, phenylmethylsulfonyl fluoride (PMSF), Endoproteinase Asp-N, Endoproteinase Glu-C (*Staphylococcus aureus* V8 protease), and thermolysin (protease type X from *Bacillus thermoproteolyticus rokko*) Sigma.

The hER-LBD at 3.13×10^{-5} M (1 mg/mL), in 25 mM Tris, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 5 M urea, and with 62 μM low specific activity (0.25 Ci/mmol)E_2, was stored in liquid N_2 (−196 °C) (Seielstad et al., 1995). Prior to proteolysis, this material was thawed and then diluted 1:10 into an appropriate buffer, resulting in a final urea concentration of 0.5 M. In a separate study, urea concentrations up to 5 M were found to slow the action of the proteases, but did not affect the specificity nor the resulting proteolytic pattern (data not shown, J.A.K. and K.E.C., unpublished results).

Proteolysis. Dithiothreitol (1 mM) was maintained in the buffer at all times. The hER-LBD was diluted with 10 mM Tris, pH 7.4 at 25 °C to give 3 nmol/mL ER and 0.5 M urea for the trypsin, chymotrypsin, and endoproteinase Asp-N protease studies; for the thermolysin digests, $CaCl_2$ was added to be 10 mM. For endoproteinase Glu-C (*Staphylococcus aureus* V8 protease), where the protease specificity could be affected by the buffer (Houmard & Drapeau, 1972), the protein was diluted as above, except into final concentrations of buffer of 0.1 M NH_4HCO_3 pH 7.8 (glutamyl specific) or 0.1 M sodium phosphate pH 7.8 (glutamyl and aspartyl specific). Trypsin and chymotrypsin were dissolved in 0.001 M HCl, thermolysin in 50 mM Tris, 100 mM NaCl, 10 mM $CaCl_2$ pH 7.5, Asp-N in 10 mM Tris, pH 7.4, and V8 protease in H_2O. Protease digestions were performed using the following conditions; trypsin (100 μg/mg protein; 0 °C; 10, 30, 60, or 120 min), chymotrypsin (10 μg/mg protein; 25 °C; 1 h, or 50 μg/mg protein; 25 °C; 2 h), thermolysin (50 μg/mg protein; 25 °C; 10 min), V8 protease (100 μg/mg protein; 25 °C; 1 h), and Asp-N (100 μg/mg protein; 25 °C; 2 h). After incubation with the protein at the concentrations and times indicated, the enzyme action was stopped: by the addition of 10 mM EDTA to the thermolysin digests, or for the other proteases, by the addition of 1 mM PMSF or by direct injection onto the HPLC column at the appropriate time interval for mass spectrometry studies, or 1% SDS and heating for 5 min in a boiling water bath for SDS-PAGE. The peptides were then immediately separated on SDS polyacrylamide gels or analyzed by mass spectrometry.[3]

SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE). Electrophoresis was performed using a 10% T:7.3% C acrylamide gel, following the protocol of Schagger and von Jagow (Schagger & von Jagow, 1987). *O,O*-Diallyltartramide (DATD) was substituted for bis-acrylamide in the running gel (Anker, 1970). Gels were either stained with Coomassie Blue G (Schagger & von Jagow, 1987) or electroblotted to PVDF membrane for N-terminal sequencing. Molecular weight prestained standards were used: myosin (217000 Da), β-galactosidase (130000 Da), BSA (72000 Da), carbonic anyhdrase (42200 Da), soybean trypsin inhibitor (31300 Da), lysozyme (18100 Da), and aprotinin (7100 Da). Additional low molecular weight standards, cytochrome C (11700 Da) and insulin B (3500 Da) were sometimes included.

N-Terminal Microsequencing. Gel patterns were electroblotted to Pro-Blott PVDF membranes using a BioRad Mini Trans-Blot Electrophoretic Transfer Cell and stained with Coomassie Blue R, following the recommendations of the manufacturer for the Pro-Blott membrane (Applied

[3] A referee made the helpful suggestion that the use of a consistent protocol to terminate proteolytic digestion prior to analysis by SDS-PAGE/Edman degradation and HPLC/ESI-MS would remove any potential ambiguity in comparing results from both methods. For the proteases used in this study, acidification to pH 2 with HCl or CF_3-COOH (Wilkinson, 1986) would be effective.

279 *M D P* S A G D M R A A N L W P S P L M I K R 300

301 S K K N S L A L S L T A D Q M V S A L L D A E P P I L Y S E Y D P T R P F S E A S M M G L L T N L A 350

351 D R E L V H M I N W A K R V P G F V D L T L H D Q V H L L E C A W L E I L M I G L V W R S M E H P G 400

401 K L L F A P N L L L D R N Q G K C V E G M V E I F D M L L A T S S R F R M M N L Q G E E F V C L K S 450

451 I I L L N S G V Y T F L S S T L K S L E E K D H I H R V L D K I T D T L I H L M A K A G L T L Q Q Q 500

501 H Q R L A Q L L L I L S H I R H M S N K G M E H L Y S M K C K N V V P L Y D L L L E M L D A H R L H 550

551 A P T S R G G A S V E E T D Q S H L A T A G S T S S H S L Q K Y Y I T G E A E G F P A T V 595

FIGURE 1: Sequence of the major species in the human estrogen receptor hormone binding domain (hER-LBD, *MDP*S_{282}–A_{571}). The three italicized N-terminal residues, *MDP* arise from the plasmid; the hER sequence begins with S_{282}. As discussed in the text, there are also minor species with C-termini at A_{569}, S_{575}, and V_{595} (full length, Seielstad et al., 1995).

Biosystems, Inc). Bands of interest were excised with a razor blade and N-terminal sequenced by Edman chemistry on an Applied Biosystems Model 477A protein sequencer coupled to a Model 120A on-line PTH analyzer, by Dr. K.-L. Ngai and the personnel in the Genetic Engineering Facility of the Univ. of Illinois.

Liquid Chromatography (HPLC). HPLC analyses were performed, as described previously (Seielstad et al., 1995), on microbore reversed-phase columns from Brownlee-Applied BioSystems (1.0 mm i.d. C-18, 7 μm particles, 300 Å pore size, either 100 mm or 250 mm long), at a flow rate of 36 μL/min. Compound linear gradients of a binary solvent system were run over a 60-180 min period. Solvents and reagents used for the HPLC and ESI-MS analyses were HPLC grade, and water was Milli-Q grade distilled prior to use. The binary solvent system used in the analysis of protein and peptide fragments was comprised of the following: solvent A, H_2O, 0.06 vol % CF_3COOH, and solvent B, CH_3CN, 0.05 vol % CF_3COOH. Intact receptor was analyzed using a steep uniform gradient (1.5% solvent B/min), while proteolytic fragments were analyzed using a complex linear gradient changing from 0.25% solvent B/min to 1.5% solvent B/min.

Electrospray Ionization Mass Spectrometry (ESI-MS). A VG Quattro (quadrupole-hexapole-quadrupole, QHQ) mass spectrometer system (Fisons Instruments, VG Analytical; Manchester, UK) was used for the analysis of the receptor protein and protease fragments. Data aquisition and processing were controlled by the VG MassLynx (version 2.0) data system; MaxEnt (Maximum Entropy) software was used for the processing and analysis of zero charge state electrospray data for the receptor and large protein fragments. The VG MassLynx program was used to select peptide fragments from the known sequence of hER and match them with the masses observed for the hER peptides. Mass searches of the sequence afforded several possible protein components within a fairly generous window about the specified mass. These lists could be substantially pared by identifying those fragments possessing both N- and C-termini arising from likely protease recognition. Finally, all fragments identified from a given protease digestion and identified by LC/MS were matched to avoid overlapping termini. For all proteases and all digest times used, all peptides from the ER sequence could be identified.

A microbore HPLC/ESI-MS interface, operating at 36 μL/min was split [2:3] such that ~12 μL/min HPLC eluent was directed to the ESI-MS probe for nebulization. A sheath flow probe was used to improve protein sensitivity, with a sheath liquid comprised of [3:1] methoxyethanol:1-propanol. HPLC sample injections of 5–20 μL were comprised typically of 15–30 pmol/μL receptor or proteolysis fragments in buffer solution (0.5–5.0 M urea). Using the sheath flow probe configuration, it is estimated that as little as 10 pmol of protein or peptide fragments could be detected and resolved by the ESI-MS. While this value was not rigorously determined *per se*, its estimate results from knowledge of protein amounts injected, split ratio, and signal intensities observed both by SDS-PAGE and ESI-MS for identical samples.

RESULTS

Analysis of the Size of the Proteolytic Fragments of the Ligand Binding Domain of the Human Estrogen Receptor (hER-LBD) by SDS-PAGE. For all of the proteinase studies, we have used hER-LBD expressed in *E. coli* and purified to near homogeneity, as a complex with estradiol; this material is composed principally of the sequences *MDP*S_{282}...A_{571} and *MDP*S_{282}...A_{569},[4] together with small amounts of species ending at S_{575} and V_{595} (Seielstad et al., 1995) (cf., Figure 1 and Discussion). The progression of the protease digestion was monitored by SDS-PAGE, and conditions were selected which generated discrete fragments that could be isolated for subsequent Edman N-terminal analysis.

Trypsin digests performed using 100 μg trypsin/mg receptor ([1:10] ratio) initially generate transiently observed protein fragments of *26 kDa* and *24 kDa*.[5] These fragments form rapidly (less than 10 min), even though digests are maintained at 0 °C. Within 30 min, appreciable amounts of *14 kDa* and *7 kDa* protein fragments are generated, with concomitant loss of the original *33 kDa* protein and the disappearance of the *26* and *24 kDa* transient species (Figure 2, lane 4; see also Figure 5 and Table 1).

Incubation of chymotrypsin with receptor, at low concentration [1:100] and short incubation times (15–60 min) at 0 °C, produces a trace quantity of a transient protein fragment of *29 kDa*. Similar to the behavior seen with trypsin, however, further digestion of hER-LBD by chymotrypsin

[4] The first three italicized residues *MDP* arise from the plasmid construct and are not part of the hER sequence, which begins at S_{282}.

[5] The specificity of three of the proteinases used in this study: trypsin, chymotrypsin, and *S. aureus* V8, is determined by the P_1 residue (Schecter & Burger, 1967), that is, the residue on the amino terminal side of the scissile bond. In contrast, endoproteinases Asp-N and thermolysin have P_1' specificity, that is, determined by the residue on the carboxy terminal side of the scissile bond. For simplicity of discussion, the P_1 residue will be used to designate the site of cleavage, regardless of whether the protease displays P_1 or P_1' specificity.

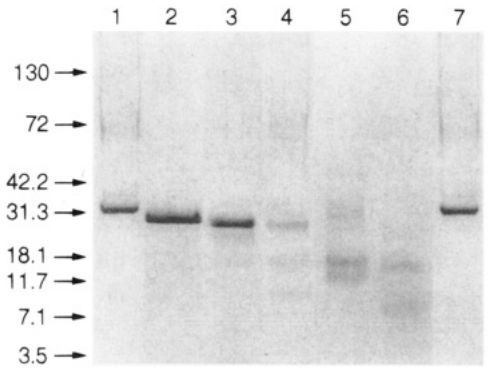

FIGURE 2: SDS-PAGE analysis of the intact hER-LBD (lanes 1 and 7) and peptide fragments generated by treatment with V8 proteinase (lane 2), Asp-N endoproteinase (lane 3), trypsin (lane 4), and chymotrypsin (lane 5 (1:100) and lane 6 (1:20)). The concentration of the proteinases, times, and temperatures of treatment are given in the methods. The major bands were excised and analyzed by N-terminal sequencing.

([1:20], up to 120 min, 0 °C) generates two predominant species of *19 kDa* and *12 kDa*. A *6 kDa* protein fragment is observed to accumulate late in the digestion, the production of which coincides with the proteolytic consumption of the *12 kDa* protein (Figure 2, lanes 5 and 6).

Thermolysin displays a general sequence specificity toward hydrophobic residues (Matsubara, 1970) that is similar to that observed for chymotrypsin. Consequently, protein fragments resulting from thermolysin digests of hER-LBD display, upon SDS-PAGE analysis, a pattern nearly identical to that observed for chymotrypsin digests (data not shown). Incubation of hER-LBD with thermolysin ([1:20], 25 °C, 10 min) initially results in a transiently observed protein band corresponding to *30 kDa*. Accumulation of fairly stable protein fragments of *17 kDa* and *11 kDa* appear in a manner similar to that observed for *19 kDa* and *12 kDa* fragments from chymotrypsin digestion. Likewise, the strong *11 kDa* band observed during thermolysin proteolysis subsequently yields to the appearance of a *7 kDa* band in a fashion similar to the chymotrypsin progression from the *12 kDa* to *6 kDa* protein fragments.

The specificity of *S. aureus* V8 proteinase may be tailored by the choice of buffer. In ammonium buffer, V8 proteinase cleaves selectively at the carboxyl side of glutamyl residues, while in phosphate buffer, it cleaves at the carboxyl side of both glutamyl and aspartyl residues (Houmard and Drapeau, 1972). Incubation of the hER-LBD using *S. aureus* V8 proteinase in ammonium buffer (glutamyl specific) had no apparent effect (even when used at [1:5] and incubated for 120 min at 25 °C). However, incubation of hER-LBD using the same proteinase in phosphate buffer (glutamyl or aspartyl specific) generates a single *30 kDa* protein fragment (Figure 2, lane 2) that is distinguishable from the original protein. No additional fragmentation is observed either by use of excess proteinase or following extended incubations.

Endoproteinase Asp-N, which cuts on the amino-terminal side of aspartic acid (Drapeau, 1980), likewise has limited effect on the hER-LBD. It produces a single *29 kDa* fragment (Figure 2, lane 3). Trace amounts of both *18 kDa* and *6 kDa* fragments are observed, but could not be produced in sufficient quantities for further studies.

Edman N-Terminal Sequencing of the Major Proteolytic Fragments of the hER-LBD. Analysis of proteolytic digest fragments for N-terminal sequence by Edman degradation is limited to those protein fragments which accumulate in sufficient quantity so that they can be isolated as distinct, reasonably pure bands from SDS-PAGE. Since the information gleaned by this method is limited (i.e., C-terminal sequence must still be inferred from PAGE-estimated molecular weight), Edman degradation was performed on only a few of the most abundant and stable protein fragments. These N-terminal protein sequences were used to corroborate the major protein fragment assignments made using the substantially improved resolution available through the ESI-MS technique.

The sequencing results are summarized in Table 1.[5] Three hER-LBD fragments generated by trypsin ([1:10], 0 °C, 30 min) proteolysis, namely *26 kDa*, *14 kDa*, and *7 kDa*, were sequenced. The *26 kDa* protein band was found to consist of at least two unique N-terminal sequences, corresponding to trypsin cleavage sites following K_{302} and K_{303}. Thus, the *26 kDa* band is comprised of a mixture of protein fragments of closely related sizes with multiple N-termini. Edman degradation of the stable *14 kDa* and *7 kDa* protein fragments, which result following extended incubation of hER-LBD with trypsin, identify single N-terminal sequences for protein fragments beginning at N_{304} and S_{468}, respectively. The nature of these fragments and the relationships between them were eludicated further by LC/ESI-MS (*vide infra*).

In contrast to trypsin, chymotrypsin, a protease with more general substrate specificity (Haschemeyer & Haschemeyer, 1973), creates five discrete protein fragments from the hER-LBD (Figure 2, lanes 5 and 6). The largest of these fragments, estimated at *31 kDa* and *29 kDa*, are transient in nature. Sufficient material, however, was isolated from the *29 kDa* protein band following electroblotting to unambiguously identify M_{297} as its N-terminal residue. The apparently stable *19 kDa* band resulting from more extensive proteolysis was found to comprise a mixture of proteins with numerous N-termini whose sequences could not be resolved by this method. An accessible internal cleaveage site following L_{466} was identified from the K_{467} N-terminal sequence of the *12 kDa* fragment. Finally, the *6 kDa* fragment, resulting from prolonged proteolysis and associated with the disappearance of the *12 kDa* protein fragment, was also found to have the same K_{467} N-terminus as the *12 kDa* protein.

V8 proteinase and Asp-N endoproteinase each generated only single large protein fragments of *30* and *29 kDa* (Figure 2). The protein generated from V8 proteinase digestion had the same N-terminal sequence as intact hER-LBD (Table 1), suggesting that a short peptide was cleaved from the C-terminus. Endoproteinase Asp-N, however, generated a protein fragment whose N-terminus was D_{285} (Table 1). As our ability to accurately and unambiguously identify the protein components generated by proteolysis using the ESI-MS technique improved, less reliance was placed on corroborating data using the Edman procedure. Consequently, the protease fragments generated from the hER-LBD using thermolysin were not characterized by the Edman N-terminal sequencing technique.

Analysis of Proteolytic Fragments from hER-LBD by Liquid Chromatography/Electrospray Ionization Mass Spectrometry (LC/ESI-MS). The results of analysis of the proteolytic fragments of the hER-LBD by LC/ESI-MS are summarized in Table 1 (see also Figure 5). In every case, the sequencing results and fragment identity derived from the LC/ESI-MS data were consistent with those obtained

Table 1: Summary of Core Proteolysis Fragments of the hER-LBD

	SDS-PAGE		ESI-MS				
protease	*kDa*	N-terminal sequence[a]	kDa	Da[b]	(Δ)[c]	t_R (min)[d]	predicted sequence
trypsin	*26*	N_{304}SLALSLTAD[e]	29	not observed			$N_{304}-R_{555}$[f]
		K_{303}NSLALSLTA[e]		28880	(−1.1)	102	$K_{303}-R_{555}$
				29098	(+1.6)		$S_{301}-R_{555}$
	24	*h*		not observed			
	14	N_{304}SLALSLTAD	18	18518	(−0.8)	104	$N_{304}-K_{467}$
				18645	(−1.9)		$K_{303}-K_{467}$
				18866	(+3.7)		$S_{301}-K_{467}$
	10	*h*	10	10252	(−1.2)	85	$S_{468}-R_{555}$
				10285	(31.7)[i]		$S_{468}-R_{555}$
				9489	(−1.3)		$S_{468}-R_{548}$
	7	S_{468}LEEKDHIHR	7	7264	(+0.4)	84	$S_{468}-K_{529}$
				7494	(+1.8)		$S_{468}-K_{531}$
			2	2010	(1.3)		$N_{532}-R_{548}$
chymotrypsin	*29*	M_{297}IKRSKKNSL	31	31203[j]	(+24)	106	$M_{297}-A_{571}$
			29	29380[j]	(−1.9)	108	$M_{297}-T_{553}$
	19	*k*	19	19132	(−1.5)	109	$I_{298}-L_{466}$
				19263	(−1.6)		$M_{297}-L_{466}$
	12	K_{467}SLEEKDHI	12	11764	(−0.9)	89	$K_{467}-A_{569}$
				11935	(+0.5)		$K_{467}-A_{571}$
			10	9762	(−0.2)	91	$I_{298}-W_{383}$
				9890	(+2.5)		$M_{297}-W_{383}$
	6	K_{467}SLEEKDHI		not observed			
thermolysin	*30*	*h*	31	30696	(+5.0)	112	$I_{298}-H_{567}$
	17	*h*	19	19144	(−4.6)	114	$M_{297}-T_{465}$
				19014	(−3.4)		$I_{298}-T_{465}$
	11	*h*	12	11693	(+1.4)	96	$L_{466}-H_{567}$
				11806	(+1.2)		$L_{466}-L_{568}$
				12053	(+5.0)		$L_{466}-A_{571}$
	7	*h*	9	9403	(+0.8)	98	$L_{469}-R_{548}$
				9732	(+1.4)		$L_{466}-R_{548}$
		h	2	1979	(−1.3)	57	$L_{549}-H_{567}$
V-8	*30*	M_{279}DPSAGDMRA	32	32385	(+2.8)	98	$M_{279}-D_{564}$
Asp-N	*29*	D_{285}MRAANLxPS[l]	31	31715	(+7.6)	101	$D_{285}-T_{563}$
	18	*g*		not observed			
	6	*g*		not observed			

[a] N-Terminal residue is indicated with the number of additional residues that provided a distinct sequence. [b] Observed mass in daltons. [c] Deviation between observed mass and mass of the predicted fragment. [d] Retention time on HPLC. [e] Sequences for two N-termini were observed. [f] The $N_{304}-R_{555}$ species was not observed by ESI-MS in this 29 kDa preparation. [g] Minor species, not sequenced. [h] Not sequenced. [i] Mass deviation suggests methionine sulfone modification. [j] Weak. [k] A mixture of proteins with numerous N-termini whose sequences could not be resolved by this method. [l] Ambiguity in sequence position 8.

from Edman N-terminal sequencing, therefore having established the validity of the LC/ESI-MS technique, no further analyses by Edman degradation were necessary, since the LC/ESI-MS data were very precise, more complete, and substantially more revealing.

Trypsin. The most complete analysis of hER-LBD proteolysis has been made with trypsin. A typical HPLC trace from hER-LBD cleaved by trypsin ([1:10] for 30 min at 25 °C) is shown in Figure 3. The elution position of the major peaks we have identified (cf., Table 1) are indicated on the trace. A representative ESI mass spectrum for the 18 kDa fragments is shown in Figure 4. It is evident from the spectrum itself (Figure 4A) that this HPLC isolated peak consists of more than one component; deconvolution to the zero-charged molecular mass species (Figure 4B) reveals components of 18518, 18645, and 18866 Da which correspond to the protein fragments $N_{304}-K_{467}$, $K_{303}-K_{467}$, and $S_{301}-K_{467}$, respectively (cf., Table 1).

The string of basic residues located at the putative junction of domains D and E, $K_{299}RSKK_{303}$, represents a site of rapid cleavage by trypsin. The initial cut appears to follow R_{300}; thereafter, continued incubation results in an erosion of the N-terminal residues with cleavage following K_{302} and K_{303}. We believe R_{300} to be the site of initial cleavage for the following reason: We observe only a single N-terminal peptide fragment composed of $M_{279}-R_{300}$, while we find that the 29 kDa and 18 kDa core protein fragments have N-termini at S_{301}, K_{303}, and N_{304}. As a result of the transient nature of the 29 kDa fragment, analysis by ESI-MS did not result in observation of all three N-terminal species; multiple analyses of the 18 kDa digest products resulting from the 29 kDa intermediate, however, did result in observation of all three N-terminal species.

A progressive series of cleavages is observed by LC/ESI-MS at the C-terminus of hER-LBD. The first protein fragments produced (29 kDa), consisting of a mixture of proteins with N-termini of S_{301}, K_{303}, and N_{304}, display a single C-terminal cleavage at R_{555}. Additional incubation with trypsin over the 10−120 min period yields progressively more internal cleavage sites at R_{548}, K_{531}, and K_{529}. After prolonged trypsin incubation, these three sites are found as C-terminal residues in the 29 kDa (R_{555}), 10 kDa (R_{555}, R_{548}), and 7 kDa (K_{531}, K_{529}) trypsin-generated fragments.

A sole site of trypsin proteolysis that is much more internal in the hER-LBD sequence is observed at residue K_{467} almost immediately upon trypsin treatment. This trypsin site results in rapid conversion of the original 33 kDa protein into stable 18 kDa and 12/10 kDa proteins corresponding to $N_{304}-K_{467}$ and $S_{468}-A_{571}/S_{468}-R_{548}$, respectively, via the initial 29 kDa intermediate. It is interesting to note that even upon

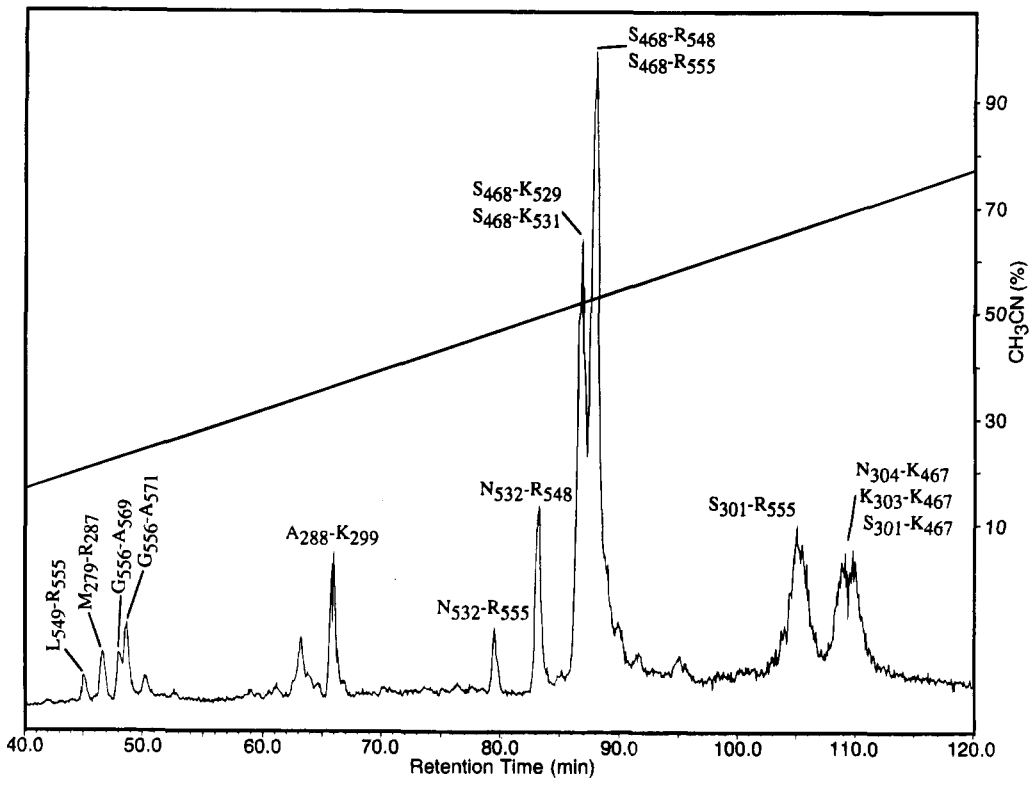

FIGURE 3: Reversed-phase high performance liquid chromatography (RP-HPLC) trace of protein fragments generated by proteolysis of hER-LBD with trypsin ([1:10] protease:receptor, 30 min, 25 °C). Trace represents total ion chromatogram (TIC) resulting from scans from 300 *m/z* − 2160 *m/z* over 4 s intervals. Elution gradient is overlaid for clarity. The protein and peptide components for the major peaks are indicated.

prolonged treatment, no trypsin cleavage is observed at the nearby K_{472} site. We have also identified two C-terminal limit digest fragments, N_{532}−R_{548} and L_{549}−R_{555}, formed during the conversion of the 10 kDa to the 7 kDa fragment.

Chymotrypsin. A pattern similar to trypsin digestion emerges upon analysis of the fragments of hER-LBD generated by chymotrypsin treatment. The *29 kDa* species seen on SDS-PAGE, is shown by LC/ESI-MS to be a mixture of 31 kDa and 29 kDa proteins, both of which have suffered N-terminal cleavage at L_{296} and M_{297}. These residues represent the primary specificity sites for chymotrypsin which are situated closest to the region of initial trypsin cuts (residues 300−303). While the C-terminus of the 31 kDa species extends to the end (A_{571}), the shorter 29 kDa species has been cut at T_{553}, a position close to where trypsin also makes an initial cut (R_{555}). Threonine represents a known site of secondary specificity for chymotrypsin cleavage (Haschemeyer and Haschemeyer, 1973).

Chymotrypsin, like trypsin, also demostrates ready access to an internal region of the LBD sequence, cleaving the native protein at residue L_{466} (adjacent to the trypsin clip at K_{467}). This cut within the initial 31 kDa chymotrypsin fragment produces an N-terminal 19 kDa and a C-terminal 12 kDa species. A *6 kDa* fragment, observed by SDS-PAGE and having an N-terminus at K_{467} (as determined by Edman degradation), appears to arise from the 12 kDa protein. We have so far been unable to locate this 6 kDa fragment in the HPLC trace by ESI-MS.

An additional 10 kDa species, observed at very low levels by ESI-MS, appears to be most consistent with the protein fragment I_{298}−W_{383}. The expected chymotryptic fragments flanking this purported cleavage product, however, have remained obscured in the ESI-MS background. Thus, it is presently not clear whether cleavage at W_{383} represents a chymotrypsin recognition site in native hER-LBD, or whether this 10 kDa protein fragment represents a transient intermediate generated during degradation of a denaturing LBD.

Thermolysin. The general, hydrophobic-selective endoproteinase thermolysin (Matsubara, 1970),[5] like trypsin and chymotrypsin, generated a rich pattern of discrete and reproducible protein fragments from the native hER-LBD. Rapid, initial thermolysin cleavage sites were identified within both the flanking N-terminal region (at N_{290} and M_{297}) and C-terminal region (at L_{568} and H_{567}). Additionally, two internal sites of protease accessibility at T_{465} and S_{468} are revealed by thermolysin in the region already identified as an internal recognition site by trypsin and chymotrypsin, K_{467} and L_{466}, respectively. Proteolysis by thermolysin initially generates a 31 kDa protein fragment (I_{298}−H_{567}), which is rapidly degraded by proteolysis at the T_{465} and S_{468} sites to generate 19 kDa and 12 kDa proteins. Prolonged treatment results in degradation of the 12 kDa C-terminal species by proteolysis at residue R_{548} to generate 9 kDa and 2 kDa protein fragments.

V8 and Asp-N Endoproteinases. These two proteinases[5] afford cleavages exclusively outside the N_{304}−K_{529} ligand binding core. In each case, the cleavage sites observed within the residues flanking the hER-LBD core are consistent with the known selectivity of the relevent proteinase (cf., Table 1).

In studies using the nonspecific proteinase subtilisin, we have been unable to detect the formation of any *discrete* intermediate protein species arising from the degradation of hER-LBD. Instead, a diffuse series of protein bands appear

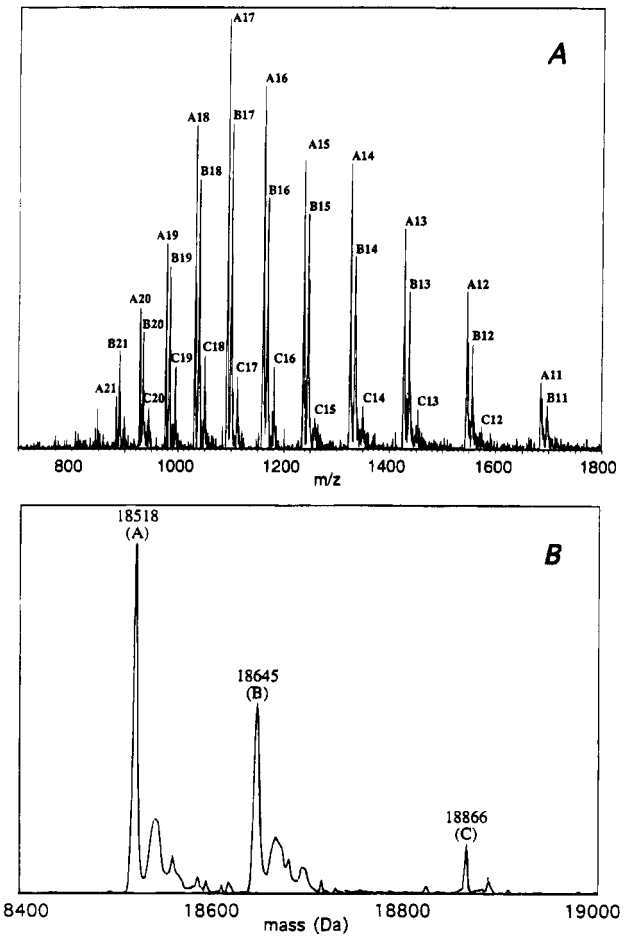

FIGURE 4: Electrospray ionization mass spectrum of the 18 kDa trypsin fragment of hER-LBD. This peak elutes at 110 min in the chromatogram shown in Figure 3. **Panel A.** ESI-MS displays three major ion series (A, B, and C; numbers represent multiply-charged state for each peak). **Panel B.** Zero-charge mass deconvolution of the ESI MS shown in panel A, revealing that the 18 kDa peak consists of three major mass species that correspond to three unique trypsin-generated species: 18518 ($N_{304}-K_{467}$), 18645 ($K_{303}-K_{467}$), and 18866 ($S_{301}-K_{467}$).

by SDS-PAGE, and an insufficient protein signal was obtained when the lysate was analyzed by LC/ESI-MS.

DISCUSSION

We have used protease mapping to analyze the structure and conformation of the ligand binding domain of the human estrogen receptor (hER-LBD) expressed in *E. coli* and purified as a complex with estradiol. The hER-LBD preparation used in these studies consists primarily of the sequences $MDPS_{282}-A_{571}$ and $MDPS_{282}-A_{569}$,[4] with minor amounts of species ending at S_{575}, and V_{595} (full length construct, Seielstad et al., 1995). As we (Katzenellenbogen, B. S., et al., 1987; Elliston & Katzenellenbogen, 1988; Seielstad et al., 1995) and others (Thole, 1993; Fritsch et al., 1993; Thole & Jungblut, 1994) have noted, the ligand binding domain of ER is relatively resistant to proteolysis when occupied by estradiol. Some of the proteinases (V8, Arg-C, and Asp-N) fail to cleave internal to this domain, despite prolonged incubation times. The most active proteases, trypsin, chymotrypsin, and thermolysin, however, effect a progression of cleavages within the ligand binding domain. We have identified the most prominent of these fragments, using the LC/ESI-MS to measure their masses accurately and then fitting these masses to the possible fragments of the known hER sequence. In some cases, we have established the time course of their generation. From these analyses emerges the following picture for the structural nature of this domain.

Flexibility of the N- and C-Termini of the Human Estrogen Receptor Ligand Binding Domain Probed by Proteolysis. Our hER-LBD construct begins with the sequence $MDPS_{282}$[4] and thereby contains 25 residues that are N-terminal to K_{303}. This residue is generally considered to represent the N-terminus of the ligand binding domain of ER on the basis of sequence homology with other nuclear steroid receptors (Parker, 1991). Four of the proteases reported, namely trypsin, chymotrypsin, thermolysin, and Asp-N endoproteinase, effect cleavage in this 282-303 region (cf., Table 1): Asp-N at G_{284}, thermolysin at M_{297}, and chymotrypsin at L_{296} and M_{297}. However, trypsin effects the deepest cleavages into the N-terminus of our construct, cleaving first at R_{300}, but then progressing rapidly to K_{302} and K_{303}, leaving N_{304} as a stable N-terminus. Thus, as probed by proteolysis, the N-terminus of the hER-LBD appears to begin with a tightly wound, protease-resistant region at N_{304}.

The C-terminus of our hER-LBD preparation is heterogeneous due to three cleavages (at A_{569}, A_{571}, and S_{575}) that are presumably caused by the *E. coli* signal leader peptidase during its isolation and purification from inclusion bodies (Seielstad et al., 1995). Thus, the predominant species present contains 18−22 residues beyond T_{553}, which, also by sequence homology analysis, is generally considered to represent the C-terminus of the ligand binding domain in ER (Parker, 1991). All of the proteases reported here effect cleavages near the C-terminus, although the cuts are less localized and more progressive than those observed at the N-terminus. Initial cuts produced by the proteases thermolysin at H_{567}, V8 at D_{564}, Asp-N at T_{563}, trypsin at R_{555}, and chymotrypsin at T_{553}, are consistent with T_{553} being the C-terminus of the hER-LBD. In contrast with proteolytic resistance observed at the N-terminus, however, continued incubation of the hER-LBD with thermolysin affords an additional cut at R_{548}, while trypsin affords three further cleavages, at R_{548}, K_{531}, and K_{529}. These final two trypsin cleavages represent the ultimate removal of 23−25 residues from the C-terminal end of domain E. Thus, the C-terminus of the LBD is less precisely defined in terms of protease accessibility than is the N-terminus of this domain.

The increased proteolytic susceptibility of the 529−553 C-terminal region of the hER-LBD, as compared to the N-terminus, is noteworthy and may be related to ligand binding and other functions of this domain. Since protease invasion from the C-terminal end stops at K_{529}, this may represent the termination of a tightly-wound region associated with the structural integrity of this domain (similar in character to N_{304} at the N-terminus of this domain) and its function in ligand binding. In fact, hER, trypsinized to effect cuts at N_{304} and K_{529}, still retains bound estradiol (Carlson & Katzenellenbogen, unpublished; see also, Fritsch et al., 1993; Thole & Jungblut, 1994). Sequences that immediately preceed K_{529} have been identified through mutational analysis as being necessary for ligand binding, while those that follow C_{530} appear to be largely dispensible (Fawell et al., 1990). This is supported by the finding that the truncated ER mutants hER 1−530 and mER 120−538 (the latter of which corresponds to hER 116−534) retain 5% and 30% of the

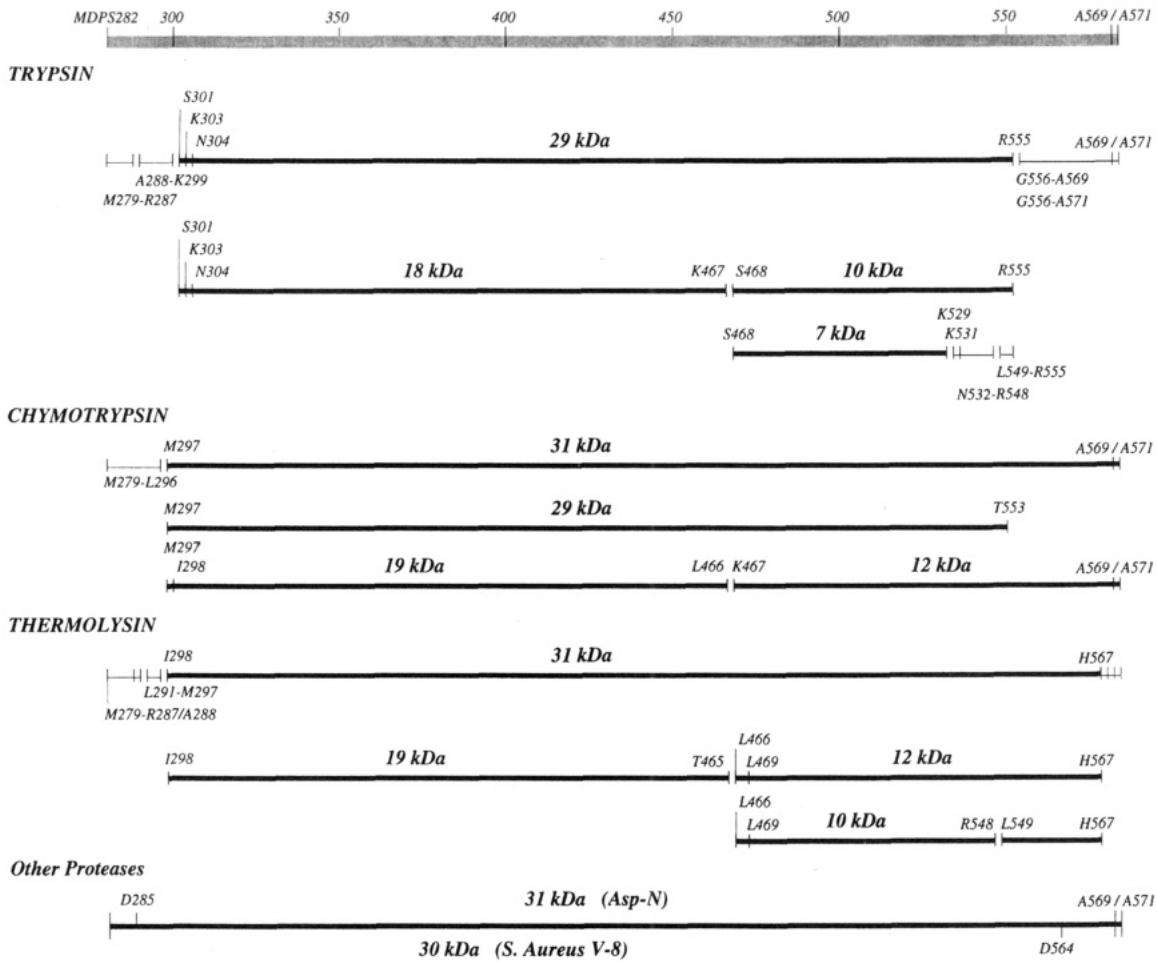

FIGURE 5: Summary map of the proteolysis sites of the hER-LBD by trypsin, chymotrypsin, thermolysin, V8 proteinase, and Asp-N endoproteinase. Bold lines indicate core fragments: light lines are limit digests.

affinity of full length estrogen receptor for estradiol, respectively (Lees et al., 1989; Wooge et al., 1992; Katzenellenbogen, B. S., et al., 1993; Wrenn & Katzenellenbogen, 1993). Thus, the portion of the hER-LBD extending from N_{304} to K_{529} appears to contain the essential core for binding estradiol. The C-terminal region of the progesterone receptor appears to play a related role in ligand binding: When significant C-terminal portions of the progesterone receptor, homologous to the 529−553 region of hER, are deleted, this receptor no longer binds progesterone; the binding of the antiprogestin RU486, however, is unaffected by these deletions, although it acts through these truncated receptors as an agonist (Vegeto et al., 1992).

While the region of domain E between 529 and 553 does not appear to play an essential role in ligand binding, sequence homology in this region is very high across seven reported ER sequences, as well as in the analogous regions of other steroid (type I) and non-steroid (type II) nuclear receptors (Figure 6 and legend). This region of ER encompasses a sequence (534−548) onto which the ligand-modulated transcriptional activation function (AF-2) has been localized by mutational analysis (Danielian et al., 1992); the homologous C-terminal regions in other nuclear receptors also encompass this function (Vegeto et al., 1992; Barettino et al., 1994; Baniahmad et al., 1995; Tone et al., 1994). As transcription activation presumably involves interactions with coactivators or corepressors, or with elements of the basal transcription apparatus, it seems reasonable that the sequence responsible for this activity be surface accessible and thereby subject to proteolytic attack.

This AF-2 region of nuclear hormone receptors has been proposed to adopt a short amphipathic helical structure (Danielian et al., 1992), and at least in some systems it appears able to fold as an autonomous unit, as this region retains its transcription activating function when it is fused to other DNA-binding proteins (Barretino et al., 1994; Baniahmad et al., 1995). The cuts we see at 529, 531, and 548 lie outside of the proposed helix: Those at 529 and 530 in hER precede a region where all of the nuclear receptors have one or two proline residues; the 548 site in hER is at or close to the C-terminal end of several of the receptors, and in the others it is followed by a proline within one or two residue positions (Figure 6). While there are no potential trypsin cleavage sites between K_{531} and R_{548}, there are many sites where several of the other proteases might have cleaved. So, lack of cleavage in the 531−548 sequence suggests that it is a tightly structured region.

AF-2 represents a region of ligand-modulated activity, and thus the structure of the 529−553 region that accommodates this function and its exposure to proteolysis could indeed be regulated by bound ligand. While we have investigated in this study the proteolytic sensitivity of hER-LBD only when complexed with estradiol, there are several reports that the C-terminus of the ligand binding domain of several nuclear receptors, as assessed by protease sensitivity, can assume different conformations when bound with agonist,

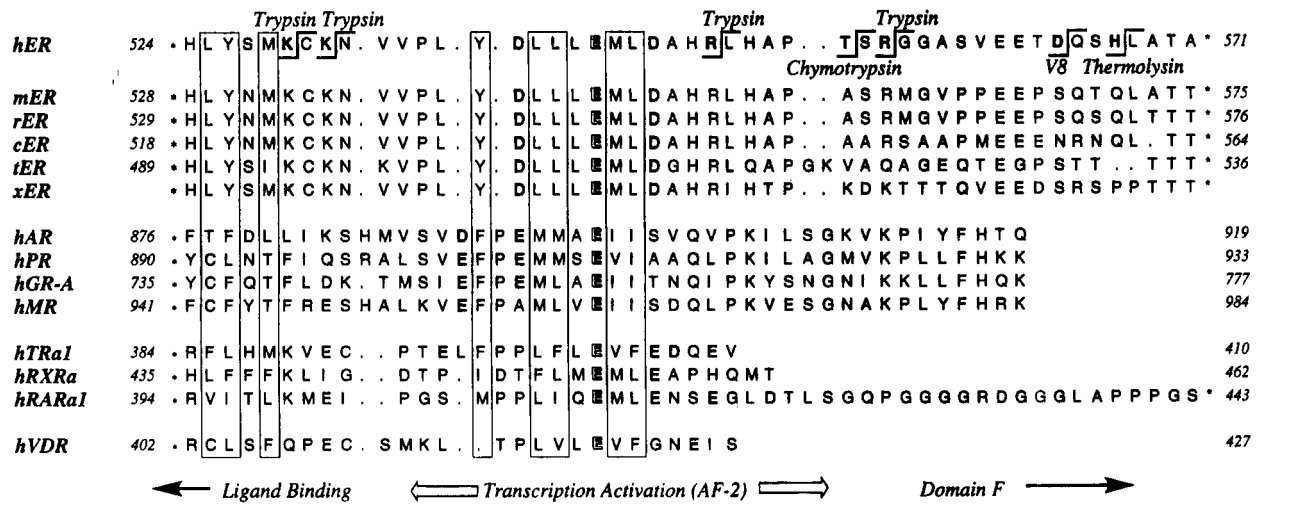

FIGURE 6: Sequence comparison of the C-terminal portion of the ligand binding domain for various nuclear receptors. Sequences of this Domain E/F junction were aligned with respect to the human estrogen receptor construct used in this study using Clustal W (Thompson et al., 1994) with some subsequent adjustments performed by hand. Residue numbers represent the first and the last residue displayed; except where indicated by an asterisk, the C-terminus of each receptor is shown. The protease sites in this region which are identified in this work are indicated on the hER sequence. The conserved glutamic acid residue implicated in transactivation functions is indicated by bold/outlined letters, while more poorly conserved acidic residues flanking this position are indicated by outlined letters. Conserved, hydrophobic sequences, which may be involved in ligand modulated activity in this region, are boxed. Sequence conservation and mutational studies suggest that the ligand binding domain and associated transactivation functions are localized to residues before R_{548} (hER). Prefix letters for sequence abbreviations are as follows: chicken (c), mouse (m), human (h), rat (r), trout (t), xenopus (x).

antagonist, or no ligand: ER (Beekman et al., 1993), progesterone receptor (Allan et al., 1992; Vegeto et al., 1992; Weigel et al., 1992), and retinoid receptors (Allan et al., 1992; Keidel et al., 1994; Leid, 1994).

A Surface Exposed Region within the Human Estrogen Receptor Ligand Binding Domain: Possible Subdomain Structure. Prior to the degradation of the C-terminus of the ligand binding domain by trypsin, chymotrypsin, or thermolysin these proteases appear to effect a cleavage deep within the sequence of the 26 kDa core structure (N_{304}–T_{553}). Rapid proteolysis is observed at K_{467} for trypsin, L_{466} for chymotrypsin, and at both T_{465} and S_{468} for thermolysin. In fact, trypsin cleavage sites at R_{548}, K_{531}, and K_{529}, as well as the thermolysin cuts at H_{567} and R_{548}, and the chymotrypsin cleavage of the *12 kDa* fragment into the *6 kDa* protein fragment may all require prior cleavage at this more internal 465–468 region. Indeed, the endoproteinases Asp-N and V8, while possessing numerous recognition sites within the C-terminus of domain E, should not and do not produce any internal cleavage in the 465–468 region, nor do they display any C-terminal degradation. This region within the hER-LBD that is accessible to proteases, T_{465}–S_{468}, does not correspond to any function of the ER that has as yet been determined by chemical probe or mutagenesis studies. However, the fact that three different proteases can effect cleavage here indicates that this region of the LBD is potentially quite flexible or randomly ordered, and that the entire LBD N_{304}–T_{553} may, in fact, consist of two subdomains.

Both structural studies and computational modeling of sites in protein domains that are cleaved by serine proteases indicate that some 10–12 residues surrounding the site of cleavage (i.e., residues F_{461}–K_{472} in the case of the internal cuts we observe) must be at or near the surface and able to adopt an extended conformation to be accessible to proteolysis (Hubbard et al., 1991, 1994). In our recently published model for nuclear ligand hormone binding domains, based on the subtilisin fold (Goldstein et al., 1993), the internal protease sensitive region (465–468) falls at a surface-accessible strand-turn-helix conformation that could uncoil to adopt an extended strand suitable for recognition by trypsin, chymotrypsin, or thermolysin. In a different model for the thyroid hormone receptor, based on a TIM barrel (Cheng et al., 1994), this protease sensitive region is in a similar strand-turn-helix region. It will be interesting to investigate whether the unusual intradomain flexibility of this region, which permits cleavage by certain proteases, is related to ligand binding specificity or to other noteworthy functions of the ligand binding domain.

From proteolysis studies on tamoxifen aziridine-labeled ER, we (Katzenellenbogen, B. S., et al., 1987; Elliston & Katzenellenbogen, 1988), Gorski (Fritsch et al., 1993), and Thole (Thole, 1993; Thole & Jungblut, 1994) have identified a *7–10 kDa* fragment that includes the site of labeling by tamoxifen aziridine, namely C_{530} (Harlow et al., 1989). By Edman sequencing, Thole identified the N-terminus of this trypsin-generated fragment as S_{468}, the result of cleavage at K_{467} (Thole & Jungblut, 1994). By ESI-MS, we identify this 7 kDa trypsin fragment as a mixture of the fragments S_{468}–K_{529} and S_{468}–K_{531}. Whether this 7 kDa fragment *alone* can bind ligand, as was initially proposed (Thole, 1993), remains unanswered. In ER, both Thole (Thole & Jungblut, 1994) and Gorski (Fritsch et al., 1993) have demonstrated that the LBD clipped by trypsin can still bind estradiol and is labeled on the *7 kDa* fragment by tamoxifen aziridine. Under these conditions, however, the *7 kDa* fragment remains noncovalently associated with the larger *19 kDa* fragment (Fritsch et al., 1993; Thole & Jungblut, 1994). By ESI-MS, we characterize this larger 19 kDa partner as being N_{304}–K_{467}. Thus, together with the 7 kDa fragment (S_{468}–K_{531}), these species constitute nearly the whole of domain E and represent what we term a 26 kDa proteolytically resistant core. Similarly, Simons has described a subdomain fragment of mass *16 kDa*, found upon trypsin proteolysis of the *unliganded* rat glucocorticoid receptor (rGR). This fragment can be covalently labeled with

dexamethasone mesylate subsequent to proteolysis, but proteolytic cleavage is blocked by prelabeling rGR with the same agent (Chakraborti et al., 1992). While it has been suggested that the *16 kDa* fragment from rGR *alone* may be capable of binding ligand, it has not been established that this fragment does so as an independent protein fragment. Indeed, it may still remain noncovalently associated with its C-terminal partner under the assay conditions, as is observed with the *19 kDa* and *7 kDa* trypsin-generated fragments of ER (Fritsch et al., 1993; Thole & Jungblut, 1994).

Protease mapping, by the selective cleavage of exposed regions of proteins in their native state, coupled with definitive identification of the peptide fragments thereby generated by LC/ESI-MS, is a powerful technique for probing the topology of proteins of unknown structure. When applied to the hER-LBD, we have been able to establish that the structural N-terminus of the hormone binding domain begins with a tightly wound region at N_{304}, coincident with the onset of high conservation of sequence homology. The C-terminus is less clearly defined by this method; cleavages occur progressively inward from 564 to 553, nominally considered to be the end of domain E, and continue further to 548, 531, and 529. The region up to K_{529} may represent the terminus of a tightly-wound region associated with ligand binding, while the region 530−553, which encompasses AF-2, appears to contain surface-exposed elements for interaction with other factors and is thus accessible to protease cleavage. An internal, protease sensitive site is also found at 465−468; such a site is also found in the glucocorticoid receptor, suggesting that the ligand binding domain of steroid hormone receptors may, in general, consist of two subdomains.

ACKNOWLEDGMENT

We are grateful to Dr. Richard Milberg for assistance with ESI-MS and to Dr. K.-L. Ngai for helpful discussions. N-Terminal analysis was performed by the Genetic Engineering Facility at the University of Illinois. G.L.G. gratefully acknowledges the technical assistance of Chris Hospelhorn and Lin Cheng.

Note Added in Proof: While this work was under review, a report appeared (Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., & Moras, D. (1995) *Nature 375*, 377−382) in which the novel eleven α-helical sandwich structure of the ligand binding domain of the RXR-α receptor without ligand (RXR-LBD) was determined by X-ray crystallography. The sequence homology between ER-LBD and RXR-LBD is sufficient (26% identity, 46% similarity) for us to evaluate our ER-LBD topology mapping results in the context of this new structural information.

The N-terminus of the RXR-LBD begins ordered structure at S_{225}; this corresponds to S_{309} in ER-LBD, close to where we proposed the domain structure to start, based on protease resistance. The C-terminus of RXR-LBD displays residues including the transactivation domain (AF-2) as an exposed helix-strand element; this region corresponds to the C-terminal sequence D_{545} of the ER-LBD, a region that we found to undergo progressive proteolysis. The sequence alignment between hER and RXR-α is fairly good in the region where ER-LBD undergoes internal proteolysis (hER 465−468), but is characterized by several additional residues in the ER sequence. This suggests that the protease accessible site in hER is located on a larger coiled region immediately preceding helix 9 in the RXR-LBD structure. Thus, the topological conclusions we made from the results of our protease sensitivity studies on ER-LBD are consistent with the topology of this domain in the RXR-α. A further report by Thole [Thole, H. H., Maschler, I., & Jungblut, P. W. (1995) *Eur. J. Biochem.*, 510−516] on surface mapping of porcine estrogen receptor by limited proteolysis using SDS-PAGE/Edman sequencing methods provides results similar to ours. Their characterization of the cleavage sites near the C-terminus of the ligand binding domain is less complete, but they identify additional cleavage sites internal to the domain when it is in a monomeric state.

REFERENCES

Aebersold, R. (1993) *Curr. Opin. Biotechnol. 5*, 77−84.

Aitken, A., Geisow, J. J., Findlay, J. B. C., Holmes, C., & Yarwood, A. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 43−68, IRL Press, Oxford.

Allan, G. F., Leng, X., Tsai, S. Y., Weigel, N. L., Edwards, D. P., Tsai, M.-J., & O'Malley, B. W. (1992) *J. Biol. Chem. 267*, 19513−19520.

Anker, H. S. (1970) *FEBS Lett. 7*, 293.

Baniahmad, A., Leng, X., Burris, T. P., Tsai, S. Y., Tsai, M.-J., & O'Malley, B. W. (1995) *Mol. Cell. Biol. 15*, 76−86.

Barettino, D., Vivianco Ruiz, M. d.M., & Stunnenberg, H. G. (1994) *EMBO J. 13*, 3039−3049.

Beato, M. (1989) Gene regulation by steroid hormones, *Cell 56*, 335−344.

Beekman, J. M., Allan, G. F., Tsai, S. Y., Tsai, M.-J., & O'Malley, B. W. (1993) *Mol. Endocrinol. 7*, 1266−1274.

Carrey, E. A. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 157−144, IRL Press, Oxford.

Chait, B. T. (1994) *Structure 2*, 465−467.

Chakraborti, P. K., Garabedian, M. J., Yamamoto, K. R., & Simons, S. S., Jr. (1992) *J. Biol. Chem. 267*, 11366−11373.

Cheng, S., Ransom, S. C., McPhie, P., Bhai, M. K., Mixson, A. J., & Weintraub, B. D. (1994) *Biochemistry 33*, 4319−4326.

Danielian, P. S., White, R., Lees, J. A., & Parker, M. G. (1992) *EMBO J. 11*, 1025−1033.

Drapeau, G. R. (1980) *J. Biol. Chem. 255*, 839−840.

Elliston, J. F., & Katzenellenbogen, B. S. (1988) *J. Steriod Biochem. 29*, 559−569.

Evans, R. M. (1988) *Science 240*, 889−895.

Fawell, S. E., Lees, J. A., White, R., & Parker, M. G. (1990) *Cell 60*, 953−962.

Fritsch, M., Anderson, I., & Gorski, J. (1993) *Biochemistry 32*, 14000−14008.

Goldstein, R. A., Katzenellenbogen, J. A., Luthey-Schulten, Z. A., Seielstad, D. A., & Wolynes, P. G. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9949−9953.

Gross, M. L. (Ed.) (1993) J. *Am. Soc. Mass. Spectrom. 4*, 523−684.

Harlow, K. W., Smith, D. N., Katzenellenbogen, J. A., Greene, G. L., & Katzenellenbogen, B. S. (1989) *J. Biol. Chem. 264*, 17476−17485.

Haschemeyer, R. H., & Haschemeyer, A. E. V. (1973) *Proteins*, p 83, John Wiley & Sons, New York.

Houmard, J., & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3506−3509.

Hubbard, S. J., Campbell, S. F., & Thornton, J. M. (1991) *J. Mol. Biol. 220*, 507−530.

Hubbard, S. J., Eisenmenger, F., & Thornton, J. M. (1994) *Protein Science 3*, 757−768.

Katzenellenbogen, B. S., Elliston, J. F., Monsma, F. J., Jr., Springer, P. A., Ziegler, Y. S., & Greene, G. L. (1987) *Biochemistry 26*, 2364−2373.

Katzenellenbogen, J. A., Carlson, K. E., Heiman, D. F., Robertson, D. W., Wei, L. L., & Katzenellenbogen, B. S. (1983) *J. Biol. Chem. 258*, 3487−3495.

Katzenellenbogen, B. S., Bhardwaj, B., Fang, H., Ince, B. A., Pakdel, F., Reese, J. C., Schodin, D., & Wrenn, C. K. (1993) *J. Steroid Biochem. Mol. Biol. 267*, 39−48.
Keidel, S., LeMotte, P., & Apfel, C. (1994) *Mol. Cell. Biol. 14*, 287−298.
Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornert, J.-M., & Chambon, P. (1986) *EMBO J. 5*, 891−897.
Lees, J. A., Fawell, S. E., & Parker, M. G. (1989) *Nucleic Acids Res. 17*, 5477−5488.
Leid, M. (1994) *J. Biol. Chem. 269*, 14175−14781.
Matsubara, H. (1970) *Methods Enzymol. 19*, 642−651
Parker, M. G. (1991) *Nuclear Hormone Receptors*, Academic Press, London.
Pavlik, E. J., & Katzenellenbogen, B. S. (1980) *Mol. Pharm. 18*, 406−412.
Schagger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368−379.
Schecter, I., & Burger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157−162.
Seielstad, D. A., Carlson, K. E., Greene, G. L., Kuschner, P. J., & Katzenellenbogen, J. A. (1995) *Mol. Endocrinol. 9*, 647−658.
Sherman, M. R., Pickering, L. A., Rollwagen, F. M., & Miller, L. K. (1978) *Fed. Proc. 37*, 167−173.
Thole, H. H. (1993) *FEBS Lett. 320*, 92−96.
Thole, H. H., & Jungblut, P. W. (1994) *Biophys. Res. Commun. 199*, 826−833.
Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.
Tone, Y., Collingwood, T. N., Adams, M., & Chatterjee, V. K. (1994) *J. Biol. Chem. 269*, 31157−31161.
Vegeto, E., Allan, G. F., Schrader, W. T., Tsai, M.-J., McDonnell, D. P., & O'Malley, B. W. (1992) *Cell 69*, 703−713.
Weigel, N. L., Beck, C. A., Estes, A., Prendergast, P., Altmann, M., Christensen, K., & Edwards, D. P. (1992) *Mol. Endocrinol. 6*, 1585−1597.
Wilkinson, J. M. (1986) in *Practical Protein Chemistry* (Darbre, A., Ed.) p 127, John Wiley & Sons, New York.
Wooge, C. H., Nilsson, G. M., Heierson, A., McDonnell, D. P., & Katzenellenbogen, B. S. (1992) *Mol. Endocrinol 6*, 861−869.
Wrenn, C. K., & Katzenellenbogen, B. S. (1993) *J. Biol. Chem. 268*, 24089−24098.

BI950720W