*Bioorganic & Medicinal Chemistry Letters*, Vol.3, No.6, pp. 1095-1100, 1993
Printed in Great Britain

0960-894X/93 $6.00 + .00


# LIMITED PROTEOLYSIS AND SITE-DIRECTED MUTAGENESIS OF THE NF-κB p50 DNA-BINDING SUBUNIT

Mikiko Sodeoka,[†] Christopher J. Larson,[†] Lin Chen,[†] William S. Lane,[§] and Gregory L. Verdine*[†]

[†]*Department of Chemistry and* [§]*Microchemistry Facility, Harvard University Cambridge, Massachusetts 02138*

(*Received in USA* 11 *February* 1993)

## ABSTRACT

Tryptic digestion was used to obtain an active 41 kDa fragment of NF-κB. The major proteolysis products were comprised of residues 2-364 ($p50_{2-364}$), 2-362 ($p50_{2-362}$) and 2-366 ($p50_{2-366}$). By site-directed truncation mutagenesis, overproducers for $p50_{2-364}$ and $p50_{2-366}$ were constructed. These vectors direct high level expression of homogeneous, active protein.

## INTRODUCTION

The transcription factor NF-κB is the focus of intense interest because of its central role in controlling inflammatory, immune, and acute phase responses, and proliferation of viruses including HIV-1, cytomegalovirus, and adenovirus.[1] In most resting cell types, NF-κB is anchored in the cytoplasm bound to an inhibitory protein termed I-κB.[2] Signals from the cell surface are believed to result in phosphorylation of I-κB, thereby releasing it from NF-κB, which diffuses into the nucleus and turns on genes. The elucidation of this series of events is a major triumph for modern biology, as it represents the first case in which the extramembrane, cytopasmic, and nuclear components of a signalling pathway could be connected via known molecules.

NF-κB has generated interest not only in its biological activities, but has also been the subject of considerable intrigue with regard to structure. Both subunits of NF-κB - p50 and p65 - possess homology to the Rel proto-oncoprotein, as do a number of other proteins that appear to play important roles in immune function and carcinogenesis.[3,4] The common structural feature of these proteins, the Rel homology region (RHR), is a ~280 amino acid domain that permits dimerization and DNA binding. Point mutations and deletions throughout the RHR abrogate DNA binding, whereas changes outside the RHR have no effect; hence the entire RHR is both necessary and sufficient for binding.[5] This is unlike the situation for most DNA-binding proteins studied to date, in which the DNA-binding domain can often be localized to a stretch of fewer than 100 contiguous amino acids.[6] The large size of the RHR has hampered efforts at characterizing its structure and function by biological techniques.

In order to gain structural information on the RHR, we have begun a program of overproduction, crystallization, and biochemical studies of Rel family proteins, focusing on NF-κB. The initial target of our investigation has been the homodimeric form of p50, the overproduction of which was reported in the previous paper.[7] Recombinant p50 was found to be soluble and active for DNA binding, yet was "nibbled" by proteases at its C-terminal end.[8] Out of concern that this C-terminal homogeneity would adversely affect crystallization, we

undertook the present study to (i) define a proteolytically stable core DNA-binding domain in p50, and (ii) overproduce such proteins at high levels in *E. coli*.

## RESULTS

In an attempt to reduce the C-terminal heterogeneity of recombinant p50, we carried out limited proteolytic digestion using trypsin (Seravac, Maidenhead, England). The p50 mixture was rapidly reduced to a protein of ~41 kDa which was substantially resistant to further proteolysis (Figure 1).[5] The ~41 kDa fragment was then prepared on a large scale and purified to homogeneity (Figure 2, lane 2).[10] This preparation was shown to bind DNA with a strength and specificity indistinguishable from that of recombinant p50 (Figure 3, lane 3, and additional data not shown).[11,12]

The 41 kDa trypsin digestion product was characterized by N-terminal sequencing and electrospray-ionization mass spectrometry.[13] The N-terminus was determined to be AEDDP,[14,15] which corresponds to that of recombinant p50, thereby revealing that proteolysis had not taken place at the N-terminal end. Electrospray-ionization mass spectrometry revealed the presence of three predominant tryptic products having molecular weights 40,558 Da, 40,847 Da, and 41,102 Da (Figure 4, A and B).[16] These values correspond well to the predicted molecular weights of proteins possessing residues 2-362 ($p50_{2\text{-}362}$; $M_r$ 40,562), 2-364 ($p50_{2\text{-}362}$; $M_r$ 40,847), and 2-366 ($p50_{2\text{-}366}$; $M_r$ 41,102). Thus, proteolysis had removed the entire glycine-rich stretch at the C-terminal end of the protein, leaving fragments that terminated within the nuclear localization signal at the end of the RHR (Figure 5).

Taking advantage of the ability to generate single-stranded DNA with overproducers derived from pLM1, we carried out oligonucleotide-directed mutagenesis on the p50 overproducer (pLM1-p50) in order to create new vectors possessing stop codons at positions 365 and 367 in the p50 coding sequence; a construct having a stop codon at position 363 was not generated. The resulting vectors, pLM1-p502-364 and pLM1-p502-366, direct expression of residues 2-364 and 2-366, respectively.[14] Upon induction of protein synthesis, these overproducers proved to possess outstanding efficiency, with the recombinant protein being expressed at levels substantially higher than any of the native *E. coli* proteins (Figure 2, compare lanes 3 and 4). From these cells, soluble $p50_{2\text{-}364}$ was purified to yield ~25 mg/L of cultured cells as determined by Bradford assay; analogous results were obtained for $p50_{2\text{-}366}$. $p50_{2\text{-}364}$ and $p50_{2\text{-}366}$ were expressed and purified by the same procedure as was utilized for p50 in the previous communication.[17] N-terminal sequencing and electrospray ionization mass spectrometry on $p50_{2\text{-}366}$ (Figure 4, C and D) and $p50_{2\text{-}364}$ (Figure 4, E and F) confirmed in each case the expected composition.

Highly purified $p50_{2\text{-}366}$ and $p50_{2\text{-}364}$ were homogeneous, formed stable dimers,[12] and bound DNA with $K_d$'s that were essentially indistinguishable from each other and from p50 (Figure 3 and additional data not shown).

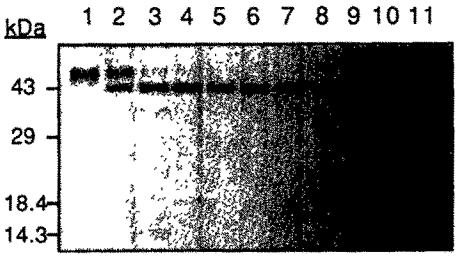


Figure 1. Digestion of p50 with trypsin. The purified p50 mixture (lane 1) was treated with trypsin (protein/trypsin: lane 2-4, 1000/1; lane 5-7, 100/1; lane 8-11, 10/1) at room temperature followed by the addition of SDS sample buffer after 1 min (lane 8), 5 min (lane 9), 20 min (lane 2, 5, and 10), 60 min (lane 3, 6, and 11), and 120 min (lane 4 and 7). Samples were analyzed by SDS-PAGE.

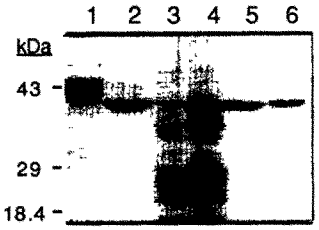


Figure 2. Induction and purification of proteins generated in this study, as analyzed by SDS-PAGE (5-20% gradient gel). Lane 1, purified p50; lane 2, product of limited tryptic digestion of p50; lanes 3 and 4, respectively, uninduced and induced *E. coli* BL21(DE3) harboring the plasmid pLM1-$p50_{2\text{-}364}$; lane 5, purified $p50_{2\text{-}364}$ lane 6, purified $p50_{2\text{-}366}$. Denoted on left are mobilities and sizes (in kDa) of molecular mass markers.

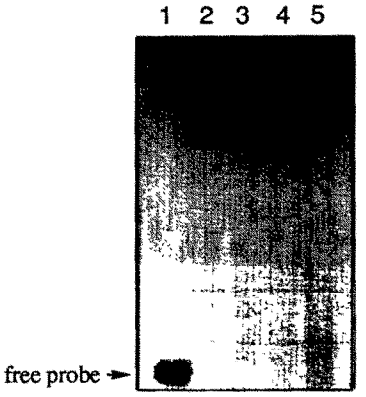


Figure 3. Specific DNA-binding activity of proteins generated in this study. Lane 1, no protein (control); lane 2, p50; lane 3, trypsin-digested p50; lane 4, $p50_{2\text{-}364}$; lane 5, $p50_{2\text{-}366}$. Competition binding experiments (not shown) showed that binding was specific in all cases.

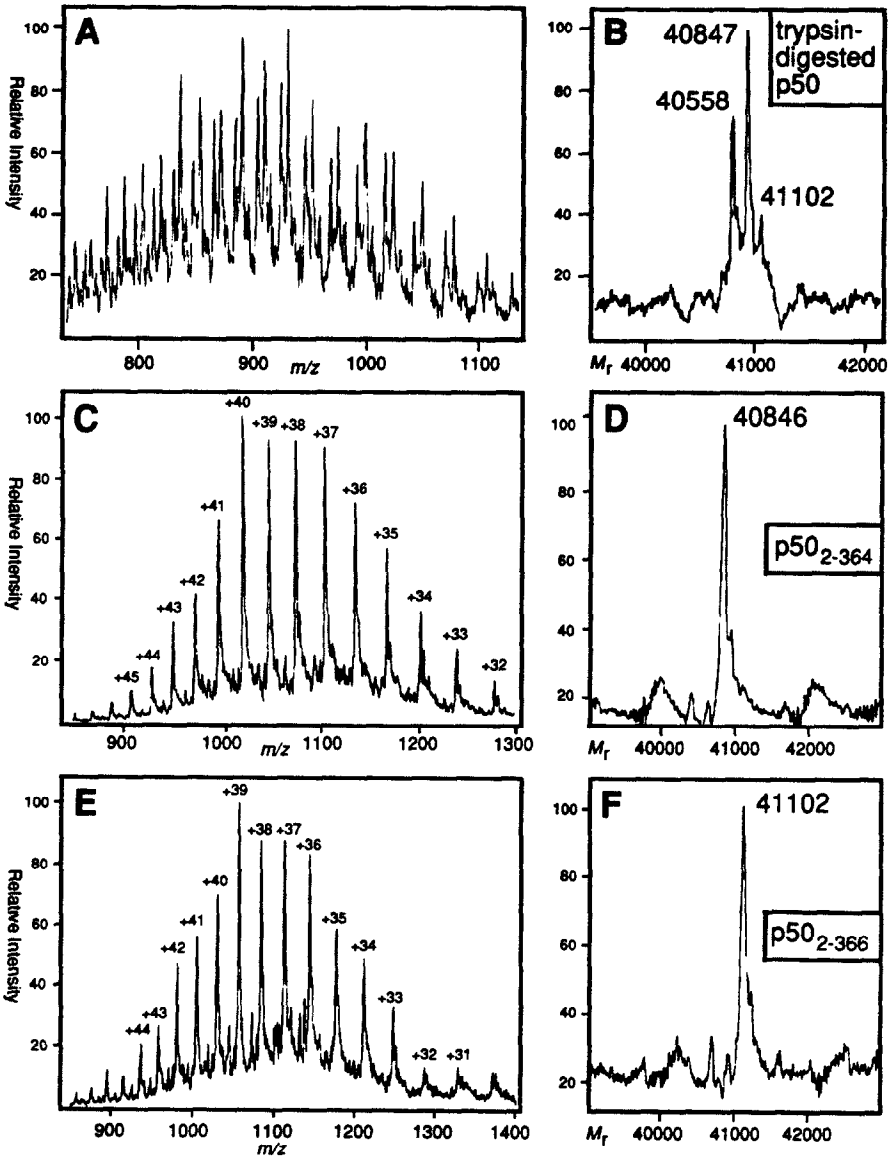


Figure 4. A, raw *m/z* spectrum of trypsin-digested p50, and B, deconvoluted relative molecular mass ($M_r$) spectrum, showing three major species; C, raw *m/z* spectrum of trypsin-digested $p50_{2\text{-}364}$, and D, deconvoluted molecular mass spectrum; E, raw *m/z* spectrum of $p50_{2\text{-}366}$, and F, deconvoluted molecular mass spectrum.

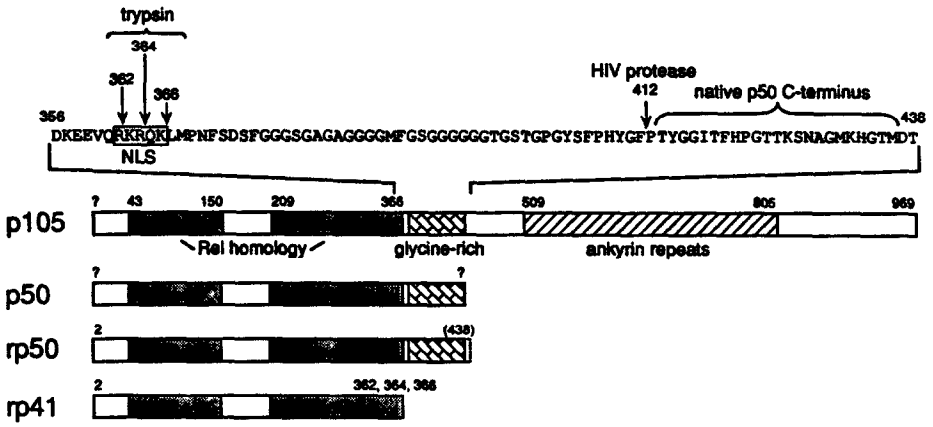


Figure 5. Schematic depiction of proteins generated by proteolysis and their relationship to native p50 and its precursor p105. The precise site of cleavage of p105 to produce p50 in human cells is not known. A human HIV protease site is shown.

## CONCLUSIONS

One object of this study, which has been achieved, was to generate homogeneous, active fragments of p50 that would be suitable for high-resolution structural analysis. From this and the preceding study we can now conclude that (i) p50 can readily be overexpressed in *E. coli*, but is susceptible to C-terminal proteolysis; (ii) p50 contains a flexible tail consisting of residues ~366 onward, which we suspect would impede crystallization; (iii) the flexible tail of p50 can be removed cleanly on a preparative scale by limited proteolysis with trypsin; and (iv) the 41 kDa core proteins containing the entire RHR can be overexpressed and purified to homogeneity on a large scale. These conclusions lead directly to an entry into crystallography studies. The exceptionally high levels of precisely engineered protein generated by the pLM1 system, as described in this and the previous communication, have also allowed us to pursue biochemical studies on the mode of DNA recognition by p50, which will be reported shortly.

On a related, technical note, this study underscores the power of combined N-terminal sequencing and EISMS analysis as a simple, rapid manner in which to determine the boundaries of proteolytic fragments.

In studies from these and other laboratories using pLM1 together with ECPCR, this vector system has shown uniformly excellent performance in overproduction, DNA sequencing, mutagenesis, and other protein expression applications.[18,19]

## REFERENCES AND FOOTNOTES

1. Lenardo, M.J., Baltimore, D. *Cell* **1989**, *58* , 227
2. Baeuerle, P.A. *Biochim. Biophys. Acta* **1991**, *1072* , 63
3. Neri, A., Chang, C.C., Lombardi, L., Salina, M., Corradini, P., Maiolo, A.T., Chaganti, R.S.K., Dalla-Favera *Cell* **1991**, *67* ,1075
4. Narayanan, R., Klement, J.F., Ruben, S.M., Higgins, K.A., Rosen, C.A. *Science* **1992**, *256* , 367
5. Ghosh, S., Gifford, A.M., Riviere, L.R., Tempst, P., Nolan, G.P., Baltimore, D. *Cell* **1990**, *62* , 1019; Kieran, M., Blank, V., Logeat, F., Vandekerckhove, J., Lottspeich, F., Le Bail, O., Urban, M.

B., Kourilsky, P., Baeuerle, P.A., Israel, A. *Cell* **1990**, *62* , 1007; Kumar, S., Rabson, A.B., Gelinas, C. *Molec. Cell. Biol.* **1992**, *12* , 3094

6. Harrison, S.C. *Nature* **1991**, *353* , 715
7. see previous communication in this issue
8. Figure 2, lane 1, and see previous communication.
9. Purified p50 mixture (40 mg in 100 mL of 50 mM ammonium bicarbonate, 1 mM $CaCl_2$, pH 8.0) was mixed with trypsin [for 1000/1 (w/w) p50/trypsin, 4 mL of 10 ng/mL; for 100/1, 40 mL of 10 ng/mL; for 10/1, 4 mL of 1 mg/mL solution in the same buffer) and incubated at room temperature. Aliquots (20 mL) were removed at 1, 5, 20, 60, and 120 min and the protease was inactivated by boiling in SDS sample buffer.
10. Cells induced to produce p50 were lysed and pelleted, and the supernatant was treated with PEI and ammonium sulfate according to the procedure described in the previous paper, except that no PMSF was added to buffer A. The ammonium sulfate pellet was redissolved in 50 mL of buffer C (20 mM HEPES buffer, pH 7.5, 1 mM DTT, 2 mM EDTA, 5% glycerol). To this solution were added 100 mL of 50 mM $CaCl_2$ and 25 mL of trypsin (1 mg/mL in 50 mM ammonium bicarbonate, 1 mM $CaCl_2$, pH 8.0). After incubation at room temperature for 50 min, the reaction was quenched by the addition of 200 mL of PMSF (200 mM in EtOH). The mixture was loaded onto a 200 mL S Sepharose (Pharmacia) column pre-equilibrated with buffer B. The column was eluted with a 800 mL linear gradient from 0 to 1.0 M NaCl in buffer B. Fractions exhibiting specific DNA binding activity eluted at ~ 0.55 M NaCl. Active fractions were stored by the addition of glycerol to 15% (v/v) at -78 °C.
11. This ruled out the possibility that the DNA-binding activity was due to a minor contaminant, since the p41 comprised more than 95% of the total protein in the sample.
12. By FPLC gel filtration analysis and glutaraldehyde crosslinking, purified p41 was determined to be essentially all dimeric.
13. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M. *Science* **1989**, *246*, 64.
14. As commonly observed, the N-terminal Met residue is cleaved post-translationally *in vivo*.
15. Protein samples were applied directly to a polybrene precycled glass fiber filter and placed in the reaction cartridge of an ABI Model 477A protein sequencer. The sample was subjected to automated Edman degradation using the standard program NORMAL-1, with the following modifications (for a faster cycle time, 36 min). The reaction cartridge temperature was elevated to 53 °C during coupling with a commensurate decrease in the three R2 delivery steps from 400 to 250 seconds. The resultant phenylthiohydantoin amino acid fractions were subsequently manually identified using an on-line ABI Model 120A HPLC.
16. Prior to analysis by electrospray-ionization mass spectrometry, samples were applied to a Vydac C18 column (2.1 mm x 150 mm) and eluted with a linear gradient of 0 to 100% acetonitrile/0.06% trifluoroacetic acid over 70 min, using a Hewlett Packard 1090/1040 system monitoring at 210 nm. Mass spectra were recorded on a Finnigan-MAT TSQ-700 (San Jose, California) triple quadrupole mass spectrometer equipped with an electrospray ion source. Approximately 5 picomole/mL of protein in 50% methanol/0.5% acetic acid was directly infused at a flow rate of 1 mL/min. The electrospray needle was operated at a voltage differential of -3.7 keV. Spectra were collected by scanning the *m/z* range of 700 to 1400 in 2 s and averaging 50 to 100 scans. Computer-aided deconvolution of the resulting spectrum of multiply charged *m/z* peaks to a plot of relative abundance vs. mass has been described (Hail, M., Lewis, S., Zhou, J., Schwalts, J., Jardine, I., & Whitehouse, C. (1990) in *Biological Mass Spectrometry*, (Burlingame, A. L. & McCloskey, J. A., Eds.), pp101-115, Elsevier, Amsterdam).
17. Active fractions were stored by the addition of glycerol to 32% (v/v) at -20 °C.
18. Cane, D.E., Wu, Z., Oliver, J.S., Hohn, T.M. *Archives Biochem. Biophys.* **1993**, *300* , 416; Chen, L., Scarr R.B., Verdine, G.L.,unpublished results; Wiley, D.C., Schreiber, S.L., personal communication.
19. This work was supported by a grant from the Hoffmann-La Roche, administered through the Institute of Chemistry in Medicine. Additional funding was provided by the Searle Scholars Program and the NSF (PYI program).