The ATP-dependent Clp Protease of Escherichia coli

SEQUENCE OF clpA AND IDENTIFICATION OF A Clp-SPECIFIC SUBSTRATE*

Susan Gottesman, William P. Clark, and Michael R. Maurizi

From the Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

The clpA gene, which codes for the ATP-binding subunit of the ATP-dependent Clp protease of Escherichia coli, has been sequenced. The coding region contains a single open reading frame for a protein of 758 amino acids; within the amino acid sequence are two consensus sequences for ATP-binding sites. The sequence of ClpA does not resemble that of other previously described ATPases or Lon, the other sequenced ATP-dependent protease of E. coli, except in the ATP-binding site consensus region.

The clpA gene is expressed as a monocistronic message. Primer extension experiments define a major start point of transcription at -183 relative to the start of translation. A rho-independent terminator is located 23 bases beyond the end of the coding region.

The ClpA protein is degraded in vivo in a Clp-dependent fashion (t1/2 ~ 60 min). A fusion protein containing the first 40 amino acids of ClpA fused in frame to β-galactosidase is degraded very rapidly in a clpA+ host (t1/2 ~ 3 min) but not in a clpA- host. This fusion protein is the first Clp-specific substrate described.

The rapid degradation of specific regulatory proteins is an important aspect of cellular control mechanisms (1-4). In addition, abnormal and damaged proteins as well as many foreign proteins introduced in cells by cloning or infection are rapidly degraded intracellularly. The turnover of these short-lived proteins in Escherichia coli and other organisms is frequently energy dependent (1-8). It is increasingly evident that the energy dependence of protein degradation in E. coli in vivo is largely attributable to the action of proteases that are either totally dependent on or highly activated by ATP (9-12). Identifying the ATP-dependent proteases found in cells, studying their mechanisms of action, and determining the unique features of in vivo substrates of these proteases should help our understanding of this mode of physiological regulation.

The best understood ATP dependent protease is the Lon protease of E. coli (9, 10). Lon substrates in vivo include the cell division inhibitor SulA, the λ antiterminator N, and the positive regulator of capsule synthesis RcsA (13-16). Also, the degradation of many abnormal proteins is at least partially dependent on functional Lon protease in vivo (7, 8, 17-20). In vitro, purified Lon protease directly cleaves multiple peptide bonds in a variety of denatured proteins and in purified λ N protein; maximal activity requires the continuous presence of ATP or an analog of ATP (9, 10, 15, 21). In vitro studies imply that under physiological conditions, proteolysis by Lon is accompanied by hydrolysis of two ATPs/peptide bond cleaved (22). The deduced amino acid sequence of Lon protease reported recently by Goldberg and co-workers (23) contains a sequence motif identical to the nucleotide-binding sites found in many ATP-binding proteins and ATPases. No other recognizable features in common with other proteases were noted.

A second ATP-dependent protease of E. coli, which we have called Clp (called Ti by Chung and co-workers), has been identified and purified to homogeneity (11, 12, 24, 25). Clp protease also directly cleaves peptide bonds in various proteins in a process that requires ATP hydrolysis. Clp differs from Lon in structure, in its in vivo substrates, and in the regulation of synthesis of the gene products. Lon is a tetramer composed of identical 87-kDa subunits, whereas the Clp protease consists of two dissimilar subunits, ClpA (81 kDa) and ClpP (21 kDa). ClpA has been shown to have ATPase activity and to bind ATP (24, 25). ClpP has been shown to be labeled by the serine protease inhibitor diisopropyl fluorophosphate and has low endopeptidase activity against small peptides in the absence of ClpA (26). The Lon protease is regulated by htpR as part of the heat shock response (27, 28). ClpA is clearly not a heat shock protein (24), but its regulation has not been described. clpA mutants do not share the properties of lon mutants, although ClpA seems to contribute somewhat to the degradation of abnormal proteins in the absence of Lon. A comparison of these two energy-dependent proteases may give us the first opportunity to understand the essential elements of an energy-dependent protease system.

We report here the sequence of the clpA gene and its regulatory region. We have found recently that ClpA contains two regions highly homologous with proteins from prokaryotic and eukaryotic cells; each of these domains contains a consensus sequence for a nucleotide-binding site. Sequence similarities between Clp and Lon protease are restricted to a very short domain (<=50 amino acids) centered on the nucleotide ATP-binding motif. We have used translational fusions of Lac to ClpA to define a substrate fusion protein specifically degraded by Clp.

**EXPERIMENTAL PROCEDURES**

Subcloning and Sequencing of clpA—pWP3 contains a 3.0-kilo-base BamHI-PstI fragment containing the clpA* gene; cells carrying

---

* The previous paper in this series was Ref. 24. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M31045.

† To whom correspondence should be sent: Bldg. 37, Rm. 4103, Laboratory of Molecular Biology, NCI, NIH, Bethesda, MD 20892. Tel.: 301-496-3524; Fax: 301-496-0260.

1 M. R. Maurizi and S.-H. Kim, manuscript in preparation.
2 Gottesman, S., Squires, C., Pichersky, E., Carrington, M., Hobbs, M., Martick, J. S., Dalrymple, B., Kuramitsu, H., Shirazu, T., Foster, T., Clark, W. P., Ross, B., Squires, C., and Maurizi, M. R. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, in press.
Fig. 1. Restriction map and sequencing strategy for clpA. Boxed area indicates open reading frame for ClpA. Hatched areas show two conserved domains, as determined from comparisons with other ClpA-like proteins (Footnote 2). Heavy vertical lines are locations of part A of nucleotide-binding domains. Arrows below the line indicate sequenced regions from different primers. Rb, Ribosome. Restriction enzyme cleavage sites are abbreviated as follows: B, BamHI; H, HindIII; S, SmaI; N, NruI; V, EcoRV; R, EcoRI.

this plasmid synthesizes ClpA at high levels (24).

The BamHI-PstI fragment was cloned from pWPC3 to both M13mp18 (M13WPC-A) and M13mp19 (M13WPC-7) (29). The single-stranded M13 DNA was used as template for dideoxy sequencing of clpA, using Sequenase (United States Biochemical Corporation) (30). The DNA sequence for the amino-terminal portion of the clpA structural gene, determined from a part of the cloned clone, was sequenced previously (24); it confirmed the amino-terminal sequence of the purified protein. Beginning from this amino-terminal segment, sequence was determined using synthetic internal oligonucleotides as primers, as indicated in Fig. 1. Sequence was determined from both strands, using the two M13 isolates, for the full open reading frame (24). The nucleotide sequence synthesized on Applied Biosystems model 380A DNA synthesized and purified using Applied Biosystems cartridges as described by the manufacturer. The GenBank accession number for the sequence of clpA is M31045.

Bacterial Strains—Strains SG21136 and the isogenic derivatives SG21138 and 22013 are AclpA derivatives of the lac strain MC4100 (31). SG21138 carries the clpA510::kan mutation (24), and SG2013 carries clpA::CM" N4956 is an r" derivative of C600.

pWPC4 carries a deletion of the Nru-EcoRV fragment of clpA (Fig. 1); we estimated previously that the deletion should remove the amino terminus of ClpA and a region of about 300 base pairs upstream of the translation start point. clpA" cells carrying pWPC4 show no significant synthesis of ClpA (24). The deletion, called AclpA194, was transferred from the plasmid to a λ derivative by crossing SV80 (imm λ clb57 clpA'8) with pWPC4. clpA" phage (SV81) derivatives were infected with anti-ClpA antibody (24). The deletion was transferred from the phage into appropriate hosts by lysogenizing clpA194::kan hosts with SV81 at low temperature (selecting imm λ) and screening for high temperature derivatives for those that lose both kanamycin resistance and phage λ immunity. SG21146 is a ΔclpA164 derivative of the Δlac strain SG20250 (7) constructed in this manner.

Construction of clpA-lac Fusions—The BamHI-HindIII fragment of pWPC3 (Fig. 1) was inserted into pATC591 (32). This translational fusion vector is a derivative of pBS891 (33) containing a HindIII linker in the Smal site. This construction creates an out-of-frame fusion of clpA to the lacZ gene; Δlac strains carrying this plasmid (pBC 0) remain Lac-. The plasmid was cut with Smal and either a 10-mer or 12-mer HindIII linker inserted at the Smal site. The ligated plasmids, which were still Lac-, were cut with HindIII and religated. Only the 10-nucleotide linker yielded Lac- plasmids after HindIII restriction and ligation. This is consistent with the known frame of the open reading frame for ClpA and LacZ. The sequence at the fusion point was confirmed by sequencing from a lac primer through the joint in the plasmid. The resulting protein fusion contains the first 40 amino acids of ClpA, a linker of Pro-Lys-Leu-Gly-Gly-Asn-Ser-Asp-Pro, followed by LacZ sequences starting at amino acid 9. This ClpA-LacZ fusion plasmid was called pBC101. The fusion allele will be referred to as clpA84.

A transcriptional fusion of clpA to lacZ was made by cloning the BamHI-Smal fragment from pWPC3 to pRS526 (33) and screening for Lac" colonies. This transcriptional fusion will be referred to as clpA45.

λ-Transducing phage carrying the clpA-lac Z fusions were isolated by growing λR456 in hosts carrying the fusions plasmids and screening the resulting phage lysates for Lac+ recombinants (33). Purified Lac- phage were used to isolate lysogens in appropriate hosts for assay of β-galactosidase activities (24). SBS44 carries the imm31 clpA84 lacZ transductional fusion; SBS5 carries the clpA84 transcriptional fusion.

Rates of In Vivo Degradation of ClpA and ClpA-LacZ Fusion Proteins—Cells were grown at 37 °C in minimal salts medium M50 (14) supplemented with 0.4% glucose and 50 μg/ml each of all amino acids except methionine and cysteine. At different times during exponential growth (usually when the A600 reached 0.4), 25 μCi/ml of [35S] methionine was added; after 1 or 2 min, 50 μg/ml nonradioactive L- methionine was added. Samples (0.3-0.6 ml) were removed and mixed with an equal volume of hot 2% SDS in 50 mM Tris (pH 7.5) and 1 mM EDTA and placed at 100 °C for 5 min. Alternatively, samples from the culture were treated with trichloroacetic acid (final concentration 5%) on ice for 30 min, washed twice with 1 ml of cold acetone, dried, and dissolved in 0.5 ml of 0.5% SDS buffer. As described above, for immunoprecipitation, aliquots of the SDS-treated samples were added to 50 mM Tris (pH 7.5), 1 mM EDTA, 0.5% Nonidet P-40 (final SDS concentration was 0.05-0.2%) containing protease inhibitors (24), and either 5 μl of anti-β-galactosidase antibody (Cooper Biomedical) or 60 μl of negatively absorbed anti-ClpA antiserum (see below) was added. After shaking for 1-2 h in the cold, the cells were collected by centrifugation, washed in 1 ml of Tris/EDTA/Nonidet P-40, then in the same buffer plus 1 mM KCl, and again in the same buffer without KCl. The immunoprecipitated proteins were solubilized with 50 μl of SDS buffer and, after heating at 100 °C for 40 min, loaded/lane of a 10% acrylamide gel. Gel electrophoresis was run according to Laemmli (35). Radioactive protein bands were visualized by autoradiography after fixing the gel in 25% methanol, 10% acetic acid, treating the gel with Autoradiol (National Diagnostics), and drying. Half-lives of ClpA-LacZ fusion protein were estimated by densitometry of autoradiograms. Radioactivity in ClpA protein bands was measured by cutting the bands from the dried gel, rehydrating in 100 μl of water, and soaking in 0.80 ml of a 9:1 mixture of NCS tissue solubilizer (Amersham Corp.) and water. After 24 h, 10 ml of toluene containing 0.8% 2,5-diphenyloxazole was added, and the gel pieces were shaken at room temperature for 24 h before counting.

Western Blotting of SDS Extracts of E. coli—Growing cells (1-2 ml) were treated with trichloroacetic acid (5%) on ice, washed twice in 1 ml of cold acetone, and dissolved in SDS sample buffer. Amounts of cell culture and SDS sample buffer were adjusted to give the protein equivalent of 1 ml of cell culture at A600 = 0.5 in 0.1 ml of sample buffer, and 30 μl was loaded in each lane. Gel electrophoresis, blotting, and immunochromatological detection of ClpA were done as described previously (7, 24), except that the antibody was negatively adsorbed against 3 volumes of crude extract of ΔclpA E. coli cells (approximately 20 mg/ml protein).

Isolation of mRNA and Primer Extension—Messenger RNA was prepared from cells grown in TB medium (tryptone, 10 g/liter; NaCl, 5 g/liter; 0.1 μg/ml ampicillin). The equivalent of 100 ml of cells at A600 = 0.4 was harvested, washed in cold 40 mM sodium acetate, and thereafter treated as described by Aiba et al. (36), except that the RNA was precipitated twice from 1 ml LiCl with 75% ethanol. Primer extension experiments were carried out as described by Maniatis et al. (37) using 25-50 μg of RNA and a 28-base synthetic oligonucleotide primer whose sequence matched the region of clpA preceding the Smal restriction site (nucleotides 94-121 in Fig. 2). The oligonucleotide was prepared on an Applied Biosystems model 380A DNA synthesizer and purified by reverse phase chromatography as described previously (24). The oligonucleotide primer was radioactively labeled at the 5' end with [γ-32P]ATP and T4 DNA kinase as described (24). The products of the primers reactions were run on 15% polyacrylamide gel along with sequencing reactions run using the primer. The position of the product corresponds to the start point of transcription indicated in the sequencing lanes.

Amino Acid Analysis—Purified ClpA (24) was hydrolyzed in 6 N HCl at 115 °C for 60 and 20 min, the dried hydrolysates were derivatized with phenyl isothiocyanate, and the products were separated on a Beckman C18 Ultrasound reverse phase column as described by Heinrickson and Meredith (38). Because yields of hydrophobic amino acids were low and inconsistent (60-80%), a hydrolysate of E. coli glutamate dehydrogenase (purchased from the authors, M. R. M.) was prepared in parallel with ClpA hydrolysates and used as the standard mixture of amino acids to calculate the amino acid composition of ClpA. Cysteine was deter-
Sequence of E. coli clpA

Fig. 2. Sequence of clpA structural gene and regulatory regions. Sequence is numbered from ATG start of translation; A is numbered 1. Possible Shine-Delgarno sequence is overlined. Start points of transcription as defined by primer extension are indicated by asterisk and number above sequence; underlined sequences are possible -10 and -35 sequences for transcription start 1. Amino acids in the core of part A and part B of nucleotide-binding sites are underlined, the limits of the conserved domains are indicated by arrows (*). The DNA sequence that encodes a possible transcription termination site at the end of clpA is underlined. Note that the T before the initial ATG is numbered 0, therefore, all negative numbers are actually 1 base pair further from the ATG than the numbers would suggest.

RESULTS

Sequence of clpA—In our previous paper (24) we reported the cloning of ClpA and showed that the amino acid sequence of the amino-terminal portion of ClpA determined by protein sequence analysis agreed with that predicted from partial sequencing of the DNA of the cloned gene. The remainder of clpA has now been sequenced from two M13 clones, each carrying one of the strands of clpA. The internal primers used for sequencing are shown in Fig. 1. The entire clpA gene and surrounding regulatory region were sequenced from both strands; the sequence is shown in Fig. 2. In this sequence, the open reading frame beginning at base pair 1 codes for a protein of 758 amino acids of a molecular mass of 83,875 daltons. This is in good agreement with the estimated size of 81 kDa determined for the purified protein by SDS-acrylamide gel electrophoresis.
Sequence of E. coli clpA

| Amino acid | Prediction from sequence | Prediction from amino acid analysis |
|------------|--------------------------|-----------------------------------|
| Asp/Asn    | 10.0                     | 9.7                               |
| Glu/Gln    | 13.2                     | 13                                |
| Ser        | 5.54                     | 5.9                               |
| Gly        | 6.73                     | 7.3                               |
| His        | 2.77                     | 2.7                               |
| Thr        | 5.01                     | 5.0                               |
| Ala        | 8.97                     | 9.3                               |
| Pro        | 3.30                     | 3.6                               |
| Arg        | 6.60                     | 6.8                               |
| Tyr        | 1.45                     | 1.7 (1.5)*                        |
| Val        | 8.18                     | 8.2                               |
| Met        | 1.98                     | 2.2                               |
| Ile        | 6.07                     | 4.9                               |
| Leu        | 10.0                     | 9.7                               |
| Phe        | 3.56                     | 3.2 (3.4)*                        |
| Lys        | 5.81                     | 5.6                               |
| Cys        | 0.40                     | 0.38*                             |
| Trp        | 0.40                     | 0.40*                             |

* Amino acids except for tryptophan and cysteine measured after hydrolysis of purified ClpA in 6 N HCl for 30 min at 155°C.

Values for tryptophan and values in parentheses for tyrosine and phenylalanine determined from the second derivative of the UV absorbance spectrum (40).

Determined with 5,5'-dithiobis(nitrobenzoic acid) (39).

---

After hydrolysis of pure ClpA protein and that calculated from the open reading frame. The amino acid composition reported here differs substantially from that reported previously for the same protein, called Ti by Chung and co-workers (25). The sequence of two internal peptides, determined by Chung for purified Ti and provided to us, are found within our predicted amino acid sequence. This agreement confirms that their purified Ti protein is in fact identical to ClpA and that disagreement in amino acid composition probably results from an error with their earlier preparation. The UV absorbance spectrum of ClpA in buffer B (24) at pH 7.5 showed a maximum absorbance at 278 nm and an absorption coefficient of 0.40 ± 0.02 (mg/ml)-1. The pI of ClpA calculated from the open reading frame. The amino acid composition reported after hydrolysis of ClpA protein fragments (data not shown).

Transcription of clpA—The in vivo start points of clpA transcription from both the chromosome and a plasmid-borne gene were determined by primer extension, as described under "Experimental Procedures." Three relatively strong bands were detected which would predict transcription start points within 200 base pairs of the start of translation when chromosomal RNA was the template (Fig. 3, lanes a, b, and d). All three bands were increased about 20-fold when RNA from a transformant carrying multiple copies of the clpA gene was used (Fig. 3, lane e), and they were all absent when the RNA used was extracted from the AcclpA164 mutant (Fig. 3, lane c). The first of the putative start sites begins at -183 from the translation start and is preceded by -10 and -35 regions that are reasonably close to consensus (underlined in Fig. 2).

23 nucleotides downstream from the termination of translation is a region capable of encoding a stable 14-base pair stem and 4-base pair loop, consistent with a transcription termination signal (Fig. 4).

Predicted Structure for the ClpA Protein—ClpA has been shown to have ATPase activity in vitro and to interact with ClpA to activate ATP-dependent proteolysis (24, 25). Therefore, ClpA would be expected to have an ATP-binding site. Examination of the sequence of ClpA revealed that the sequence Gly-X-Y-Gly-Val-Gly-Lys-Thr occurs twice, at amino acid residues 214-221 and 495-502 (underlined in Fig. 2). This or a closely related version of the sequence motif is found in nearly all ATP-binding proteins examined to date (23). According to Walker, part B of the consensus has 3 or 4 hydrophobic residues (Φ) followed by aspartate or glutamate ([(Φ)Asp/Glu]) and appears 50-100 amino acids to the carboxyl-terminal side of part A. In ClpA, part B sequences are found at amino acid residues 281-286 and 560-564, about 60 amino acids away from their respective parts A (Fig. 2, under-
Fig. 5. Comparison of amino acid sequences around the consensus ATP-binding sites of ClpA and other nucleotide-binding proteins from *E. coli*. The two-part consensus was derived from Walkers (41) as modified by Chin et al. (23) and this paper. Part A is ([Gly/-])X(-Gly-Lys-Thr[Asp/Glu]), where X is any amino acid, and corresponds to the following amino acid numbers in the primary sequence: ClpAl (200-299), ClpAZ (483-578), Lon (343-436), Rho (160-249), DnaB (218-326), NtrC (156-253), and the first of the regions is bounded in the plant genome sequence with the corresponding regions in the different genes shown (except for Lon and Clp) are from the National Biomedical Research Foundation database located at Georgetown University and correspond to the following amino acid numbers in the primary sequence: ClpAl (200-299), ClpAZ (483-578), Lon (343-436), Rho (160-249), DnaB (218-326), NtrC (156-253). The sequence for Lon is from Chin et al. (23).

Data base analyses by Chin et al. (23) indicate that both part A and B of the consensus only occur in proteins that bind ATP. Thus, the occurrence of two such consensus motifs in ClpA strongly suggests that ClpA has two binding sites for ATP.

We have reported recently that the ATP-binding consensus sequences in ClpA are prominent features in two regions of the primary protein sequence defined by very close homology to a second *E. coli* Clp-like protein, ClpB, and to a group of proteins found in other bacteria, lower eukaryotes, and plants. Because each of these regions shows conservation of sequence with the corresponding regions in the different genes and the first of the regions is bounded in the plant genome by introns, we refer to these as domain 1 and domain 2. Domain 1 of ClpA (amino acid residues 183-415, coded for by nucleotides 547-1245), shares 54% identical and 88% similar amino acids with ClpB, and domain 2 (residues 420-609, coded for by nucleotides 1258-1827), shares 53% identical and 89% similar amino acids with ClpB. Conservation between ClpA and the Clp-like proteins from other organisms is virtually the same as that between ClpA and ClpB of *E. coli*. In contrast, homology between the sequences of the two domains of ClpA is limited to the two relatively short regions immediately surrounding the two parts of the ATP-binding consensus sequence (Fig. 5). Domain 1 is longer than domain 2 and contains elements found in the β subunit of *E. coli* F1 ATPase not found in domain 2. Thus, although it is likely that both domains bind ATP, the differences between them suggest that the two domains have functionally distinct roles in the enzyme.

As reported by Chin et al. (23), Lon protease has a single ATP-binding consensus sequence. Remarkably, there are no extensive sequence homologies between Lon protease and ClpA outside of the narrow region around the ATP-binding consensus sequence. An alignment of the consensus sequences in ClpA, Lon protease, and several other ATPases from *E. coli* is shown in Fig. 5. There are no absolute conservations outside of the core consensus ([Gly/Ala]X(-Gly-Lys-Thr/Ser)-space-Φ,(Asp/Glu)), although groups of proteins have identical amino acids in certain positions. All of the proteins have 2-3 hydrophobic amino acids within the 4 amino acids immediately preceding the first glycine/alanine in part A, and most of them have hydrophobic amino acids at positions 2-3, 5-6, and 8 following the threonine. It is worth noting that the basic amino acid often included in the consensus 5-8 amino acids before the first glycine (23) is not found in several of the proteins (and therefore may not be a necessary feature of such a site).

More extensive analyses of the sequences of ClpA and Lon have revealed few similarities that might reflect the common enzymatic properties of the two proteases. The spacings between part A and part B are similar in both domains of ClpA and in Lon protease, but this is not a unique feature of the proteases, since the spacing is similar also in DnaA, RecA, and NtrA. The region between the Gly-Lys-Thr and the Φ Asp in Lon is very basic (as it is in RecD, UvrD, and helicase) but is neutral in both domains of ClpA. Sequence alignments between ClpA and the other proteins in Fig. 5 were calculated by either BESTFIT or by SEQHP (42) using 140-160-amino acid long regions centered about the ATP-binding consensus sequences. The region in domain 2 of ClpA shows a better quality alignment with Lon than with other ATPases and aligns better with Lon than does the corresponding region of domain 1. Domain 2 of ClpA would thus appear to be evolutionarily more closely, albeit still quite distantly, related to Lon protease.

Secondary structure predictions for the regions around the ATP-binding consensus sequences in both domains of ClpA, RblA, a highly conserved homolog of ClpA found in *Rhodo pseudomonas blastic* (43), Lon protease, and the β subunit of F1 ATPase are shown in Fig. 6. The ATP-binding consensus sequences would be expected to be found in structures known to form the elements of a nucleotide-binding pocket or Rossman fold (44), the essential elements of which are β-sheet-Gly-Lys-Thr-loop-α-helix in part A, and α-helix-loop-Φβ-sheet in part B. Domain 1 of ClpA and RblA conform reasonably well to the equivalent segments of the β subunit of F1 ATPase (Fig. 6). Domain 2 shows some significant differences compared with domain 1 and appears to resemble more closely Lon protease in predicted structure, particularly around part B, where the predicted Φβ-sheet is very short and is followed by an α-helix terminating in a strong turn. The positions of predicted turns are quite similar in Lon and domain 2 of ClpA, which partially reflects the locations of proline residues in the primary sequence. The prolines in this region of ClpA

*Fig. 6. Secondary structural predictions for the region around ABC sequences. The PEPPlot program (42), which uses the algorithm of Chou and Fasman to calculate α-helix (■) and β-sheet (□) propensities, was used to predict possible secondary structures in ClpA, RblA, a conserved homolog of ClpA from *R. blastic* (43), Lon protease, and the β subunit of F1 ATPase from the respective amino acid sequences. Regions for which α-helices or β-sheets are equally predicted are shown as lightly shaded (□) and regions for which no preferred structure was predicted were assumed to be coils (—). Turns (A) are indicated wherever strong predictions or several weak predictions of turns were made.*
are absolutely conserved and occur in ClpA and RblA as Pro-X-Pro-Thr-Oly-Val-Oly-Lys-Thr-X-Pro-Pro-X-Pro-His-Ala-Val-Leu-Leu-Leu-Asp-X-Pro; the same region is in Lon has Pro-X-Pro-Pro-Oly-Oly-Lys-Thr-X-Pro-Pro-Leu-Phen-Leu-Leu-Leu-Asp-X-Pro. These similarities between domain 2 and Lon protease further suggest these sites have equivalent functions in the respective enzymes.

Domain 1 of ClpA, on the other hand, has a number of features in common with the β-subunit of F1 ATPase, in addition to the structural similarities mentioned above. The positions of prolines are similarly spaced in the primary sequence, and most of the predicted turns are located in comparable locations (data not shown). Domain 1 shows a slightly better quality amino acid alignment with the β-subunit of F1 ATPase than does domain 2. As reported elsewhere, beyond the γi-Asp (part B of the consensus sequence) 2 tyrosines known to be located in the ATP-binding pocket of the β-subunit in F1 ATPase are found in the arrangement, Tyr-Xs-Thr-X1,A-Tyr, at positions +80 in domain 1 of ClpA and +88 in F1 ATPase. The location of these residues in domain 1 suggests, as with the β-subunit of F1 ATPase, ATP hydrolysis occurs at this site.

A site very similar to part B of the consensus sequence is found at about +115, measuring from the α-Asp in ClpA, and at +133 in the β-subunit of F1 ATPase. It is followed by an α-helix and a region rich in basic residues. The occurrence of a second part B (Ile-Asp-Val-Ile-Asp in ClpA) was first noted by Craig Squires for the family of ClpA-like proteins. That this second part B is highly conserved in ClpA-like proteins in other bacteria and in higher organisms implies that it is important for the integrity of ClpA. We have also found it in UvrB, UvrD, DnaB, and RecD, but it is not found in Lon protease, domain 2 of ClpA, the α-subunit of F1 ATPase, RecA, Rep helicase, or NtrA. The more extensive similarities between domain 1 of ClpA and the β-subunit of F1 ATPase serve to underscore the differences between the two domains of ClpA and support the conclusion that they probably have functionally distinct although interdependent roles in the activity of Clp protease.

ClpA-LacZ Protein Fusions—Our previous work had demonstrated that clpA, in contrast to lon, is not regulated by the heat shock σ factor, hspR (24). In order to study changes in expression of ClpA under different physiological conditions, we constructed an in-frame fusion between the 5′-terminal 1122 amino acids of ClpA and a g-amino acid linker joined to the 9th amino acid of the α-subunit of F1 ATPase, RecA, Rep helicase, or NtrA. The more extensive similarities between domain 1 of ClpA and the β-subunit of F1 ATPase serve to underscore the differences between the two domains of ClpA and support the conclusion that they probably have functionally distinct although interdependent roles in the activity of Clp protease.

ClpA-LacZ Protein Fusions—Our previous work had demonstrated that clpA, in contrast to lon, is not regulated by the heat shock σ factor, hspR (24). In order to study changes in expression of ClpA under different physiological conditions, we constructed an in-frame fusion between the 5′-terminal region of ClpA (up to base pair 121 of Fig. 2) and lacZ. This fusion encodes the amino-terminal 40 amino acids of ClpA followed by a g-amino acid linker joined to the 9th amino acid of β-galactosidase. A transcripitional fusion of clpA to lacZ, carrying the same fragment of clpA, was also constructed. Both fusions contain 1000 base pairs in front of the ClpA translation start codon. Both fusions were transferred by homologous recombination to λ RS45; these λ derivatives were used to construct single copy lysogens of the clpA-lac fusions for the study of clpA synthesis and accumulation.

Initial tests suggested a major difference between the expression of the translational fusion, SB84, in clpA+, and clpA− hosts. Lysogens of the clpA− host were visibly more Lac+ than those in the clpA+ host. The transcriptional fusion, SB85, did not show the same clpA-dependent difference. Such a difference could reflect either ClpA-dependent translational regulation of ClpA synthesis or the specific degradation of the ClpA-LacZ protein fusion by the Clp protease. Protein turnover experiments with the fusion demonstrate that the second possibility is true (Fig. 7), although translational regulation may also contribute to the difference. The ClpA-LacZ fusion is degraded with a half-life of about 4 min in clpA+ hosts but with a half-life greater than 20 min in a clpA− host. lon mutants have no effect on the accumulation of ClpA-LacZ (data not shown). Therefore, this fusion, which carries only the first 40 amino acids of ClpA, is degraded in a Clp dependent fashion. Since anti-β-galactosidase antibody was used to detect the fusion protein in the experiment of Fig. 7, it can be concluded that the entire LacZ portion of the protein is degraded, and no large intermediates accumulate.

Instability of ClpA in Vivo—Given the instability of the ClpA-LacZ fusion and in vitro observations on the instability of ClpA activity,1 it seemed possible that ClpA is itself a substrate for Clp-dependent degradation. The half-life of ClpA protein in vivo was determined by pulse labeling and immunoprecipitation of ClpA followed by gel electrophoresis and autoradiography (see “Experimental Procedures”). In wild-type cells, ClpA was degraded with a t1/2 of approximately 1 h (Fig. 8A). Although this rate of degradation is not as rapid as some of the regulatory degradation that occurs in E. coli, it is sufficient to remove almost half of the protease from the
cell during each generation. Moreover, ClpA was stable in vivo in a mutant lacking the proteolytic component of Clp protease, ClpP (Fig. 8A). Thus, active Clp protease is required for ClpA degradation in the cell. The half-life of ClpP was also determined by the same method; ClpP is not degraded in vivo (data not shown). In experiments with purified ClpA and ClpP, excess ClpA protein was rapidly degraded in an ATP-dependent manner in a reaction that required both active ClpA and active ClpP. Thus, it is likely that modulation of ClpA levels in the cell is accomplished at least in part by autodegradation of free ClpA subunits.

Accumulation of ClpA in cells was measured by running equivalent amounts of cell extract from cells grown to low, moderate, or high density on SDS-polyacrylamide gels, blotting, and immunochromatographic detection of ClpA. ClpA amounts/unit of cells did not appear to vary more than 10–20% during exponential growth (Fig. 8B). Examination of the accumulation of β-galactosidase in ClpA-lac fusions suggests that the synthesis of ClpA increases when the cell density reaches an A_{600} = 0.5 (data not shown). Therefore, the combination of any changes in synthesis rate with growth and degradation of ClpA results in a relatively constant amount of ClpA in the cell.

Given the instability of both ClpA and the ClpA-lacZ fusion, it seems reasonable to suggest that the amino terminus of ClpA acts as a recognition region for Clp-dependent degradation. The degradation may be prevented when ClpA is in a proper complex with ClpP; the ClpA-lacZ fusion protein, which is presumably unable to form such a complex, may be rapidly degraded because it is always free.

**DISCUSSION**

The ClpA structural gene and regulatory elements have been sequenced and the start points of transcription for the gene determined. The gene does not seem to be part of a larger operon. This also seems to be true of lon, another ATP-dependent protease of *E. coli*, clpP, the gene for the second component of the Clp protease, has been mapped by us to a region on the *E. coli* chromosome, far from clpA.

The two consensus nucleotide-binding sites of ClpA resemble the single site in Lon and other ATPases from *E. coli*; explicit amino acid homologies to Lon do not extend significantly beyond these sites, except for the disposition of proline residues noted. Nevertheless, we would tentatively conclude that domain 2 of ClpA is more likely to have structural and functional similarity to Lon. Although secondary structure predictions based on sequence alone are not entirely reliable, the consistencey with which the features found in nucleotide binding domains are predicted for the ATPases discussed in this paper makes it more likely that the agreements in structure between Lon and domain 2 of ClpA and between β-subunit of F1 ATPase and domain 1 of ClpA are significant. What the functional equivalence of domains 1 and 2 to these other proteins means in mechanistic terms is not yet clear.

ClpA possesses a basal ATPase activity and a substrate-activated ATPase activity. Preliminary results in vitro indicate that it is possible to inhibit protease activity and protease-stimulated ATPase activity without affecting basal ATPase activity. In light of the two putative ATP-binding sites identified by sequencing, it is possible that basal ATPase activity occurs in one domain and protease-stimulated ATPase activity occurs in the other. Alternatively, one domain could contain the catalytic site, and the second domain could have an allosteric ATP-binding site, occupancy of which is required for ATPase activity at the other site. Although both models are equally possible, we favor the former, inasmuch as both F1 ATPase and Lon protease, which appear by sequence analysis to be functionally analogous to domains 1 and 2, respectively, each possesses intrinsic ATPase activity. Experiments are in progress to demonstrate the binding of ATP to both sites in ClpA and, by site-directed mutagenesis, to alter the activity of each of the possible active sites.

The finding that clpA does resemble another gene in *E. coli* as well as genes in a variety of prokaryotic and eukaryotic organisms suggests that the clpA organization may in fact turn out to be a more generally used motif for ATP-dependent proteases than the lon organization.

The cellular function of the Clp protease is not yet clear. The widespread conservation of Clp-like protease genes may suggest that this protease is responsible for a fairly general and central housekeeping function rather than for the degradation of specific substrates that may be unique to specific organisms. It is fairly unusual to find families of duplicated genes in *E. coli*. The other examples thus far described include the ribosomal RNA genes and tufA and tufB, essentially identical genes for translation initiation factor (45). ClpA and ClpB differ for the amino-terminal 200 amino acids and in a central domain of 180 amino acids that clpB but not clpA contains. It seems possible that the diverged amino termini of these proteins reflect different targets. If the sensitivity of the fusion protein containing the first 40 amino acids of ClpA to Clp-dependent degradation is due to recognition of this amino-terminal sequence, Clp interaction with its substrates may depend on amino-terminal sequences. Recognition of the amino-terminal amino acids is one component of the recognition of substrates for the ubiquitin-dependent degradation of substrates in eukaryotic cells (46). Since ClpA retains an amino-terminal methionine, the recognition of Clp substrates must be charted in sequence beyond the amino-terminal amino acid.

ClpA synthesis does not increase on heat shock as lon synthesis does (24). Instead, the pattern of expression of clpA-lacZ translational fusions may suggest that accumulation increases when cells are at midlogarithmic growth, when oxygen levels begin to fall (data not shown). The conditions for optimum ClpA synthesis, and possibly for optimum Clp activity, are apparently very different from the heat shock, aerobic conditions that favor the synthesis of Lon and the increased degradation of Lon substrates, and may provide some hint of the role of Clp for *E. coli*.

**Acknowledgments**—We thank C. H. Chung for the sequences of two internal peptides from Ti (ClpA). We thank Craig and Cathy Squires for discussions on the structure of ClpA and their comparisons of ClpA with other ClpA-like proteins.

**REFERENCES**

1. Gottesman, S. (1989) *Annu. Rev. Genet.* 23, 163–198
2. Murray, A. W., and Kirschner, M. W. (1989) *Nature* 339, 275–280
3. Rechsteiner, M. (1987) *Annu. Rev. Cell Biol.* 3, 1–30
4. Bond, J. S., and Butler, P. E. (1987) *Annu. Rev. Biochem.* 56, 333–364
5. Goldberg, A. L., and St. John, A. C. (1976) *Annu. Rev. Biochem.* 45, 747–803
6. Olden, K., and Goldberg, A. L. (1978) *Biochim. Biophys. Acta* 542, 385–398
7. Maurizi, M. R., Trisler, F., and Gottesman, S. (1985) *J. Bacteriol.* 164, 1124–1135
8. Strauss, D. B., Walter, W. A., and Gross, C. A. (1986) *Genes & Dev.* 2, 1851–1858
9. Charette, M., Henderson, G. W., and Markovitz, A. (1981) *Proc. Natl. Acad. Sci. U. S. A.* 78, 4728–4732
10. Chung, C. H., and Goldberg, A. L. (1981) *Proc. Natl. Acad. Sci. U. S. A.* 78, 1931–1935
11. Katayama-Fujimura, Y., Gottesman, S., and Maurizi, M. R.
Sequence of E. coli clpA

(1987) J. Biol. Chem. 262, 4477–4485
12. Hwang, B. J., Park, W. J., Chung, C. H., and Goldberg, A. L. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 5560–5564
13. Mizusawa, S., and Gottesman, S. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 358–362
14. Torres-Cabassa, A. S., and Gottesman, S. (1987) J. Bacteriol. 169, 981–989
15. Maurizi, M. R. (1987) J. Biol. Chem. 262, 2696–2703
16. Gottesman, S., Gottesman, M., Shaw, J., and Pearson, M. L. (1981) Cell 24, 225–233
17. Keller, J. A., and Simon, L. D. (1985) Mol. Microbiol. 2, 31–41
18. Bukhari, A. I., and Zipser, D. (1973) Nature New Biol. 243, 238–241
19. Gottesman, S., and Zipser, D. (1978) J. Bacteriol. 133, 844–851
20. Kowit, J. D., and Goldberg, A. L. (1977) J. Biol. Chem. 252, 8350–8357
21. Goldberg, A. L., and Waxman, L. (1985) J. Biol. Chem. 260, 12029–12034
22. Menon, A. S., Waxman, L., and Goldberg, A. L. (1987) J. Biol. Chem. 262, 722–726
23. Chin, D. T., Goff, S. A., Webster, T., Smith, T., and Goldberg, A. L. (1988) J. Biol. Chem. 263, 11718–11726
24. Katayama, Y., Gottesman, S., Pumphrey, J., Rudikoff, S., Clark, W. P., and Maurizi, M. R. (1988) J. Biol. Chem. 263, 15226–15230
25. Hwang, B. J., Woo, K. M., Goldberg, A. L., and Chung, C. H. (1988) J. Biol. Chem. 263, 8727–8734
26. Woo, K. M., Chung, W. J., Ha, D. B., Goldberg, A. L., and Chung, C. H. (1989) J. Biol. Chem. 264, 2088–2091
27. Goff, S. A., Casson, L. P., and Goldberg, A. L. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 6647–6651
28. Phillips, T. A., VanBogelen, R. A., and Neidhardt, F. C. (1984) J. Bacteriol. 159, 283–287
29. Messing, J. (1983) Methods Enzymol. 101, 20–78
30. Tabor, S., and Richardson, C. C. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 4767–4771
31. Casadaban, M. J., and Cohen, S. N. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 4530–4533
32. Stout, V., and Gottesman, S. (1990) J. Bacteriol. 172, 659–669
33. Simons, R. W., Housman, F., and Kleckner, N. (1987) Gene (Amst.) 53, 85–96
34. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
35. Laemmli, U. K. (1970) Nature 227, 680–686
36. Aiba, H., Adhya, S., and de Crombrugghe, B. (1981) J. Biol. Chem. 256, 11905–11910
37. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
38. Helmrickson, R. L., and Meredith, S. C. (1964) Anal. Biochem. 136, 65–74
39. Haseeb, A. F. S. A. (1972) Methods Enzymol. 25, 457–464
40. Levine, R. L., and Federici, M. M. (1982) Biochemistry 21, 2600–2606
41. Walker, J. E., Sarasti, M., Runswick, M. S., and Gay, N. S. (1982) EMBO J. 1, 945–950
42. Dervieux, J., Haerberli, P., and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395
43. Tybulewicz, V. L. J., Falk, G., and Walker, J. E. (1984) J. Mol. Biol. 179, 185–214
44. Senior, A. E. (1988) Physiol. Rev. 68, 177–231
45. Riley, M., and Krawiec, S. (1987) in Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E., eds) pp. 967–981, American Society for Microbiology, Washington, D. C.
46. Bachmann, A., and Varshavsky, A. (1988) Cell 56, 1019–1032
The ATP-dependent Clp protease of Escherichia coli. Sequence of clpA and identification of a Clp-specific substrate.
S Gottesman, W P Clark and M R Maurizi

*J. Biol. Chem.* 1990, 265:7886-7893.

Access the most updated version of this article at [http://www.jbc.org/content/265/14/7886](http://www.jbc.org/content/265/14/7886)

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at [http://www.jbc.org/content/265/14/7886.full.html#ref-list-1](http://www.jbc.org/content/265/14/7886.full.html#ref-list-1)