AMINO-TERMINAL SEQUENCE IDENTITY OF UBIQUITIN AND THE NONHISTONE COMPONENT OF NUCLEAR PROTEIN A24

Lois T. Hunt and Margaret O. Dayhoff
National Biomedical Research Foundation
Georgetown University Medical Center
3900 Reservoir Road, N.W.
Washington, D.C. 20007

Received November 29,1976

SUMMARY: The amino-terminal 37 residues of bovine and human ubiquitin and of the nonhistone component from bovine nuclear protein A24 are identical. The probability that the two proteins are unrelated is infinitesimal. That they share such a highly conserved sequence indicates that they perform a cell function of basic importance and have a common evolutionary origin. They may even be products of the same gene. We make several suggestions concerning their structural and functional relationships based on a correlation of existing information.

We find that the amino-terminal 37-residue sequences of ubiquitin and the nonhistone component of the nuclear protein A24 complex are identical (see Figure 1). The complete sequences of bovine and human ubiquitin (1, 2) and the partial sequence of the nuclear protein A24 complex from calf thymus cells (3) have recently been published, but the correspondence between these proteins was not recognized until now. Our correlation of this information may contribute to an understanding of the function and evolution of ubiquitin and of nuclear protein A24.

Ubiquitin has been demonstrated by radioimmunoassay in a variety of mammalian tissues and in many organisms (including bacteria, yeast, higher green plants, invertebrates, and vertebrates), and it is thought to be present in all living cells (4). This 74-residue sequence is very highly conserved compared with other proteins (5), as the human and bovine sequences are identical (1, 2). The wide distribution and the conserved structure suggest that ubiquitin has an ancient, important function in all cells, prokaryotic and eukaryotic (4); this function is still unknown.

Copyright © 1977 by Academic Press, Inc.
All rights of reproduction in any form reserved.

Nonhistone Chromosomal Protein	M Q I F V K T L T G K T I T L E V E P S D T I E N V K A K I Q D K E P I P/
Ubiquitin	M Q I F V K T L T G K T I T L E V E P S D T I E N V K A K I Q D K E P I P
	P D Q Q R L I F A G K Q L E D G R T L S D Y N I Q K E S T L H L V L R L R

Figure 1. Comparison of the amino-terminal 37 residues of bovine nonhistone chromosomal protein from nuclear protein A24 (3) and the complete sequence of bovine ubiquitin (1). See Ref. 5 for the one-letter amino acid abbreviations.

Nuclear protein A24 has been found in several mammalian tissues, including rat liver and calf thymus, in 0.4N H_2SO_4 extracts of nucleoli and nuclei; it was absent from ribosomes and nucleolar ribonucleoproteins (6). It appears to be a complex of at least two polypeptide chains; one is probably histone H2A and the other is a nonhistone chromosomal protein (3). The nature of the association between the chains is not yet known.

What is the probability that the two sequences could be identical over a region of 37 consecutive residues? Are they products of the same gene? We can obtain one estimate of this probability by using our computer program SEARCH to examine our current protein sequence data collection for any similar 37-residue segments. This data base contains 889 sequences, with 64,784 37-residue segments. The 1976 mutation data matrix of amino acid pair scores with a bias parameter of 2 was used (5). For each segment comparison, a search score was computed by summing the pair scores for each pair of amino acids matched in the comparison. Typically, in such a search, the distribution of scores from unrelated sequence segments is approximately normal. The scores for related segments that are less than 60% different are found above the upper tail of this distribution, whereas scores for more distantly related sequence segments are found in the upper tail. The program produces a statistical analysis of the scores based on a normal distribution; the analysis includes the top 50 z values. A z value is the number of standard deviation units between any score and the mean, and it has an associated probability (5).

In the search of the bovine NHCP* segment, only one sequence segment,

* NHCP, nonhistone chromosomal protein.

the amino-terminal 37 residues of bovine ubiquitin, gives the maximum possible score of 239. The distribution of the other 64,783 segments has a mean score of 48.367 with a standard deviation of 14.816. The z value of 12.87 for the score 239 has an associated probability of less than 1×10^{-37} that the two segments are unrelated. Similarity of this degree has been found only in sequences that share a common evolutionary origin or are the products of a single gene. Convergent evolution of sequences because of similar function produces at most a very small effect. None of the other segments in the collection has a score that is significantly above the normal distribution of scores. The scores of the next highest 31 segments range downward from 119 to 100; none of these segments has more than 12 residues identical with those of bovine NHCP.

Another argument that supports the hypothesis that ubiquitin is a degradation product of NHCP is based on the highly conserved nature of the ubiquitin sequence. Its rate of mutation acceptance is similar to the lowest known, those of glutamate dehydrogenase and of histones H2A, H3, and H4 (5). For these slowly changing proteins the probability of observing a difference between human and bovine sequences in a 37-residue segment is less than 0.5. The single amino acid difference found in the first eight residues of celery ubiquitin further confirms this slow rate of change (4). It is most likely that ubiquitin is produced by the same gene as NHCP and is a degradation product of it. Otherwise we must postulate a gene duplication of considerable antiquity leading to two different genes with protein products of complex but different functions (one nuclear and one non-nuclear) having the same severe selective constraints operating on them throughout the evolution of this change in function. Such conservation has not been observed in any other proteins. Therefore, we predict that the sequence of residues 38-74 of bovine NHCP will be found to be identical with the rest of the bovine ubiquitin sequence.

How does the known structure of bovine ubiquitin correspond to the structure of the nuclear protein A24 complex? Ubiquitin has 74 residues and a molecular

weight of 8,451 daltons (1, 2). Bovine histone H2A has 129 residues, at least one acetyl moiety, and an approximate molecular weight of 14,000 (3, 7). Nuclear protein A24 has an approximate molecular weight of 27,000, at least two polypeptide chains, and a composition that includes all the residues present in histone H2A (3, 7, 8) and ubiquitin. Analyses of the undissociated complex gave a single 7-residue carboxyl-terminal sequence identical with that of histone H2A (3, 7) and a single 37-residue amino-terminal sequence unlike that of any known histone (3). Tryptic and chymotryptic peptide maps contained all of the corresponding peptides of histone H2A, including its acetylated amino-terminal tryptic peptide (7, 8, 9). This amino-terminal tripeptide sequence is found only in histones H2A and H4, and the carboxyl-terminal heptapeptide sequence is unique to histone H2A (10). The probability of observing the heptapeptide sequence in the data collection, based on the frequency of occurrence of the amino acids that compose it, is less than 1×10^{-4}. Thus, one of the major chains of the A24 complex certainly appears to be histone H2A, as the authors propose (3, 8).

The difference between the combined molecular weights of ubiquitin and histone H2A and that of nuclear protein A24 (6, 8) indicates that the nonhistone component of the complex could include up to 35-40 additional residues, having an approximate molecular weight of 4,550 (see Figure 2). From the estimated total residues and the percentage composition of the A24 complex (6, 8) and from the combined compositions of ubiquitin and histone H2A, we derived a possible 37-residue composition for this hydrophilic segment:

3 Ala	5 Lys	4 Ser
3 Gly	2 Arg	4 Thr
3 Asx	1 His	6 Pro
5 Glx	1 Ile	

The simplest hypothesis is that this segment follows the ubiquitin segment as the carboxyl region of a single polypeptide chain. If one of the prolines were the carboxyl-terminal residue, it would block carboxypeptidase digestion and only one carboxyl-terminal sequence would be found for the A24 complex.

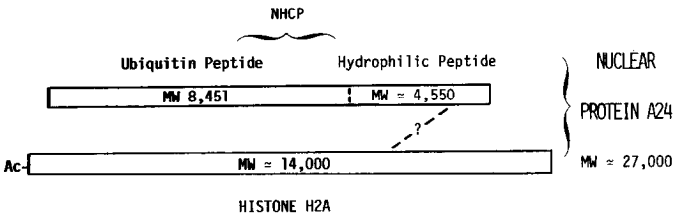

Figure 2. A proposed structure for nuclear protein A24. One of the two chains in the complex is histone H2A, with a blocked amino end (3, 8). The other chain, NHCP, is composed of ubiquitin plus a hydrophilic segment and has a carboxypeptidase-resistant carboxyl end.

These suggestions are compatible with the published experimental evidence (1, 3, 6, 8). Determination of the complete primary structure of NHCP should reveal whether it is indeed a single chain of approximately 110 residues, with residues 1-74 corresponding to ubiquitin.

Does the difference in length between NHCP and ubiquitin correlate with the difference in location in the cell or in function? Ubiquitin was isolated from the non-nuclear fraction, whereas nuclear protein A24 was found in nuclear and nucleolar chromatin. It is possible that the interchain association or binding between histone H2A and NHCP depends on the postulated hydrophilic carboxyl third of NHCP, that an arginine esterase releases ubiquitin from NHCP, and that ubiquitin subsequently diffuses from the nucleus into the cytoplasm.

The in vivo, intracellular role of ubiquitin has not yet been determined. Although in vitro it stimulates adenylate cyclase by binding to a β-adrenergic receptor, there is no evidence that it is normally secreted from the cell (1, 4, 11) or that its normal function involves this reaction. It was suggested that the function of nuclear protein A24, which exhibits histone-like tight binding to chromatin, might be to inhibit transcription of ribosomal cistrons in the nucleolus, because the level of A24 was observed to decrease as ribosomal RNA synthesis increased (6). That is, A24 would act as a ribosomal RNA gene repressor (6, 12) and would have to be inactivated in some way in order for ribosomal RNA synthesis to resume (12). If derepression involved

partial degradation of the repressor protein, then such degradation products (including ubiquitin?) might diffuse or be transported to other areas of the cell and might even perform some other function. Which features, if any, of this mechanism of gene repression might occur in ribosomal RNA synthesis in prokaryotes is not known.

Histones H2A, H3, and H4, which interact with genetic material as well as with each other (13), have the most highly conserved sequences observed (5). That a similar degree of sequence conservation occurs in ubiquitin and probably in NHCP supports the proposal of a role with histone H2A in the regulation of gene expression.

ACKNOWLEDGMENTS: This research was supported by NIH grant GM-08710 from the National Institute of General Medical Sciences.

REFERENCES

1. Schlesinger, D.H., Goldstein, G., and Niall, H.D., *Biochemistry* 14: 2214-2218, 1975
2. Schlesinger, D.H., and Goldstein, G., *Nature* 255: 423-424, 1975
3. Olson, M.O.J., Goldknopf, I.L., Guetzow, K.A., James, G.T., Hawkins, T.C., Mays-Rothberg, C.J., and Busch, H., *J. Biol. Chem.* 251: 5901-5903, 1976
4. Goldstein, G., Scheid, M., Hammerling, U., Boyse, E.A., Schlesinger, D.H., and Niall, H.D., *Proc. Nat. Acad. Sci. USA* 72: 11-15, 1975
5. Dayhoff, M.O., ed., *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 2, Nat. Biomed. Res. Found., Washington, D.C., 1976
6. Goldknopf, I.L., Taylor, C.W., Baum, R.M., Yeoman, L.C., Olson, M.O.J., Prestayko, A.W., and Busch, H., *J. Biol. Chem.* 250: 7182-7187, 1975
7. Yeoman, L.C., Olson, M.O.J., Sugano, N., Jordan, J.J., Taylor, C.W., Starbuck, W.C., and Busch, H., *J. Biol. Chem.* 247: 6018-6023, 1972
8. Goldknopf, I.L., and Busch, H., *Biochem. Biophys. Res. Commun.* 65: 951-960, 1975
9. Goldknopf, I.L., Olson, M., James, T., Mays, J., and Guetzow, K., *Fed. Proc.* 35: 1722, 1976
10. Dayhoff, M.O., Hunt, L.T., Barker, W.C., and Orcutt, B.C., *Protein Segment Dictionary 76*, Nat. Biomed. Res. Found., Washington, D.C., 1976
11. Goldstein, G., *Ann. N.Y. Acad. Sci.* 249: 177-185, 1975
12. Crippa, M., *Nature* 227: 1138-1140, 1970
13. Kornberg, R.D., *Science* 184: 868-871, 1974