

Structure of Actin Molecule

Tatsuo OOI and Sho TAKAHASHI*

Received November 20, 1984

Information on various aspects of actin molecules is summarized to obtain a possible structure of actin. Amino acid sequences of 18 actins from various sources are aligned and location of possible secondary structures is predicted according to methods developed by Chou and Fasman, and by Robson. Profiles of the distribution of hydrophobic and hydrophilic residues along the sequence suggest that the molecule is composed of several domains. Furthermore, the search of homologous segments to those in proteins of known tertiary structures gives partial information on possible spatial structures of the segments. By utilizing the information on several characteristic sites of actin such as divalent cation binding site, the structure of an actin molecule in relation to polymerization sites and other interaction sites is discussed.

KEY WORDS: Structure of Actin/ Prediction of Secondary Structure/
Sequence Homology/ Hydrophobic Profile/

I. INTRODUCTION

A living organism has to maintain its life by consuming chemical energy through a metabolic machinery keeping dynamic equilibrium with the environment. Chemical energy stored as chemical substances is converted to energy of various forms, *e.g.*, chemical energy and mechanical energy. In the step of energy conversion, the machinery of motility seems to be essential for an organism to live, because both nutrient and waste have to be carried in and out to corresponding locations by some means. Some motile system is provided for that purpose, such as cytoplasmic flow in a cell and muscle contraction in a higher organism.

During past two decades, actin and myosin have been accepted as the major proteins which are involved in the motile system in general. Particularly, actin exists in various organisms from plants to muscles of animals and probably even in bacterium. It is believed that this protein plays an essential role in the motile system interacting with another essential protein, myosin, which is an enzyme of ATPase. Nevertheless, three dimensional structures of these proteins are not explored yet, because of the difficulty of crystallization. At the present stage the molecular mechanism of motility and muscle contraction is not known. Once a primary structure is determined, the three dimensional structure would be deduced from the sequence in principle. Unfortunately it is impossible that a correct tertiary structure could be predicted only from its amino acid sequence, although many attempts for this purpose have been performed recently.

One of promising techniques to determine the three dimensional structure of a

* 大井龍夫, 高橋 敏: Laboratory of Physical Chemistry of Enzyme, Institute for Chemical Research, Kyoto University, Uji 611, Japan.

protein is the X-ray crystallography, if the crystallization of the protein is successful. As far as actin is concerned, only a complex of G-actin and another substance such as DNase I was found to be crystallized, and as mentioned in the previous paper crystallization of F-actin would be hopeless because F-actin itself is regarded as a result of the linear crystallization. These are the reason why we have performed experiments of chemical crosslinking by using bifunctional reagents.^{1,2,3} We must have as much information as possible about the proximity of residues in the chain. Also, there are several approaches to the prediction of higher structures from the sequence since data on three dimensional structures of globular proteins have been accumulated.⁴ In this paper, we will try to analyze the primary structures of actin obtained from various sources and profile of distribution of amino acid residues along the chain, and draw a possible structure of actin molecule.

II. AMINO ACID SEQUENCES

A number of amino acid sequences of actin are available from various organisms. Some of them have been determined by sequencing on proteins extracted from an organism and some have been deduced from the DNA sequences which correspond to actin genes. Figure 1 shows the sequences determined so far, which are aligned to give the best match at the corresponding positions. The species and organisms are as follows; 1) rabbit skeletal alpha actin,⁵ 2) rat skeletal alpha actin,⁶ 3) human skeletal alpha actin,⁷ 4) bovine cardiac muscle actin,⁸ 5) human cardiac muscle actin,⁹ 6) chicken gizzards smooth muscle actin,⁸ 7) bovine brain cytoplasmic beta actin,⁵ 8) bovine brain cytoplasmic gamma actin,⁵ 9) amoeba actin,¹⁰ 10) *Dictyostelium discoideum* actin,¹¹ 11) *Physarum polycephalum* actin,¹² 12) *Drosophila melanogaster* actin at 79B locus,¹³ 13) *Drosophila melanogaster* actin at 88F locus,¹⁴ 14) yeast actin 1,¹⁵ 15) yeast actin 2,¹⁶ 16) soybean actin,¹⁷ 17) maize actin,¹⁸ 18) *Oxytricha fallax* actin.¹⁹ In this figure the residue number is written including deleted positions, so that the total number is 378. This numbering is used in this section (in the next three sections, we used the numbering of 375 residues of rabbit skeletal muscle).

The amino acid sequences shown in Fig. 1 indicate that the sequence of actin is strikingly conservative, *i.e.*, more than 70% residues are identical when we exclude the last one, which has a long deletion sequence from 68th to 85th. This alignment suggests that the tryptophan residue at the 76th of actin from chicken gizzards smooth muscle shown in the 6th row would be at the 81st position, since the replacement of the 76th residue to the 81st position gives rise to one residue shift and thus the perfect alignment of the sequences in the neighborhood of the residues. In addition, the same revision was made on the sequence of rabbit skeletal muscle actin.⁵

The sequence of actin from the ciliated protozoan shown in the last row of Fig. 1 deviates from the rest of them particularly near the 50th, 230th and 270th position. Nevertheless the homology of the sequence seems to be very good from the N-terminus to the C-terminus, implying that this protein might have a unique structure in order to fulfil its motile function, or the change in the primary structure might cause such a fatal effect that the evolution of this protein could not have been achieved.

Structure of Actin Molecule

	10	20	30	40	50	60	70	80	90	100
1	DEDETTALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
2	MCDEDETTALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
3	MCDEDETTALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
4	DEDETTALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
5	MCDEDETTALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
6	EEETALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
7	DDIAALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
8	EEIEAALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
9	MGDEVQALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
10	DGEDVQALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
11	EGEDVQALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
12	MCDEEASALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
13	MCDDAGALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
14	MOSEVAALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
15	MOSEVAALVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
16	MADAEDIEPLVCDNGSGLVKA	GAFAGDDAPRAVFP	SIVGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
17	MAD EDIQICVCDNGSGLVKA	GAFAGDDAPRAVFP	SI GRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
18	MSDQTCVINDNGSGLVKA	GAFAGDDAPRAVFP	SI VGRPRHQGVMVGMG	GKDSYVGD	EAGSKRGLLTKYPI	IEHGIITNWDDMEK	IWHHTFYNELRVAP			
	110	120	130	140	150	160	170	180	190	200
1	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
2	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
3	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
4	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
5	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
6	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
7	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
8	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
9	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
10	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
11	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
12	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
13	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
14	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
15	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
16	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
17	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
18	EEHPTLLTEAPLNPKANREKMT	QIMFETFNVPAMYVA	IQAVLSLYASGRTT	GI	VLDSDGGVTHNVP	IEGYALPHAIMRLD	LAGRDLTDYLMKIL	TERGY		
	210	220	230	240	250	260	270	280	290	300
1	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
2	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
3	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
4	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
5	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
6	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
7	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
8	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
9	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
10	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
11	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
12	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
13	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
14	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
15	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
16	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
17	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
18	SVFTTAAERIVRDI	KEKLCYVALDFENEMATA	AASSSLEKSYELPDGGVIT	IGNERFRFCPEL	FQPS	FIGMESAGIHETT	TYSIMKCDIDIRKDY	YANN		
	310	320	330	340	350	360	370	380		
1	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
2	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
3	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
4	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
5	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
6	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
7	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
8	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
9	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
10	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
11	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
12	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
13	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
14	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
15	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
16	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
17	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				
18	VMSGGTTFMYPGI	ADRMRKKEIT	ALAPSTMKIKII	IAPPERKYSVWIGGS	ILASLSTFQQMWIS	KQYDEAGPSIVHRKCF				

Fig. 1. Alignment of 18 actin sequences reported so far. 1) rabbit skeletal alpha actin, 2) rat skeletal alpha actin, 3) human skeletal alpha actin, 4) bovine cardiac muscle actin, 5) human cardiac muscle actin, 6) chicken gizzards smooth muscle actin, 7) bovine brain cytoplasmic beta actin, 8) bovine brain cytoplasmic gamma actin, 9) amoeba actin, 10) *Dictyostelium discoideum* actin, 11) *Physarum polycephalum* actin, 12) *Drosophila melanogaster* actin at 79B locus, 13) *Drosophila melanogaster* actin at 88F locus, 14) yeast actin 1, 15) soybean actin, 16) maize actin, and 18) *Oxytricha fallax* actin.

Since the role of actin in motility seems to be very important, the actin gene or a part of the gene would be expected to be transposed to some other protein genes, we looked for homologous DNA sequences to the actin after converting the amino acid sequence to the DNA sequence using the codon table. The conversion could be made by utilizing two parallel sequences; one is the nucleotide sequence converted by the use of one-to-one correspondence except for those extra codons of Ser, Leu, and Arg, and the other is used for the sequence of the extra codons. The result of the comparison with 3.3 million bases compiled in DNA Data Bank, *i.e.*, GenBank release 25, was that actin genes are the only ones which contain homologous segments longer than 40 bases; *i.e.*, actin genes or partial genes in bovine, human, mouse, rabbit, rat, chicken, quail, amoeba, nematode, fruit fly, sea urchin, maize, slime mold, and protozoan. The homologous one was found in the gene product of the v-gfr oncogene,²⁰ which has a homologous segment of 360 bases long. The segment corresponds to the actin sequence from 10th to 130th residue.

III. PREDICTION OF SECONDARY STRUCTURES

Once we have an amino acid sequence of a protein, propensities to form secondary structures can be calculated by using parameters inherent to every amino acid residues. Figure 2 illustrates those patterns obtained by the parameters of Chou and Fasman method.²¹ Here, an average value over 5 residues before and after a given residue (*i.e.*, the average of values of 11 residues) is plotted against the residue number of rabbit skeletal muscle actin (375 residues). The mean value was taken as the zero level in the figure. The region which has a propensity higher than zero may be regarded as that of a high probability to form a corresponding secondary structure, *i.e.*, α -helix, β -conformation, or turn. The rough characteristic profile of propensities is as follows; the region from 10th to 60th are rich in turn suggesting that the chain folds into several segments, the region from 110th to 140th has high values of both α -helix and β -structure, and the succeeding region near the 160th residue shows a high propensity of turn. A high propensity of α -helix appears in the region from 210th to 230th, followed by β -structure region to the 290th residue. After a high peak of the turn propensity at 300th, α -helix and β -conformation appear to be dominant in the C-terminal part. The prediction according to this scheme is summarized that a segment from 1st to 140th is rich in β -turn followed by α -helix, that from 140th to 200th is rich in β -conformation, that from 200th to 220th is α -helix, that from 220th to 310th is rich in β -turn, and the segment from 310th to 375th would be α -helix or β -structure.

The prediction performed above, however, would not be so reliable, since we have no established method to predict correctly the secondary structures in a protein. One way to obtain a more reliable result may be the combination of several predicted results by using different methods. The result shown in Fig. 3 is such a joint prediction of secondary structures by two methods, one is the Chou-Fasman method²¹ illustrated in Fig. 2, and the other is the method of Robson.²² In the figure, the location of ionizable residues, acidic ones (aspartic acid and glutamic acid) and

Structure of Actin Molecule

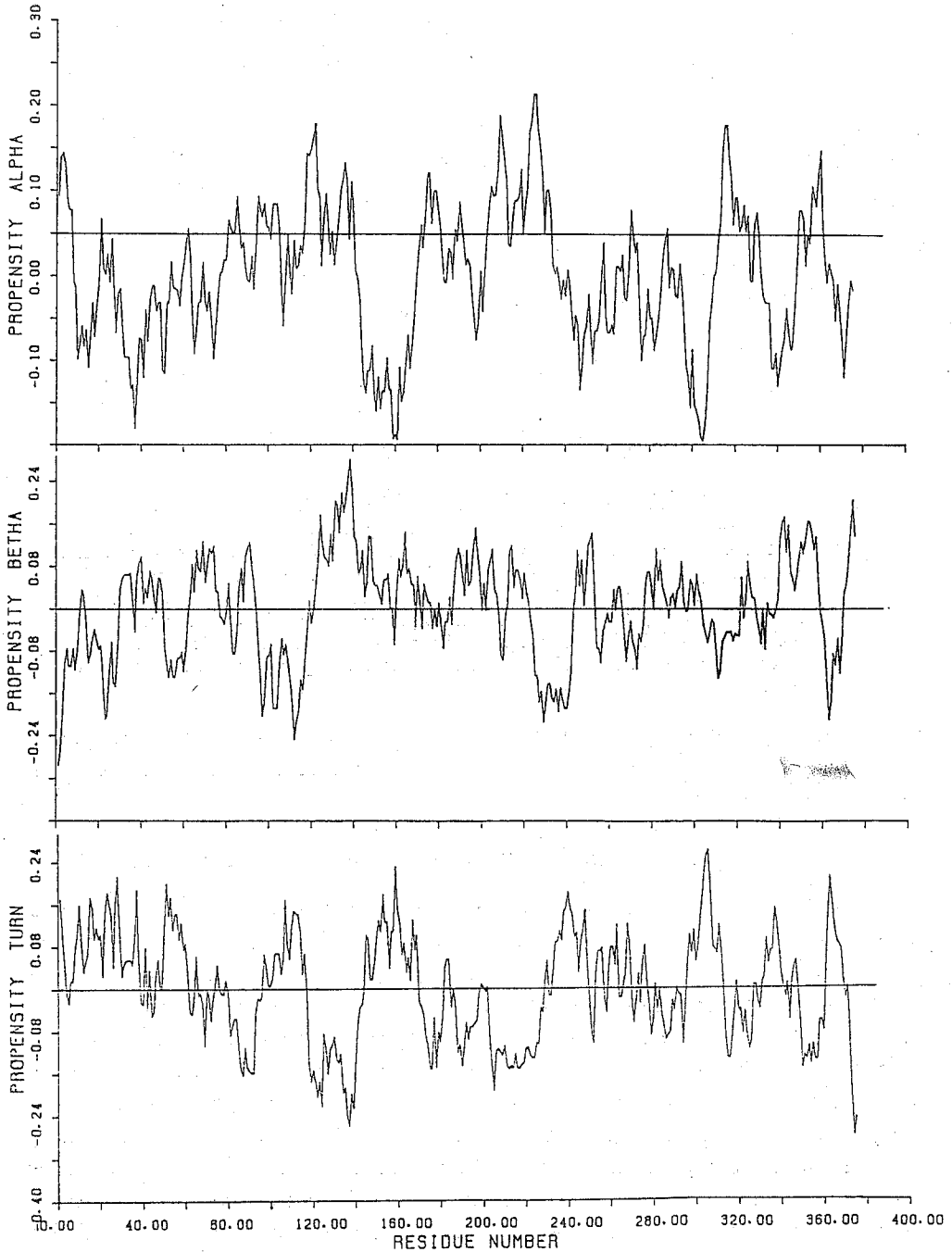


Fig. 2. Profiles of distribution of propensity of α helix (top), propensity of β conformation (middle), and propensity of turn (bottom). The parameters to compute the propensities proposed by Chou and Fasman were used. A value at every residue was the averaged value over 11 residues which include 5 residues before and after the residue.

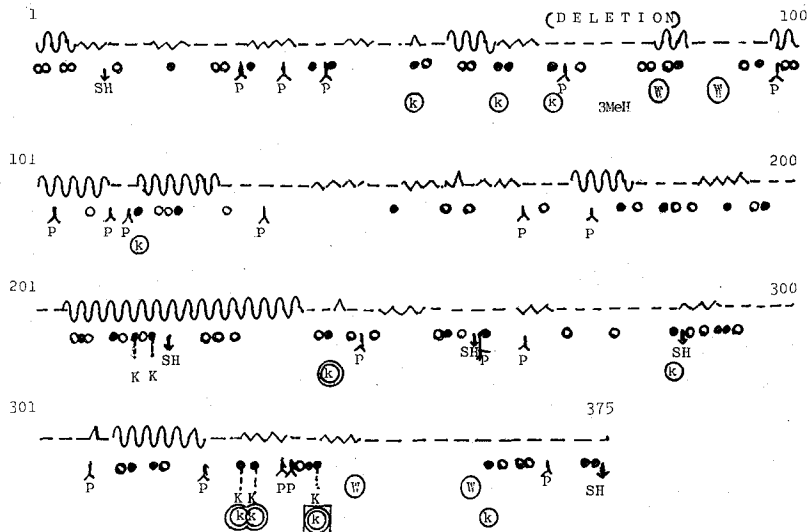


Fig. 3. The location of secondary structures predicted by the combination of the methods of Robson and of Chou and Fasman. Wave symbol stands for α helix, zig-zag β conformation, and chain line random coil. The location of acidic (open circle) and basic (filled circle) residues is also shown. Proline residues (P), cysteine residues (SH), and tryptophan residues (W) are indicated at the corresponding positions. The basic residues shown by (K) are the residues assigned by cross-linking experiments. The lysine residues which change the reactivity on polymerization (circle), tropomyosin binding (double circle) and subfragment-1 binding (square), are shown by a small letter k.

basic ones (lysine and arginine) is also shown. The helical content of this prediction is about 25%, which is a little lower than that estimated from the circular dichroic spectrum of actin in solution, 30%.

IV. PROFILE OF DISTRIBUTION OF HYDROPHOBIC RESIDUES

The folding of a nascent chain of a protein at the physiological condition must be dependent on the distribution of hydrophobic and hydrophilic residues along the chain. When an amino acid sequence is known, a profile of the distribution can be calculated by making some smoothing of a curve obtained by plotting values of every amino acid residues in the sequence. Figure 4 demonstrates the profiles of hydrophobicity,²³⁾ transfer free energy,²⁴⁾ and hydrophilicity²⁵⁾ of an actin molecule. The smoothing was made by taking an average of 11 residues as done before for the prediction of secondary structures. Since the negative value of hydrophilicity is plotted in the figure, the three patterns are similar to each other, *i.e.* a region higher than the zero level is hydrophobic and therefore tends to bury inside the molecule; on the contrary, the negative region is hydrophilic to be exposed to the outside of the molecule. The profile indicates a feature of the appearance of hydrophobic and hydrophilic regions alternatively with an interval of 10 to 20 residues.

The sequence profile over a longer range may be illustrated by plotting an average value of 21 residues (10 residues before and after a given residue) as shown in Fig. 5. The patterns for hydrophobicity and hydrophilicity were very similar to

Structure of Actin Molecule

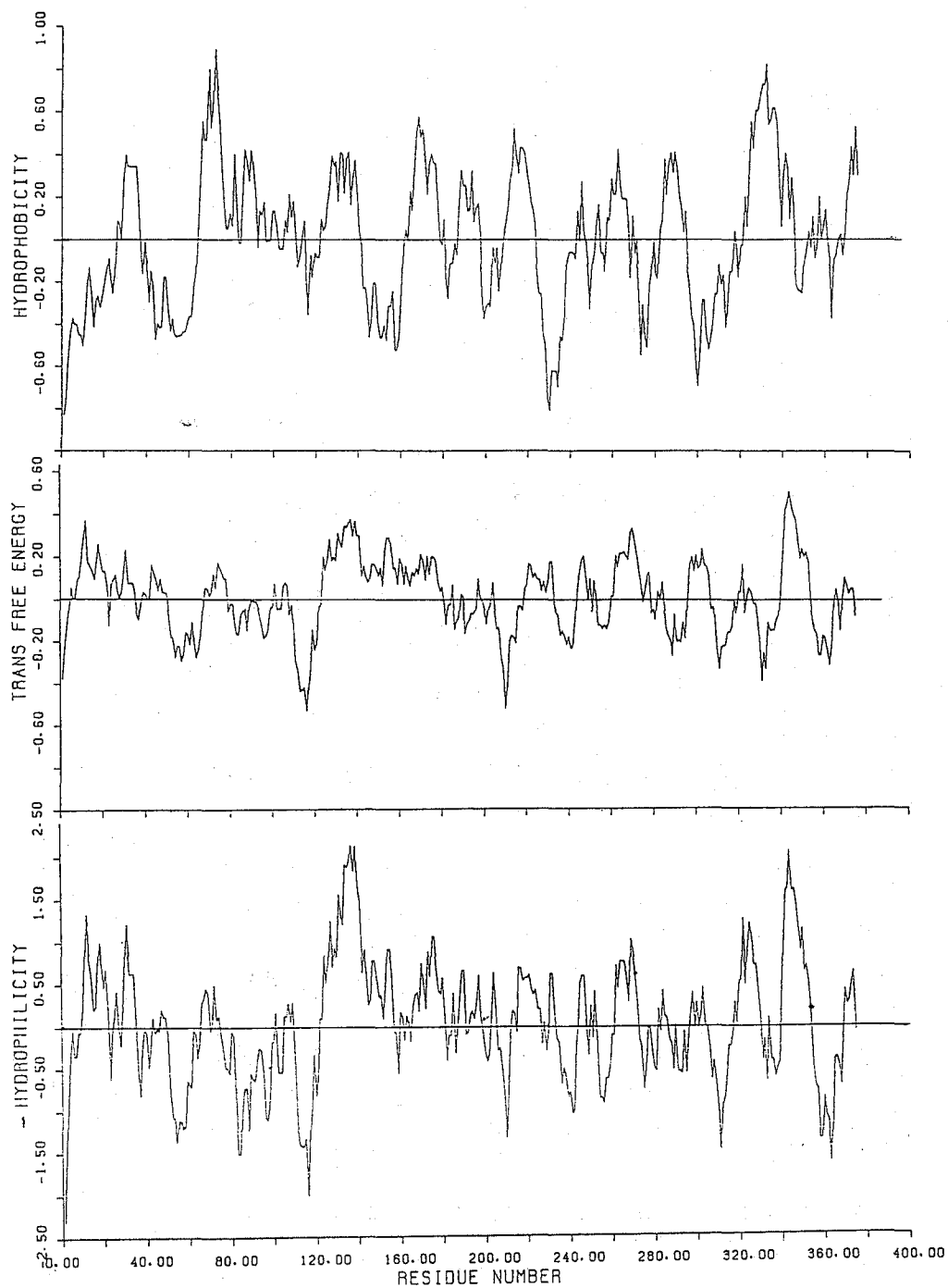


Fig. 4. Profiles of hydrophobicity (top), transfer free energy (middle), and hydrophilicity (bottom) computed along the chain are shown. Average values of 11 residues as done for the computation of those shown in Fig. 2 were adopted.

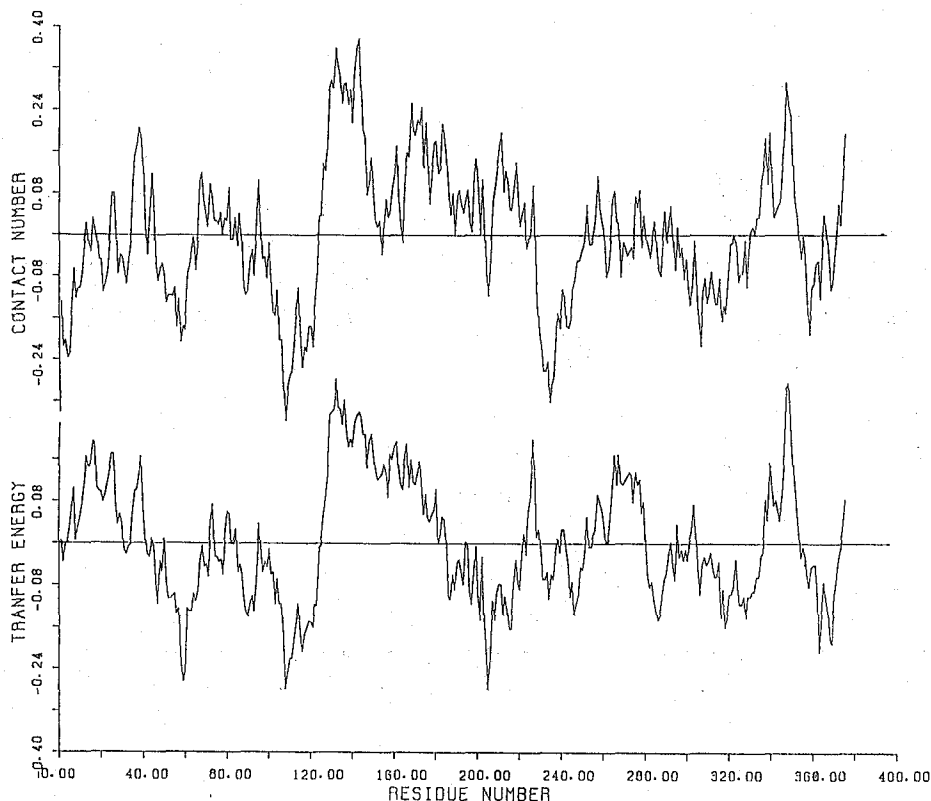


Fig. 5. Profiles of contact number (top) and transfer free energy (bottom) computed by averaging over 21 residues are shown against the residue number.

those shown in Fig. 5. The contact number used for the computation of the upper curve is an expected number of residues which contact with a given residue,²⁶⁾ *i.e.*, the greater the contact number of a residue, the more deeply the residue will be buried inside the molecule. The profiles illustrated in Fig. 5 suggest that the molecule would consist of approximately four domains, 1-110, 110-230, 230-300, and 300-375. The pattern from 110th to 140th is remarkable, *i.e.*, the curve shows a sharp increase approximately from the lowest value to the highest. This implies that this part of the sequence has the drastic change in hydrophobicity from hydrophilic residues to hydrophobic residues as recognized in the sequence shown in Fig. 1. The domain of 130-230 seems to be very hydrophobic on the average, so that this part could be assigned to be a core region in the molecule.²⁷⁾

V. HOMOLOGOUS SEGMENTS TO STRUCTURALLY KNOWN PROTEINS

The prediction of secondary structures and distribution of hydrophobic and hydrophilic residues described before are the pure deduction from the amino acid sequence. For the purpose to predict the three dimensional structure, we have been developing a method to use the information of known tertiary structures of various proteins and sequence homologies among a given protein and the proteins.²⁸⁾ When

Structure of Actin Molecule

Table I. List of homologous regions of actin against structurally known proteins.

ACTIN	149	T G I V L D S G D G V T H N V P I Y
1LH1	8	A A L V K S S W E E F N A N I P K H H H H H H H H H H H H H H H H H
ACTIN	171	L P H A I M R L D L A G R D L T D Y L M K I L T
2TAA	173	L P D D L D T T K V V K N E W Y D W V G S L V S B B B H H H H H H H H H H H H H H H H
ACTIN	162	N V P I Y E G Y A L P H A I M
2TAA	449	N V P V P M A G G L P R V L Y B B B B B B B B B B B B B B B B
ACTIN	307	P G T A D R M Q K E I T A L A P S T
1AZA	89	G G E S D S V T F D V S K L T P G E T T B B B B B B B B B T T T T T T T T
ACTIN	74	H G I I T N D D W M E K I W H H
1LZM	55	N G V I T K D D - A E K L F N Q B B B H H H H H H H H H H H
ACTIN	91	Y N E L R V A P E E H P T L L
2ATC	86	I N D Y E V V G K S R P S L P B B B B B B B B B B
ACTIN	100	E H P T L L T E A P L N P
3CAT	21	Q K P D V L T T G G G N P H H H H H
ACTIN	89	T F Y N E L R V A P E E H
2MHB	91	L H C D K L H V D P E N F H H H H H H H H H H
ACTIN	98	P E E H P T L L T E A P
5CPA	56	G S N R P A I W I D L G T T T T T
ACTIN	244	D G Q V I T I G N E R F
3FAB	108	Q G S L V T V S S A S T B B B B B B B B B B B B
ACTIN	361	E Y D E A G P S I V H R
3FAB	2	Q L E Q S G P G L V R P B B B B B B B B B B B B
ACTIN	312	R M Q K E I T A L A P S T
1REI	70	D Y T F T I S S L Q P E D B B B B B B B B B B
ACTIN	237	E K S Y E L P D G Q V I
2APP	248	D C S T N L P D F S V S B B B B B B B B
ACTIN	256	R C P E T L F Q P S F I G
2SSI	1	D A P S A L Y A P S A L V B B B B
ACTIN	350	V W I G G S I L A S L S
3TLN	48	T T L P G S L W A D A D B B B B B B B B B B T

amino acid sequences are converted into series of numerical values of some physico-chemical property such as hydrophobicity, the homology of two sequences can be computed quantitatively by the calculation of a correlation coefficient, since the sequences are replaced by numerical values. The properties have been selected so as to give high correlations both on primary structures and on three dimensional structures.²⁸⁾ By using this method, the actin sequence was compared with those compiled in Protein Data Bank⁴⁾ and other amino acid sequences available.

Examples of the homologous segments of actin to 108 different proteins in Protein Data Bank are listed in Table I. Names of proteins in Table I are, 1LH1 is leghaemoglobin, 2TAA takaamylase, 1AZA azurin, 1LZM T4 phage lysozyme, 2ATC aspartate transcarbamylase, 3CAT catalase, 2MHB methaemoglobin beta chain, 5CPA carboxypeptidase A, 3FAB immunoglobulin fragment FAB', 1REI immunoglobulin B-J fragment, 2APP acid proteinase, 2SSI subtilisin inhibitor, and 3TLN thermolysin. These segments have correlation coefficients higher than 0.55 and the segment lengths greater than 12 residues.

Since three dimensional structures of the above proteins are known, secondary structures of all the segments are also known as listed in the next row of homologous segments (H and B stand for α -helix and β -conformation, respectively, and blank represents random coil). If two sequences are identical, we expect that the three dimensional structures of both segments are the same, or very similar. Therefore, the homologous residues in actin listed in Table I are assumed to take the corresponding secondary structures. In other words, this is another assignment of secondary structures along the chain. Nevertheless, the number of proteins of known three dimensional structures is not large enough to cover whole the actin sequence with the highly homologous segments.

Table II. List of homologous regions of actin against muscle tropomyosin and A proteins of ribosomes.

ACTIN	177	R	L	D	L	A	G	R	D	L	T	D	Y	L	M	K
TM	70	K	L	E	L	A	E	K	K	A	T	D	A	E	A	D
TM	249	L	E	K	S	I	D	D	L	E	D	E	L	Y		
TM	239	A	E	F	A	E	R	S	V	T	K	L	E			
ACTIN	206	R	E	I	V	R	D	I	K	E	K	L	C	Y		
TM	61	S	E	A	L	K	D	A	Q	E	K	L	E	L		
ACTIN	166	Y	E	G	Y	A	L	P	H	A	I	M	R	L	D	L
VULGA	81	L	T	G	L	G	L	K	E	A	K	D	K	V	D	G
		A	P	S	T	L	K	E	A	V	S					
ACTIN	317	I	T	A	L	A	P	S	T	M	K	I				
VULGA	33	V	S	A	A	A	P	A	M	M	A	V				
ACTIN	1	D	E	D	E	T	T	A	L	V	C	D	N	G	S	
GRISE	55	E	Q	D	E	F	D	V	I	L	T	G	A	G	E	

The search of homologous regions of actin to other proteins could be made when their sequences are known. Table II demonstrates the results obtained by the comparison with some proteins, *i.e.*, a muscle protein, tropomyosin which is an α -helical protein, and acidic proteins of ribosomal proteins (usually called as A-proteins). The region from 177th to 181th of actin is homologous to three regions in tropomyosin, from 70th to 84th, 249th to 261th, and 239th to 250th, suggesting that the region takes an α -helical conformation. Another region from 206th to 218th would be also α -helical. The longest homologous segment was found in the A-protein of ribosomes of *Desulfovibrio vulgaris* (DvA),²⁹⁾ as shown in Table II, *i.e.*, 25 residues long. The last two homologous segments of A-proteins of DvA and *Streptomyces griseus* (SA1)³⁰⁾ have correlations greater than 0.58 and lengths longer than 13 residues. We do not know what implication these homologies have, but some structural similarity must be present in actin and A-proteins, because we cannot expect that such a high correlation would happen randomly. As far as we tried to find highly correlative segments in more than 500 proteins, there was not such a high correlative segment as shown in Table II.

Although we could not cover whole the actin sequence, the assignment of the location of secondary structures by the homology against the known proteins is as follows; 4-10 β -conformation, 75-87 α -helix, 97-102 possibly turn (although the prediction shown in Fig. 3 is α -helix), 146-155 β -conformation, 177-181 α -helix, 232-255 β -conformation rich, 260-270 β -conformation, 312-325 β -conformation, and 337-347 and 360-366 β -conformation. In this assignment, some agree with those in Fig. 3, and some do not.

VI. ASSIGNMENT OF MOLECULAR STRUCTURE OF ACTIN

Since we have many characteristic properties of an actin molecule, we may infer a molecular structure according to the analyses done before. The location of typical ionizable groups along the chain and of proline residues which have a tendency to locate at turn regions in the three dimensional structure is shown in Fig. 3. There are 19 prolines in actin, among which 16 are conserved in various actins shown in Fig. 1, suggesting the important role of the residues. The candidates of reactive lysine residues found in our crosslinking studies are shown by the capital letter K, and five cysteine residues by SH. The interesting feature of the distribution of the residues shown in Fig. 3 is that the ionizable groups are not located evenly, but cluster at several regions, *e.g.*, near 120th, 210th, 280th and 365th residue.

First of all, an actin molecule binds one mole of divalent cation, Ca^{++} or Mg^{++} strongly, removal of which gives rise to the irreversible denaturation of actin.³¹⁾ Since the divalent cation is not necessary to be Ca^{++} (*i.e.*, divalency is essential), the binding site would be composed of a cluster of negative charges as found at the Ca^{++} binding sites of parvalbumin.³²⁾ Ionizable groups of basic and acidic residues are aligned alternatively in most of the clustering regions in actin as illustrated in Fig. 3. If some clustering region of at least three acidic residues be a binding site, there are only two possibilities, one is the N-terminal region from 1st to 10th, and the other is

the region from 224th to 228th in Fig. 3 assigned as α -helix (Fig. 3). The former position has been mentioned as the candidate.³³⁾ Nevertheless, when the chain folds, two residues separated far each other along the primary structure could approach together, making a site specific for some ligand. Therefore, the candidate for the binding site of the divalent cation might not be such a region that acidic residues cluster.

Second, G-actin binds one mole of ATP strongly, and this ATP is hydrolyzed to ADP on the transformation from G-actin to F-actin, and the ADP is tightly bound in a monomer in F-actin. The bound ADP does not exchange easily to free nucleotides in solution. As for the binding site of ATP or ADP, we have to consider two sites, one is the nucleoside or adenine base binding site and the other is a binding site of the phosphate group which is negatively charged. Since three dimensional structures of proteins which bind nucleotide molecule(s) such as dehydrogenases and ribonuclease have been determined by X-ray crystallography, we expected to find a homologous region in actin to those proteins. Since any homologous region with a high correlation coefficient could not be found, the binding mode of nucleotide in actin would be different from those found so far.

On the other hand, the binding of the phosphate group attached to 5' end of the ribose ring gives a clue of the site, *i.e.*, some clustering region of positive charges or basic residues is a candidate. Such candidate regions are not many, one is a cluster of Arg 30, Arg 39 and Arg 41, and the other is that of Lys 63, Arg 64, and Lys 70 (Fig. 1), when we look for a region consisting of more than three basic residues. Therefore, the N-terminal domain is one of candidates of the nucleotide binding region. If clustering of the basic residues in space is considered, another region could be found. Since the search of such possible regions in the three dimensional structure is too complicated, regions formed by α -helix or β -structure have been looked for. Fig. 6 shows the alignment of residues from 206th to 230th when an α -helix is formed. An interesting feature of this alignment is that clusters of three positive and negative residues are separated by hydrophobic residues which appear every three or four residues. Thus, this region is also one of the candidates of phosphate binding. Experiments on changes of the reactivity of lysine residues

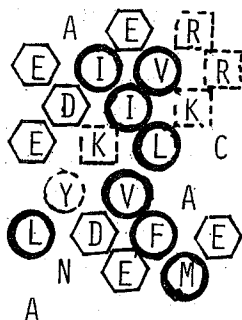


Fig. 6. Spatial distribution of amino acid residues on the surface of a possible α helix. Residues from Ala 207 to Ala 223 in Fig. 1 are aligned.

on actin polymerization³⁴⁾ suggest that the enhanced reactivity of Lys 335 (339 in Fig. 1) is attributed to loss of the γ phosphate of ATP.

Third, we have the experimental evidence that some aromatic residues are involved in polymerization of G-actin to F-actin, *i.e.*, circular dichroic spectrum in the absorption region at about 280 nm changes markedly upon polymerization,¹⁾ and flow dichroism at the same region also suggests that aromatic rings are aligned in F-actin.³⁵⁾ The aromatic groups involved would be tryptophan residues and the location of four tryptophan residues (W) is illustrated in Fig. 3. Interestingly these residues are located only at two regions, one near 85th (Trp 81 and Trp 88) and the other near 350th (Trp 343 and Trp 359) (the numbering in parentheses is based on that in Fig. 1). At least one of these region should be involved in the contact of the neighboring actin monomers. Results of change in reactivities of lysine residues^{34,36)} indicate that lysine residues in the N-terminal domain, Lys 50(52), 61(63), 68(70), and Lys 113(115), in addition to Lys 283(287) and Lys 290(294) become less reactive (numbers in parentheses are those in Fig. 1) on polymerization.

Fourth, the intramolecular cross-linking experiment suggests that a region from 210th to 220th might be close to that from 330th to 340th. The intermolecular cross-linking experiment implies that the region of 210–220 of one monomer is close to the C-terminal domain of another monomer. That is, both regions would be in the proximity of the polymerization sites. The photo-crosslinking experiment¹⁷⁾ suggests the same contact, *i.e.*, crosslinking from Cys 373 to Lys 218. Since no ionizable residue is present between 340th and 362th residues, the peptides of the region would be buried inside the molecule, and both regions extended from this part are assumed to be one of the contact sites to another actin monomer.

Fifth, the binding site of tropomyosin is inferred to be some α -helical part in the molecule, because tropomyosin is a typical coiled-coil molecule of two α -helices, and the binding mode of parallel α -helices might be the most favorable interaction between two molecules. As listed in Table II, the region from 179th to 193th residue (Fig. 1) meets the requirement. Experimentally, Lys 335 (339 in Fig. 1) reduces the reactivity on binding of tropomyosin.³⁸⁾ Modification of Lys 237 (240 in Fig. 1) prevents the binding of tropomyosin but when troponin or Mg^{++} or Ca^{++} is added, the binding is restored.³⁹⁾ Thus, the proximity of tropomyosin binding sites is found in the last half of the molecule. Possible binding sites of subfragment-1 and heavy meromyosin^{38,40)} are located near the C-terminal domain, *e.g.*, in the proximity of Lys 335.

We described the possible sites located on the actin molecule, and summarized as follows; the first domain from N-terminus to 110th is one of candidates of binding of divalent cation and nucleotide, and one of the contact sites in an actin monomer, the region from 100th to 110th might form a turn conformation because of the presence of a number of proline residues. In the second domain, the chain from 110th residue runs into the inside of the molecule forming a core of the molecule followed by an α -helical region near 180th residue and another α -helical region at about 220th residue. The third domain from 230th to 300th would form some β -sheet structure because the distribution of ionizable residues is rather even and the pro-

pensity to form β -conformation is high. There is almost no experimental clue in the above two domains. The last domain from 300th to C-terminus would be one of the contact sites of polymerization and include binding sites of tropomyosin and subfragment-1.

Although we have a considerable knowledge about the molecule, still it is impossible to construct the detailed three dimensional structure of actin. According to the molecular structure described here, we hope that the construction of the molecular structure becomes possible by computing detailed interactions among the residues.

This work was supported by the research grant from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- (1) O. Ohara, S. Takahashi, T. Ooi, and Y. Fujiyoshi, *Biochem. Biophys. Res. Commun.*, **100**, 988-994 (1981).
- (2) O. Ohara, S. Takahashi, T. Ooi, and Y. Fujiyoshi, *J. Biochem.*, **91**, 1000-2012 (1982).
- (3) O. Ohara, S. Takahashi, and T. Ooi, *J. Biochem.*, **93**, 1547-1556 (1983).
- (4) Protein Data Bank, Brookhaven National Laboratory, Upton, New York, U.S.A.
- (5) J. Vanderkerckhove and K. Weber, *Eur. J. Biochem.*, **90**, 451-462 (1978).
- (6) R. Zakut, M. Shani, D. Givol, S. Neuman, D. Yaffe, and U. Nudel, *Nature*, **298**, 867-859 (1982).
- (7) A. Hanauer, M. Levin, R. Heilig, D. Daegelen, A. Hahn, and J.L. Mandel, *Nucl. Acid. Res.*, **11**, 3503-3516 (1983).
- (8) J. Vanderkerckhove and K. Weber, *FEBS Lett.*, **102**, 219-222 (1979).
- (9) H. Hamada, M.G. Petrino, and T. Kakunaga, *Proc. Natl. Acad. Sci. USA*, **79**, 5901-5905 (1982).
- (10) W. Nellen and D. Gallwitz, *J. Mol. Biol.*, **159**, -118 (1982).
- (11) J. Vanderkerckhove and K. Weber, *Nature*, **282**, 475-477 (1980).
- (12) J. Vanderkerckhove and K. Weber, *Nature*, **276**, 720-721 (1978).
- (13) E.A. Fyberg, B.J. Bond, N.D. Hershey, K.S. Mixter, and N. Davidson, *Cell*, **24**, 107-116 (1981).
- (14) F. Sanchez, S.L. Tobin, U. Rdest, E. Zulauf, and B.J. McCarthy, *J. Mol. Biol.*, **163**, 533-551 (1983).
- (15) D. Gallwitz, F. Perrin, and R. Seidel, *Nucl. Acid Res.*, **9**, 6339-6350 (1981).
- (16) R. Ng and J. Abelson, *Proc. Natl. Acad. Sci. USA*, **77**, 3912-3916 (1980).
- (17) D.M. Shah, R.C. Hightower, and R.B. Meagher, *Proc. Natl. Acad. Sci. USA*, **79**, 1022-1026 (1982).
- (18) D.M. Shah, R.C. Hightower, and R.B. Meagher, *J. Mol. Appl. Genet.* **2**, 111-126 (1983).
- (19) P. Kaine and B.B. Spear, *Nature* **295**, 430-432 (1982).
- (20) G. Naharro, K.C. Robbins, and E.P. Ready, *Science*, **223**, 63-66 (1984).
- (21) P.Y. Chou and G.D. Fasman, *Adv. Enzymol.*, **47**, 45-148 (1978).
- (22) B. Robson and E. Suzuki, *J. Mol. Biol.*, **107**, 327-356 (1976).
- (23) D.D. Jones, *J. Theore. Biol.*, **50**, 167-183 (1975).
- (24) J. Janin, *Nature*, **277**, 491-492 (1979).
- (25) J. Kyte and R.F. Dolittle, *J. Mol. Biol.*, **157**, 105-132 (1982).
- (26) K. Nishikawa and T. Ooi, *J. Peptide Protein Res.*, **16**, 19-32 (1980).
- (27) K. Mihashi, and T. Ooi, *Biochemistry*, **4**, 805-813 (1965).
- (28) Y. Kubota, K. Nishikawa, S. Takahashi, and T. Ooi, *Biochim. Biophys. Acta*, **701**, 242-252 (1982).
- (29) T. Itho, and E. Otaka, *Biochim. Biophys. Acta*, in press.
- (30) T. Sugiyama, and K. Higo, *Biochim. Biophys. Acta*, **701**, 164-172 (1982).
- (31) K. Mihashi and T. Ooi, "Molecular Biology of Muscle Contraction", Igakushoin, Tokyo (1963).
- (32) R.H. Kretsinger and C.E. Nockolds, *J. Biol. Chem.*, **248**, 3313-3326 (1973).
- (33) D. Mornet and K. Ue, *Proc. Natl. Acad. Sci. USA*, **81**, 3680-3681 (1984).
- (34) R.C. Lu, and L. Szilagyi, *Biochemistry*, **20**, 5914-5919 (1981).

Structure of Actin Molecule

- (35) S. Higashi and F. Oosawa, *J. Mol. Biol.*, **12**, 843-865 (1965).
- (36) S.E. Hitchcock-De Gregori, S. Mandala, and G.A. Sachs, *J. Biol. Chem.*, **257**, 12573-12580 (1982).
- (37) K. Sutoh, *Biochemistry*, **23**, 1942-1946 (1984).
- (38) L. Szilagyi and R.C. Lu, *Biochim. Biophys. Acta*, **709**, 204-211 (1982).
- (39) S.C. El-Saleh, R. Thieret, P. Johnson, and J.D. Potter, *J. Biol. Chem.*, **259**, 11014-11021 (1984).
- (40) K. Sutoh, *Biochemistry*, **21**, 3654-3661 (1982).