

## II. 2. RELATIVE PROBABILITIES OF ISOMERS IN CYSTINE-CONTAINING RANDOMLY COILED POLYPEPTIDES

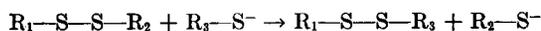
Walter Kauzmann\*

The Frick Chemical Laboratory, Princeton University, Princeton, New Jersey

I. Introduction . . . . .	93
II. Method of Calculating Probabilities . . . . .	94
III. Applications . . . . .	96
IV. Summary . . . . .	104

### I. Introduction

Many globular proteins are cross-linked intramolecularly through the disulfide bonds of cystine, and there is evidence that in all molecules of a given native protein the arrangement of these cross-links is the same. A molecule containing  $N$  cystine cross-linkages ( $2N$  half-cystine residues) can, however, be cross-linked intramolecularly in  $(2N)!/(2^N N!)$  different ways. If a protein is dissolved in a denaturing solvent such as urea, which converts the polypeptide chain into something approximating a random coil, and if a trace of thiol-containing substance is added at a somewhat alkaline pH, the disulfide groups will undergo interchange by the mechanism of Huggins *et al.* (1),



If the protein concentration is low, intermolecular cross-links can be avoided (2) and eventually a new form of the protein will be produced in which there is a random distribution of all of the  $(2N)!/(2^N N!)$  possible arrangements of the cross-linkages. The arrangements, or isomers, produced in this way will in general have considerably different probabilities. It is the purpose of this paper to compare these probabilities. The method of calculating probabilities that is developed makes it possible to determine the thermodynamic probability of the arrangements that are actually known to be present in such native proteins as ribonuclease (3, 4) and insulin (5).

\* This work was supported by a grant from the National Science Foundation.

## II. Method of Calculating Probabilities

According to the statistical theory of randomly coiled polymers (6), a linear, noncross-linked polymer chain containing  $n$  flexible links, whose ends are separated by a distance that lies between  $L$  and  $L + dL$ , can exist in

$$N_n(L) dL = K_n [\exp(-k^2 L^2/n)] L^2 dL \quad (1)$$

configurations. The constant  $k$  in this expression is of the order of magnitude of the reciprocal of the length of a single linkage, its precise magnitude depending on the geometrical relationships between successive linkages and on the rigidity of the links themselves. The constant  $K_n$  depends on  $n$  but is independent of  $L$ ; its numerical value can be found in the following way. Suppose that the addition of a single link to the chain increases the number of configurations of the chain by a factor  $C$ . Then the total number of configurations accessible to the chain is  $C^n$  and it must be true that

$$\int_0^{L_{\max}} N_n(L) dL = C^n \quad (2a)$$

$$= K_n \int_0^{L_{\max}} [\exp(-k^2 L^2/n)] L^2 dL \quad (2b)$$

$$= K_n \left\{ \int_0^{\infty} [\exp(-k^2 L^2/n)] L^2 dL - \int_{L_{\max}}^{\infty} [\exp(-k^2 L^2/n)] L^2 dL \right\} \quad (2c)$$

where  $L_{\max}$  is the length of the fully extended chain. That is,  $L_{\max} = nl$ , where  $l$  is the length of a single link. The maximum value of the integrand in the second integral in (2c) is  $\exp(-k^2 L_{\max}^2/n) = \exp(-nk^2 l^2)$ . Since  $kl$  is of the order of unity, this integrand will be very small if  $n$  is greater than, say, 4 or 5. Therefore the second integral in (2c) will be neglected. Since

$$\int_0^{\infty} [\exp(-k^2 L^2/n)] L^2 dL = (n/k^2)^{3/2} \sqrt{\pi}/4$$

we find

$$K_n = C^n (k^2/n)^{3/2} (4/\sqrt{\pi}) \quad (3)$$

Suppose that two chains, one containing  $n_1$  links and the other containing  $n_2$  links, are joined at *one end only*. The total number of configurations will, of course, be

$$C_{\text{chain}}(n_1 n_2) = C(n_1) C(n_2) = C^{n_2 + n_1} \quad (4)$$

Let us next find the number of configurations,  $C_{\text{ring}}(n_1 + n_2)$ , when *both ends* of the two chains are joined forming a ring containing  $n_1 + n_2$  links.

This is done by requiring that the free end of one chain be within a volume  $v$  of the free end of the other chain,  $v$  being of the order of  $l^3$ . Then the probability that the ends of the doubly joined chains are separated by a distance that lies between  $L$  and  $L + dL$  is

$$N_{n_1 n_2}^{\text{ring}}(L)dL = vK_{n_1}K_{n_2}[\exp(-k^2L^2/n_1)][\exp(-k^2L^2/n_2)]L^2dL \quad (5)$$

and the total number of configurations available to the ring is

$$\begin{aligned} C_{\text{ring}}(n_1 + n_2) &= \int_0^{L_{\text{max}}} N_{n_1 n_2}^{\text{ring}}(L)dL = vK_{n_1}K_{n_2} \int_0^{\infty} \exp[-k^2L^2(1/n_1 + 1/n_2)]L^2dL \\ &= AC^{n_1+n_2}/(n_1 + n_2)^{3/2} \end{aligned} \quad (6)$$

where  $A = 4k^3v/\sqrt{\pi}$ . Since  $k^3v$  is of the order of unity, the constant  $A$  is also of the order of unity. Thus the formation of a ring from an open chain containing  $n = n_1 + n_2$  links reduces the number of configurations accessible to the chain by a factor given by

$$a(n) = A/n^{3/2} \quad (7)$$

This result has been derived previously by Kuhn and Majer (7).

Starting with formulas (4), (5), and (6) it is possible to compute the number of configurations corresponding to more complex ways of joining chain segments to each other. Arrangements of the cross-links in a long chain can be built up conceptually by successive combinations of pairs of subpatterns. Each subpattern may be assigned two *end points*,  $\alpha$  and  $\beta$ , through which it is joined to other subpatterns. For each subpattern it is necessary to find the *effective number of links*,  $N_e$ , defined by the distribution function for the distance  $L$  between the end points  $\alpha$  and  $\beta$  of the subpattern:

$$N_{N_e}(L)dL = K_{N_e}[\exp(-k^2L^2/N_e)]L^2dL \quad (8)$$

Two basic procedures may be considered for combining a pair of subpatterns to give another, larger pattern. For each procedure there is a simple rule for finding the value of  $N_e$  for the resulting pattern in terms of the values of  $N_e$  of the two subpatterns. There is also a simple rule for finding the number of configurations of the resulting pattern from the number of configurations of the subpatterns. These rules are as follows.

*Procedure I.* Let the two subpatterns,  $a$  and  $b$ , be joined at only one end point (see Fig. 1). If subpattern  $a$  has  $C_a$  configurations and  $N_a$  effective links and if subpattern  $b$  has  $C_b$  configurations and  $N_b$  effective links, then the number of configurations of the resulting combined pattern is

$$C = C_a C_b \quad (9)$$

and the effective number of links between the two free end points of the combined pattern is

$$N_e = N_a + N_b \quad (10)$$

*Procedure II.* If both end points of the subpatterns are joined to give the equivalent of a ring (see Fig. 1), then the number of configurations of

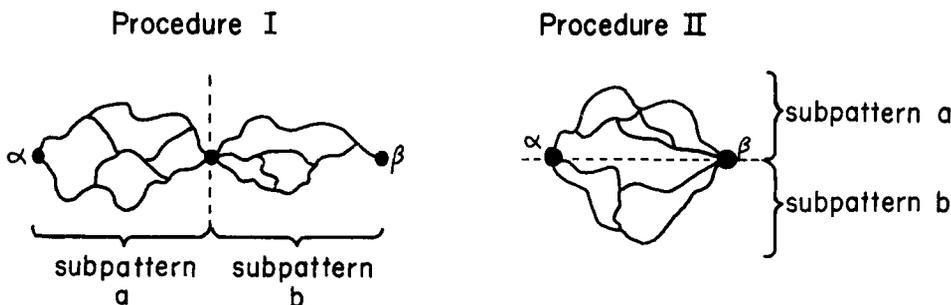


FIG. 1. Basic procedures for constructing cross-linkage patterns from subpatterns.

the combined pattern is

$$C = AC_a C_b / (N_a + N_b)^{3/2} \quad (11)$$

and the effective number of links between the fused end points is

$$N_e = [(1/N_a) + (1/N_b)]^{-1} \quad (12)$$

Figure 2 gives the values of  $N_e$  and  $C$  for various elementary patterns, determined in most instances by means of these rules. For example, pattern  $c_3$  is formed by procedure II from subpattern  $c_2$  and a segment with  $n_3$  links. Pattern  $b_2$  is formed by procedure I from two subpatterns of type  $c_2$ . Pattern  $d_2$  is formed by procedure II from subpattern  $b_2$  and a segment with  $n_5$  links. Values of  $N_e$  and  $C$  for a great many complex patterns are derivable from the information given in Fig. 2.

### III. Applications

These results will now be applied to some particular examples. Let us consider first the three isomers of a chain containing four half-cystine residues, which will be numbered in sequence, 1, 2, 3, 4, starting from one end of the chain (Fig. 3). Let  $n_1$  be the number of links between the end of the chain and half-cystine No. 1,  $n_2$  be the number of links between cystine No. 1 and half-cystine No. 2, etc. Then it is found that the relative numbers of configurations (and hence the relative probabilities in a solvent that produced random coils) of the 1-2, 3-4 isomer, the 1-3, 2-4 isomer and

the 1-4, 2-3 isomer are in the ratio  $(n_2n_4)^{-3/2} : (n_2n_3 + n_3n_4 + n_2n_4)^{-3/2} : [n_3(n_2 + n_4)]^{-3/2}$ . If the half-cystines are equally spaced from each other ( $n_2 = n_3 = n_4$ ) then this ratio is  $1:3^{-3/2}:2^{-3/2} = 1:0.191:0.352$ , and the 1-2, 3-4 isomer is more probable than the others. If  $n_3$  is larger than  $n_2$  and  $n_4$ , the 1-2, 3-4 isomer becomes even more favored. On the other hand, small values of  $n_3$  tend to favor the 1-4, 2-3 isomer. The 1-3, 2-4 isomer does not seem to be especially favored by any set of values of  $n_2$ ,  $n_3$ , and  $n_4$ .

Figure 4 shows the 15 possible isomers of chains containing three cystine cross-linkages, along with formulas from which one can find the relative probability of each isomer when the numbers of links between the half-cystines is known. Also given are the numerical values of the relative probabilities when the half-cystines are equally spaced along the chain (i.e., when  $n_1 = n_2 = n_3 = n_4 = n_5$ ). In this case the 1-2, 3-4, 5-6 isomer is considerably more probable than any of the others. This result can be generalized: *if the half-cystine residues in a chain are approximately equally spaced along the chain, then the most probable arrangement is that in which neighboring cystines are bonded to each other*. The reason for this is as follows: if a given number of rings is introduced into a chain (through cross-linking), the number of configurations of the chain is reduced by a larger factor when the rings are large. The least probable isomers are invariably those in which the cross-linking is between half-cystines that are separated by the greatest number of links along the chain (e.g., 1-3, 2-5, 4-6 and 1-4, 2-5, 3-6).

The methods described above make it possible to calculate the relative probabilities of the 12 isomers of insulin, in which two polypeptide chains

TABLE I  
RELATIVE PROBABILITIES OF ISOMERS OF INSULIN<sup>a</sup>

Isomer	Relative probability	Isomer	Relative probability
2-1', 1-3, 4-2' (native)	1.0	3-1', 1-2, 4-2'	8.0
2-1', 1-4, 3-2'	0.25	3-1', 1-4, 3-1'	0.25
2-1', 1-2', 3-4	0.61	3-1', 1-2', 2-4'	0.25
1-1', 2-3, 4-2'	0.93	4-1', 1-2, 3-2'	8.0
1-1', 2-4, 3-2'	0.26	4-1', 1-3, 2-2'	1.0
1-1', 2-2', 3-4	0.61	4-1', 2-3, 4-1'	0.93

<sup>a</sup> The numbering of the half-cystine residues corresponds to that in Fig. 5.

are held together by a pair of cystines, and a third cystine is present in one of the chains. The results are shown in Fig. 5 and Table I. Native insulin (which has the pattern 1-3, 2-1', 4-2') is neither one of the most probable nor one of the least probable isomers (5).

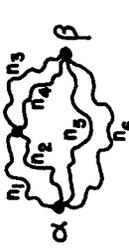
<u>Symbol</u>	<u>Pattern</u>	<u>Ne</u>	<u>C</u>
$a$		$n_1 + n_2$	$C^{n_1+n_2}$
$c_2$		$(m_1 + m_2)^{-1}$	$c_2(n_1+n_2) = A C^{n_1+n_2}/(n_1+n_2)^{3/2}$
$c_3$		$(m_1 + m_2 + m_3)^{-1}$	$c_3(n_1, n_2, n_3) = \frac{A^2 C^{n_1+n_2+n_3}}{(n_1 n_2 + n_2 n_3 + n_3 n_1)^{3/2}}$
$c_N$		$(\sum_{i=1}^N m_i)^{-1}$	$c_N(n_1, n_2, \dots, n_N) = \frac{A^{N-1} C^{\sum n_i}}{(\prod n_i \sum m_i)^{3/2}}$
$b_2$		$(m_1 + m_2)^{-1} + (m_3 + m_4)^{-1}$	$b_2(n_1, n_2, n_3, n_4) = \frac{A^2 C^{n_1+n_2+n_3+n_4}}{[(n_1+n_2)(n_3+n_4)]^{3/2}}$
$b_N$		$\sum_{i=1}^N (m_{2i-1} + m_{2i})^{-1}$	$b_N(n_1, n_2, \dots, n_N) = \frac{A^N C^{\sum n_i}}{\prod (n_{2i-1} + n_{2i})^{3/2}}$
$d_2$		$[(m_1 + m_2)^{-1} + (m_3 + m_4)^{-1} + m_5]^{-1}$	$d_2(n_1, n_2, n_3, n_4, n_5) = \frac{A^5 C^{n_1+n_2+n_3+n_4+n_5}}{[(n_1 n_2 n_3 n_4)(m_1 + m_2 + m_3 + m_4) + n_5(n_1 + n_2)(n_3 + n_4)]^{3/2}}$



$d_N$

$$d_N(n_1, n_2, n_3, n_4, \dots, n_{2N}, n') = \frac{A^{N+1} C^{\sum n_i + n'}}{[\prod (n_{2i-1} + n_{2i}) [n' + \sum (n_{2i-1} + m_{2i})^{-1}]]^{3/2}}$$

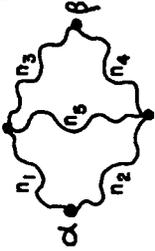
  



$g$

$$g(n_1, n_2, n_3, n_4, n_5, n_6) = \frac{A^4 C^{n_1 + n_2 + n_3 + n_4 + n_5 + n_6}}{[n_1 n_2 (n_3 + n_4) (n_5 + n_6) + n_3 n_4 (n_1 + n_2) (n_5 + n_6) + n_5 n_6 (n_1 + n_2) (n_3 + n_4)]^{3/2}}$$



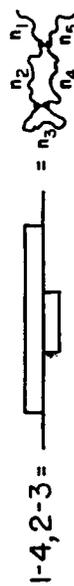
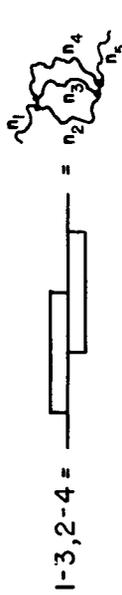
$e$

$$e(n_1, n_2, n_3, n_4, n_5) = \frac{A^2 C^{n_1 + n_2 + n_3 + n_4 + n_5}}{[n_5 (n_1 + n_2 + n_3 + n_4) + (n_1 + n_2) (n_3 + n_4)]^{3/2}}$$

FIG. 2. Effective number of links,  $N^*$ , between end points,  $\alpha$  and  $\beta$ , and total number of configurations for various simple patterns. Note that  $m_i = 1/n_i$ .



### Arrangement of cross-links



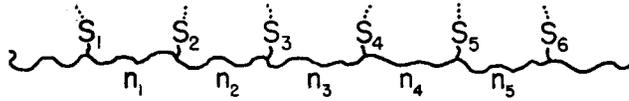
### Number of configurations

$$C_1^{n_1+n_3+n_5} C_2^{n_2} C_3^{n_4} = \frac{A^2 C_1^{n_1+n_2+n_3+n_4+n_5}}{(n_2 n_4)^{3/2}}$$

$$C_1^{n_1+n_5} C_3^{(n_2 n_3 n_4)} = \frac{A^2 C_1^{n_1+n_2+n_3+n_4+n_5}}{(n_2 n_3 + n_3 n_4 + n_4 n_2)^{3/2}}$$

$$C_1^{n_1+n_5} b_2 \left(\frac{n_3}{2}\right)^2 n_2 n_4 = \frac{A^2 C_1^{n_1+n_2+n_3+n_4+n_5}}{[n_3 (n_2 + n_4)]^{3/2}}$$

FIG. 3. Isomers of chains containing two cystine cross-links.



Cross-linking arrangement	Number of configurations	Rel. prob. ( $n_1=n_2=n_3=n_4=n_5$ )
1-2, 3-4, 5-6	$c_2(n_1)c_2(n_3)c_2(n_5)C^{n_2+n_4}$	1.00
1-4, 2-3, 5-6	$b_2(\frac{n_2}{2}, \frac{n_2}{2}; n_1, n_3)c_2(n_5)C^{n_4}$	} 0.353
1-2, 3-6, 4-5	$c_2(n_1)b_2(\frac{n_4}{2}, \frac{n_4}{2}; n_3, n_5)C^{n_2}$	
1-3, 2-4, 5-6	$c_3(n_1, n_2, n_3)c_2(n_5)C^{n_4}$	} 0.192
1-2, 3-5, 4-6	$c_2(n_1)c_3(n_3, n_4, n_5)C^{n_2}$	
1-6, 2-3, 4-5	$c_2(n_2)c_2(n_4)c_2(n_1+n_3+n_5)$	} 0.125
1-6, 2-5, 3-4	$b_3(\frac{n_3}{2}, \frac{n_3}{2}; n_2, n_4; n_1, n_5)$	
1-5, 2-3, 4-6	$c_2(n_2)c_3((n_1+n_3), n_4, n_5)$	} 0.090
1-3, 2-6, 4-5	$c_2(n_4)c_3(n_1, n_2, (n_3+n_5))$	
1-5, 2-6, 3-4	$c_2(n_3)c_3(n_1, (n_2+n_4), n_5)$	} 0.054
1-6, 2-4, 3-5	$c_4((n_1+n_5), n_2, n_3, n_4)$	
1-3, 2-5, 4-6	$d_2(n_1, n_2; n_4, n_5; n_3)$	} 0.044
1-4, 2-5, 3-6	$d_2(n_1, n_4; n_2, n_5; n_3)$	
1-5, 2-4, 3-6	$d_2(n_1, n_4; n_2, n_3; n_5)$	
1-4, 2-6, 3-5	$d_2(n_2, n_5; n_3, n_4; n_1)$	

Fig. 4. Isomers of chains containing three cystine cross-links.

The arrangement of the four disulfide cross-linkages in ribonuclease is also now known (3, 4), as are the number of amino acid residues between each pair of half-cystines (4). The relative probability of the native isomer of ribonuclease can therefore be compared with that of other possible isomers. Since there are 105 isomers altogether, however, this will not

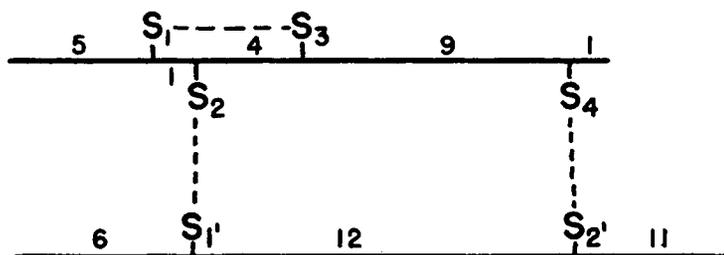


FIG. 5. Locations of half-cystine residues in insulin. The numbers indicate the number of amino acid residues between successive half-cystines.

be done here; instead a few typical isomers are compared in Fig. 6 and Table II. Again we find that the native isomer is neither one of the most probable nor one of the least probable isomers.

Bovine serum albumin contains 17 disulfide cross-linkages, whose ar-

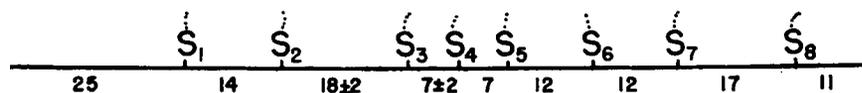


FIG. 6. Locations of half-cystine residues in ribonuclease. The numbers indicate the number of amino acid residues between successive half-cystines.

rangement within the native molecule is as yet not known. It is interesting, however, to make a few calculations for various possible isomers assuming that the half-cystines are equally spaced along the chain. Some results are shown in Fig. 7. It is clear that in accordance with the general state-

TABLE II

RELATIVE PROBABILITIES OF SOME ISOMERS OF RIBONUCLEASE<sup>a</sup>

Isomer	Relative probability
1-2, 3-4, 5-6, 7-8	12.2
1-6, 2-8, 3-7, 4-5 (native)	1.0
1-5, 2-6, 3-7, 4-8	0.13
1-5, 2-6, 3-8, 4-7	0.08

<sup>a</sup> The numbering of the half-cystine residues corresponds to that in Fig. 6.

ment made earlier, by far the most probable isomer is that in which each half-cystine is joined to its next neighbor along the chain. It is interesting to note that this isomer has the highest reduced viscosity of all possible isomers in solvents such as urea and guanidinium chloride. Kolthoff\* has

\* I. M. Kolthoff, personal communication, 1958.

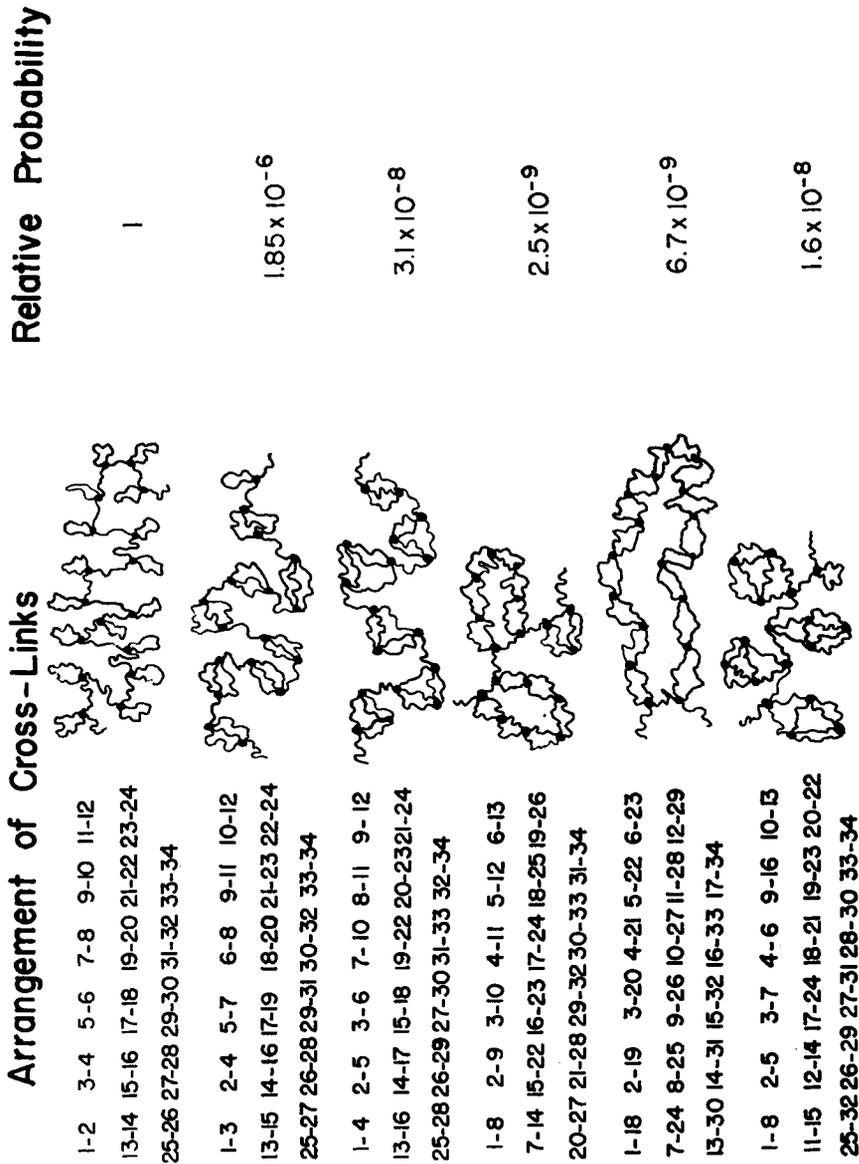


Fig. 7. Relative probabilities of some typical cross-linking arrangements of bovine serum albumin (34 half-cysteine residues).

recently shown that when the disulfide linkages are ruptured in a 0.5% solution of bovine serum albumin which is 4 *M* in guanidinium chloride and then reformed (presumably at random) in the same denaturing medium, the reduced viscosity becomes the same as that of the original protein in guanidinium hydrochloride. This indicates that in native serum albumin most of the half-cystines are joined to near neighbors along the chain. That is, the native molecule must be one of the more probable isomers.

#### IV. Summary

A polypeptide containing more than one cystine residue can exist in several isomeric forms differing in the arrangement of the cross-linkages between the half-cystine residues. If the polypeptide is dissolved in a solvent that favors the randomly coiled form of the polypeptide, the probabilities of the different isomers can be calculated from statistical principles commonly used in polymer chemistry. The results of such calculations are presented and the cross-linking in ribonuclease, insulin, and bovine serum albumin is discussed. The isomers in native ribonuclease and native insulin are neither among the most probable nor among the least probable, whereas native serum albumin may be one of the more probable isomers.

#### REFERENCES

1. C. Huggins, D. F. Tapley, and E. V. Jensen, *Nature* **167**, 592 (1951).
2. W. Kauzmann and R. G. Douglas, Jr., *Arch. Biochem.* **65**, 106 (1956).
3. A. P. Ryle and C. B. Anfinsen, *Biochim. et Biophys. Acta* **24**, 633 (1957).
4. C. H. W. Hirs, W. H. Stein, and S. Moore, in "Symposium on Protein Structure" (A. Neuberger, ed.). Wiley, New York, 1958.
5. A. P. Ryle, F. Sanger, L. F. Smith, and R. Kitai, *Biochem. J.* **60**, 541 (1955).
6. P. J. Flory, "Principles of Polymer Chemistry," p. 402. Cornell Univ. Press, Ithaca, New York, 1953.
7. W. Kuhn and H. Majer, *Makromol. Chem.* **18/19**, 239 (1956).

#### Discussion

VELICK: The experiment on the reduction and reoxidation has been done. A number of years ago Stern and White reduced and reoxidized insulin and did bio-assays for activity. As I recall, they did not get any activity whatsoever.

KAUZMANN: Insulin contains two polypeptide chains held together by disulfide bonds. On reduction the two chains would be separated and on reoxidation one would expect to obtain a mixture of molecules containing one, two, three, and more chains. It is hardly surprising that no activity was found.

HAUROWITZ: Should you not be able to differentiate these different forms by measurement?

KAUZMANN: Yes. Certain of the isomers would be much more viscous and others would be less viscous than the native isomer dissolved in urea. As far as I know, no one has ever measured the viscosity of a protein in urea on reoxidation after reduction of the disulfide groups.

HAUROWITZ: You do not find any change on reduction?

KAUZMANN: One would expect a big change on reduction, and such a change has been observed. If you take serum albumin in urea, measure the viscosity, and then reduce the disulfides, you will find a big increase in viscosity—it almost doubles. This should occur with any disulfide-containing protein dissolved in a solvent such as urea. On reoxidation there should be a drop in viscosity, but whether the viscosity attained will be above or below that observed before reduction depends on the way in which the disulfides were hooked up in the original protein.

EDSALL: I would like to ask how many different isomers you get per serum albumin.

KAUZMANN: There are 17 disulfide groups in serum albumin, so there are  $34!/(2^{17} 17!)$  or about  $10^{10}$  isomers.

MORALES: In the treatment there is the assumption that the chains do not occupy space.

KAUZMANN: Yes.

MORALES: Intuitively I have a feeling that such an assumption might favor the first distribution of SS bonds (more extended isomers) even more because at least the segments do not rub against one another as they would in some of the more condensed distributions.

KAUZMANN: Yes, I agree.

MORALES: I do not know whether it is in your treatment or not, but on what criterion do you assume the segment is long enough?

KAUZMANN: It will not work if  $n$  equals one or two. If  $n$  equals five, it would work.

MORALES: Toward which distribution would the short segment cause a bias?

KAUZMANN: It would be in the direction of having no flexibility at all.

MORALES: So that factor might bias it?

KAUZMANN: It depends on how you treat it. You can put  $n$  equals one or you can neglect it altogether. The true answer is somewhere in between. I hope that no one is going to take these calculations as anything other than an indication of the order of magnitude of the probabilities of the different isomers.

MORALES: The first factor would favor the extended isomers, though.

KAUZMANN: A discussion has been going on for many years about the importance of that factor and it is still not decided. It is an extremely difficult thing to decide whether or not it is important.

LINDLEY: I have been very interested in Dr. Kauzmann's paper which seems to me to have a very direct bearing on some aspects of hair synthesis by the animal. Just before the formation of the fully hardened fibre there seems to be a pouring in of protein which is very rich in thiol groups and the final stage of keratinization is accompanied by the oxidation of these groups to disulfide bonds. However at this stage practically all evidence of cellular organization has disappeared so it may well be that the processes involved are largely uncontrolled and Dr. Kauzmann's statistical treatment of the problem may have direct relevance.

HAUROWITZ: The results of Karush are not at all surprising because he starts with 17 SS bonds in serum albumin and he ends up with 17 SS bonds in free cystine. It seems to prove that the configurational changes do not affect the rotation very much, and it depends chiefly on SS.

KAUZMANN: There is no question that configuration affects the rotation because proteins with little or no SS show very large changes in rotation when they are denatured.

JENSEN: Before we leave serum albumin and go to fibrinogen, Professor Kolthoff has some comments on methods in this field that I think will be valuable to hear.

KOLTHOFF: Mr. Chairman, I will abuse this opportunity to tell a story in connection with sulfhydryl in albumin. During the discussion I could not help thinking of a story about the philosopher who was going to talk to a group of theologians. He was being introduced by a theologian who gave the following definition of a philosopher, "He is like a blind man whom you put in a dark room looking for a black cat which is not there." So when the philosopher started his talk he said, "That was a very fine definition of a philosopher. But he has not made clear what is the difference between a theologian and a philosopher." He said, "If you put a theologian in a dark room looking for a black cat which is not there, he will find it."

I would like to talk a little about reactive disulfide in albumin, which depends on the experimental conditions under which you work. Yesterday Dr. Lindley and Dr. Swan talked in fairly great detail about the reaction between SS bonds and mercaptide or sulfite, and it was stated already that you get an equilibrium. Dr. Swan mentioned already that when we are dealing with low molecular weight compounds like cystine, we make use of the equilibrium reaction with sulfite to titrate the disulfide by titrating the mercaptide. What you do in such a titration is that you displace the equilibrium to the right and this enables one to determine SS quantitatively. So we applied this method to serum albumin first in the native state. If you do that you find no reactive disulfide with the native albumin.

On the other hand, if you denature and dissolve it in 8 *M* urea (we did most of our work in 4 *M* guanidine hydrochloride) and apply the above titration you find an average (of some 100 determinations by different methods which I am not going to discuss) of about 17.2 disulfide bonds in the albumin. You only find it in the denatured albumin. We then decided to determine the equilibrium in the reaction of disulfide in denatured albumin and sulfite. It has taken us a long time to find a method which gave equilibrium values, but we finally applied a very simple method which was given for low molecular weight compounds by Cecil. We did most of our early work with a pH of 6.5. Later we went over the whole pH scale. A given concentration of sulfite is added and the pH adjusted. After various times you freeze the equilibrium by simply bringing the pH to 2. Then you can determine how much SH is formed under these equilibrium conditions. So after waiting for various times until you get a constant value, you can determine the rate of reaction that way. We titrate the SH with mercuric chloride at pH 2 which takes less than one minute. I might tell you about the simple electrode we use for that. We use a rotated mercury pool electrode, composed of a very small cup with a platinum wire sealed in. The cup is joined to a glass tube which can be rotated. The small cup is filled with mercury and the electrode is rotated no faster than 60 r.p.m. because otherwise you run into trouble when the mercury starts to vibrate, and you get results which are not so easy to interpret. The whole secret of the success of the electrode is a layer of a water-repellant substance on the inside glass of the cup; we use silicone. The electrode is much simpler

and ten times more sensitive than the dropping mercury electrode and is at least as sensitive as the rotating platinum electrode.

At a pH of 6 in 4 M guanidine hydrochloride the number of SS bonds of bovine serum albumin reacted, increases with increasing sulfite concentrations up to about 11 SS bonds per mole. In the native protein at pH 6 no reactive SS bonds are found even with 0.1 M sulfite within a reaction period of 24 hours. This is not due, as we thought at first, to inaccessibility of the SS bonds to the sulfite, since, if the reaction mixture also contains 20 equivalents of mercuric chloride, 17 SS bonds react within 5 hours.

With the native albumin it takes longer to complete the reaction than with denatured albumin. The results show that there must be a tremendous difference in the bond strength between the SS in the native state and the denatured state.

There was another interesting thing which came out. When we determined the viscosity or the change of the viscosity with the number of SS bonds broken, by taking different concentrations of sulfite and plotting the reduced viscosity versus the number of SS bonds broken, the result was almost a straight line in 4 M guanidine hydrochloride. With no SS bonds broken the reduced viscosity is 0.19; after 11 SS bonds are broken the viscosity is 0.48; when from 11 to 17 SS bonds are broken, the viscosity does not change at all any more. This result makes it tempting to state that the first 11 SS bonds may be interhelix bonds where the molecule can stretch further and the other 6 SS bonds do not contribute further to the viscosity and are bonds possibly within the helix where the breaking would not affect it very much. I am just presenting a kind of hurried picture of the results obtained.

There is one more thing I would like to mention. I think it is somewhat connected with some questions which came up in Dr. Kauzmann's paper. We thought it would be interesting (most of this work is not in the finished state except this part of the equilibrium which will be published in the near future)\* to see if one can reverse the sulfite-disulfide reaction and get the original protein back. We tried all kinds of things.

Let me summarize what we did. We did two things but one is far better than the other although we got the same result. At pH 2 the reaction still takes place, but it is extremely slow. The equilibrium value is reached after 48 or 60 hours waiting. So what we did was to establish equilibrium at pH 5. We broke 15 of the SS bonds and then we adjusted the pH to 2 and in the complete absence of oxygen we removed the SO<sub>2</sub> from the solution simply by bubbling hydrogen through under reduced pressure. We cannot reverse the reaction completely because the equilibrium concentration of reacted protein disulfide is not affected by the removal of SO<sub>2</sub>. The SH was exactly the same before and after the removal of the SO<sub>2</sub>. The pH was now adjusted to 5 to give a new equilibrium at a much smaller sulfite concentration than originally. The pH was readjusted to pH 2, SO<sub>2</sub> was removed and the process repeated once more. If that is done, you find no SH. In other words, the reaction has been reversed quantitatively. That is, the chemical reaction. The interesting part to know is whether the original protein was formed again, which would be a miracle in itself. You have broken all kinds of SS bonds, and a loop here is being stretched, and an SH group is situated here, and an SSO<sub>2</sub> group there. In order to see if the original molecule was formed again, we determined the viscosity. All this was done in 1% solution.

\* This work has now appeared: I. M. Kolthoff, A. Anastasi, and B. H. Tan, *J. Am. Chem. Soc.* **80**, 3235 (1958).

In 4 *M* guanidine hydrochloride, the reduced viscosity is 0.19. When 11 SS bonds were broken, the reduced viscosity was 0.48. This is the constant value in 1% albumin. We removed the SO<sub>2</sub> and we got the reaction chemically completely reversed, and the intrinsic viscosity was something of the order of 0.27 to 0.28 but was definitely higher than the value in the original solution. This was attributed to cross-linking which should become less with decreasing albumin concentration. We repeated the experiments in 1.5% and 0.5% albumin. In the most dilute solutions, we got back the original value of the viscosity. We would like to do it in still more dilute solutions, but we do not know how to measure that viscosity.

BROWN: Dr. Kolthoff's comments on the reaction of disulfides with bisulfites are most interesting to us. We have studied the same types of systems in hair waving. We measured the effects of bisulfite solutions containing urea or guanidine salts on hair and found that changes in cystine content, swelling, and tensile properties were at a maximum in a pH region of 5-6. Therefore, with his chemical system, denatured proteins and keratin in hair show similar types of reactivity, and this is a point which Dr. Lindley made yesterday.

BENESCH: I would like to ask about your electrode which looks like an immense improvement. How often do you have to change the mercury?

KOLTHOFF: In these titrations very few times because we do not get a foreign metal (amalgam) deposited. We get no contaminant at the surface of the electrode with another metal. We may do it once a week or every day or so. There is very little mercury in it.

BENESCH: Do you use it for the reduction of inorganic compounds? How often would you have to change it then?

KOLTHOFF: I could not answer you. If you take an inorganic compound which is reduced, I would be a little afraid of using the same mercury. I would prefer to refill. If you refill, you can reproduce the results. In these amperometric titrations, I do not think it matters whether the surface is dirty or not. All you want is the straight line.

BENESCH: It has the same sensitivity as the rotating platinum electrode, although the surface is much smaller?

KOLTHOFF: Yes. We have had trouble with that disulfide titration with silver. I do not know whether you tried it. We cannot do it at a lower pH than 9. We have to use mercury. We can now do it with the rotated mercury electrode to great advantage.

BENESCH: Have you published on this analytically?

KOLTHOFF: We described this in the first paper on the reactivity of the disulfide in one brief paragraph. I think we gave the dimensions of the cup. It is a little cup and there is a platinum wire to make the contact, and all you have to do is to coat the thing with silicone inside.