*Proc. Nat. Acad. Sci. USA*
Vol. 70, No. 3, pp. 697–701, March 1973

# Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins

**(protein structure/chain continuity/independent regions/self-assembly)**

DONALD B. WETLAUFER

Department of Biochemistry, The Medical School, 227 Millard Hall, University of Minnesota, Minneapolis, Minn. 55455

*Communicated by John T. Edsall, December 19, 1972*

**ABSTRACT** **Distinct structural regions have been found in several globular proteins composed of single polypeptide chains. The existence of such regions and the continuity of peptide chain within them, coupled with kinetic arguments, suggests that the early stages of three-dimensional structure formation (nucleation) occur independently in separate parts of these molecules. A nucleus can grow rapidly by adding peptide chain segments that are close to the nucleus in aminoacid sequence. Such a process would generate three-dimensional (native) protein structures that contain separate regions of continuous peptide chain. Possible means of testing this hypothesis are discussed.**

For some decades thermodynamic determination of three-dimensional protein structure has been a basic guiding principle. However, in the past few years there has been a growing awareness of the possibility that kinetic factors may play a significant role in determining protein structure (1–5). In particular, the argument was advanced by Levinthal (refs. 6, 7, see also *Appendix*) that the time for a random search of all possible structures would be unrealistically long for even a small protein, and that something like a nucleation event must occur to permit structure formation in biologically feasible time. Studies from our laboratory have led to experimental evidence for nucleation in physiologically feasible protein-folding experiments (8–10). Nucleation has also been inferred from kinetic evidence by Tsong *et al.* (11) in the refolding of ribonuclease with intact disulfide crosslinks. Arguments for the plausibility of nucleation-initiated protein folding have begun to appear in the literature (12, 13).

In studying several known protein structures, I have noted the existence of distinct structural regions in several single-chain proteins, and less distinct regions in others. Such observations are not original, having been made in almost every case by the investigators who determined the structure of the particular proteins.

In defining what I mean by structural regions, I limit our discussion to protein models that present only the course of the peptide backbone. A "region" is a section of peptide chain that can be enclosed in a compact volume. A continuous region can be completely surrounded by a closed surface (a closed line if two-dimensional, as in Fig. 1), and is characterized by possession of two terminal points. I define the terminal points of a protein region as those points where the peptide chain crosses the enclosing envelope, and also those points where the peptide chain terminates (amino or carboxyl terminus). A region having more than two terminal points is discontinuous. Fig. 1 illustrates continuous and discontinuous regions with several two-dimensional examples. Each of the C regions is seen to have two terminal points, and each of the D regions has four terminal points. The aminoacid residues in a continuous region are one-dimensionally contiguous; the condition of contiguity does not obtain for a discontinuous region. It remains to say how compactness is determined. For the present work, compactness has been decided subjectively, but a quantitative index of compactness is suggested in the *Discussion*.

Examination of protein structures was facilitated by the construction of three-dimensional peptide chain models. Bent-wire models were built with the Rubin–Richardson bending tool (14). As a general test of fidelity the physical models were compared with the stereoscopic images of Dickerson and Geis (15) and those in ref. 16.

### Examination of known proteins for continuous regions

*Chymotrypsin.* Referring to the hydrogen-bonding map (Fig. 2), Blow writes, "The pattern of zigzag lines is drawn to emphasize the existence of two folded units in the molecule from residues 27–112 and from residues 133–230" (17). These are both "continuous" regions. On inspection of a Kendrew skeletal model of chymotrypsin, it was seen that a newspaper could be inserted far into the model, almost completely bisecting it into the structural regions noted above.

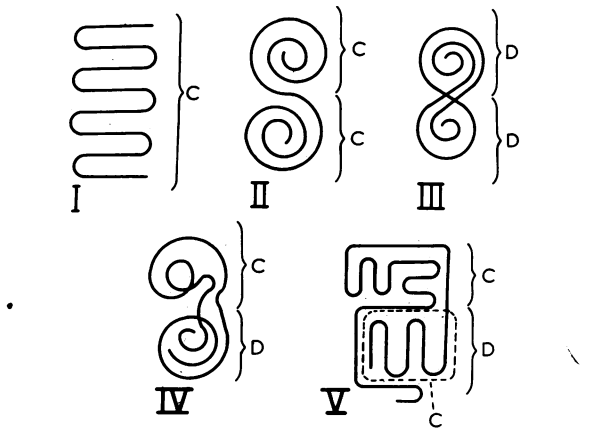


FIG. 1. Schematic peptide chain arrangements to illustrate continuous (*C*) and discontinuous (*D*) structural regions. Scheme I represents a one-region protein, which is by definition continuous. II represents a two-region protein, both regions of which are continuous. In III the two regions represented are both discontinuous. In IV the upper region is continuous, the lower discontinuous. In V the upper region is continuous and, taken as a whole, the lower region is discontinuous. A continuous region in the lower part of Scheme V is enclosed by a *dashed line*.

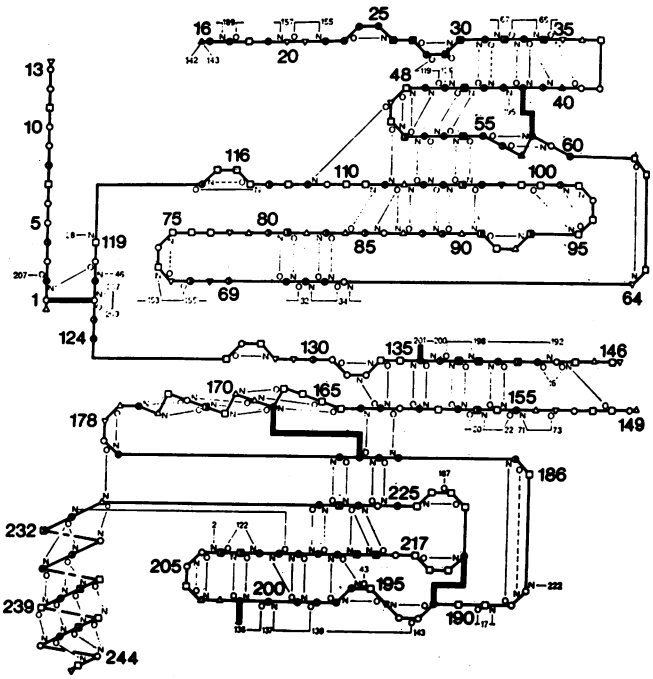


FIG. 2. Hydrogen bonds between main-chain atoms in $\alpha$-chymotrypsin. The *thin lines* representing hydrogen bonds are *dashed* when uncertain. The *broad black bars* represent disulfide bridges [from Birktoft and Blow (47)]. Reproduced with permission of the authors and publisher.

*Trypsin.* "The trypsin molecule folds up in two halves, each of which contains a pseudo-cylindrical arrangement of hydrogen bonds between adjacent antiparallel extended chains similar to that described by Blow for $\alpha$-chymotrypsin. The two termini of the molecule form transmolecular straps which terminate in opposite halves of the molecule..." (18). Thus, trypsin, like chymotrypsin, contains two large "continuous" regions; however, when the small additional segments of polypeptide chain (the "straps") are considered, both regions become discontinuous.

*Elastase.* "...one can see that as in chymotrypsin the molecule appears to be divided into two halves composed of residues 27–127 plus the C-terminal helix (234–245) in the upper left hemisphere and residues 128–230 plus the N-terminal sequence (residues 16–26) in the lower right hemisphere" (19). Our remarks for trypsin can be repeated for elastase.

*Subtilisin.* Referring to photographs of a model of the peptide chain of subtilisin, Kraut writes, "It is immediately obvious that the backbone chain has folded into three distinct and potentially separable pieces" (20). These regions are 1–100, 101–176, and 177–275; all three regions are "continuous."

*Papain.* "A characteristic feature of the papain molecule is its binuclear nature. It is constructed around two hydrophobic cores..." (21). The two continuous regions are comprised of residues 10–111 and 113–207. The two chain ends overlap into the opposite halves of the molecule to make both halves of the whole molecule "discontinuous."

*Lysozyme.* "...residues 1–86 constitute a structure with two wings. Residues 88–100 form a rather irregular $\alpha$-helix, again with a partially hydrophobic surface, which bridges the gap between these two wings... Finally, the last 30 residues are folded around the globular unit built up by the first 40" (22). Thus, there are two "continuous" regions; 1–38 and 39–87. Residues 88–100 could be equally well assigned to either of these regions. Considering the whole molecule, region 39–87 remains "continuous," while a "discontinuous" region is composed of residues 1–38 plus 101–129, arranged in a fashion similar to Fig. 1, *V*.

*Lactate Dehydrogenase.* "The interpretation of the 2.8 Å map led us to the conclusion that the LDH subunit could be described as two halves.... There are now (at higher resolution) only two lengths of chain of more than 10 residues which are not involved in either $\beta$ or helical structures, and one is in each half of the subunit" (23). These two regions appear to be composed of residues 20–161 and 162–231. The latter of these may possibly be further subdivided at residue 265. The regions are "continuous."

*Malate Dehydrogenase.* The peptide chain contour of this enzyme is so similar to that of lactate dehydrogenase that it is easier to describe it in terms of a few differences than in terms of the many similarities (23a).

*Phosphoglycerate Kinase.* "The single peptide chain of 3-phosphoglycerate kinase is folded into two distinct globular units, only one of which seems to be involved in substrate binding." (24) We cannot tell from the data at 6 Å resolution whether the two large globules are "continuous," nor whether any further subdivisions may exist.

*Thermolysin.* "There is a deep groove or cleft across the middle of the molecule...with the zinc atom situated in the bottom of this cleft.... Immediately behind the zinc, running parallel to the cleft, and through the center of the molecule is a helix which continutes two ligands to the zinc. Apart from its carboxyl terminal residues this helix is totally internal. The only covalent connection between the upper and lower halves of the molecule is through this central helix." (25) Both of the two regions are "continuous."

*Bovine-Plasma Albumin.* On the basis of analysis of the products of limited proteolysis and CNBr cleavage (26, 27) of this protein and the physical behavior of certain protein fragments, Foster has proposed a structural model. Quoting Foster, "A most interesting characteristic of this model is its subunit-like structure. At least four regions of the molecule are found to be *intralinked* by disulfide bonds, but connected to each other only through the peptide backbone" (26). The available data do not permit a decision on the "continuity" issue.

*Immunoglobulins.* "The repetition of homologous (amino-acid sequence) regions, the linear, periodic arrangement of intrachain disulfide bonds, and the location of inter-chain bonds combine to suggest an unusual quasiperiodicity.... If one considers the homologies among the various regions of the molecule, one plausible arrangement of the chains is a T-shaped series of compact domains, each of which is formed by a separate V-homology region or C-homology region. In this arrangement, each domain is stabilized by a single, intrachain disulfide bond and is linked to neighboring domains by less tightly folded stretches of the polypeptide chain..." (28). X-ray diffraction studies by Sarma *et al.* (29) give results at 6-Å

resolution that can be interpreted either as a T-shaped or a Y-shaped molecule, showing distinct structural regions. We must remember that IgG is a 4-chain protein, so the possibility exists that structural regions seen by x-ray diffraction at 6 Å may not be intrachain.

*Ribonuclease A.* This protein has two major regions, both of them discontinuous. One is composed of residues 1–10, 52–76, and 105–124, inclusive (30). The other is composed of residues 11–49 and 80–103, inclusive. We do not see any "continuous" regions.

*Myoglobin.* There are two "continuous" regions, composed of residues 1–79 and 80–153. A problem arises in the existence of a large interface between the two regions, involving most of helixes G and H of the second region with helixes A, B, and E of the first region (31). With such a large interface, and the accompanying extensive interactions, it is hard to view the two regions as structurally independent. We do not know how to deal with the role of heme in the structure. Apomyoglobin contains less helix than metmyoglobin (32), but the structure of apomyoglobin is unknown.

*Carp Parvalbumin.* There are two "continuous" regions, composed of residues 1–34 and 35–107. However, as in metmyoglobin, there are large surfaces of close approach between the regions—residues 15–25 and residues 65–75 form two rather extended loops of peptide chain (33) that appear to be structurally interdependent.

*Staphylococcal Nuclease.* There are no *distinct* regions in this protein. Residues 97–142 form what might be called a "continuous" region, with the remainder of the polypeptide chain draped and looped about in no compact region. Even in 97–142 the quality of compactness is poorly realized (34).

*Pancreatic Trypsin Inhibitor.* This protein, containing 58 aminoacid residues, shows but a single "continuous" region (35).

*Rubredoxin.* Like pancreatic trypsin inhibitor, this small protein (54 residues) shows but one region (36).

## RESULTS AND DISCUSSION

Nine of the first ten proteins listed above have two or three "continuous" regions (phosphoglycerate kinase continuity is not established). The last six proteins show either "discontinuous" regions, dubious regions, or one region. The data presently available for bovine plasma albumin and IgG, although strongly suggestive of "continuous" regions, are still insufficient. The size of a "continuous" region in the present collection ranges from about 40 to 150 residues. In pancreatic trypsin inhibitor and rubredoxin, the size of the protein is near the lower limit of size range of "continuous" regions found in other proteins. Perhaps multiple nucleation does not occur in such small proteins.

Native proteins that carry components that are not amino acids, such as heme, flavin, and pyridine nucleotides, that are required for the self-assembly process appear to pose additional considerations and are not as easily handled at our present stage of analysis. Likewise proteins such as insulin and carboxypeptidase A, which derive from precursors of uncertain structural relationship to the active proteins, must be examined with reserve. I have considered chymotrypsin, trypsin, and elastase to be structurally closely related to their respective zymogens, on the basis of the evidence relating chymotrypsin to chymotrypsinogen, and the close structural homology of the three active enzymes (19).

The existence of compact intrachain regions has been noted for individual proteins, but does not seem to be recognized as a general pattern of protein structure. Further, the character of chain continuity in most of these regions has not been previously recognized. I believe that chain continuity is an important structural feature because it is just what one would expect of a process of nucleation followed by rapid growth. The most rapid growth on a nucleus will occur by addition of nearby residues onto the nuclear structure. Distant residues will require longer times for diffusion to, and orientation on, the nuclear structure. It is evident that those residues that are one-dimensionally near a nuclear structure must be near in three dimensions. By the same reasoning, I expect that discontinuous regions, composed of one-dimensionally distant chain segments, will require considerably longer assembly times. Therefore, the formation of continuous regions will be favored over discontinuous regions.

The size of the regions I have described in proteins ranges from about 40 to 150 aminoacid residues. I do not equate these regions with nuclei, but rather view a region as comprising the nucleus plus structure added to the nucleus by the growth processes described in the preceding paragraph. I roughly estimate the size of a nucleus to be of the order of 8–18 aminoacid residues. The upper limit of this estimate is set by the requirement for rapid assembly on a biological time scale. The absolute time requirements will obviously vary among organisms and tissues. Using the Poincaré argument (*Appendix*), which is admittedly oversimplified, to estimate the time $t$ for nucleation by a random search mechanism, I find that $t = 10^n/10^{14}$ for an oligopeptide of $n$ residues. By this argument, a nucleus of 18 residues will require an average of $10^4$ sec to form. Since biosynthetic processes involving steps longer than $10^4$ sec seem quite improbable, we set 18 residues as our upper limit. The lower limit of our estimate is based on the well-established idea that there is a minimum size for stable three-dimensional structures in peptides (37, 38). What this minimum number will be for a particular peptide clearly depends on the aminoacid sequence and the solvent. Several observations of nonrandomness exist for peptide fragments of proteins in aqueous solvents (39–42). The optical criteria for structure in this sampling of peptides show from 5% to 20% "helix." Although the helixes and other structures in these peptides appear to be of marginal stability, only marginal stability is required for a nucleus. On the basis of these studies (39–42), we tentatively conclude that a nucleus may be as small as 8 residues.

Another consequence of the marginal stability of oligopeptide structures is that they might be distorted from the structure formed at the instant of nucleation, or unfolded or "dissolved" at some stage of structure or growth. The foregoing possibilities appear to be real, though perhaps less likely than stabilization and "locking-in" of the nuclear structure in the nucleation-propagation region.

As noted above, continuous regions appear to range in size from 40 to 150 aminoacid residues. Does some constraint(s) operate to limit the size of the structural regions? It may be that a relationship between nucleation time and growth time is dependent on the size of a region in such a way as to limit the

size of regions and favor (in larger proteins) the formation of multiple-region proteins.

I do not feel that we have completely adequate criteria to define protein regions, but enough to make a start. It appears that we could measure the compactness of a region by the use of a surface/volume ratio, $A/V$, normalized by dividing by the surface/volume ratio $A_0/V$ of a sphere of equal volume. In analogy to hydrodynamic frictional ratios, $f/f_0$, $(A/V)/(A_0/V) = (A_0/A)$, a relative surface area, that could have real values ranging from 1.00 upward, forming a numerical scale on which the more compact structure will have smaller values. Experimental evaluations of $A$ and $V$ should be possible by means of programs such as those used (43) to measure the static solvent-accessibility of globular proteins.

H-bonding maps of the sort shown in Fig. 2 for chymotrypsin have been published for a few proteins. The regions seen in the three-dimensional peptide chain model are the same as those seen in the H-bonding map. This result also obtains with subtilisin (44) and lactate dehydrogenase (23). On the other hand, the H-bonding map of ribonuclease S (45) does not show continuous regions, consistent with examination of the peptide backbone model. The H-bonding map of myoglobin (31) is largely uninformative, since about 75% of the residues are in helical structures. Thus, very few chain segments interact via H-bonding with three-dimensionally adjacent chain segments, so the map does not show which chain-segments are closely juxtaposed.

Further exploration of correlations between H-bonding maps and 3-dimensional peptide chains regions would clearly be of interest. I hope that such maps will soon be available for additional proteins. In the same vein, it will be of great interest to test for correlations between protein regions and three-dimensional distribution of nonpolar residues. Determination of the three-dimensional distribution of nonpolar side chains has been begun by Kuntz (46) with carboxypeptidase A.

I should also like to explore the intramolecular interfaces between regions to test whether any unusual compositional or structural characteristics exist, and find it very suggestive that all of the 13 water molecules in the interior of chymotrypsin are located at the contact surface of the two regions of that molecule (47).

One of the most searching experimental tests for an independent continuous region would be to demonstrate self-assembly of just that region, with a high degree of fidelity. It appears that such a test may be experimentally feasible for putative regions of some proteins.

Proteins with discontinuous regions may result from nucleation and folding pathways in which the nucleus is distorted so as to give a final structure without recognizable continuous regions. They could also arise from several rather than a few nucleation events. Or they may be genuine exceptions to our present attempt to generalize folding mechanisms. In no event can we ignore them.

How does this hypothesis relate to the "domain hypothesis"? The domain hypothesis (28) claims that for the immunoglobulins there are structurally separate regions of the molecule, and asserts that independent genetic control is exerted over these regions. My present hypothesis applies to the much larger class of globular proteins, postulates that rapid self-assembly is a major reason for structurally independent regions, and makes no claims relating to genetic control of the regions.

## APPENDIX

If we calculate, by the Poincaré recursion argument, the average time for a chain polymer of 100 links to randomly sample the possible structures and return to the initial structure, the result is $10^{85}$ sec. The recursion time for a tetradecapeptide, in contrast, is only 0.04 sec. To make this calculation, we assume a polymer chain of $n$ links, each of which has two rotatable bonds with three energetically equal structures for each bond. Then the total number of structures is $3^{2n}$, which can be approximated for convenience to $10^n$. Assume that each of the internal rotations occurs independently at a frequency of $10^{13}$ sec$^{-1}$, then $2 \times 10^{13}\ n$ structures are sampled per second. The time t required to sample each conformation m times is $t = (m\ 10^n \text{ sec}/2 \times 10^{13}n)$. This argument was first pointed out to me by Prof. V. Bloomfield (personal communication, 1968); a similar argument was suggested by Levinthal in a generally unavailable publication (7). Of course such a model is greatly oversimplified, ignoring excluded volume effects, sidechain flexibility, correlated motions, attraction between different parts of the chain, etc. Nonetheless, it is highly doubtful that accounting for these effects could reduce the random search time to a biologically feasible value, say in the range of $10^{-1}$–$10^3$ sec. The argument was originally advanced to show the necessity for some preliminary event (nucleation) that could reduce the number of structures searched and the search time. If this is a reasonable argument for a small protein, then multiple regions of nucleation seem reasonable for larger proteins. Indeed, this argument can be added to those commonly made for the existence of conventional subunit proteins (48).

I gratefully acknowledge the generous cooperation of Dr. B. Rubin and Prof. Jane Richardson for providing plans for construction of their model-building instrument, and related information. I also thank the whole community of protein x-ray crystallographers, too numerous to mention individually, for providing data and preprints, and acknowledge useful discussions with Prof. I. D. Kuntz, who also provided preprints of his publications. I thank W. L. Anderson, E. Johnson, and S. Saliterman for constructing several models. This work was supported by grants from the University of Minnesota Graduate School and by USPHS Grant 1R01 GM 18814-01.

1. Mirsky, A. E. & Pauling, L. (1936) *Proc. Nat. Acad. Sci. USA* **22**, 439–447.
2. Linderstrøm-Lang, K. (1950) *Cold Spring Harbor Symp. Quant. Biol.* **14**, 117–125.
3. Kauzmann, W. (1959) *Advan. Protein Chem.* **14**, 1–63.
4. Tanford, C. (1970) *Advan. Protein Chem.* **24**, 2–97.
5. Scheraga, H. A. (1971) *Chem. Rev.* **71**, 195–217.
6. Levinthal, C. (1968) *J. Chim. Phys.* **65**, 44–45.
7. Levinthal, C. (1969) in *Mössbauer Spectroscopy in Biological Systems* (proceedings of a meeting held at Allerton House, Monticello, Ill.), pp. 22–24.
8. Saxena, V. P. & Wetlaufer, D. B. (1970) *Biochemistry* **9**, 5015–5023.
9. Saxena, V. P. (1971) *Fed. Proc.* **30**, 1287.
10. Ristow, S. S. & Wetlaufer, D. B. (1973) *Biochem. Biophys. Res. Commun.*, in press.
11. Tsong, T. Y., Baldwin, R. L. & McPhie, P. (1972) *J. Mol. Biol.* **63**, 453–475.
12. Epstein, H. G., Schecter, A. N., Chen, R. F. & Anfinsen, C. B. (1971) *J. Mol. Biol.* **60**, 499–508.
13. Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2293–2297.
14. Rubin, B. & Richardson, J. (1972) *Biopolymers*, in press.
15. Dickerson, R. E. & Geis, I. (1969) Stereosupplement to *The Structure and Action of Proteins* (Harper and Row, New York).
16. Two-Color Stereo Diagrams (Plates I-XXIX) (1972) in *Cold Spring Harbor Symp. Quant. Biol.* **36**.

17. Blow, D. (1971) in *The Enzymes*, ed. Boyer, P. D., (Academic Press, New York), 3rd ed., Vol. 3, pp. 185–212.
18. Stroud, R. M., Kay, L. M. & Dickerson, R. E. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 125–140.
19. Hartley, B. S. & Shotton, D. M. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York), 3rd ed., Vol. 3, pp. 323–373.
20. Kraut, J. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York), 3rd ed., Vol. 3, pp. 547–560.
21. Drenth, J., Jansonius, J. N., Koekoek, R. & Wolthers, B. G. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York), 3rd ed., Vol. 3, pp. 485–499.
22. Phillips, D. C. (1967) *Proc. Nat. Acad. Sci. USA* **57**, 484–495.
23. Rossman, M. G., Adams, M. J., Buehner, G. C., Hackert, M. L., Lentz, P. J., Jr., McPherson, A. Jr., Schevitz, R. W. & Smiley, I. E. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 179–191.
23*a*. Hill, E., Tsernoglou, D., Webb, L. & Banaszak, L. J. (1972) *J. Mol. Biol.*, in press.
24. Blake, C. C. F., Evans, P. R. & Scopes, R. K. (1972) *Nature New Biol.* **235**, 195–198.
25. Matthews, B. W., Jansonius, J. N., Colmar, P. M., Schoenborn, B. P. & Dupourque, D. (1972) *Nature New Biol.* **238**, 37–41.
26. Pederson, D. M. & Foster, J. F. (1969) *Biochemistry* **8**, 2357–2365.
27. King, T. P. & Spencer, M. (1970) *J. Biol. Chem.* **245**, 6134–6148.
28. Cunningham, B. A., Gottlieb, P. D., Pflumm, M. N. & Edelman, G. M. (1971) in *Progress in Immunology*, ed. Amos, B. (Academic Press, New York), pp. 3–24.
29. Sarma, V. R., Davies, D. R., Labaw, L. W., Silverton, E. W. & Terry, W. D. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 413–425.
30. Kartha, G., Bello, J. & Harker, D. (1967) *Nature* **213**, 862–865.
31. Watson, H. C. (1968) in *Progress in Stereochemistry*, eds. Aylett, B. J. & Harris, Margaret M. (Butterworth & Co., London), Vol. 4, pp. 299–333.
32. Harrison, S. C. & Blout, E. R. (1965) *J. Biol. Chem.* **240**, 299–303.
33. Kretsinger, R. H., Nockolds, C. W., Coffee, C. J. & Bradshaw, R. A. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 217–220.
34. Cotton, R. A., Bier, C. J., Day, V. W., Hazen E. E., Jr. & Larsen, S. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 243–255.
35. Huber, R., Kukla, D., Rühlmann, A. & Steigemann, W. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 141–150.
36. Herriott, J. R., Sieker, L. C., Jensen, L. H. & Lovenberg, W. (1970) *J. Mol. Biol.* **50**, 391–406.
37. Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg, Ser. Chim.* **29**, 230–259.
38. Goodman, M., Verdini, A. S., Toniolo, C., Phillips, W. H. & Bovey, F. A. (1969) *Proc. Nat. Acad. Sci. USA*, **64**, 444–450.
39. Bodanszky, A., Ondetti, M., Mutt, V. & Bodanszky, M. (1969) *J. Amer. Chem. Soc.* **91**, 944–949.
40. Brown, J. E. & Klee, W. A. (1971) *Biochemistry* **10**, 470–476.
41. Shearer, W. T., Brown, R. K., Bryce, G. F. & Gurd, F. R. N. (1966) *J. Biol. Chem.* **241**, 2665–71.
42. Epand, R. M. (1972) *J. Biol. Chem.* **247**, 2132–38.
43. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
44. Drenth, J., Hol, W. G. J., Jansonius, J. N. & Koekoek, R. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 107–116.
45. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970) *J. Biol. Chem.* **245**, 305–328.
46. Kuntz, I. D. (1972) *J. Amer. Chem. Soc.* **94**, 8568–8572.
47. Birktoft, J. J. & Blow, D. M., (1972) *J. Mol. Biol.* **68**, 187–240.
48. Sund, H. & Weber, K. (1966) *Angew. Chemie, Int. Ed. Engl.* **5**, 231–245.