Biophysical Chemistry 19 (1984) 39–47
Elsevier

BPC 00824

A NEW THEORETICAL APPROACH TO THE INVESTIGATION OF THE SYMMETRY OF PROTEIN OLIGOMERS WITH BIFUNCTIONAL REAGENTS

M.J. SCULLEY, G.B. TREACY and P.D. JEFFREY

Department of Physical Biochemistry, The John Curtin School of Medical Research, Australian National University, Canberra, ACT 2601, Australia

Received 29th June 1983
Accepted 13th September 1983

Key words: Cross-linking; Protein oligomer symmetry; Hexamer; Hemocyanin; Urease

The use of bifunctional reagents to form cross-links between subunits in protein oligomers and subsequent disruption of noncovalent interactions with SDS allows comment upon the number of subunits and the symmetry in the original assembly. In existing treatments the number of equations needed to describe theoretically the proportions of all the cross-linked species that can be formed as a function of time in this way makes the analysis of the system unmanageable for proteins with more than four subunits. A method is presented that allows the required equations for any oligomer to be formulated as an algorithm suitable for solution by computer. Its application is illustrated with reference to experimental results obtained with two protein hexamers, *Jasus* hemocyanin and α-urease from jack bean.

1. Introduction

Valuable insights into protein quaternary structure are potentially available through the use of bifunctional reagents that form linkages between pairs of subunits. To interpret the distribution of the variously cross-linked species that one observes after disruption of the noncovalent interactions in the original structure, the experimenter needs to know how the patterns obtained can be interpreted in terms of specific given models. Two groups of workers [1,2] have addressed the problem of predicting these patterns. Both theoretical treatments, of which that due to Hajdu et al. [1] is the more general, refer to tetrameric assemblages and cannot be extended directly to oligomers made up of more than four subunits. The reason is that they require the formulation of equations which contain, explicitly, terms for all of the kinds of cross-linked species that can be formed and which may contribute to the observed dissociation pattern.

In general, the number of different cross-linked species that can be formed in a given oligomer by a bifunctional reagent is 2^n, where n is the number of different ways of forming cross-links between pairs of subunits. For example, for a six-subunit protein there are 15 possible cross-links and 2^{15}, or more than 32000 differently cross-linked species that can be generated. Clearly, a four-subunit array (2^6 species) is the most complex case to which it is feasible to apply the existing analysis procedure. However, protein hexamers are not uncommon, indeed, even in the absence of a detailed theory, the cross-linking method has been applied to glutamate dehydrogenase [3] and leucine aminopeptidase [4] and it has also been explored in relation to even more elaborate structure [5,6]. It therefore seemed desirable to devise a methodology capable of application to an oligomer containing any number of subunits.

In this paper we present such a procedure developed in terms of a six-subunit model. We illustrate its use by analysis of the experimental results obtained with two proteins for which there

is substantial evidence of hexameric constitution, the 17 S aggregate of an arthropod hemocyanin and α-urease from jack bean.

2. Materials and Methods

2.1. Materials

The 17 S aggregate from the hemolymph of an arthropod (*Jasus* sp.) was prepared from samples of hemolymph donated by Dr. C. Parish, which had previously been through a recrystallization procedure. The samples were then dialysed against 0.05 M Tris-HCl at pH 7.8 containing 0.03 M $CaCl_2$, followed by 0.05 M Tris-HCl, pH 7.8, and then against 0.025 M NaH_2PO_4 adjusted with NaOH to pH 7.5. Polyacrylamide gel electrophoresis at pH 7.8 showed that the protein was now all in the form of 17 S material.

Purified urease solutions in phosphate buffer were donated by Miss E. Owen, and showed a single band on polyacrylamide gel electrophoresis at pH 7.8.

Fluka glutaraldehyde, 25% in water, was diluted with 0.01 M phosphate buffer to the required concentration.

2.2. Cross-linking

In order to minimise intermolecular cross-linking, the concentration of protein was kept low, and a suitable concentration of glutaraldehyde was found, which gave a complete reaction time of less than 24 h. However, the molarity of the glutaraldehyde must be high enough to ensure a large excess over the molarity of the subunit in the reaction mixture.

At various times, aliquots were treated with hot SDS solution containing dithiothreitol (DTT), and the extent of cross-linking at that point could then be estimated by running polyacrylamide gels to see which cross-linked species had been formed.

Protein solutions, typically 400 μl of approx. 0.2% in 0.025 M phosphate, pH 7.5, were incubated at 20°C, and at zero time an aliquot of glutaraldehyde solution (40 μl of 1%) was added, and the mixture vibrated for 0.5 min. At intervals samples, usually 40 μl, were removed to stoppered tubes containing 100 μl of 3% SDS and 10μl of 1% fresh DTT solution, already heated in a boiling water bath for 1 min. After 2 min further heating the tubes were cooled and 50 μl of 40% glycerol added in readiness for later SDS-polyacrylamide gel investigation.

Examination of negatively stained preparations of crosslinked *Jasus* hemocyanin and urease hexamers in the electron microscope showed specimens indistinguishable in appearance from the native molecules. Thus, the cross-linking produced no gross distortion of the structures, nor was there any evidence of intermolecular cross-linking. The molecular weights of the bands measured in SDS on polyacrylamide gels also established that only cross-linked species up to and including hexamers were produced under the conditions used for cross-linking.

2.3. Polyacrylamide gel electrophoresis and densitometry

All experiments were carried out at 20°C in a water-cooled Buchler apparatus using columns 9.5 cm in length, and a continuous buffer system containing 0.05 M Tris adjusted to pH 7.8 with HCl. The SDS concentration in the gels and electrode buffer was 0.1% and the concentration of acrylamide was 3.5%, with a ratio of acrylamide to *N,N'*-methylenebisacrylamide of 40 : 1. Gels were polymerised with 0.1% *N,N,N',N'*-tetramethylenediamine and 0.067% ammonium persulfate.

After the gel run, the gels were immersed in 25% isopropanol containing 10% acetic acid to fix the protein and elute SDS. After several hours the fixer was replaced by the same solution containing 0.05% Coomassie blue to stain the protein. Destaining the next day was done in a Comalco horizontal destainer and the gels were then allowed to equilibrate overnight in 7% acetic acid.

The proportions of protein in the different bands observed were determined by scanning the gels at a wavelength of 540 nm in a Schoefl spectrodensitometer SD 3000. The areas under each peak of the scan, proportional to the amount of protein present, were traced for each component, and the areas cut out and weighed. The

results of this process obtained by two independent observers were found to be in good agreement, the percentage of the total attributed to a given peak being assessable to within better than ±5% in all cases.

2.4. Computation

A computer program which uses the results of section 3 to calculate the concentration of each band as a function of the duration of cross-linking was written in Fortran V. The program operates in the following way: for any protein molecule, only a finite number of possible arrangements of cross-links exists. The probability of occurrence of each one of these arrangements at a particular time is calculated using the equations giving probability of occurrence of a cross-link in terms of cross-linking time. For each of these arrangements, the program calculates how many monomers, dimers, trimers, etc., will result from treatment with SDS. The program does this by tracing the cross-link arrangement through the molecule and finding which subunits are connected to each other via cross-links. The total concentration of each species is obtained by summing these numbers, weighted by their probability of occurrence, over all the possible cross-link arrangements. This procedure, repeated for different cross-linking durations, gives the theoretical time dependence of the strength of each band.

3. Theory

A cross-linking study of any given protein oligomer may tell the experimenter no more than the number of subunits it contains. This, of course, is useful information and is hard to determine unambiguously by any other technique. However, it is important to realise that information about the spatial organization of the subunits will be reflected in the pattern of cross-linked species only when the length of the bifunctional reagent and the relative dispositions of the reactive groups on the surfaces of the protein subunits are such that the formation of cross-links between some subunits is preferred over others. In order to interpret such a pattern in terms of the theory it is assumed that during the cross-linking reaction the following conditions apply: (1) no association or dissociation of protein oligomers occurs; (2) the formation of a cross-link does not affect subsequent reactions; (3) no intermolecular cross-links are formed; (4) all of the subunits are spatially indistinguishable; (5) the second function of the reagent reacts considerably faster than the first so that cross-linking can be described solely by the rate constant for the first binding reaction. These conditions do not differ from those used and enunciated in previous treatments [1,2] but they should be noted so that the basis of the technique is clear and its limitations are appreciated.

Consider a set of six subunits denoted a–f, as in fig. 1a. The 15 possible cross-links between pairs of subunits are also shown. Each subunit can make up to five different cross-links with the other subunits and the kinds of sites, or binding domains, involved are identified for convenience, following the same nomenclature as that adopted by Hajdu et al. [1] as p–t. First, considering the reaction of p-type sites only, let [B] be the concentration of bifunctional reagent and k_p the rate constant governing the reaction of the reagent with p-type sites. Now, let p be the fraction of subunits with the p site occupied (the probability that a p site has reacted). It follows that $u = 1 - p$ is the fraction of subunits with the p site unoccupied and, if N_0 is the number of subunits in the system then the concentration of unreacted p sites is uN_0. The kinetic behavior of the p-type sites is governed by the equation [7]:

$$\delta(uN_0)/\delta t = -k_p[B](uN_0) \quad (1)$$

and by analogy with Hajdu et al. [1]

$$\delta u/\delta\tau = -k_p u$$

where

$$\delta\tau = [B]\delta t; \qquad \tau = \int_0^t [B]dt$$

Therefore

$$u = e^{-k_p\tau} \quad (2a)$$

$$p = 1 - e^{-k_p\tau} \quad (2b)$$

Eq. 2 provides expressions for the fraction of

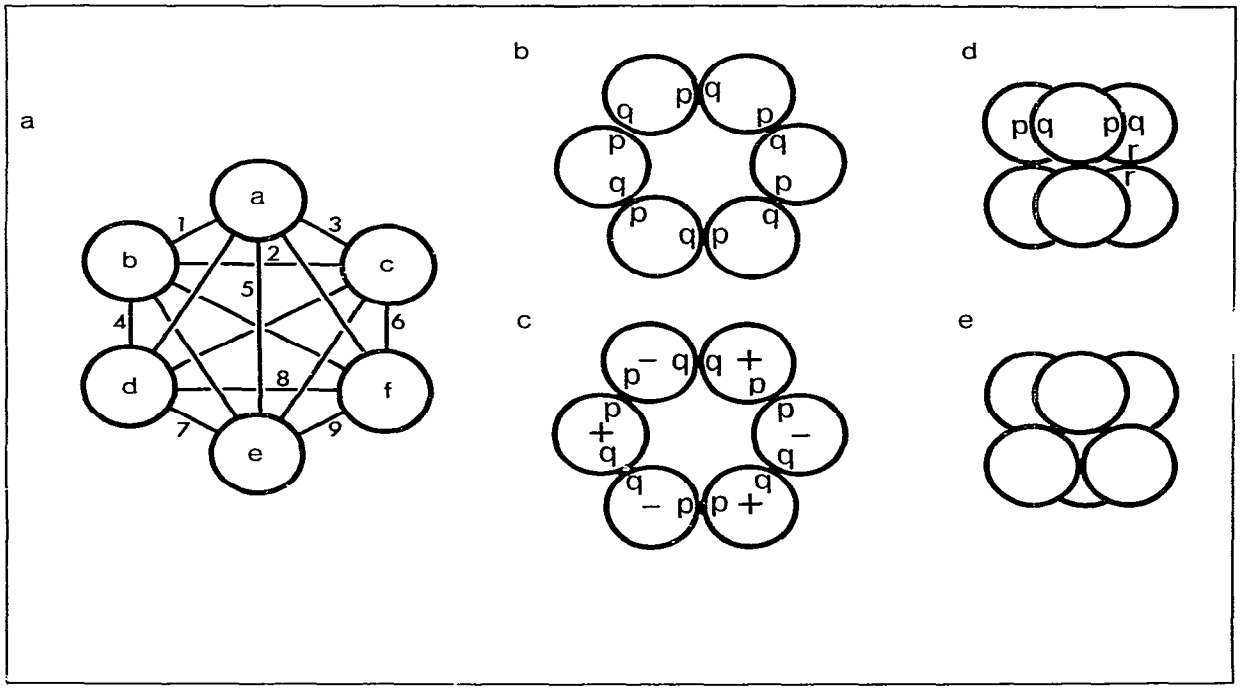

Fig. 1. Cross-links and symmetry in hexamers. (a) The 15 possible cross-links (only nine are numbered) between six subunits. (b) Cyclic point group symmetry, six heterologous (pq) interactions are involved. (c) Cyclic arrangement with dihedral point group symmetry. Two sets (pp and qq) of isologous interactions are involved. (d) A trigonal prism with dihedral point group symmetry. Each ring of three subunits involves heterologous (pq) interactions. There are isologous (rr) interactions between the rings. (e) A trigonal antiprism with dihedral point group symmetry. As in d there are three heterologous (pq) interactions around each ring and isologous (rr) interactions between the rings. There is also the possibility of a second set (ss) of isologous interactions between the rings.

p-type sites that will be occupied, or not, as a function of time, in terms of the rate constant for the reaction of bifunctional reagent with this kind of site and its concentration. It may be noted that τ is directly proportional to t, the time of reaction, only if [B] is constant. Following the same reasoning we may now write

$$\begin{aligned} q &= 1 - e^{-k_q\tau}; \quad & u &= e^{-k_q\tau} \\ r &= 1 - e^{-k_r\tau}; \quad & w &= e^{-k_r\tau} \\ s &= 1 - e^{-k_s\tau}; \quad & x &= e^{-k_s\tau} \\ t &= 1 - e^{-k_t\tau}; \quad & y &= e^{-k_t\tau} \end{aligned} \tag{3}$$

where k_q–k_t are the rate constants governing the reactions of q–t-types sites, respectively, and q–t and u–y are the corresponding probabilities that a site of a given type has or has not reacted, respectively.

Using eq. 3 we may now calculate the fraction of any species containing a specified set of cross-links by simple multiplication of the relevant probabilities of reaction of the groups on the subunits. For example, the fraction of material having no cross-links in the six-subunit array is given by

$$P_0 = uuuvvvwwwxxxyyy \tag{4}$$

Therefore, using eq. 3, we find the time dependence of this fraction to be

$$P_0 = e^{-3(k_p+k_q+k_r+k_s+k_t)\tau} \tag{5}$$

Probabilities of the type given by eq. 5 are not

yet suitable for comparison with experimental data. This requires conversion of the information embodied in eq. 5 into the actual fractions of monomer–*n*-mer that will be generated when the noncovalent interactions in the native *n*-mer, which has been chemically cross-linked for different times, are dissociated in SDS. This may be straightforwardly performed by computer, since essentially it only involves for any specified pattern of cross-links assessing how many of the subunits are connected to each other. This procedure gives theoretically identical results to that of Hajdu et al. [1], but has the advantage of not requiring the solution of sets of differential equations involving intractable numbers of terms, while at the same time avoiding the drawback inherent in the probabilities of reaction by Hucho et al. [2] which did not allow for the possibility of the change with time of the relative propensity of cross-linking of different domains of bonding [1]. The method is applicable to molecules with any number of subunits by generalization of eqs. 3 and 4.

This treatment allows the investigator to compute theoretical curves of the fraction of monomer–*n*-mer as a function of time expected from cross-linking the subunits in the system by specifying values of the rate constants for the formation of the different types of cross-links. In practice, the rate constants are expressed as ratios and a number of them may be set equal to zero. This process is equivalent to specifying a particular arrangement or restricted set of arrangement of subunits and has the result, when a satisfactory correspondence between computed and theoretical curves is obtained, of selecting the simplest model capable of adequately describing the experimental results. More details of the process are given in the next section.

4. Results and Discussion

The time course of formation of dimers, trimers, tetramers, pentamers and hexamers by intrasubunit cross-linking, as measured experimentally, is shown in fig. 2a for *Jasus* hemocyanin and in fig. 2c for urease. There are clear differences, the most noteworthy being the extremely high value reached by the dimer fraction in the cross-linking of the hemocyanin. The value, above 50%, may be contrasted with that for the urease cross-linking reaction where the proportion of dimer is never more than 25% of the total cross-linked protein. Another significant difference between the two proteins occurs in the rate of formation of cross-links leading to tetramers. The fraction of tetramer quickly rises to a high value in *Jasus* hemocyanin and remains above that for trimer formation for the duration of these experiments. In the case of urease the trimer is always a larger fraction of the total than the tetramer.

In order to interpret such results in terms of a native oligomer with its subunits arranged in symmetrical fashion, it is necessary to specify a set of models for the quaternary structure that provides a sufficiently simple and plausible framework for such interpretation. Many workers have discussed the symmetry principles governing the arrangement of a finite, small, number of identical protein subunits arrayed in spatially equivalent positions to form a discrete oligomer (for example, see refs. 8 and 9) and the treatments will not be repeated in detail here. Briefly, for hexamers in which only symmetrical forms are allowed, i.e., where an isologous interaction excludes a heterologous one, and vice versa, there are only a few possibilities, as may be appreciated by reference to fig. 1. The simplest (fig. 1b) is with the six subunits arranged in a planar ring with true cyclic point group symmetry: six identical heterologous (pq) interactions are involved. Alternatively, the subunits may be in a planar ring but with dihedral point group symmetry so that two sets (pp and qq) of three identical isologous interactions are generated (fig. 1c). Finally, the six subunits may form two superimposed layers of three subunits each: the two extremes are the trigonal prism (eclipsed configuration) (fig. 1d) and the trigonal antiprism (staggered configuration) (fig. 1e). In both of the latter, which also belong to the dihedral point group, there are three heterologous (pq) interactions in the plane of each ring and new sets of isologous interactions possible (rr in the former case, rr and ss in the latter) between the rings. In the eclipsed configuration each protomer interacts with only three others, whereas in the staggered arrangement (an oc-

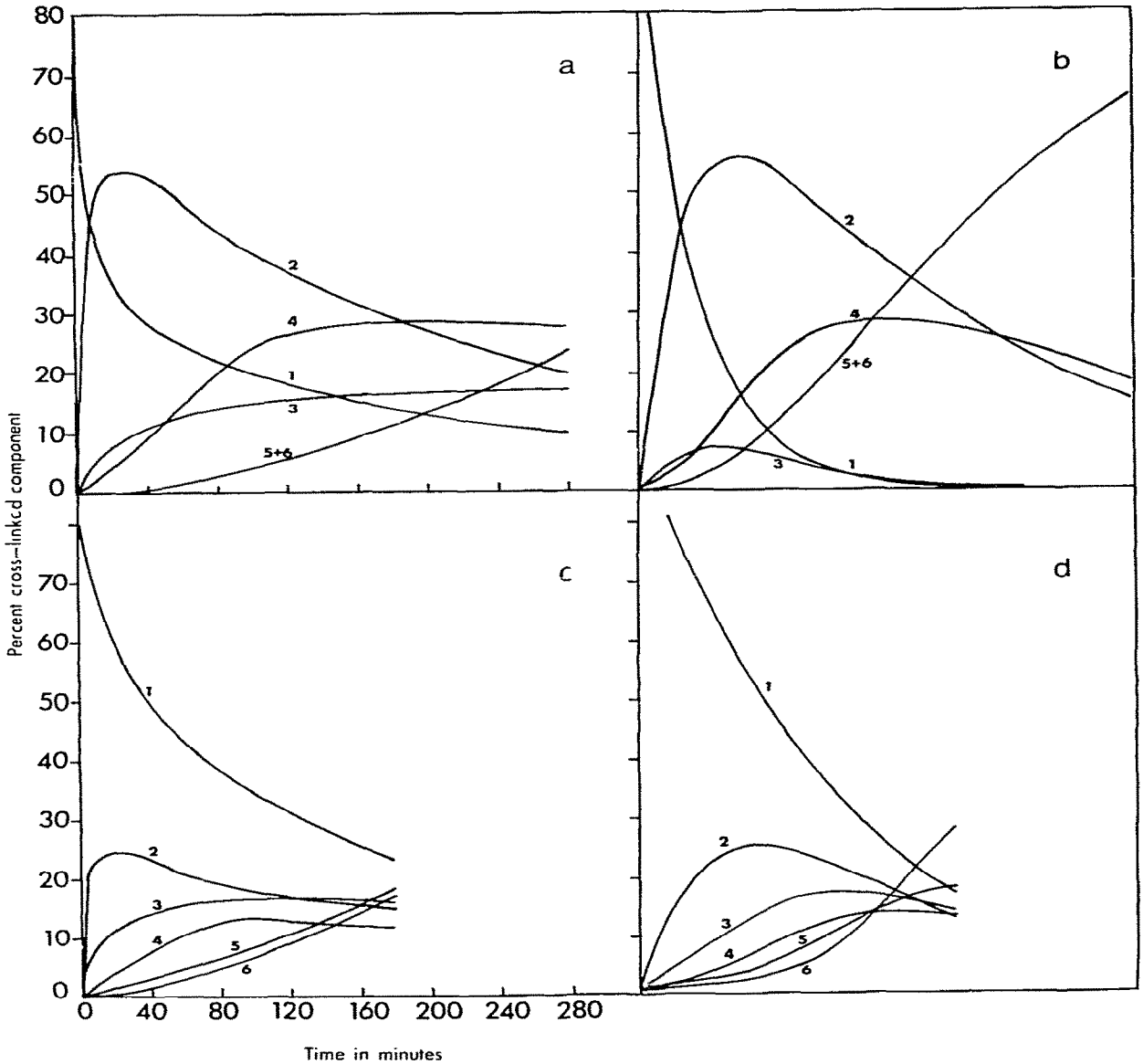

Fig. 2. Formation of glutaraldehyde cross-linked components of two hexameric proteins as a function of time. Monomers–hexamers are identified by the numbers 1–6, respectively, on the curves. Panels a and b show the experimental and theoretical curves, respectively, for *Jasus* hemocyanin. The sum of the bands measured at each time is plotted for the pentamer and hexamer because it was not possible to resolve these two bands satisfactorily on gels. The theoretical curves in b were generated as described in the text with two rate constants in the ratio 17.3:1. Panels c and d show the experimental and theoretical curves, respectively, for α-urease from jack bean. The theoretical curves were generated with two equal rate constants. No time scale is given on the theoretical curves because of the unknown relationship between t and τ (see text). Theoretical percentages of the different cross-linked components were calculated at equally spaced arbitrary times and the axis scaled up to match approximately the time range used experimentally.

tahedron when the subunits are spherical, or nearly so) each protomer interacts with four others. Not all of these interactions are necessary for the integrity of the structure. The subunit arrangement and relative bond strengths in any given oligomer dictate its dissociation pattern and are reflected, to some extent, in the ability of the reactive groups to form cross-links between protomers. The rationale for the cross-linking procedure is based on this. Thus, one should be able to comment on the arrangement of subunits in the native oligomer in terms of the types of model defined in fig. 1 by comparison of the distribution of different cross-linked species in the experimental dissociation patterns with those expected theoretically from consideration of different rates of reactivity between pairs of subunits.

The curves that best fit the experimental results with hemocyanin and urease are shown in fig. 2b and d, respectively. For *Jasus* hemocyanin, equally good descriptions of the progress of the cross-linking reactions are provided by distribution patterns calculated with only two rate constants. A cyclic structure with rate constants in the ratio 8.65 : 1 or a layered structure with the two constants in the ratio 17.3 : 1, as plotted in fig. 2b, were both judged to give satisfactory fits. In the former case the two constants refer to two different types of interaction alternating around the ring, while in the layered model, the higher rate constant refers to interactions between layers, the lower to the interactions around the rings. For urease, a layered structure with two equal rate constants unequivocally gave the best fit. In attempting fits employing cyclic arrangements of subunits it proved impossible ever to construct a model in which the maximum proportion of dimer falls below 30% of the total. To specify a cyclic structure in the computation all rate constants except those referring to interactions 1, 4, 7, 9, 6, 3 in fig. 1a were set equal to zero. For a layered arrangement rate constants corresponding to interactions around the rings, i.e. 1–3 and 7–9 in fig. 1a are set equal while those corresponding to interactions between the two layers are given some fraction or multiple of this value or set equal to zero. Thus, for example, to define an eclipsed arrangement it is only necessary to assign a value to rate constants for cross-links 4–6 in fig. 1a which correspond to interaction rr in fig. 1d, while for a staggered arrangement, values would also be given to the diagonal cross-links in fig. 1a to take into account the extra interactions possible when a subunit is in contact with four others as in fig. 1e.

It could be noted that a direct fit to the experimental curves of fraction of cross-linked species versus time is not practicable because the time-related independent variable in the theoretical treatment is not simply time but the quantity, τ, defined in section 3, which involves the variation with time of the concentration of bifunctional reagent. Thus, the abscissa in the theoretical plots is only linearly related to the horizontal time axis employed in experiments if [B] is constant. As a consequence of not knowing the explicit relationship between t and τ, a satisfactory fit, as judged by comparison of curves like those shown in fig. 2, has been assessed by inspection, with particular attention being paid to salient features like those referred to above. Since the rate constants are expressed as ratios in the procedure suggested here, the arbitrary nature of the time axis is not a problem in correlating the parameters eventually accepted from the fitting process with model structures.

Returning to the discussion of the interpretation of these results in terms of models for the two proteins studied as examples we may draw the following conclusions. First, for *Jasus* hemocyanin it is seen that the subunits must be oriented with respect to each other such that the groups on them reactive to the bifunctional reagent fall into two very distinct classes. We infer two types of isologous interactions or a mixture of isologous and heterologous bonding in the original structure. Combination of this information with the appearance of negatively stained electron micrographs of *Jasus* hemocyanin which, in common with the 17 S aggregate of other arthropod hemocyanins [10] characteristically show hexagonal and square or rectangular projections, would strongly indicate the staggered configuration shown in fig. 1(e) as being most likely for *Jasus* hexamers. This organization is in agreement with the latest picture of the hexameric structure arrived at by combining electron microscopy, computer modelling and

low-resolution X-ray analysis of data from *Limulus polyphemus* and *Panulirus interruptus* hemocyanins [11]. In the present context it seems important to emphasize that the contribution of the cross-linking experiments to the result is essential. It could not have been arrived at from consideration of electron micrographs alone, for example, as the history of research into these structures clearly demonstrates.

For urease, the architecture is evidently such that the fitting process itself eliminates any kind of planar ring as a possibility, but this still leaves open the question of the relative juxtaposition of the subunits, now expected to be arranged as two rings of three subunits each. It seems intuitively more likely that an orientation closer to the eclipsed than the staggered configuration is the correct one, since such a very simple set of rate constants (two equal) suffices to describe the cross-linking reactions. In agreement with others [12] we observe two kinds of projections in the electron microscope, roughly equilateral triangles and squares or rectangles apparently consisting of two layers. These observations, taken together, define a trigonal prism of six subunits as the favored structure for α-urease. Again, this agrees with current ideas about the quaternary structure of this protein proposed by workers using a variety of other techniques [12,13]. We should not lose sight of the fact that one of the most important pieces of information that cross-linking experiments can give is the number of subunits in an unknown oligomer. Counting the number of bands on an SDS-polyacrylamide gel after intersubunit cross-linking and dissociation of noncovalent bonds is the most direct and unambiguous way of establishing this datum, basic to the analysis of oligomeric structure. It is by no means easy to decide by other methods, especially when n is greater than about 3 or 4. Although the value of n was already fairly secure in the case of urease and more so for hemocyanin, the present study serves to strengthen the case that in these two proteins we are dealing with hexamers. This knowledge would have been extremely useful in eliminating false leads earlier in their study.

Another kind of information is potentially available by the use of reagents of different chain length, as discussed by Hajdu et al. [1]. In the experimental part of this study our principle aim was to use the two different hexamers to illustrate the application of the theory we have developed, consequently, we have not explored the distribution near the subunit contact surfaces of amino acid side chains that react with bifunctional reagents. Although some preliminary experiments indicate that differences of this kind may exist between *Cherax destructor* and *Jasus* hemocyanin hexamers we have not yet followed this up. Cross-linking with glutaraldehyde provided sufficient discrimination between structures for our present purposes, as illustrated by the results with urease and *Jasus* hemocyanin. In general, we endorse the recommendations of Hajdu et al. concerning the use of reagents of different chain length and indeed the other recommendations and suggestions in their valuable discussion of the method [1].

The treatment described here, in a sense, combines the principal elements of the two different theoretical treatments for tetrameric assemblages proposed by Hajdu et al. [1] and Hucho et al. [2], respectively. The latter work, although elegant in its utilization of probabilities to distinguish between different site reactivities, had some limitations, as was pointed out subsequently [1]. These arise from its inability to allow for the variation with time of the relationships between the probabilities of cross-linking pairs of subunits. The kinetic approach taken by Hajdu et al. [1] generates exactly the same kind of information as the present one, account having been taken of the variation of the reaction rate with time. The limitation in this case is the need to solve a large number of differential equations. Indeed, this starts to become serious, even at the level of four subunits, and Hajdu and his colleagues found it convenient to streamline their task by omitting consideration of some of the possible ways of forming cross-linked tetramers from their final equations. For oligomers containing more than four subunits the system of equations that would be required is intractable. The advantage of the methodology we have devised is that it avoids the problem by beginning with a set of equations expressing the probability of forming a particular cross-link as a function of time. The resulting set of equations is

limited in number and readily arranged to allow calculation by computer of the fractions of all cross-linked species as a function of time for arrays containing any designated number of subunits. We believe this provides a satisfyingly general approach to the analysis of protein quaternary structure by procedures employing cross-linking with bifunctional reagents and hope that it will prove to be useful to other workers interested in this subject.

References

1 J. Hajdu, F. Bartha and P. Friedrich, Eur. J. Biochem. 68 (1976) 373.
2 F. Hucho, H. Müllner and H. Sund, Eur. J. Biochem. 59 (1975) 79.
3 F. Hucho and M. Janda, Biochem. Biophys. Res. Commun. 57 (1974) 1080.
4 F.H. Carpenter and K.T. Harrington, J. Biol. Chem. 247 (1972) 5580.
5 R.J. Siezen, J.G. Bindels and H.J. Hoenders, Eur. J. Biochem. 107 (1980) 243.
6 L.C. Packman and R.N. Perham, Biochemistry 21 (1982) 5171.
7 P.J. Flory, Principles of polymer chemistry (Cornell University Press, Ithaca, 1953).
8 B.W. Matthews and S.A. Bernhard, Annu. Rev. Biophys. Bioeng. 2 (1973) 257.
9 A.J. Cornish-Bowden and D.E. Koshland, Jr, J. Biol. Chem. 246 (1971) 3092.
10 P.D. Jeffrey, Biochemistry 18 (1979) 2508.
11 J. Lamy, P.-Y. Sizaret, J. Frank, A. Verschoor, R. Feldmann and J. Bonaventura, Biochemistry 21 (1982) 6825.
12 W.N. Fishbein, W.F. Engler, J.L. Griffin, W. Scurzi and G.F. Bahr, Eur. J. Biochem. 73 (1977) 185.
13 N.E. Dixon, J.A. Hinds, A.K. Fihelly, C. Gazzola, D.J. Winzor, R.L. Blakeley and B. Zerner, Can. J. Biochem. 58 (1980) 1323.