Thermodynamic Considerations in the Synthesis and Assembly of Biological Macromolecules

Leonard Peller

Cardiovascular Research Institute, University of California, San Francisco, California 94143. *Received December 3, 1979*

ABSTRACT: Some thermodynamic arguments are applied to polymerizations of biological interest with examples chosen from protein assembly reactions and the de novo synthesis of double strand polynucleotides. The effect of myosin fragments on the critical concentration of G-actin in its conversion to the F-actin bihelical aggregate is examined. For the bifunctional species heavy meromyosin, a maximum depression in this concentration is predicted at the midpoint of its binding isotherm and is strongly dependent on the extent of freedom of the second binding element following the attachment of the first. The influence of temperature and pH on the association of the cyclic trimer of the monomer coat protein of tobacco mosaic virus to yield successively a two turn disk or helical section and ultimately helical aggregates is discussed. A transition temperature can be defined below which no large helical species can arise. Above this temperature the mean size of the helical product depends critically on the ratio of an equilibrium constant for propagation to that for nucleation. A similar situation is encountered in the de novo synthesis of complementary homopolymers of double strand nucleic acids. The energetics of covalent bond formation are relatively unimportant in determining the mean size of the double helical species in equilibrium with monomeric nucleoside triphosphates. Only when the synthesis is joined to the catalyzed hydrolysis of the pyrophosphate byproduct does the role of "high energy phosphate bonds" become manifest. The utility of coupled most probable distributions in describing linear polymerizations in vitro exhibiting cooperative behavior is stressed.

At the core of contemporary biochemistry and molecular biology are the investigations of the mechanisms of synthesis of macromolecules. These polymerizations are broadly speaking of two kinds. One category comprises the synthesis of linear polymers, nucleic acids, and proteins catalyzed by enzymes with covalent bond formation as the dominant concern. The spontaneous association of these species to yield assemblages such as nucleoproteins and homo- and heteroprotein aggregates constitutes the second category. Here secondary forces rather than covalent bonds provide the organizing impetus.

Thermodynamic considerations have received far more attention in the latter instance where the notion of self-assembly specifically precludes the intervention of enzymes. However, to the extent that both classes of polymerization are studied in vitro in *closed* systems thermodynamic factors cannot be ignored for either process.

The fundamental concepts of polymer statistics as they apply to the synthesis of linear polymers were developed beginning four decades ago by Flory.[1-3] The invalidity of the central assumption—the reactivity of a given functional group being independent of chain length—often interferes with the utilization of this approach for enzymes interacting with complex copolymers. Most striking is the chain size variability to reaction of even homo-oligomers stands in the way of the easy extension of these results to kinetic descriptions of enzyme catalyzed polymerization.

A happier situation prevails with respect to the thermodynamic treatment of these processes. Here the division of the polymerizations or assemblies into a few discrete sets of species *within* which single standard free energies or equilibrium constants for association obtain seems quite secure. It is from the connection between these sets that some new and often cooperative features for linear polymerization emerge.

In this article three macromolecular systems are discussed. The de novo synthesis of double helical DNA from two complementary nucleoside triphosphates comes from the first category above. The other two examples involve the self-association of a single-coat protein of tobacco mosaic virus and the heteroassociation of the muscle protein actin with fragments of the second principal muscle protein component myosin. These belong to the second category.

Association of Myosin Fragments with Fibrous Actin

As an initial example of the application of thermodynamic considerations to the formation of biological macromolecules, we examine the polymerization of the globular protein G-actin to yield its bihelical product fibrous or F-actin. The latter comprises the thin filaments of muscle. Historically, this was the first system of protein aggregation to be analyzed in thermodynamic terms by Oosawa and Kasai.[4] The terminology introduced by them fails to emphasize the essential equivalence of their results to those of the most probable distribution of Flory.

We are here concerned with the possible modification of this polymerization by the association of the actin both in its monomeric form and more importantly in its bihelical fibrous form with two hydrolytic fragments of the other principal muscle protein, myosin. The polymerization is accompanied by the hydrolysis of an adenosine triphosphate molecule strongly bound to each G-actin which we here ignore as has been the customary practice in analyzing this process.[4,5]

Part a of Figure 1 depicts the addition of a monomer to one end of the growing polymer. We do not here consider the establishment of a critical nucleus as was done by Oosawa and co-workers[4,6] and will be discussed below in our second example. Consequently, a single equilibrium constant K_a suffices to describe the synthesis with a mass action expression

$$[A_i]/\{[A_{i-1}][G\text{-}A]\} = K_a \quad (1)$$

This growth or propagation association constant relates the equilibrium concentrations of i-mer, $[A_i]$, and its precursor $i-1$-mer, $[A_{i-1}]$, to that of unreacted monomer, [G-A]. An equilibrium probability of chain extension or growth can be defined by

$$p = [G\text{-}A]K_a \quad (2)$$

As in all such processes, the condition for obtaining high polymers is for p to approach unity.[3]

Oosawa and Kasai[4] recognized that starting with pure monomer a critical initial concentration would exist below which high polymers could not be formed. This concentration was given by

$$[G\text{-}A]_{0,c} = 1/K_a \quad (3)$$

0024-9297/80/2213-0609$01.00/0

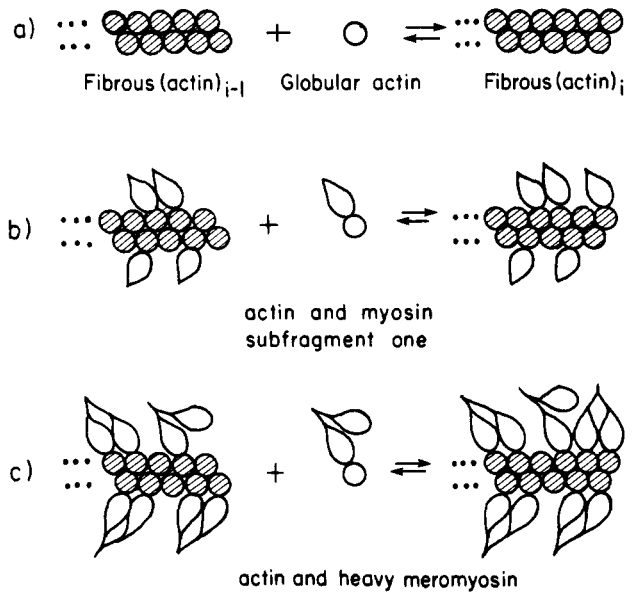

Figure 1. The addition of glubular actin monomers (G-actin; 4.3×10^4 daltons) to yield duplex fibrous actin (F-actin): (a) in the absence of myosin hydrolytic cleavage product; (b) in the presence of myosin subfragment 1 (S-1; 1.1×10^5 daltons); and (c) with heavy meromyosin (HMM; 3.5×10^5 daltons) both singly and doubly attached.

This condition is essentially one which follows from the most probable distribution statistics requirement that $p \rightarrow 1$ as [G-A] at equilibrium must clearly be less than its initial concentration $[G\text{-}A]_0$.

However, such a critical or perhaps more appositely a threshold concentration must arise in any polymerization in which the stoichiometry of reaction is as in part a. This phenomenon thus is *not* confined to polymerizations in which cooperative aspects such as nucleation arise. However, for most syntheses of macromolecules from low molecular weight monomers a critical concentration of say 10^{-5} M corresponding to a quite favorable equilibrium constant for reaction occurs at a level far below that of practical interest. With protein assembly when the building block may have a molecular weight from 100 to 1000 times higher than the conventional monomers of synthetic high polymer chemistry such a molar concentration corresponds to a quite ponderable *weight* concentration. Moreover, the fibrous product of the protein assembly will owing to its highly asymmetric structure dramatize its presence above the critical concentration by contributing a large increment to the solution viscosity. Thus the focus on this property for these systems is not surprising.

The actin monomer when incorporated in the F-actin bihelix will associate with the myosin hydrolytic cleavage products depicted in parts b and c—the monofunctional subfragment one (S-1) and the bifunctional heavy meromyosin (HMM). As indicated in the legend of Figure 1 the myosin fragments are substantially larger than the actin adsorbent. Nonetheless these adsorbates seem capable of fully saturating the latter as revealed by among other methods electron microscopy.[7] The removal of the distal tail portion of myosin makes this process a kind of graft copolymerization as opposed to formation of a network (vide infra).

The actin conjoining to myosin or its fragments leads to an enhancement of the rate of hydrolysis of ATP by these species and to contraction when the intact fibrous actin and the tail aggregated myosin filaments are laterally aligned either in nature in muscle or in artificial preparations.[8,9] Consequently, there has been considerable interest in the simpler association equilibria of the myosin fragments with actin in vitro.

The analyses of these equilibria have usually relied on the simple Langmuir adsorption isotherm for *both* monofunctional and bifunctional species. The author earlier pointed out that a bifunctional or two-headed species in the muscle parlance will bind in a fashion not describable in the same way as a monofunctional element and suggested an isotherm[10] of a type encountered with diatomic molecules bound to a linear lattice. A whole class of such models had been previously introduced in the context of divalent ion association with polymers by Hill.[11] These have been most recently applied to the HMM-actin equilibrium by Hill.[12]

If both elements of myosin and HMM are presumed to be identical as some studies suggest (though there are counter views on this point[13]) and attachment of one to F-actin does not produce a strain on the accommodation of the second at a nearby site (a highly questionable assumption), any difference in association constants for HMM and S-1 can be interpreted in terms of an effective local concentration for the second element once the first is attached.[10] To be more precise let K_s be the association constant for S-1 to an actin site in the fibrous form, then $2(K_s)(K_s[M_L])$ would be the effective equilibrium constant for binding an HMM species by *both* appendages to an *isolated pair* of adjoining actin sites. The factor for the first appendage is $2K_s$ (chosen in two equivalent ways for the bifunctional element) and $K_s[M_L]$, where $[M_L]$, the local concentration defined above, is contributed by the second appendage. This condition is encountered in antibody–antigen association as well.[14]

The previous discussion serves as a preamble to considering the actin polymerization in the presence of the two myosin fragments. We can write the mass action expression for the addition of an actin when subfragment 1 can adsorb to the polymer as

$$[A_i]'/[A_{i-1}]'[G\text{-}A]' = K_a(\Gamma_i'/\Gamma_{i-1}'\Gamma_1') \qquad (4)$$

In this equation the superscript prime refers to all the actin species with variable amounts of the myosin fragment bound. The weighing factors are $\Gamma_i' = (1 + K_s[S\text{-}1])^i$, and [S-1] is the concentration of *free* subfragment one.

Under these conditions the critical initial concentration of G-actin is given by

$$[G\text{-}A]_{0,c}' = (1/K_a)(\Gamma_1')(1 + K_s[S\text{-}1])^{-1} \qquad (5a)$$

If the adsorbate has the same affinity for G-actin as for a monomeric element of the fibrous product $\Gamma_1' = (1 + K_s[S\text{-}1])$ and the critical concentration is not influenced by the presence of the associating subfragment. On the other hand, if rather than the picture in part b of Figure 1 the affinity of [S-1] for the isolated monomer (which has bound ATP not ADP) is not significant, $\Gamma_1' = 1$ and

$$[G\text{-}A]_{0,c}' = (1/K_a)(1 + K_s[S\text{-}1])^{-1} \qquad (5b)$$

and is expected to drop monotonously with the concentration of adsorbate. Present experimental evidence[15] would favor the former over the latter. The effect of simple equilibria of small molecule association on the critical concentration has been considered previously by Oosawa and Higashi.[16]

The situation with respect to HMM binding as depicted in part c can be analyzed in a similar fashion. The isotherm proposed earlier can be developed by arguments familiar from the treatment of one-dimensional lattices employing a matrix **M** of statistical weights given by

$$\mathbf{M} = \begin{pmatrix} 1 & 1 & 1 \\ \alpha & \alpha & \alpha \\ 0 & \beta & 0 \end{pmatrix} \quad (6)$$

Here $\alpha \equiv 2K_s[\text{HMM}]$, with [HMM] the concentration of unbound adsorbate, and $\beta = K_s[M_L]$, if the assumptions regarding the freedom of access of the second element to an actin site are accepted. Otherwise β can simply be viewed as an experimentally measurable parameter.

The effect of HMM binding on the critical concentration of actin can be readily determined by noting that in the limit of high degrees of polymerization Γ_i'', which is a partition function for the adsorption process, approaches proportionality to λ_{max}^{i-1} where λ_{max} is the only positive nonzero eigenvalue of **M**. Hence

$$[A_i]''/[A_{i-1}]''[\text{G-A}]'' = K_a(\Gamma_i''/\Gamma_{i-1}''\Gamma_1'') \rightarrow K_a(\lambda_{max}/\Gamma_1'') \quad (7)$$

or

$$[\text{G-A}]_{0,c}'' = (1/K_a)(\Gamma_1''/\lambda_{max}) \quad (8)$$

with $\lambda_{max} = \{(1 + \alpha) + [(1 + \alpha)^2 + 4\alpha\beta]^{1/2}\}/2$. If a single element of the HMM binds to G-actin with the same affinity as to F-actin, then $\Gamma_1'' = 1 + \alpha$ and

$$[\text{G-A}]_{0,c}'' = (2/K_a)\{1 + [1 + (4\alpha\beta)/(1 + \alpha)^2]^{1/2}\}^{-1} \quad (9a)$$

At both low concentration levels of ligand HMM ($\alpha << 1$) and very high levels ($\alpha >> 1$), the quantity in braces in eq 9a approaches $^1/_2$ with the critical concentration of G-actin unaffected by the presence of HMM. The critical concentration is not influenced at such high levels of bifunctional ligand where the preponderant mode of binding is by one element not two. The critical concentration is a minimum where $\alpha = 1$ and is given approximately by

$$[\text{G-A}]_{0,c}'' \approx (1/K_a)(2/\beta^{1/2}) \quad (9b)$$

($\alpha = 1$ is the midpoint of the isotherm; the midpoint for the approximate isotherm where binding is exclusively by both appendages occurs at substantially lower concentration;[10] namely, $\alpha = 3/4\beta$.) This maximum depression of the critical concentration occurs at a level of ligand where the passage from two-appendage binding to single-appendage attachment is occurring.

Most measurements of the relative affinity constants of S-1 and HMM for F-actin suggest that the latter is a factor of 10–20 times larger than the former.[17,18] However, one recent report utilizing isotherms taking account of the need for neighboring pairs of vacant actin sites for strong binding of HMM claims that a factor of the order of 600 is appropriate.[19]

Without specifying the origin of the enhanced affinity, we can define $\beta = K_{HMM}/2K_s$ (vide supra). Hence $K_{HMM}/K_s \approx 600$ leads to a maximum depression of the critical concentration in accordance with eq 9b to a fraction of 0.12 of its value in the absence of the bifunctional adsorbate. This is possibly detectable in light of the precision in determining this concentration level. By contrast a factor of 10–20 in the two affinities would probably result in a diminution of the critical concentration so slight as to escape verification.

Another protein species, the largely α-helical tropomyosin, occupies seven contiguous sites on association with F-actin[20] and unsurprisingly exhibits no tendency to attach to a single G-actin.[15] One might anticipate that this protein would show a more marked effect in diminishing the critical concentration.

Finally, the phenomenon of "superprecipitation" involves the interaction of F-actin with filaments of myosin self-aggregated through the tail moieties[21] hydrolytically

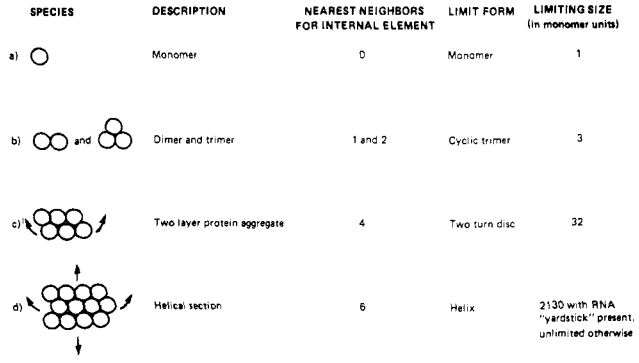

Figure 2. Some of the aggregated states derived (a) from the monomeric coat protein (1.7×10^4 daltons) of tobacco mosaic virus (TMV) classified according to increasing size. (b) The cyclic species (A-protein) is taken to be the predominant trimeric form. (c) The arrows for the two-layer species indicate growth confined to lateral accretion of units. (d) The helical form has lateral addition of elements to the *completed* nucleus leading to axial extension of the coat.

removed in the preparation of the fragments. This process though not commonly regarded as an equilibration seems to exhibit properties rather like gel formation in polyfunctional condensation polymerization.[22] There is the added fillip of ATP hydrolysis seeming to assist the association–dissociation process leading to the development of a three-dimensional network.

Association of the Subunits of Tobacco Mosaic Virus Coat Protein

The association of the protein subunits which provide a coat for the infectious RNA of tobacco mosaic virus (TMV) is perhaps the most thoroughly studied example of a thermodynamically dictated or self-assembly process involving biological macromolecules.[23] This association will take place in the absence of the RNA constituent and is endothermic in nature and pH dependent.[24] A great variety of forms or partial aggregates has been reported of which some are tabulated in Figure 2. Without the RNA "yardstick" to fix the length of the helical coat, a heterogeneity of sizes of polymeric products of this final type is not surprising (part d).

There have been several treatments of this protein aggregation which endeavor to account for the multiplicity of species.[24-26] One of these employs the language of condensation polymerization[25] in the form originally introduced by Flory. We consider here a limited aspect of this process utilizing this approach. Some modest modifications are made for adaptation to a situation where the character of the polymerization as well as the nature of the polymeric product change markedly with environmental variables, e.g., temperature and pH.

In aqueous solution at low temperature the dominant form of the subunit is the trimer believed to be cyclic and designated the A protein.[24] We examine on one hand an equilibrium between it and the two-turn disk (part c) (or its isomer, the first two turns of the helix) and partial aggregates intermediate in size between the two. To this is appended an equilibrium involving the final polymeric form—helical assemblages of diverse size. From column 3 of Figure 2 it is apparent that the assembly viewed in this fashion involves a passage of a typical internal monomer from two to four and finally to six nearest neighbors.[24] It is an inference from the endothermic nature of the polymerization that if all protein contacts are not too dissimilar energetically there will be a steady increase in the ΔH of reaction as one element passes through these various forms.

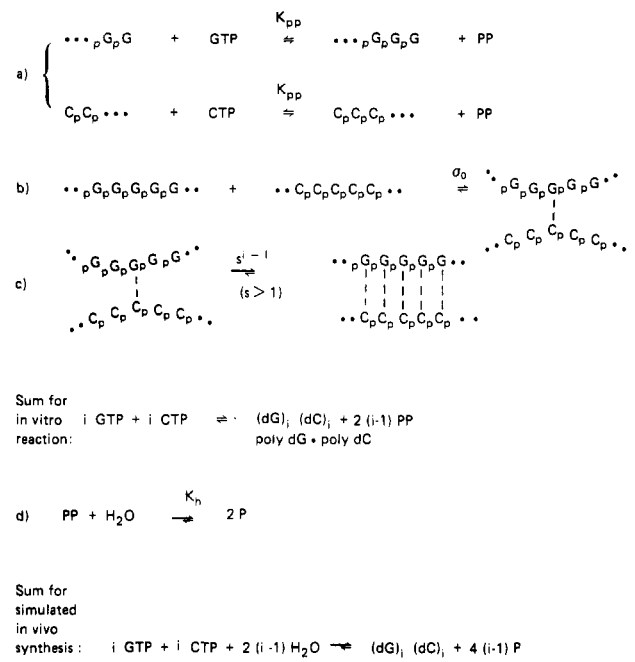

Figure 3. The de novo synthesis of poly(deoxyguanylate)·poly-(deoxycytidylate), polydG·polydC, formulated in a sequence of steps a → c readily amenable to thermodynamic analysis. The inclusion of step d with the hydrolysis of the pyrophosphate byproduct leads to an effectively irreversible polymerization. The relative lengths of the arrows for the forward and reverse reactions are intended to convey the degree of reversibility.

A consideration of these equilibria produces a conservation equation for the trimer of total concentration $[T]_0$ in the forms denoted in parts b, c, and d of Figure 3.

$$\left(\frac{p_n}{K_n}\right)\sum_{i=1}^{N-1} i p_n^{i-1} + (p_n^{N-1})\left(\frac{p_g}{K_g}\right)\sum_{i=N}^{\infty} i p_g^{N-i} = [T]_0 \qquad (10)$$

In eq 10, the probability of combination of cyclic trimers to yield species of all sizes defined in part c and the equilibrium constant for that process are designated p_n and K_n, respectively. The addition of cyclic trimer proceeds to the establishment of some critical nucleus of size N. Further accretion of trimer occurs with a different probability p_g and equilibrium constant K_g. These parameters are related to the free trimer concentration [T] by

$$p_n/K_n = p_g/K_g = [T] \qquad (11)$$

which is the same form for the reaction probability as for F-actin synthesis from its G-actin monomer.

Simple mass action considerations suffice to justify the description of TMV protein association in terms of geometric or most probable distributions. However, the representation of the process as *only* two such distributions, one cut off at an upper size limit and the other truncated at a lower size limit, is a rather drastic approximation. Short shrift is thus given to fragments such as a three-turn disk often reported to be present as well as different isomeric forms of other large (and smaller) aggregates.[24]

We persevere with this admittedly incomplete depiction because it illustrates some striking features of the thermally induced polymerization. Also, it facilitates making at least formally a quantitative connection between the effects of temperature and pH on the assembly process.

From the previous remarks about the increasing magnitude of ΔH as the number of contacts grows, we would anticipate that $\Delta H_g > \Delta H_n$ where these enthalpies are defined by the van't Hoff relations for the corresponding equilibrium constants. A temperature (T_t) might be expected where $K_n = K_g \equiv K$ and $p_n = p_g \equiv p$ with the collapse of eq 10 to

$$(p)/(1-p)^2 = K[T]_0 \qquad (12)$$

This form was employed by Lauffer and co-workers[27,28] to fit the low-temperature and high-temperature polymerization data but with *different* choices for K.

Below this crossover or transition temperature where $K_n > K_g$, it is the first term in eq 10 which is significant–characteristic of the situation where the aggregates are primarily in the form indicated in part c of Figure 2. Here the approximation represented by eq 11 is reasonably valid with the consequence that at low extents of association the number average degree of polymerization is given by

$$\langle i \rangle = (1 - p_n)^{-1} \approx (K_n[T]_0)^{1/2} \qquad (13)$$

(presuming of course that $K_n[T]_0 > 1$ or the critical concentration for nucleation is exceeded).

At temperatures greater than T_t, the second term of eq 10 is dominant and the aggregates are the helical fragments of part d of Figure 2. Now if $p_g \approx 1$, $p_n \approx K_n/K_g < 1$ from eq 11 and

$$\langle i \rangle = (1 - p_g)^{-1} \approx (K_g/K_n)^{(N-1)/2}(K_g[T]_0)^{1/2} \qquad (14)$$

For this state to be attained $K_g \approx 2K_n$ is more than sufficient. This means that a free energy difference less than RT shifts the system from small fragments to large helical products. With a value of $N = 11$ consistent with the entry in the last column for part c of Figure 2, a factor of about 45 multiplies the critical concentration condition in eq 14. This corresponds to a number average molecular weight of the order of 2.3×10^6 with $K_g[T]_0 \approx 1$.

The dependence of helical size on the ratio of a propagation to a nucleation equilibrium constant was found by Oosawa and Kasai[4] in their studies of F-actin formation. The striking effect of temperature is attributable to the large size of the nucleus here.

The analysis of the influence of temperature on the polymerization alluded to above yielded a value for ΔH_n of 30 kcal and for ΔS_n of 126 eu.[27] The molecular weight studies at higher temperature led to $\Delta H_{app} = 206$ kcal and $\Delta S_{app} = 739$ eu.[28] From eq 14 we can make the identification that $K_{app} = (K_g/K_n)^{N-1}(K_g)$ so

$$\Delta H_{app} = N(\Delta H_g - \Delta H_n) + \Delta H_n \qquad (15a)$$

$$\Delta S_{app} = N(\Delta S_g - \Delta S_n) + \Delta S_n \qquad (15b)$$

With $N = 11$, then $\Delta H_g = 46$ kcal and $\Delta S_g = 182$ eu. As four additional contacts are made for every trimer added to the growing helix, the positive enthalpy change per newly established contact over that assignable to the two-tier disk is $(\Delta H_g - \Delta H_n)/2 = 8$ kcal. This is less than the 15 kcal value for each contact in the growing nucleus.

The transition or floor temperature (T_t) below which only fragments are found is given by

$$T_t = (\Delta H_g - \Delta H_n)/(\Delta S_g - \Delta S_n) \qquad (16)$$

which is equal to 286 K or 13 °C. The rapid onset of turbidity with the appearance of large aggregates has been qualitatively judged to occur at about 14 °C.[28]

The transition temperature quoted above applies at a pH of about 6.5. The formation of helical rods is affected by pH so T_t will be dependent on this environmental variable. Recalling that the transition temperature is defined by a zero free energy difference for addition of trimer to the two-layer fragments and the helical or rodlike

species, we can employ a Clapeyron-like equation connecting T_t and pH

$$dT_t/dpH = (2.303)(RT_t)(\Delta n_{H^+})/(\Delta S_g - \Delta S_n) \quad (17)$$

$$dT_t/dpH = (2.303)(RT_t^2)(\Delta n_{H^+})/(\Delta H_g - \Delta H_n) \quad (18)$$

In equations 17 and 18 Δn_{H^+} is the difference in the number of protons bound by an element in the two polymeric forms. It will depend on pH and the pKs of the protonation loci. Expressions of this type have been of utility in describing conformational changes in nucleic acids[29] and proteins.[30]

The increasing tendency of the subunits of TMV to polymerize as the pH is decreased has been attributed initially by Caspar[24] and by others[31,32] to the protonation of two carboxyl groups with abnormal pKs in each monomer. If these protons are taken up by the two additional contacts indicated in the passage from the species in part c to those in part d of Figure 2, then Δn_{H^+} is at a maximum equal to four. With the previously quoted value of $\Delta S_g - \Delta S_n$, $dT_t/dpH \approx 94$ °C. A 0.1 of pH unit drop should shift the equilibrium about 10 °C lower—semiquantitatively in accord with experimental findings.[31] This provides a further parallel with the cooperative transconformational changes of biological macromolecules.

It has been asserted[24,25] that most probable distributions are incompatible with cooperative transformations. For the assembly process described here this has led to the conclusion that treatments based on such distributions are inappropriate[24] and antinomally to the view that the association of TMV subunits is not cooperative at least as regards the steep rise of aggregate size with temperature.[25] However, the geometric progression characterizing these distributions is preserved in those coupled processes which manifest cooperative behavior as delineated above. Even the truncation of two such distributions at a common place is not a crucial requirement for their representation of cooperative behavior as will be seen in the next section. The essential feature is the coexistence of at least two serial *or* parallel equilibria jointly describing the association process.

De Novo Synthesis of Double Helical Polynucleotides

It was stated in the introduction that a thermodynamic approach to the spontaneous association of protein elements has commended itself for some years. By contrast there has been scant examination of enzyme catalyzed polymerizations from this standpoint.

We here review and extend some thermodynamic arguments for the synthesis of two-strand polynucleotides.[33–35] The de novo synthesis of these species with pairs of complementary nucleoside triphosphates leads to a sequence of either (1) regularly alternating pyrimidine–purine bases in each strand[36] or (2) to two strands one of which is composed exclusively of the pyrimidine nucleotide and the other of the complementary purine monomeric element.[37] There are several known representatives of both types which are the products of DNA polymerase action as well as more recently found examples of RNA polymerase catalysis.[38,39]

The mechanism advanced to account for their occurrence rests on the assumption that the enzyme possesses an undetectably small amount of an oligonucleotide impurity which serves as a template-primer progenitor of the long chains.[40] This scheme has been rather fully elaborated for the first example of this type of polymerization—the monotonously alternating poly(dAdT) where an oligomer of alternating adenylate (dA)–thymidylate (dT) monomers is the required precursor.[41] This explanation is more difficult to sustain[34,35] for the homopolymeric paired chains such as poly(guanylic acid)·poly(cytidylic acid), polydG·polydC, where there perforce would need to be present in all enzyme preparations an impurity comprising *two* homooligomers. It is the thermodynamics governing the synthesis of this type of product which will be discussed here.

Commencing with deoxyguanosine triphosphate (here designated dGTP) and deoxycytidine triphosphate (dCTP), the duplex DNA product can be envisioned for *thermodynamic purposes* as arising in three steps. The first in part a of Figure 3 involves the synthesis of the two homopolymeric chains with the elimination of a pyrophosphate moiety (PP) with each phosphodiester bridge formed in the polymer. Part b describes the first contact between the two strands which is a bimolecular association step. Subsequently in part c the two tenuously hydrogen bonded chains "zip up" to yield the final double helix. The sum of these three steps comprises the de novo synthesis of a bihelix of specified size from the equivalent number of monomers.

Each of these steps makes a different contribution to the overall equilibrium. Three separate association parameters must be introduced which are fortunately determinable independently of studies of bihelix synthesis.[33] For step a studies of single strand polyribonucleotide synthetic equilibria from nucleoside *diphosphates* as monomers yielding orthophosphate as a byproduct coupled with information on the hydrolysis free energies of ATP and ADP suffice to define K_{pp}.[33] In fact this quantity varies with the purine and or pyrimidine base of the monomer in ways which can be rationalized by differing extents of base stacking in the polymer.[42] Such variation can here be absorbed in the parameter s which is raised to the power of the number of base pairs formed minus one. This quantity and σ_0 are derivable from studies of the bihelix to coil conversion[43] as a function of temperature. The nucleation association constant σ_0 plays a significant role only in the thermal transition of complementary oligonucleotides.[43,44]

In the language of polymer statistics reactions a are characterized by a growth probability[45] p/K_{pp} = [dGTP]/[PP] = [dCTP]/[PP] ≡ [NTP]/[PP] where the initial levels of the complementary nucleoside triphosphates are taken to be the same. (For reactions of this stoichiometry there is a critical concentration ratio for polymerization if product is initially present but not a critical or threshold monomer level.) After the first contact illustrated in b, the ensuing stitching up in c involving the bihelix stabilization constant s to the appropriate power requires the introduction of a second probability equal to p^2s. The latter can be construed as applying to the growth of the duplex by a nucleotide in *each* strand further augmented by the base-pairing affinity.

Omitting the algebraic details, we can examine the equilibrium state in terms of the average size of single strands and double strands and the mean size of *all* species. This examination is here confined to the presence of long duplex chains where $s > 1$ which is a necessary condition for the helix–coil equilibrium to favor the former.[46] In this instance $p^2s \approx 1$ and the conventional expressions for the number average degrees of polymerization cited above are given by

$$\langle i \rangle_{ss} = (1 - p)^{-1} \approx (1 - s^{-1/2})^{-1} \quad (19)$$

$$\langle i \rangle_{ds} = 2(1 - p^2s)^{-1} \approx 2K_{pp}\{s/\sigma_0[NTP]_0\}^{1/2} \quad (20)$$

$$\langle i \rangle_T \approx (K_{pp}s^{1/2})/2 \quad (21)$$

with $[NTP]_0$ being the initial concentration of each triphosphate.

From these relations and reasonable estimates of the three parameters a number of simple conclusions emerge. First for $s \approx 10$ which represents a reasonable degree of helix stabilization or a temperature significantly below the "melting" temperature of the double helix ($s = 1$), $\langle i \rangle_{ss} \approx 1.5$. That is to say unreacted monomer constitutes the bulk of the single strand species present at equilibrium. A larger value of s would push $\langle i \rangle_{ss}$ closer to unity. This is in accord with the experimental situation in vitro.[36,37]

With $\sigma_0 \approx 10^{-2}$ (a value derived from some oligonucleotide association studies[43]), $K_{pp} \approx 100$ (with a provenance outlined above), and an initial concentration of monomer = 10^{-3} M, we find that $\langle i \rangle_{ds} \approx 2 \times 10^5$ in nucleotides corresponding to a molecular weight of 60×10^6 daltons. This is a very high degree of polymerization in the world of macromolecules *outside* of the realm of naturally occurring DNAs. The second point about this process concerns the nature of the driving force for the reaction. The first factor, K_{pp}, in eq 20 refers to the formation of covalent bonds or more precisely the exchange of phosphoanhydride bridges in the monomers for phosphodiester linkages in the product. A variety of lines of evidence inter alia the demonstrable reversibility of this reaction[47,48] suggest that the establishment of such bonds is not an overwhelmingly favorable process. This is reflected in the magnitude of the equilibrium constant chosen to represent step a.

Now the second factor in eq 20 can be viewed as the ratio of a propagation to an initiation equilibrium constant for the development of secondary structure in the product duplex. It is the large magnitude of this ratio or more specifically the low probability of nucleation of the double helix which is responsible for the high value of the calculated degree of polymerization. Here the resemblance to the cooperative protein association reactions[4] is patent. This is the second striking feature of this polymerization.

Lastly, the estimate of the degree of polymerization of all species in eq 21 employing the same values for the parameters is of the order of 150. The small value of this quantity is a consequence of the measurable amount of unreacted monomer present at equilibrium. The combination parameter $K_{pp}s^{1/2}$ can be regarded as an equilibrium constant for the incorporation of *one* nucleotide at the growing end of the duplex.[34] This expression for $\langle i \rangle_T$ displays a dependence on the tendency for covalent bond formation and not on the nucleation constant which regulates the relative *number* populations of single and double strand species.

A biochemical view of this synthesis has the origin of long chain polynucleotide ascribed faute de mieux to the participation of "high energy" monomeric precursors.[49] This bit of conventional wisdom is plainly without foundation here. The free energy of covalent bond formation is *not* very negative and the presence of macromolecules arises from causes like those encountered in cooperative protein association reactions. More generally the appearance of high molecular weight polynucleotides in vitro either of single or double strands whether template directed or not is always attributable to a limitation on the initiation step. These limitations may be provided by the presence of an oligonucleotide primer from which all chains issue, a kinetically diminished rate of initiation,[45] or a thermodynamic inhibition to nucleation as discussed here.

Molecular weights in the range of tens to hundreds of millions of daltons while of the order of viral DNAs are still well below those estimated for the nucleic acid component of chromosomes. More recently developed techniques for examination of these species where the hazards of breakage are mitigated reveal that tens of *billions* is the appropriate scale.[50]

In vitro studies of nucleic acid synthesis either of de novo duplex formation referred to here or of replication of naturally occurring DNAs or RNA transcription of DNA invariably are conducted without utilizing the strongest driving force available for the process. The byproduct pyrophosphate is hydrolyzed to orthophosphate by the enzyme inorganic pyrophosphatase in all cells at virtually all times.[51-53] This reaction is the most irreversible step in nucleic acid synthesis as it is in the metabolic pathways leading to the formation of many small molecules. An appropriate value of the equilibrium constant K_h for step d in Figure 3 under reasonable cellular conditions[54] is about 10^5. When the polymerization is coupled to the hydrolytic action of inorganic pyrophosphatase, the equilibrium concentration of the pyrophosphate species falls by a factor of approximately $4[NTP]_0/K_h$—a quite small number. This is the only place where the biochemical dogma of the crucial role of the "high energy phosphate bond" can be credibly invoked for nucleic acid synthesis.

With this reaction coupled to polynucleotide synthesis the concentration level of pyrophosphate is depressed to immeasurable levels and the polymerization is driven to unimaginable (and in fact unattainable) extents. The previous equilibrium expressions are altered by the factor of $K_h/4[NTP]_0$ mentioned above to yield

$$\langle i \rangle_{ds} \approx 2K_{pp}(s/\sigma_0[NTP]_0)^{1/2}(K_h/4[NTP]_0) \quad (22)$$

$$\langle i \rangle_T \approx (K_{pp}s^{1/2})(K_h/4[NTP]_0) \quad (23)$$

As this factor is of the order of 10^8–10^9, it is apparent that the new degree of polymerization for double strands is far above that corresponding to the DNA in the chromosomes of higher cells. While the value of $\langle i \rangle_{ss}$ cannot be much further lowered, the greatly elevated value of $\langle i \rangle_T$ reflects the virtually complete disappearance of nucleoside triphosphate into double strand product.

Studies of these reactions have focused on the enzymes responsible for the catalysis of phosphodiester bond formation and ancillary proteins constituting putative apparatuses for the process.[55-57] Consequently the question of the relevance of thermodynamic calculations of limiting extents of de novo polynucleotide synthesis to the problem of nucleic acid replication and transcription in vitro immediately arises.

First, inasmuch as the replication (as well as transcription) takes direction from a template but leaves the template unaltered beyond acquiring two new complementary chains bound to the parental strands, this must be viewed as equivalent *thermodynamically* to the generation of paired new chains. Second, while synthesis in the cell takes place with a substantial amount of unreacted monomer present this does not represent a partial extent of reaction in a closed system (as with in vitro studies) but rather a replenishment of the triphosphates by oxidative metabolism in an effectively open system.

For these two reasons we have long felt that the estimated marginal degrees of polymerization in the absence of pyrophosphate hydrolysis should be disturbing.[33] Finally it should be noted that the polydispersity of chain length for in vitro replication of all but the smallest viral DNAs may also reflect the onset of a reverse or pyrophosphorolysis reaction. In one case, the polydispersity of the template directed replication product[58] could indeed be fitted to a most probable distribution.[34] This is possibly a sign of the effect of pyrophosphorolysis in broadening

the size spectrum as well as limiting the attainable extent of polymerization. Both of these undesirable consequences can be annulled by coupling the synthesis to pyrophosphate removal.

Conclusion

The equilibrium states of the polymerizing system discussed here are characterized by one or two coexistent most probable distributions and thus encompass a wide range of macromolecular sizes. This situation sometimes ineluctable for in vitro studies can provide insights into the nature and limitation of the driving forces leading to polymerization.

However, the products of such reactions in vivo are generally narrowly defined as to size. For the three examples taken up here the circumstances conducive to this behavior are in some respects rather different.

For the formation of F-actin a hypothetical initiator species from which all polymers could be engendered is one requirement. The irreversibility of monomer accretion as gauged by the critical concentration is probably sufficient to ensure the attainment of a Poisson distribution. Flory pointed out many years ago that this size distribution is so narrow as to be indistinguishable at high degrees of polymerization from the generation of a single discrete species.[3]

With regard to TMV protein association, the presence of the viral RNA "yardstick" must serve to stabilize the aggregation at a given size. Furthermore, the RNA–protein interaction will provide an additional thermodynamic thrust to the process.

For polynucleotide synthesis there is a directing template serving to delimit chain length. Also there is the presence of a byproduct removal agent, inorganic pyrophosphatase, whose action serves two purposes. It ensures that the thermodynamic size ceiling is far beyond that dictated by the template. Also, by effectively eliminating the reverse reaction it safeguards the system from the size dispersion that depolymerization would cause.

Acknowledgment. The author is indebted to Professor Paul J. Flory for providing an introduction and a continuing education in the subject of polymer physical chemistry through his many pioneering publications. The author also is grateful to him for a willingness to hear and critically discuss many of the arguments presented above.

Support for this work came from the Cardiovascular Research Institute and the National Science Foundation, Grant No. MCS77-11714 (principal investigator Professor Hugo Martinez).

References and Notes

(1) P. J. Flory, *J. Am. Chem. Soc.*, **58**, 1877 (1936).
(2) P. J. Flory, *Chem. Rev.*, **39**, 137 (1946).
(3) P. J. Flory, "Principles of Polymer Chemistry", Cornell University Press, Ithaca, N.Y., 1953, Chapter 8.
(4) F. Oosawa and M. Kasai, *J. Mol. Biol.*, **4**, 10 (1962).
(5) R. Cooke, *Biochemistry*, **14**, 3250 (1975).
(6) F. Oosawa and S. Asakura, "Thermodynamics of the Polymerization of Protein", Academic Press, London, 1975, Chapter 3.
(7) P. B. Moore, H. E. Huxley, and D. J. DeRosier, *J. Mol. Biol.*, **50**, 279 (1970).
(8) Y. Tonomura and F. Oosawa, *Annu. Rev. Biophys. Bioeng.*, **1**, 159 (1972).
(9) Y. Tonomura, "Muscle Proteins, Muscle Contraction and Cation Transport", University Park Press, Baltimore, 1973, Chapter 6, p 180.
(10) L. Peller, *J. Supramol. Struct.*, **3**, 169 (1975).
(11) T. L. Hill, *J. Polym. Sci.*, **231**, 549 (1957).
(12) T. L. Hill, *Nature (London)*, **274**, 825 (1978).
(13) Reference 9, p 188.
(14) D. M. Crothers and H. Metzger, *Immunochemistry*, **9**, 341 (1972).
(15) P. D. Chantler and W. B. Gratzer, *Biochemistry*, **15**, 2219 (1976).
(16) F. Oosawa and S. Higashi, *Prog. Theor. Biol.*, **1**, 79 (1967).
(17) E. W. Taylor in "Critical Reviews in Biochemistry", Vol. 7, G. D. Fasman, Ed., CRC Press, Cleveland, 1979, p 101.
(18) S. Highsmith, *Biochemistry*, **17**, 22 (1978).
(19) L. E. Greene and E. Eisenberg, *Biophys. J.*, **25**, 21a (1979).
(20) S. E. Hitchcock, H. E. Huxley, and A. G. Szent-Györgyi, *J. Mol. Biol.*, **80**, 825 (1973).
(21) Reference 9, Chapter 7.
(22) Reference 3, Chapter 9.
(23) J. D. Watson, "The Molecular Biology of the Gene", 3rd ed., W. A. Benjamin, Menlo Park, Calif., 1976, Chapter 15.
(24) D. L. D. Caspar in "Advances in Protein Chemistry", Vol. 18, C. B. Anfinsen, Jr., M. L. Anson, and J. T. Edsall, Eds., Academic Press, New York, 1963, p 37.
(25) M. A. Lauffer in "Subunits in Biological Systems", Vol. 5, S. N. Timasheff and G. D. Fasman, Eds., Marcel Dekker, New York, 1971, Part A, p 149.
(26) A. C. H. Durham and A. Klug, *J. Mol. Biol.*, **67**, 315 (1972).
(27) K. Banerjee and M. A. Lauffer, *Biochemistry*, **5**, 1957 (1966).
(28) C. E. Smith and M. A. Lauffer, *Biochemistry*, **6**, 2457 (1967).
(29) V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., "Physical Biochemistry of Nucleic Acids", Harper and Row, New York, 1974, Chapter 6, p 335.
(30) L. Peller, unpublished calculations.
(31) M. A. Lauffer and C. L. Stevens in "Advances in Virus Research", Vol. 13, K. M. Smith and M. A. Lauffer, Eds., Academic Press, New York, 1968, p 1.
(32) A. C. H. Durham and J. T. Finch, *J. Mol. Biol.*, **67**, 307 (1972).
(33) L. Peller, *Proc. Natl. Acad. Sci. U.S.A.*, **55**, 1025 (1966).
(34) L. Peller, *Biochemistry*, **16**, 387 (1977).
(35) L. Peller in "Horizons in Biochemistry and Biophysics", Vol. 5, E. Quagliariello, F. Palmieri, and T. P. Singer, Eds., Addison-Wesley, Reading, Mass., 1978, p 183.
(36) H. K. Schachman, J. Adler, C. M. Radding, I. R. Lehman, and A. Kornberg, *J. Biol. Chem.*, **235**, 3242 (1960).
(37) C. M. Radding, J. Josse, and A. Kornberg, *J. Biol. Chem.*, **237**, 2869 (1962).
(38) D. A. Smith, R. L. Ratliff, D. L. Williams, and A. M. Martinez, *J. Biol. Chem.*, **242**, 590 (1967).
(39) J. S. Krakow, K. Daley, and E. Fronk, *Biochem. Biophys. Res. Commun.*, **32**, 98 (1968).
(40) A. Kornberg in "Evolving Genes and Proteins", V. Bryson and H. I. Vogel, Eds., Academic Press, New York, 1965, p 403.
(41) A. Kornberg, "DNA Synthesis", W. A. Freeman, San Francisco, 1974, p 104.
(42) L. Peller, *J. Phys. Chem.*, **80**, 2462 (1976).
(43) J. Applequist and V. Damle, *J. Am. Chem. Soc.*, **87**, 1450 (1965).
(44) P. N. Borer, B. Dengler, I. Tinoco, Jr., and O. C. Uhlenbeck, *J. Mol. Biol.*, **86**, 843 (1974).
(45) L. Peller and L. Barnett, *J. Phys. Chem.*, **66**, 680 (1962).
(46) B. H. Zimm, *J. Chem. Phys.*, **33**, 1349 (1960).
(47) D. Beyersmann and G. Schramm, *Biochim. Biophys. Acta*, **159**, 64 (1968).
(48) M. P. Deutscher and A. Kornberg, *J. Biol. Chem.*, **244**, 3019 (1969).
(49) A. Kornberg, "Enzymatic Synthesis of DNA", Wiley, New York, 1961, Chapter 1.
(50) R. Kavenoff and B. H. Zimm, *Chromosoma*, **41**, 1 (1973).
(51) A. Kornberg in "Horizons in Biochemistry", M. Kasha and B. Pullman, Eds., Academic Press, New York, 1962, p 251.
(52) A. L. Lehninger, "Bioenergetics", W. A. Benjamin, New York, 1965, p 182.
(53) Reference 41, p 41.
(54) R. A. Alberty, *J. Biol. Chem.*, **244**, 3290 (1969).
(55) R. Schekman, A. Weiner, and A. Kornberg, *Science*, **186**, 987 (1974).
(56) B. Alberts and R. Sternglanz, *Nature (London)*, **269**, 655 (1977).
(57) S. H. Wickner, *Annu. Rev. Biochem.*, 1163 (1978).
(58) R. B. Inman, C. L. Schildkraut, and A. Kornberg, *J. Mol. Biol.*, **11**, 285 (1965).