Kinetic analysis for optimization of DNA ligation reactions

Dennis Revie⁺, Douglas W. Smith^{*} and Thomas W. Yee[§]

Department of Biology B-022, University of California at San Diego, La Jolla, CA 92093, USA

Received October 16, 1987; Revised and Accepted October 6, 1988

ABSTRACT

Kinetic equations describing ligation of DNA to circular recombinant forms were developed and solved for four types of reactions: (a) a homogeneous population of singly restricted DNA fragments, (b) insertion of singly restricted insert into vector, (c) forced directional insertion of doubly restricted insert into vector, and (d) insertion of singly restricted insert into phosphatased vector. The effects of varying vector and insert sizes, starting concentrations, and phosphatase treatment on the yield of circular 1:1 recombinants were analyzed. Selected theoretical predictions were experimentally tested and verified. Our suggestions on optimizing ligation reactions in several cases are at variance with common practice. For example, optimum conditions in case (b) and (d) ligations are best specified as individual insert and vector concentrations rather than as insert/vector molar ratios, except in case (d) ligations involving very small insert size. In case (c) ligations, highest efficiencies are obtained when both vector and insert are at relatively low concentration.

INTRODUCTION

Ligation of DNA fragments to form circular recombinants is fundamental to most cloning procedures. DNA concentrations are chosen to optimize recombinant circle formation; however, reported values for these DNA concentrations vary considerably. Initial theoretical analysis of DNA ligation reactions (1), based on the polycondensation theory of Jacobson and Stockmayer (2), focused on single fragment ligation. More recently, Legerski and Robberson (3) extended this analysis to multiple fragment ligations. These papers assume that the ligation reaction yields an equilibrium mixture of linear and circular concatamers. However, ligation of DNA is an irreversible kinetic process, and a kinetic analysis of ligation should yield more realistic results. Here we derive and solve kinetic equations modeling four common ligation situations; the results show several previously undescribed features of the ligation process. Several of the less intuitive predictions of this kinetic analysis were experimentally tested and verified. Suggestions for optimal initial DNA concentrations are in some cases at variance with common practice. The following ligation cases were modeled:

Case A: Ligation of Identical Singly-Restricted DNA Fragments. Although this simplest case does not correspond to any usual ligation experiment, it nevertheless exhibits surprising properties analogous to those observed in more complex ligations.

Case B: Ligation of Singly-Restricted Insert into Singly-Restricted Vector. Reported optimal concentrations of insert are commonly given as insert-vector molar ratios. We find, however, little relationship between vector concentration and optimized insert concentration.

Case C: Forced Directional Ligation of Doubly-Restricted Insert into Doubly-Restricted Vector. Although confirming the practical observation that this is a nearly foolproof procedure for obtaining high recombinant yields, our analysis suggests better results may be obtained at lower concentrations of insert than are commonly used.

Case D: Ligation of Singly-Restricted Insert into Phosphatased Singly-Restricted Vector. Treatment of vector with alkaline phosphatase is widely used to limit self-ligation (4). The potential efficiency of ligation with phosphatased vector is considerably higher than ligation with unphosphatased vector. This potential is frequently unrealized: unsatisfactory results have been attributed to difficulties in inactivating alkaline phosphatase as well as to nuclease contamination (5). We find in certain situations that use of DNA concentrations optimized for unphosphatased vectors can also yield unsatisfactory results.

MATERIALS AND METHODS

Computer Simulation

Numerical solution of the kinetic equations for the different ligation cases is based on an Adams-Moulton predictor-corrector with variable stepsize (6). Symbols and mathematical terms used are in Table 1, and the kinetic equations are included in a brief Appendix. Programs were written in Turbo Pascal 3.0 (8087 version) and run on an IBM PC-XT with an 8087 math coprocessor chip. For each plot, 42 data points were calculated. Intermediate points were generated by 6 by 7 point Lagrangian interpolation (6). Case A, B, and C simulations were run to 99% completion (less than 1% remaining of ligatable ends); Case D simulations were run to 99.9% completion.

Methods

Procedures for growth of cells, DNA isolation, agarose gel electrophoresis, and restriction endonuclease digestions were as described (5,7), with some minor modifications. Electrophoresis was in 1% agarose-Tris-acetate gels containing 0.5 $\mu\text{g/ml}$ ethidium bromide (EtBr) and tracking dyes. Ligations were in 10 μl at 12°C in 50 mM Tris-HCl pH 7.4, 10 mM MgCl_2 , 1 mM ATP, 10 mM DTT, and 1.0 units/ μl T4 ligase overnight. Reactions were stopped by addition of EDTA to 25 mM. Transformation was into *E. coli* LE392 (5) or JM101 (8) cells prepared by CaCl_2 treatment.

RESULTS

In ligation of DNA, bimolecular concatemerization competes with unimolecular cyclization (1). The bimolecular reaction rate is dependent on the parameter i (see Table 1 for nomenclature summary), the bulk concentration of reactive ends, while the cyclization rate is governed by

TABLE 1: NOMENCLATURE

S	Size of DNA restriction fragments, in kilobase pairs
b	Statistical segment length
l	DNA contour length
i	Concentration of DNA ends
j	Effective concentration of one end of linear DNA fragment in vicinity of other (Jacobson-Stockmayer factor)
k_f	Bimolecular forward reaction rate constant for the ligation of a pair of normally phosphorylated DNA ends.
L_n	Concentration of linear DNA concatamers formed of n identical subunits.
C_n	Concentration of n-meric DNA circles.
j_n	Jacobson-Stockmayer parameter associated with L_n .
$L_{m,n}$	Concentration of (m+n)-meric linear DNA concatamers composed of m vector subunits and n insert subunits.
$C_{m,n}$	Concentration of (m+n)-meric circular DNA molecules.
$W_{m,n}$	Molar fraction of $C_{m,n}$ recombinants lacking head-to-head inverted repeats, considered to be the fraction of recombinant molecules viable in transformation.
$j_{m,n}$	Jacobson-Stockmayer parameter associated with $L_{m,n}$.
$L_{m,n,2}$	Concentration of (m+n)-meric linear DNA concatamers composed of m phosphatased vector subunits and n normally phosphorylated insert subunits, arranged in such fashion that both terminals are insert subunits.
$L_{m,n,1}$	Concentration of (m+n)-meric linear DNA concatamers arranged such that one terminal subunit is a phosphatased vector, while the other is a normally phosphorylated insert.
$L_{m,n,0}$	Concentration of (m+n)-meric linear DNA arranged in such a fashion that both end subunits are phosphatased vector subunits.
R	Ligation rate ratio of a hemi-phosphorylated pair of compatible DNA ends versus a bis-phosphorylated pair.
g	Concentration of phosphorylated DNA ends.
h	Concentration of phosphatased DNA ends.

the Jacobson-Stockmayer factor j , the effective concentration of one end in the vicinity of the other (9). The random coil model for DNA predicts:

$$j = \left(\frac{3}{2\pi bl} \right)^{3/2} \quad (1)$$

In this model, the DNA molecule is considered as a series of segments of length b connected by universal joints. For moderately large molecules, this model provides a reasonable estimate of j . For DNA of very short contour lengths, Gaussian statistics loses its applicability

since the DNA becomes rodlike in its behavior. The random coil model will then overestimate the value of j and the tendency of DNA to cyclize. Further, a marked ten base pair periodicity appears in the measured values of j (10,11). In the case of very large molecules, excluded volume effects reduce the true value of j relative to the random coil prediction (12). Recent theories of DNA ring-closure modeling the observed fluctuations in j have been developed (13-15). The value of b , the DNA statistical segment length, is dependent on salt, pH, and temperature. Neglecting excluded volume effects, the majority of measurements of b yield values of 1200 ± 200 A (12).

In the following text, concentrations will be expressed in terms of dimensionless i/j ratios rather than absolute concentrations. The following equation relates DNA concentration to DNA size S (in kb) and i/j :

$$[\text{DNA}] = \frac{21.4(i/j)}{\sqrt{S}} \mu\text{g/ml} = \frac{32.4(i/j)}{S^{3/2}} \text{nM} \quad (2)$$

This equation is derived from equation [1], using $b=1200$ A, 3.46 A between base pairs in DNA, and average molecular weight of 660 for a base pair in DNA.

For a given fragment size, i/j is directly proportional to DNA concentration. The above equation is approximately valid for DNA fragments as small as 0.8 kb. The upper limit of validity is unclear, since the extent of excluded volume effects remains controversial (12).

Case A: Ligation of a single fragment species, all ends identical

Figure 1(a) plots, at varying starting i/j ratios, the percentage of DNA mass ligated into unit length C_1 circles and various multimers. The fraction of DNA mass in C_1 circles decreases rapidly with increasing i/j , such that when initial $i/j = 1$, less than half of the product mass consists of C_1 circles. The fraction of mass in C_2 circles shows a maximum of 17.8% at $i/j = 0.60$.

A large fraction of the mass is ligated into huge concatamers even at modest i/j ratios. At $i/j = 1$, nearly 10% appears in 11-mers and higher. At $i/j = 4$, more than half of the product consists of 51-mers and higher. This result, not generally realized, is relevant to all ligation reactions.

There are a number of reasons why massive concatamerization has never been noticed as a problem in practical ligation situations. First: Figure 1(a) is a mass ratio plot. A number ratio plot for the same data, Figure 1(b), shows a plateau in the number fraction of concatameric DNA products. At $i/j = 4$, even though 75% of the DNA mass ends up in trimeric and higher DNA forms, on a molar basis only 13% of the product DNA is in such large forms. Within the plateau region of Figure 1(b), increasing the starting concentration of linear DNA (proportional to i/j for a fixed size; see Eqn [2]) causes little change in the molar ratios of the product circular forms. Second: large DNA transforms *E. coli* less efficiently than small DNA (16). Third: plasmids with large head-to-head inverted repeats will not stably transform *E. coli* (17). These effects drastically reduce the number of large forms observed after transfor-

Figure 1: Case A ligation. (a) Weight percentages of product circular forms calculated at varying initial i/j . The curves are plotted as cumulative percentages totaling 100%. That is, the percentage concentrations of the different circular species is represented by the areas bound by the curves rather than the curves themselves. (b) Number percentages of product circular forms calculated at varying initial i/j . As with part (a), the curves are plotted as cumulative percentages totaling 100%. (c) Molar concentration of product circular C_1 relative to j_1 calculated at varying initial i/j .

Figure 2: Experimental Tests of Predictions of Case A Simulations. Lane (i) contains a molecular weight ladder (1 Kb ladder, Bethesda Research Labs). Lanes (a) through (h) were ligated at, respectively, 1.6, 3.1, 6.2, 12.5, 25, 50, 100, and 200 $\mu\text{g/ml}$ (i/j values of 0.16 to 19.5). To minimize uneven lighting of the gel from irregularities in the visible-absorbing, UV-transparent filter employed in the short-wave UV transilluminator, the gel was mounted on the bottom surface of a glass plate held 2.5 cm above the filter. Although effective in eliminating "striation artifacts" previously noted as an inherent problem in UV transilluminators (28), a hazard of this procedure, as evident here, was occasional accidental loosening of the gel. An LKB Ultrosan XL laser densitometer was used to quantitate the ligation forms.

mation. The combination of these effects means that if cells are transformed with a ligation mix of concatamerized vector, the fraction of transformed cells carrying C_1 circles will never be less than about 90%, even if the ligation is done at extremely high concentration.

When initial DNA concentrations are increased, absolute yields of all circular forms increase; they however approach an asymptotic limit at high initial concentrations. Figure 1(c) plots the yield of C_1 circles (relative to j_1) for a starting i/j from 0.1 to 100. A direct analysis of the ligation equations shows that the final concentration of C_1 approaches $j/2$ as i/j approaches infinity (analysis not shown). This result, which is independent of the random coil model, suggests a novel method for obtaining the value of j and hence the statistical segment length b (see Eqn [1]): ligate a linear DNA fragment at high concentration, separate the product forms electrophoretically, and measure the concentration of monomeric circles.

This prediction of an asymptotic limit at high DNA concentrations was tested using linearized plasmid pBR322 (Figure 2). In each lane, residual linear monomeric DNA is less than 0.5% of the total DNA. With increasing starting concentration of linear fragment, the concentration of product monomeric circles rises to an estimated limit of 4.8 $\mu\text{g/ml}$. If we now assume the validity of the random coil model (i.e., validity of Eqn [1]), this concentration for $j/2$ implies a statistical segment length of 1250 Å, in good agreement with other measurements of

Figure 3: Case B ligation. Individual legends beside each graph correlate gray levels with percentages. (a,b,c) Percentage of linear vector ligated into $C_{1,1}$ circular product. (d,e,f) Percentage of linear insert ligated into $C_{1,1}$ circular product. (g,h,i) Number percentage of $C_{1,1}$ relative to all viable vector-containing circular forms (i.e. not possessing head-to-head inverted repeats). (j,k,l) Number percentage of viable $C_{1,2}$ double insert recombinants relative to all viable vector-containing circular forms.

DNA statistical segment length (12). Further, the limiting mass ratio of C_2 dimers to C_1 monomeric circles is 0.42, in good agreement with the predicted ratio of 0.36. This simple method of obtaining the value of j may have applications in the study of DNA physical chemistry.

Case B: Ligation of a "large" (> 800 bp) singly restricted insert into non-phosphatased vector (all ends identical):

Efficient ligation appears to require mutually contradictory reaction conditions. To minimize the formation of multiple vector recombinants ($m > 1$ in $L_{m,n}$ and $C_{m,n}$), a low concentration of vector is typically used. To compete efficiently with the tendency of vector to self-

ligate, the insert concentration must be comparable to the j -value of the vector. On the other hand, insert concentration must be low compared with $j_{1,1}$, to avoid formation of multiple insert recombinants. These conditions are mutually contradictory, especially with large inserts ($j_{1,1}$ is inversely proportional to $(L_{1,1})^{3/2}$; Eqn [1]), resulting in the common observation that large inserts are more difficult to ligate into $C_{1,1}$ products than small inserts (18). Not only is the maximum potential yield of $C_{1,1}$ products less at higher MW_{insert} (compare Figs. 3a, 3b, and 3c), but also the yield of $C_{1,1}$ as a fraction of all viable recombinant forms is less (compare Figs. 3g, 3h, and 3i). Figure 3 presents the results of numerical simulation of case B ligations for three insert/vector molecular weight ratios. Vector and insert concentrations are given in terms of their respective initial i/j ratios. In Figure 3, percentages of products (shown by the different grey scales) for three $MW_{\text{insert}}/MW_{\text{vector}}$ ratios are given for: percent vector found in $C_{1,1}$ product (Fig. 3a, b, c); percent insert found in $C_{1,1}$ product (Fig. 3d, e, f); percent $C_{1,1}$ relative to all viable circular species (i.e. lacking head-to-head inverted repeats) (Fig. 3g, h, i); and percent $C_{1,2}$ double insert recombinants relative to all viable circular species (Fig. 3j, k, l). Although we corrected for head-to-head repeats, we did not attempt to correct for the lowered transformation efficiency of larger DNA forms.

In the context of ligation theory, "optimization" can have numerous different meanings. One usually thinks of maximizing the absolute yield of $C_{1,1}$ recombinants, either relative to the initial amount of vector employed in the ligation, or else relative to the initial amount of insert. A second consideration would be to maximize the fraction of transformable circular product which represents $C_{1,1}$ recombinants rather than undesired ligation forms such as recircularized vector, multiple insert recombinants, and multiple vector recombinants. The importance of this second consideration depends heavily on the available selection or screening procedures.

Definite optimum insert concentrations which optimize the absolute yield of $C_{1,1}$ product for a given amount of vector are evident from Figures 3a, 3b, and 3c. Since vector DNA is required to form a replicon, a measure of percentage vector in $C_{1,1}$ is an appropriate measure of recombinants capable of transformation. The recommended insert concentrations are only moderately influenced by changes in the initial vector concentration (see vertical axis, Fig 3). In most experiments, a low vector concentration is desirable to minimize formation of multiple vector recombinants. Keeping this in mind, we read along the horizontal axis (vector $i/j = 0.1$) to find, by polynomial interpolation, the following optima:

For $MW_{\text{insert}}/MW_{\text{vector}} = 0.25$, a 16.5% yield results at insert $i/j = 0.42$.

For $MW_{\text{insert}}/MW_{\text{vector}} = 1.00$, a 15.1% yield results at insert $i/j = 0.65$.

For $MW_{\text{insert}}/MW_{\text{vector}} = 4.00$, a 6.1% yield results at insert $i/j = 1.7$

Regardless of the starting concentrations of insert and vector, the most abundant circular vector form is the regenerated $C_{1,0}$ circle (Data not shown). One of our motivations for studying ligation theory was to see if a good strategy exists to maximize the molar ratio of desired $C_{1,1}$ forms while at the same time minimizing the appearance of multiple insert recom-

binants. Figures 3g to 3l reveal, disappointingly, that given an unphosphatased, singly re-restricted vector, there simply is no ligation strategy which simultaneously gives high yield of $C_{1,1}$ recombinants while avoiding appearance of multiple inserts ($C_{1,2}$ shown in Fig. 3j, 3k, 3l). On the other hand, such strategies do exist when using phosphatased vectors or when performing forced directional cloning.

Our analysis predicts a plateau in the molar ratios of the various recombinant forms. In other words, although the absolute yield of $C_{1,1}$ product suffers if insert is used in excess of optimum levels, the mole fraction of $C_{1,1}$ recombinants relative to all circular product forms (excluding nonviable forms; see Appendix) never decreases even with vast excess concentration

Figure 4: Experimental Tests of Predictions of Case B Simulations. Ligation mixtures contained 5 $\mu\text{g/ml}$ of a 7.25 Kb vector (*Eco*RI-digested M13mp10 RF I DNA) and variable concentrations of a 4.36 Kb insert (*Eco*RI-digested pBR322) ranging from 1 $\mu\text{g/ml}$ to 125 $\mu\text{g/ml}$. JM101 competent cells prepared by calcium chloride treatment were transformed, and the transformation mixes were plated onto JM101 lawns containing X-gal and IPTG. Plaques were scored as white recombinants or blue regenerated vectors. Data were pooled from two experiments. (a) Observed percent fraction of recombinant plaques/total plaques plotted against concentration of insert (○) compared with a theoretical curve predicting percent fraction of viable recombinant circles/total viable circles. (b) Observed recombinant plaques per ng vector versus concentration of insert (○) compared with a theoretical curve predicting percent fraction of initial vector incorporated into viable recombinant circles.

Figure 5. Case C ligation. Explanation of plots (a) through (i) is the same as for Figure 3.

of insert. Likewise, although increasing insert concentrations results in a steadily decreasing absolute yield of regenerated $C_{1,0}$ vector circles, our analysis further predicts that the mole fraction of these circles cannot be reduced below fixed limits (Figs. 3g, h, i, and unpublished data). This somewhat surprising result is presaged by the Case A simulations and is supported by the following experiment (Figure 4). A 7.25 Kb vector (*Eco*RI-digested M13mp10 RF I DNA), with concentration fixed at vector $i/j = 0.6$ (4.8 $\mu\text{g/ml}$), was ligated with varying concentrations of a 4.36 Kb insert (*Eco*RI-digested pBR322) at insert i/j ratios from 0.1 to 12 (1.0 to 125 $\mu\text{g/ml}$), and transformed into JM101 and plated onto X-Gal indicator plates containing 0.1 mM IPTG. Plaques were scored as white recombinants or blue regenerated vectors. Figure 4(a) compares theoretical and observed recombinant plaque/total plaque ratios. Although total recombinant yield suffers when excess concentration of insert is used, even vast excess concentrations of insert does not increase the ratio of recombinants to regenerated $C_{1,0}$ vector circles beyond fixed limits, in qualitative agreement with the kinetic analysis. We attribute the lower ratio than theoretically predicted to the lower transformation efficiency of larger DNA forms. Figure 4(b) illustrates the absolute recovery of recombinant plaques as a function of insert concentration. The maximum recovery of recombinant plaques is obtained at an insert concentration of 8 $\mu\text{g/ml}$, in close agreement with the predicted value of 9 $\mu\text{g/ml}$. However, experiment shows a significantly slower than predicted decrease in recombinant yield at high insert concentrations. Actually, we had expected a faster decrease in recombinant yield than theoretically predicted, due to saturation of the competent cells (16). Apparently, saturation of cells by excess initial insert (or vector) concentrations is not a serious problem in cells made competent by conventional calcium chloride procedures (5). However, this may depend on procedures used to render cells competent for transformation.

In the experiment illustrated in Figure 4, we deliberately chose to use as "insert" the common cloning vector pBR322 so as to illustrate this point. Had we chosen to reverse which fragment was considered "insert" and which fragment was considered "vector", and had plated onto antibiotic selective media rather than onto a phage-sensitive lawn, then a maximum recovery of M13mp10 inserts into varying recombinant forms would have been obtained using a pBR322 vector concentration of 8 $\mu\text{g/ml}$. In the case illustrated, we lacked a positive insertional selection. However, had cloning been into sites within the tetracycline resistance gene, a strong positive insertional selection would have been available (19) to make this a practical procedure.

Case C: Forced directional ligation of doubly restricted insert into doubly restricted vector:

This common scheme to achieve high cloning efficiency is easy to analyze. Optimized case C ligation requires that vector and insert both be at low concentration, so that slow bimolecular associations between the reactant species will be followed almost exclusively by unimolecular

cyclization. If we then let p and q represent respectively the mole fractions of insert and vector, the mole fractions of $C_{0,2}$, $C_{1,1}$, and $C_{2,0}$ products will be given by p^2 , $2pq$, and q^2 . Since $C_{2,0}$, a head-to-head vector dimer, will be nonviable in *E. coli* transformation, it is evident that transformation products under these conditions must consist almost exclusively of $C_{1,1}$ recombinants. A low p/q ratio can be used to achieve nearly complete ligation of insert into $C_{1,1}$ recombinants; alternatively, a high p/q ratio can be used to achieve nearly complete ligation of vector into the desired recombinants.

Numerical solution of the ligation equations for this case confirms and extends the above analysis. Figure 5 plots results for three insert/vector molecular weight ratios. The horizontal axes plot insert/vector molar ratios rather than i/j values. This change of coordinate is necessitated by the lack of interference from unimolecular recircularization in this ligation scheme. Except for this difference in choice of abscissa, parts (a) through (i) of Figure 5 are equivalent to parts (a) through (i) of Figure 3. As with Figure 3, we corrected for non-viability of head-to-head inverted repeats, but made no attempt to correct for the lowered transformation efficiency of large DNA. Forced directional cloning is remarkably tolerant of poorly chosen ligation conditions; greater than 95% of all the transformable ligation products are $C_{1,1}$ species (Figs. 5g, h, i), regardless of the starting insert and vector concentrations. This is clearly the best cloning strategy when possible.

Case D: Ligation of a "large" (> 800 bp) singly restricted insert into phosphatased vector:

The ligation equations which we derived for this case include a free parameter "R" which represents the relative efficiency of ligation of hemiphosphorylated versus bisphosphorylated end pairs. Preliminary experiments indicated that R has a value significantly less than 1, but the exact value of R depended on the conditions used during its measurement (data not shown). However, we show below that arbitrarily setting $R = 1.0$ in the ligation equations results in theoretical predictions which agree quite well with experiment.

Figure 6 shows the results of simulating case D ligations assuming $R = 1.0$. Two main scenarios requiring optimization can be imagined:

In the first optimization scenario, we assume that there is no problem in obtaining sufficient insert, and our intent is to maximize the yield of $C_{1,1}$ recombinants. Optimum insert concentrations (Figure 6a, b, c) exist which will maximize the incorporation of vector into $C_{1,1}$ recombinants. As with Case B ligations, optimum insert concentrations are only moderately affected by modest changes in vector concentration, showing that the common strategy of setting up ligations in terms of insert/vector molar ratios is inappropriate for this case. Holding vector concentration constant at $i/j = 0.1$, we find, by polynomial interpolation, the following insert optima:

For $MW_{\text{insert}}/MW_{\text{vector}} = 0.25$, a 22.5% yield results at insert $i/j = 0.39$.

For $MW_{\text{insert}}/MW_{\text{vector}} = 1.00$, a 24.0% yield results at insert $i/j = 0.57$.

For $MW_{\text{insert}}/MW_{\text{vector}} = 4.00$, a 16.7% yield results at insert $i/j = 1.4$

Figure 6: Case D ligation. Explanation of plots (a) through (l) is the same as for Figure 3.

In the second optimization scenario, we assume that insert is available only in very limited ("infinitesimal") amounts, so that we wish for a maximum recovery of insert into $C_{1,1}$ forms. In contrast to case B ligations, formation of $C_{1,1}$ recombinants free of undesired forms may in theory be achieved at low insert concentration (compare Figs. 6g, h, i with Figs. 6j, k, l). Note that the major qualitative differences between Figs. 3g, h, i and Figs. 6g, h, i are due to the inability of phosphatased vector to recircularize. To compete efficiently with insert self-ligation, vector concentration should be comparable to $j_{0,1}/R$, yet a conflicting requirement is that the vector concentration must not be large compared with the j -value of the linear $L_{1,1,1}$ intermediate. In contrast to Case B ligations, the result of vector ligating with $L_{1,1,1}$ is $L_{2,1,0}$, a relatively inert product. Hence, high vector concentration does not result in excessive production of transformable circles. Because of this, a reasonable compromise can be worked

TABLE 2: CASE D VECTOR I/J VALUES FOR OPTIMIZING RECOVERY OF INFINITESIMAL CONCENTRATIONS OF INSERT

MW _{Insert} /MW _{vector}				R = 0.1 Optimum		R = 1.0 Optimum	
	j _{1,0}	j _{0,1}	j _{1,1}	j _{0,1} /R	Vector i/j	j _{0,1} /R	Vector i/j
0.25	1.00	8.00	0.72	80.0	5.35	8.00	1.70
1.00	1.00	1.00	0.35	10.0	1.33	1.00	0.42
4.00	1.00	0.12	0.09	1.2	0.24	0.12	0.07

Tabulated values assume validity of the random coil model (see Equation [1]) and are determined for R=0.1 and for R=1.0. The value of j_{1,0} is normalized to 1.00. Optimum vector i/j values for R = 1.0 are emboldened since this setting of the free parameter gives predictions consistent with experimental results.

out between the conflicting requirements for efficient insert recovery, as summarized in Table 2. In the limit of infinitesimal insert concentration, the case D equations permit an exact analytical solution, yielding an equation for vector concentration which maximizes the fraction of insert going into C_{1,1} product:

$$\frac{i}{j} \text{ of vector} = \frac{\sqrt{j_{0,1}j_{1,1}/(2R)}}{j_{1,0}}$$

(3)

As expected, this recommended concentration represents a sort of compromise between j_{0,1} and j_{1,1}/R. Values given in the final four columns of Table 2 for optimum vector i/j, which assume the random coil model is valid, are determined for R = 0.1 and R = 1.

The somewhat unexpected notion that relatively high phosphatased vector concentrations might be used to maximize the recovery of insert into C_{1,1} product is presaged by the results of the Case B simulations. Indeed, using phosphatased vector is greatly preferable, since maximum achievable recoveries into C_{1,1} product are significantly higher, and a positive insertional selection is unnecessary. However, we had two major concerns about the applicability of this theoretical result to actual practice. First, although saturation of competent cells by high DNA concentrations was not apparent in testing of Case B ligations (Fig. 4), it seemed possible to us that this absence of apparent saturation was an effect of concatamerizing excess DNA to high molecular weight forms. Second, phosphatase treatment of vector frequently seems incomplete, resulting in background problems which would only be accentuated at high vector concentration.

To test the degree of agreement between theory and experiment, we performed the following experiment. A low, fixed concentration (2.5 µg/ml) of a 1.2 Kb (our measured length) *EcoRI* fragment from pUC4K (20) bearing a kanamycin-resistance gene was ligated with varying amounts of the 2.7 Kb vector pUC19 (8) which had been *EcoRI*-linearized and phosphatased. Starting vector i/j ratios varied from 0.06 to 15.4 (0.8-200 µg/ml). After transformation

Figure 7: Experimental Tests Predictions of Case D Simulations. Reaction mixtures totalling 10 μ l carried a constant 2.5 μ g/ml concentration of the kanamycin-resistance-bearing *Eco*RI fragment from pUC4K (20), which had a measured length of 1.2 Kb (versus the published value of 1.4 Kb), and a dilution series of the 2.7 Kb vector pUC19 (maximum concentration 200 μ g/ml) linearized with *Eco*RI and treated with calf intestinal phosphatase. After overnight ligation, the reaction mixture was transformed into *E. coli* LE392 made competent by calcium chloride treatment, and the transformation mixtures were spread on plates containing ampicillin and kanamycin. Plate counts (○) are plotted along with two theoretical curves for the percentage of fragment incorporated into viable recombinants assuming $R = 0.1$ and $R = 1.0$. The theoretical curves have each been arbitrarily scaled to facilitate comparison with the plate counts.

into *E. coli* LE392, the cells were spread on plates selecting simultaneously for kanamycin and ampicillin resistance. Results are plotted in Figure 7, along with two theoretical curves (arbitrarily scaled in the vertical dimension) which assume $R = 0.1$ and $R = 1.0$. The overall shape of the experimental curve is consistent with the theoretical curve assuming $R = 1.0$. This argues that hemiphosphorylated DNA end pairs are as efficient a substrate for T4 DNA ligase as bisphosphorylated end pairs. Theory and experiment hence agree remarkably well. To our surprise, and consistent with the results of the Case B ligations, saturation of competent cells was not apparent even at the highest starting vector concentration. Thus, high concentrations of phosphatased vector can indeed be used to recover minute amounts of insert.

Cases B and D: Ligation of "short" and "intermediate length" inserts:

Optimum concentration recommendations given above for Cases B and D are applicable when the insert fragment is of a sufficient size that the Jacobson-Stockmayer equation (Eqn [1]) is approximately valid, i.e. the minimum size insert which can be considered is about 800 bp. (Note that Case C recommendations for asymmetrical ligation are independent of length and are hence applicable to all sizes of insert fragment.) "Short" insert fragments, less than about 200 bp, are probably unable to circularize unless they carry bending sequences (21) and are also small relative to most vectors. This allows a major simplification of the ligation model. The following analysis assumes infinitesimal length inserts with no tendency to circularize.

For Case B and short inserts, there are two optimizing situations to consider: (a) If there is no limitation on available quantity of short insert, and one wishes therefore to maxi-

mize incorporation of vector into $C_{1,1}$ circles, one ligates using a low concentration of vector and a concentration of insert providing ends equivalent to 0.57 times the j -value of the vector. These conditions result in up to 22% of the vector being incorporated into $C_{1,1}$ circles. (b) If, instead, the short insert is available only in limited ("infinitesimal") amounts, then, provided that one has available a positive insertional selection, one can maximize the incorporation of insert into singly recombinant circles by using a vector $i/j = 0.62$. These conditions result in up to 35% of the short insert being incorporated into singly recombinant circles (analysis not shown).

Ligation of phosphatased vector (Case D) with short insert results in a situation very similar to Case C ligations, in that unimolecular cyclization of the precursor molecules is no longer a complicating factor. There are two optimizing situations to consider: (a) If there is no limitation on available quantity of short insert, and one wishes to maximize incorporation of vector into $C_{1,1}$ circles, one uses a low concentration of vector and a concentration of insert which results in as high an insert/vector ratio as is possible while still maintaining a total concentration of ends which is moderate relative to the j -value of the vector. These conditions result in up to 100% of the vector being incorporated into $C_{1,1}$ recombinant circles. (b) If, instead, the short insert is available only in limited amounts, then to maximize the incorporation of insert into $C_{1,1}$ circles, one adds a concentration of vector to the ligation mix which results in as high a vector/insert ratio as is possible while still maintaining a total concentration of ends which is moderate relative to the j -value of the vector. These conditions result in up to 100% of the insert being incorporated into $C_{1,1}$ recombinant circles.

"Intermediate length" inserts (between 200 - 800 bp) display complex kinetic behavior making it difficult to make general recommendations, not least because of the effects of helical phase angle, which can result in variations in the rate of cyclization over two orders of magnitude (15). The best general ligation strategy for such fragments may be simply to assume that optimum conditions are intermediate between those predicted by the large and short insert models.

DISCUSSION

The numerous factors involved in defining optimum ligation for all possible insert and vector combinations are admittedly difficult to keep in mind. Table 3 presents a simplified set of recommendations which, although not defining absolutely optimum conditions, nevertheless do define a reasonably acceptable compromise. These compromise recommendations are made possible by the demonstrated breadth of the ligation optima (i.e. being off by $\pm 30\%$ in insert starting concentration makes little difference in final yield), and are based on the suppositions that the most commonly used plasmid cloning vectors range from 2.5 to 7.5 kb, and that inserts to be cloned into circular plasmid vectors are generally smaller than 10 kb.

TABLE 3: A PRACTICAL SET OF RECOMMENDATIONS FOR LIGATION

The recommendations below define a reasonable compromise set of conditions for efficient ligation of vector into C₁,1 product, made possible by the demonstrated breadth of the ligation optima (i.e. being off by $\pm 30\%$ in insert starting concentration makes little difference in final yield). We presume here that the most commonly used plasmid cloning vectors range from 2.5 to 7.5 kb, and that inserts to be cloned are generally smaller than 10 kb.

- 1) Whenever possible, use either forced directional cloning or phosphatased vectors.
- 2) Under normal conditions, keep the vector concentration at 1 $\mu\text{g/ml}$ or less. If the primary intent is to recover with maximum efficiency "infinitesimal" amounts of insert, then higher concentrations of vector may be optimal.
- 3) Assuming symmetric ligation with unphosphatased vectors:
 - a) If the insert is large (i.e. > 800 bp), use 8 $\mu\text{g/ml}$ of insert.
 - b) If the insert is small (i.e. < 200 bp), use 1 nM insert for a 7.5 kb vector or 5 nM insert for a 2.5 kb vector.
 - c) If the insert is intermediate sized, optimum conditions are intermediate. As a starting point, try using 5 nM insert for a 7.5 kb vector or 10 nM insert for a 2.5 kb vector, increasing or decreasing this trial concentration according to whether the insert more closely resembles a large or small fragment.
- 4) Assuming asymmetric ligation (i.e. forced directional cloning):
 - a) Regardless of insert size, use a 3:1 insert:vector molar ratio.
 - b) However, when following the above rule, do not exceed 5 $\mu\text{g/ml}$ of insert.
- 5) Assuming symmetric ligation with phosphatased vector:
 - a) If the insert is large (i.e. > 800 bp), use 8 $\mu\text{g/ml}$ of insert.
 - b) If the insert is small (i.e. < 200 bp), use a 3:1 insert:vector molar ratio
 - c) If the insert is intermediate sized, optimum conditions are intermediate. As a starting point, try using 5 nM insert for a 7.5 kb vector or 10 nM insert for a 2.5 kb vector, increasing or decreasing this trial concentration according to whether the insert more resembles a large or small fragment.
- 6) If the intent of the ligation is to recover "infinitesimal" amounts of insert, note that forced directional ligation, if possible, offers the best possible yields, followed by ligation with excess phosphatased vector. The main text should be consulted for further details on optimizing insert recovery.
- 7) The above recommendations hold when the intent of the ligation is to yield circular products. For lambda and cosmid ligations, where the intent is to obtain linear concatamers, insert and vector should be ligated at a 1:1 molar ratio at as high a concentration as practical.

Hitherto published theoretical studies of the ligation process have been inadequate for several reasons. Dugaiczky et al. (1) modeled the concatamerization of a single DNA species into an equilibrium mixture of linear and circular concatamers. Their equilibrium analysis offers little insight into how formation of circular 1:1 ligation forms might be optimized. Legerski and Robberson (3) applied equilibrium analysis to the following cases (a) ligation of inserts into plasmid vectors linearized with one or two restriction endonucleases, (b) annealing of deoxyhomopolymer-tailed inserts with complementary tailed plasmids, and (c) the linear concatamerization of inserts to lambda-type and (d) cosmid-type DNA vectors. Their analysis can also be faulted for using an equilibrium approach. However, in the limit of low DNA concentra-

tion, the predicted annealing behavior of homopolymer-tailed fragments is nearly identical whether analyzed kinetically or as an equilibrium, due to the intrinsic inability of such fragments to self-anneal (Results not shown). Likewise for lambda (22) and cosmid (23) ligations, we did not feel it necessary to improve on their analysis, since the intent in these ligations is to produce linear concatemers rather than circular forms (24). Mashko et al. (25,26) published kinetic equations mathematically equivalent to our case A and case B equations, but solved the equations for only a few specific conditions. Likewise, during preparation of this manuscript, Dardel (27) published a kinetic analysis with results similar to our own. However, Dardel expressed all DNA concentrations in dimensioned units of molarity. In contrast, we found it much preferable to express DNA concentrations, whenever possible, in terms of dimensionless i/j units. This extends the utility of our analysis to many more ligation situations than covered by Dardel, permits a greatly enhanced physical understanding of the ligation process, and frees us from depending on the accuracy of any particular measured value of DNA statistical segment length.

We have intended to develop here a means of understanding both intuitively as well as quantitatively the compromises needed in optimizing a ligation reaction. It should be remembered that theory must be supplemented with practical experience, and what one considers optimum depends on the needs of the experiment.

For example, what one considers an acceptable balance between high frequency of insertion and low frequency of multiple insertion is dependent on the available selections. To maximize the efficiency of ligating a singly restricted insert available in only limited amounts, theory dictates use of a fairly high concentration of vector. Unfortunately, in the absence of selection, high concentrations of unphosphatased vector result in high numbers of non-recombinant vector transformants, while even phosphatased vectors are subject to background problems. Furthermore, although not apparent in our experiments using cells rendered competent by CaCl_2 treatment, excessive DNA concentrations may saturate the DNA uptake ability of cells made competent by certain high efficiency protocols (16). In certain cases, theory may regard as optimum unrealistically high or impractically low concentrations of vector or insert. One is then more interested in the consequences of using practical DNA concentrations rather than "optimization."

Comprehensive experimental optimization studies of the ligation process are not yet available. It is hoped that the ligation theory developed in this paper will guide experiments leading to the development of improved ligation protocols.

ACKNOWLEDGEMENTS

We thank Adella M. Garland for insightful discussions and Dr. B. H. Zimm for critically reading our manuscript and suggesting many helpful improvements. This research was supported by an NIH grant GM31839 to D.W.S.

Present addresses: [†]Chemistry Department, University of San Diego, San Diego, CA 92110 and
[§]Biology Department, FO-31, University of Texas at Dallas, Richardson, TX 75083, USA

*To whom correspondence should be addressed

REFERENCES

1. Dugaiczky, A., Boyer, H., & Goodman, H.M. (1975). *J. Mol. Biol.* **96**, 171-184.
2. Jacobson, H. & Stockmayer, W.H. (1950). *J. Chem. Phys.* **18**, 1600-1606.
3. Legerski, R.J., & Robberson, D.L. (1985). *J. Mol. Biol.* **181**, 297-312.
4. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W.J., & Goodman, H.M. (1977). *Science* **196**, 1313-1319.
5. Maniatis, T., Fritsch, E.F., & Sambrook, J. (1982). **Molecular Cloning: A Laboratory Manual.** (Cold Spring Harbor Laboratory, New York).
6. Acton, F.S. (1970). **Numerical Methods that Work.** (Harper & Row, New York).
7. Zyskind, J.W., Deen, L.T., and Smith, D.W. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 3097-3101.
8. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). *Gene* **33**, 103-119.
9. Wang, J.C., & Davidson, N. (1968). *Cold Spring Harbor Symp. Quant. Biol.* **33**, 409-415.
10. Shore, D., Langowski, J., & Baldwin, R.L. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 4833-4837.
11. Shore, D., & Baldwin, R.L. (1983). *J. Mol. Biol.* **170**, 957-981.
12. Record, M.T. Jr., Mazur, S.J., Melancon, P., Roe, J.-H., Shaner, S.L., & Unger, L. (1981). *Ann. Rev. Biochem.* **50**, 997-1024.
13. Shimada, J., & Yamakawa, H. (1984). *Macromolecules* **17**, 689-698.
14. Shimada, J., & Yamakawa, H. (1985). *J. Mol. Biol.* **184**, 319-329.
15. Levene, S.D. and Crothers, D.M. (1986). *J. Mol. Biol.* **189**, 61-72.
16. Hanahan, D. (1983). *J. Mol. Biol.* **166**, 557-580.
17. Mizuuchi, K., Mizuuchi, M., & Gellert, M. (1982). *J. Mol. Biol.* **156**, 229-243.
18. Malik, V.S. (1981). *Advances in Applied Microbiol.* **27**, 1-84.
19. Bochner, B.R., Huang, H.C., Schieven, G.L., & Ames, B.N. (1980). *J. Bacteriol.* **143**, 926-933.
20. Vieira, J. and Messing, J. (1982). *Gene* **19**, 259-268.
21. Marini, J.C., Levene, S.D., Crothers, D.M. and Englund, P.T. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 7664-7668.
22. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K., & Efstratiadis, A. (1978). *Cell* **15**, 687-701.
23. Collins, J., & Hohn, B. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 4242-4246.
24. Hohn, T., & Katsura, I. (1977). *Curr. Topics Microbiol. Immunol.* **78**, 69-78.
25. Mashko, S.V., Mironov, A.A., Sorokin, A.V., Bratus, A.S., Kozlov, Y.I., & Debebov, V.G. (1981). *Mol. Biol. (Moscow)*. **15**, 1397-1404.
26. Mashko, S.V., Rozinov, M.N., Kozlov, Y.I., Rebentish, B.A., Bratus, A.S., Mironov, A.A., Kaluzhskii, V.E., and Debebov, V.G. (1980). *Mol. Biol. (Moscow)* **14**, 1023-1038.
27. Dardel, F. (1988). *Nucl. Acids Res.* **16**, 1767-1778.
28. Yee, T. and Inouye, M. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 2723-2727.

APPENDIX:

To simplify the appearance of the systems of kinetic equations, we have omitted the use of concentration brackets. Differences between blunt end and cohesive end ligation are assumed to require modification only of the kinetic rate constant k_f .

Case A:

$$i = 2 \sum_{p=0}^{\infty} L_p$$

$$\frac{d}{dt} L_n = -k_f L_n (j_n + 2i) + 2k_f \sum_{p=0}^n L_p L_{n-p}$$

$$\frac{d}{dt} C_n = k_f L_n j_n$$

$$L_0 = 0$$

Case B:

$$i = 2 \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} L_{p,q}$$

$$\frac{d}{dt} L_{m,n} = -k_f L_{m,n} (j_{m,n} + 2i) + 2k_f \sum_{p=0}^m \sum_{q=0}^n L_{p,q} L_{m-p,n-q}$$

$$\frac{d}{dt} C_{m,n} = k_f L_{m,n} j_{m,n}$$

$$L_{0,0} = 0$$

The above equations do not distinguish between different permutations or relative orientations of insert and vector in any given (m,n) ligation form. It is presumed, however, that ligation forms carrying head-to-head inverted repeats are incapable of stable transformation. To compute ligation efficiency, it is necessary to correct for nonviable forms by multiplying the final yields of $C_{m,n}$ with the appropriate weighting factors $W(m,n)$.

To compute the weighting factors $W(m,n)$ for case B, the permutations of m vector and n insert molecules were generated. For each permutation, a partial weighting factor was found by counting the number of contiguous vector/vector or insert/insert joins and applying a factor of 1/2 for each instance (exception: minus one instance when $m = 0$ or $n = 0$). Assuming the equal likelihood of each permutation, $W(m,n)$ represents the arithmetic mean of all the partial weighting factors.

Case C: The kinetic equations for this case are very similar to those for case B, except that the bimolecular reaction rate terms are halved, and the unimolecular reaction rate terms are multiplied by the function "z(m,n)" to indicate that odd-length concatamers do not ligate:

$$\text{if } (m+n) \text{ is odd then } z(m,n) = 0 \text{ else } z(m,n) = 1$$

$$\frac{d}{dt} L_{m,n} = -k_f L_{m,n} [z(m,n)j_{m,n} + i] + k_f \sum_{p=0}^m \sum_{q=0}^n L_{p,q} L_{m-p,n-q}$$

$$\frac{d}{dt} C_{m,n} = k_f z(m,n) L_{m,n} j_{m,n}$$

$$L_{0,0} = 0$$

The weighting factors $W(m,n)$ were computed as follows: When m is not equal to n, formation of head-to-head inverted repeats is unavoidable, for which cases therefore $W(m,n) = 0$. For $m = n$, we have the following:

$$W(m,n) = \frac{2m!n!}{(m+n)!}$$

Case D:

$$g = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} (2 L_{p,q,2} + L_{p,q,1})$$

$$h = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} (2 L_{p,q,0} + L_{p,q,1})$$

$$\begin{aligned} \frac{d}{dt} L_{m,n,2} = & -k_f L_{m,n,2} (j_{m,n} + 2g + 2Rh) \\ & + k_f \sum_{p=0}^m \sum_{q=0}^n (2L_{p,q,2} L_{m-p,n-q,2} + 2RL_{p,q,2} L_{m-p,n-q,1}) \end{aligned}$$

$$\begin{aligned} \frac{d}{dt} L_{m,n,1} = & -k_f L_{m,n,1} (Rj_{m,n} + g + Rg + Rh) \\ & + k_f \sum_{p=0}^m \sum_{q=0}^n (4RL_{p,q,2} L_{m-p,n-q,0} + 2L_{p,q,2} L_{m-p,n-q,1} + RL_{p,q,1} L_{m-p,n-q,1}) \end{aligned}$$

$$\frac{d}{dt} L_{m,n,0} = -2k_f Rg L_{m,n,0} + k_f \sum_{p=0}^m \sum_{q=0}^n \left[\left(\frac{1}{2} \right) L_{p,q,1} L_{m-p,n-q,1} + 2RL_{p,q,1} L_{m-p,n-q,0} \right]$$

$$\frac{d}{dt} C_{m,n} = k_f j_{m,n} (L_{m,n,2} + RL_{m,n,1})$$

The weighting factors $W(m,n)$ were computed as follows: It is impossible to have a circular form with $m > n$. For $m \leq n$, the fraction of molecules lacking head-to-head inverted repeats is given by:

$$W(m,n) = (1/2)^{n-m}$$