[01] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

Directed Evolution (S. Liu B. Dubertret)

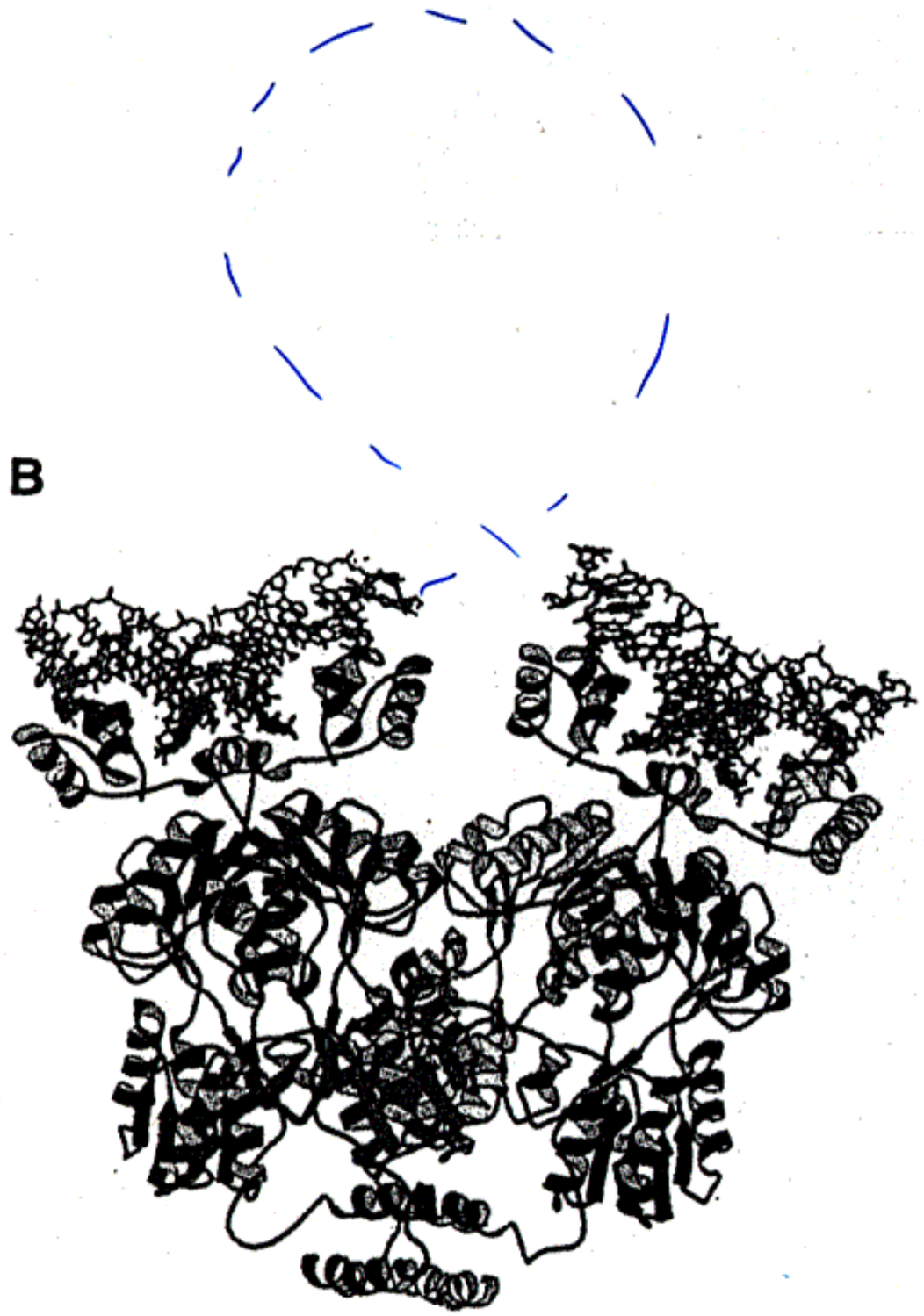
A molecular study

evolution of DNA binding site

evolution / generation of promoters
(replication origin)

evolution of proteins

NEC Inst. Princeton



ividual monomers (residues 1 to 357) are colored
gonucleotides are in blue. The orientation of the
l form in Fig. 5. (**B**) View of the tetramer which has
proximately 40° from the orientation shown in (A).

Fig. 7. View of the *lac* repressor–DNA complex illustrating that the two 21-bp fragments of deoxyoligonucleotide are separated in space and do not interact in the tetramer. The color scheme is the same as in Figs. 2 and 6.

[03] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

Dynamics of in-vitro Evolution

evolution of a DNA sequence (lacO) binding to a fixed protein lac Repressor

1)
- Start from random pool of DNA (20 bases long)
  10^{-16} mole (probability that sequence present $< 10^{-3}$)
- choose 30 base long so that solution with many phases are possible (control)
- In effect 70 base (2 primers 20 b long)
- lacR with biotin covers surface of tube

2) Follow trajectory, doing evolution cycles
at each cycle: mutation 10^{-4}/base/cycle
with 25 round of PCR at each cycle
So mutation / amplification
Selection to lacR coated tube
at each cycle 6 to 17 DNA sequenced
Solution

AATTGTGAGC GCTCACAATT

[04] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

We want

1) measure specificity — very high (1 sequ)

2) can we infer information on Protein / DNA
which bases are selected at different times

3) do the conformation of protein and DNA changes

yes 40° distortion of DNA
and some bases interact first

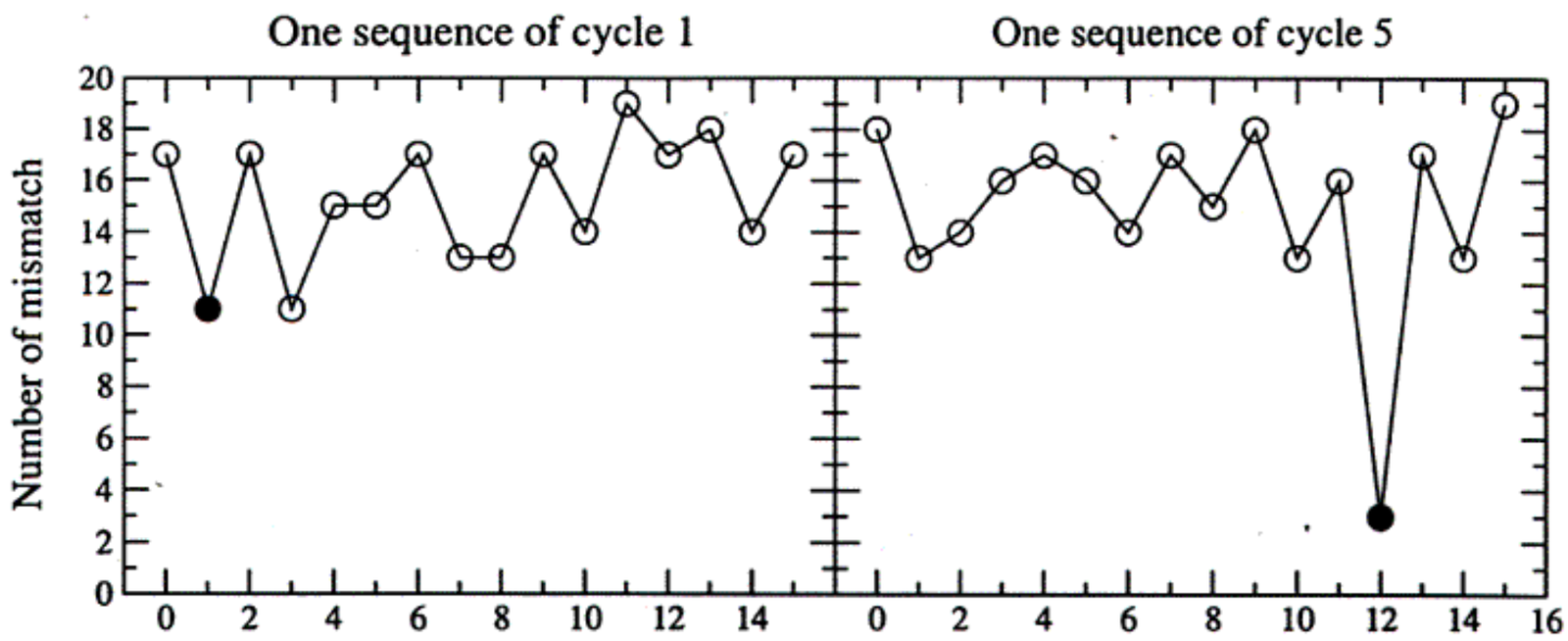


FIG. 1. Number of mismatches computed in all possible alignments between an evolving sequence and *lac*O. Alignment 0 means that the left of the two sequences coincide. Alignment 1 means that the 20 base long sequence is shifted to the right by one base, and so on. The dark circles represent the first alignment with the smallest number of mismatch. It defines the distance and the phase of the evolving sequence. The sequence of cycle 1 (left) has a phase of 1 and a distance of 11. The sequence of cycle 5 (right) has a phase of 12 and a distance of 3.

TABLES

mismatches	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
cycle 1												1	3	2		
cycle 2											1	1	5		2	
cycle 3			close to right part of LacO						1			3	4	5		
cycle 4			2	2				1				3	3	6		
cycle 5			6	4	1			3 mismatch from LacO, wild type								
cycle 6	1		5	2												
cycle 7		1	3	2												
cycle 8		2	3	5												

TABLE I. Number of sequences with m mismatches for each cycle of evolution.

phase	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
cycle 1		1			1		1		1	2						
cycle 2	1	1		1		3	1		1							1
cycle 3	1	1	1	1	3	2		1		1			1	1		
cycle 4	4		2	1	2		1	1	2	1		1	2			
cycle 5	3	1						3	1				1	2	overlap with right primer	
cycle 6	2		1					3	1			1				
cycle 7	2					1		3								
cycle 8	3						1	4	2							

3 phases

TABLE II. Number of sequences with phase p. The phase corresponds to the alignment with the smallest distance reported tableI.

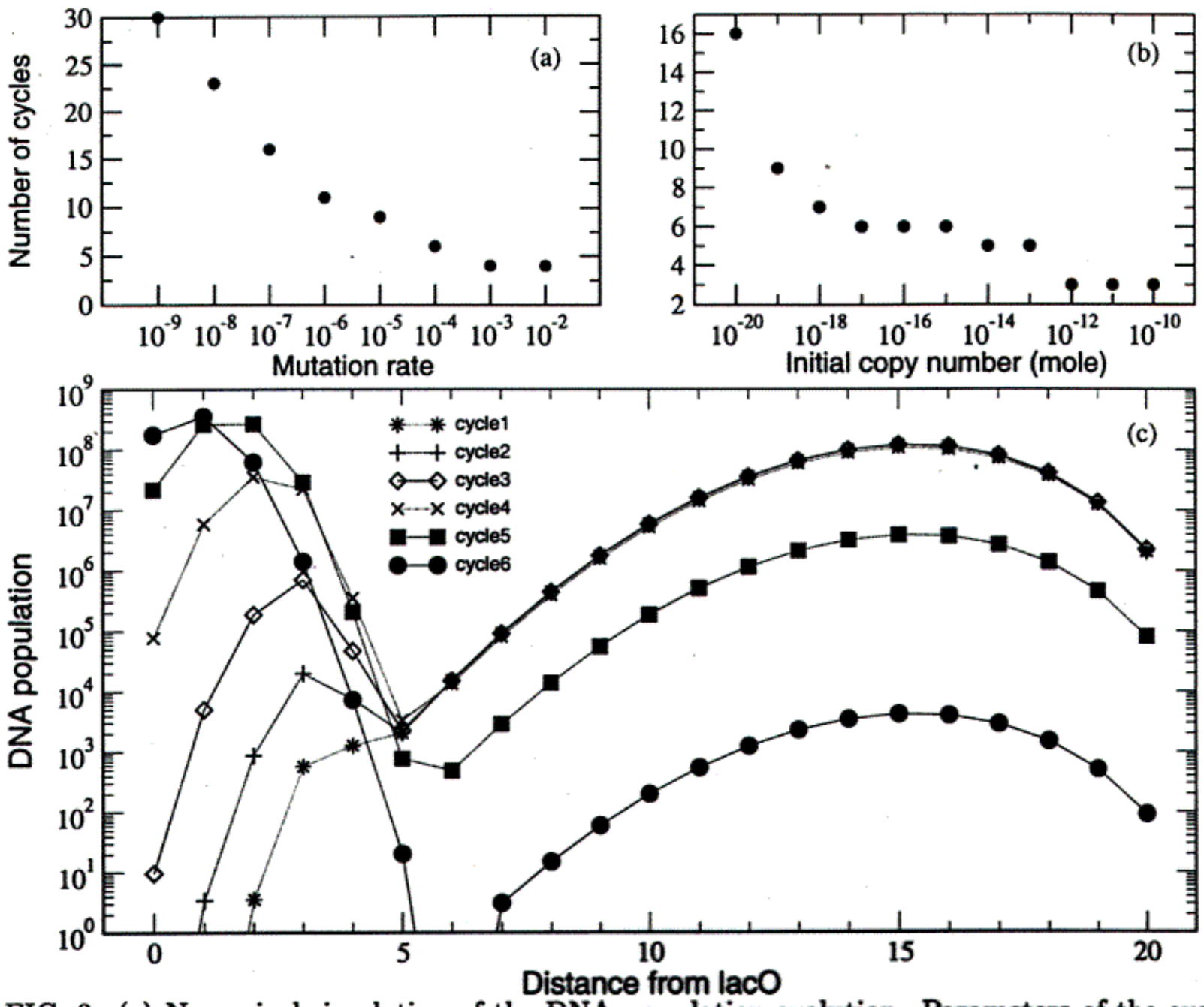


FIG. 3. (c) Numerical simulation of the DNA population evolution. Parameters of the evolution: starting number of DNA duplex: 10^{-16}mole, mutation rate: 10^{-4} per cycle per nucleotide. Influence of (a) the rate of mutation per cycle per nucleotide and (b) the initial concentration on the speed of convergence. In (a), the initial number of DNA molecule is 10^{-16}mole, in (b) the rate of mutation is 10^{-4}.

Rec A polymerization on DNA

Probe of sequence and structure
in equilibrium (ATPγS)

Amplified discrimination, proof reading
out of equilibrium (ATP)

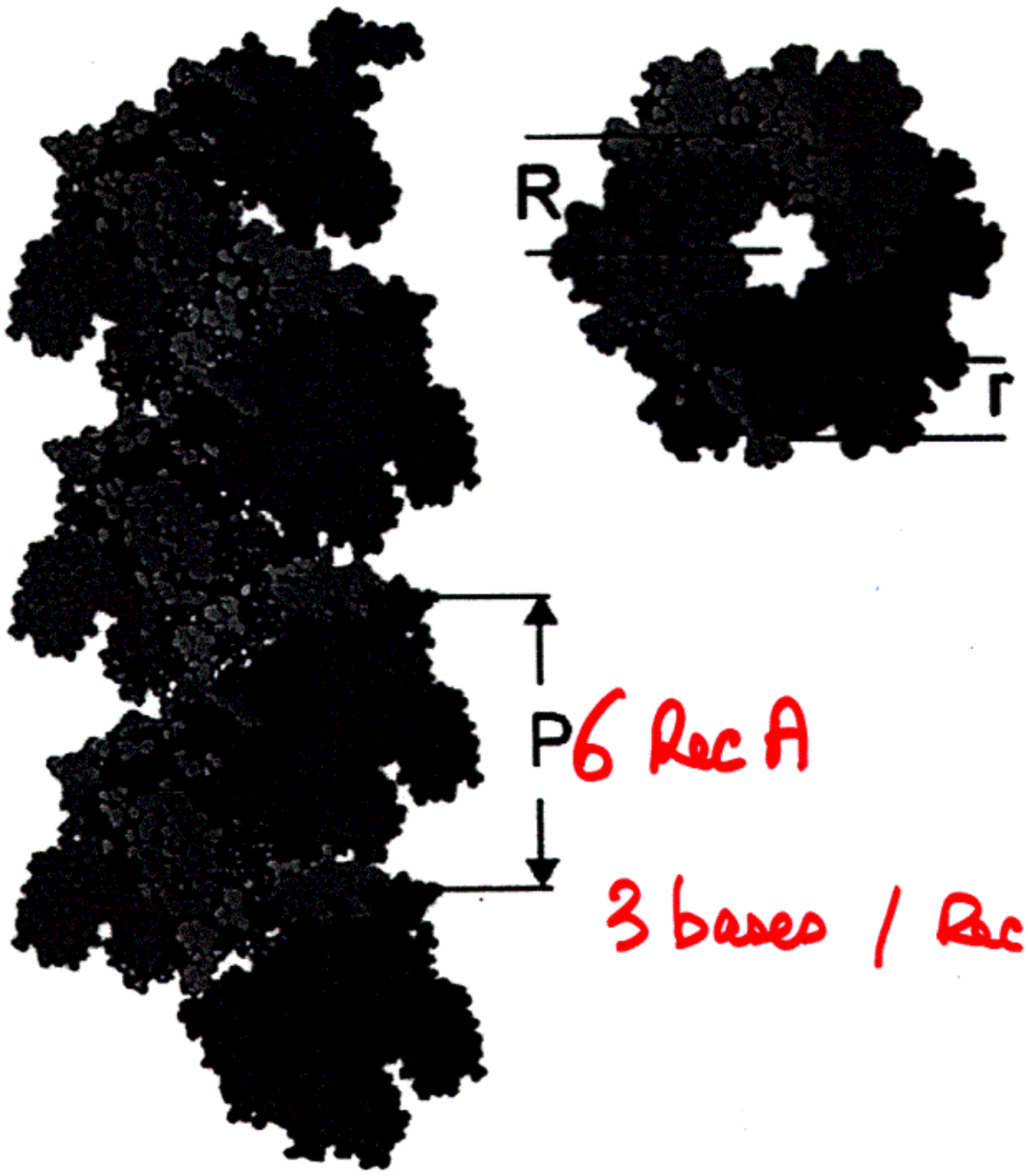
R. BAR ZIV
T. TLUSTY

Basic RecA Facts

E. Coli protein, MW=38kDa;
Plays essential role in homologous recombination and DNA repair

An Allosterically regulated DNA-binding protein
Binds DNA with 3 bases per monomer
Forms a right-handed helical filament

DNA-dependent ATPase:
RecA-DNA-ATP -> high affinity
RecA-DNA-ADP -> low affinity



[10] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

GENETICAL IMPLICATIONS OF THE STRUCTURE OF DEOXYRIBONUCLEIC ACID

By J. D. WATSON and F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge

Reprinted with permission from
Nature, Vol. 171, No. 4361, May 30, 1953
pp. 964–967.

THE importance of deoxyribonucleic acid (DNA) within living cells is undisputed. It is found in all dividing cells, largely if not entirely in the nucleus, where it is an essential constituent of the chromosomes. Many lines of evidence indicate that it is the carrier of a part of (if not all) the genetic specificity of the chromosomes and thus of the gene itself.

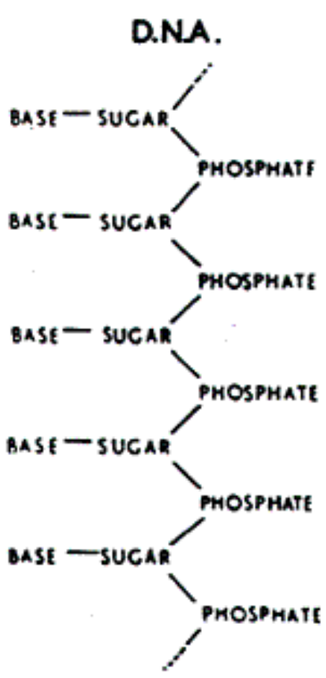


Fig. 1. Chemical formula of a single chain of deoxyribonucleic acid

Fig. 2. This figure is purely diagrammatic. The two ribbons symbolise the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

Until now, however, no evidence has been presented to show how it might carry out the essential operation required of a genetic material, that of exact self-duplication.

We have recently proposed a structure[1] for the salt of deoxyribonucleic acid which, if correct, immediately suggests a mechanism for its self-duplication. X-ray evidence obtained by the workers at King's College, London[2], and presented at the same time, gives qualitative support to our structure and is incompatible with all previously proposed structures[3]. Though the structure will not be completely proved until a more extensive comparison has been made with the X-ray data, we now feel sufficient confidence in its general correctness to discuss its genetical implications. In doing so we are assuming that fibres of the salt of deoxyribonucleic acid are not artefacts arising in the method of preparation, since it has been shown by Wilkins and his co-workers that similar X-ray patterns are obtained from both the isolated fibres and certain intact biological materials such as sperm head and bacteriophage particles[2,4].

The chemical formula of deoxyribonucleic acid is now well established. The molecule is a very long chain, the backbone of which consists of a regular alternation of sugar and phosphate groups, as shown in Fig. 1. To each sugar is attached a nitrogenous base, which can be of four different types. (We have considered 5-methyl cytosine to be equivalent to cytosine, since either can fit equally well into our structure.) Two of the possible bases—adenine and guanine—are purines, and the other two—thymine and cytosine—are pyrimidines. So far as is known, the sequence of bases along the chain is irregular. The monomer unit, consisting of phosphate, sugar and base, is known as a nucleotide.

The first feature of our structure which is of biological interest is that it consists not of one chain, but of two. These two chains are both coiled around a common fibre axis, as is shown diagrammatically in Fig. 2. It has often been assumed that since there was only one chain in the chemical formula there would only be one in the structural unit. However, the density, taken with the X-ray evidence[2], suggests very strongly that there are two.

The other biologically important feature is the manner in which the two chains are held together. This is done by hydrogen bonds between the bases, as shown schematically in Fig. 3. The bases are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other. The important point is that only certain pairs of bases will fit into the structure. One member of a pair must be a purine and the other a pyrimidine in order to bridge between the two chains. If a pair consisted of two purines, for example, there would not be room for it.

We believe that the bases will be present almost entirely in their most probable tautomeric forms. If this is true, the conditions for forming hydrogen bonds are more restrictive, and the only pairs of bases possible are :

adenine with thymine ;
guanine with cytosine.

The way in which these are joined together is shown in Figs. 4 and 5. A given pair can be either way round. Adenine, for example, can occur on either chain ; but when it does, its partner on the other chain must always be thymine.

This pairing is strongly supported by the recent analytical results[4], which show that for all sources of deoxyribonucleic acid examined the amount of adenine is close to the amount of thymine, and the amount of guanine close to the amount of cytosine, although the cross-ratio (the ratio of adenine to guanine) can vary from one source to another. Indeed, if the sequence of bases on one chain is irregular, it is difficult to explain these analytical

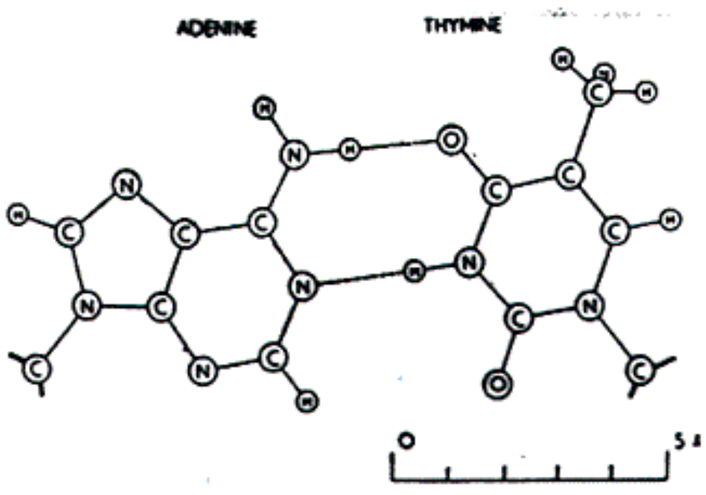


Fig. 4. Pairing of adenine and thymine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown

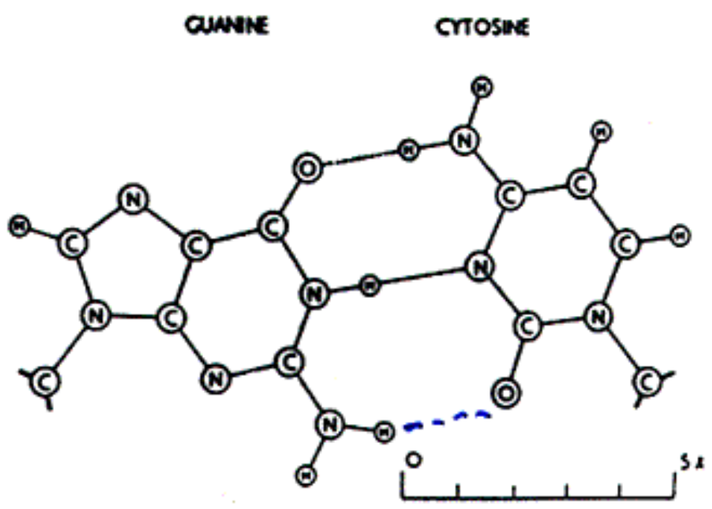
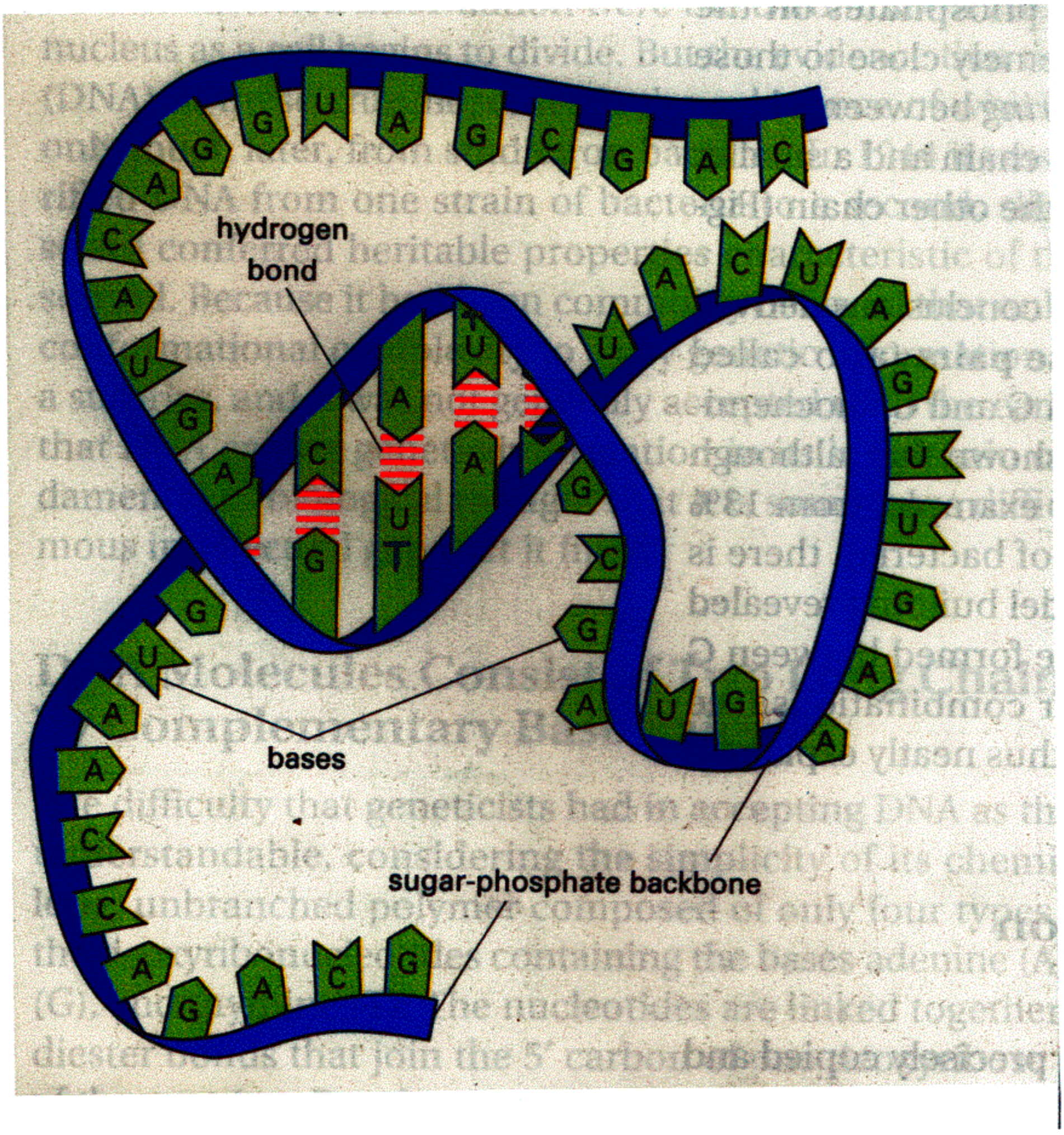


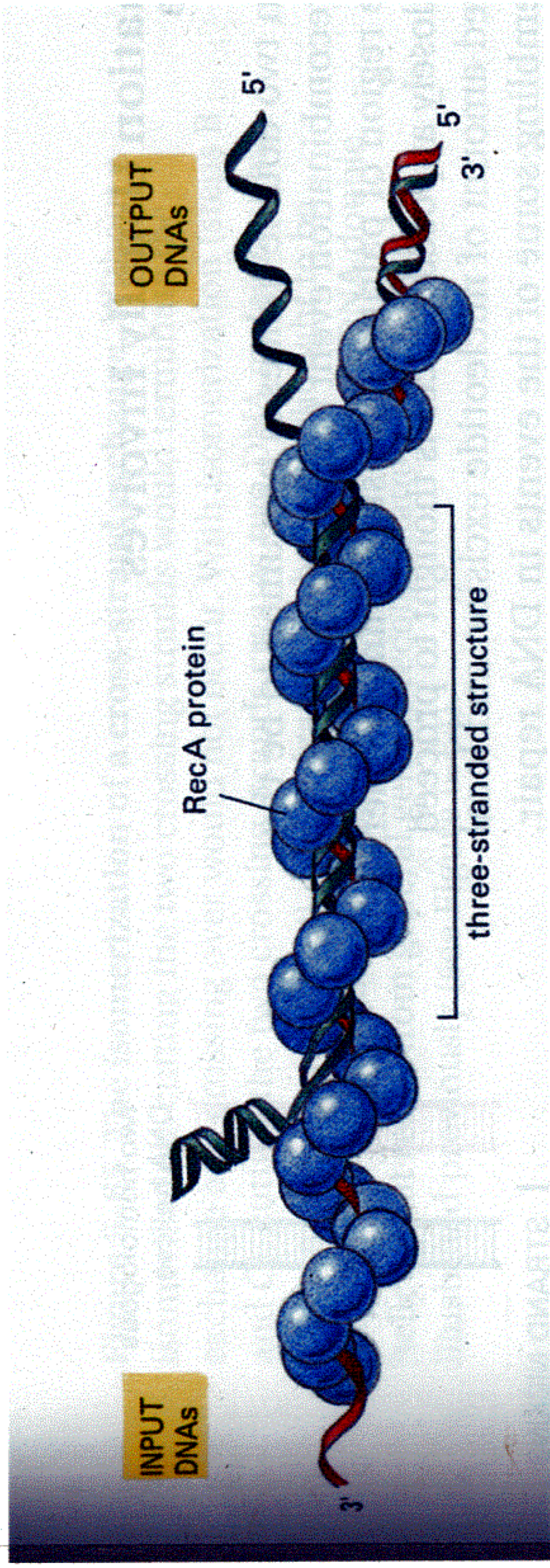
Fig. 5. Pairing of guanine and cytosine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown

bases on one of the pair of chains were given, one

[11] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

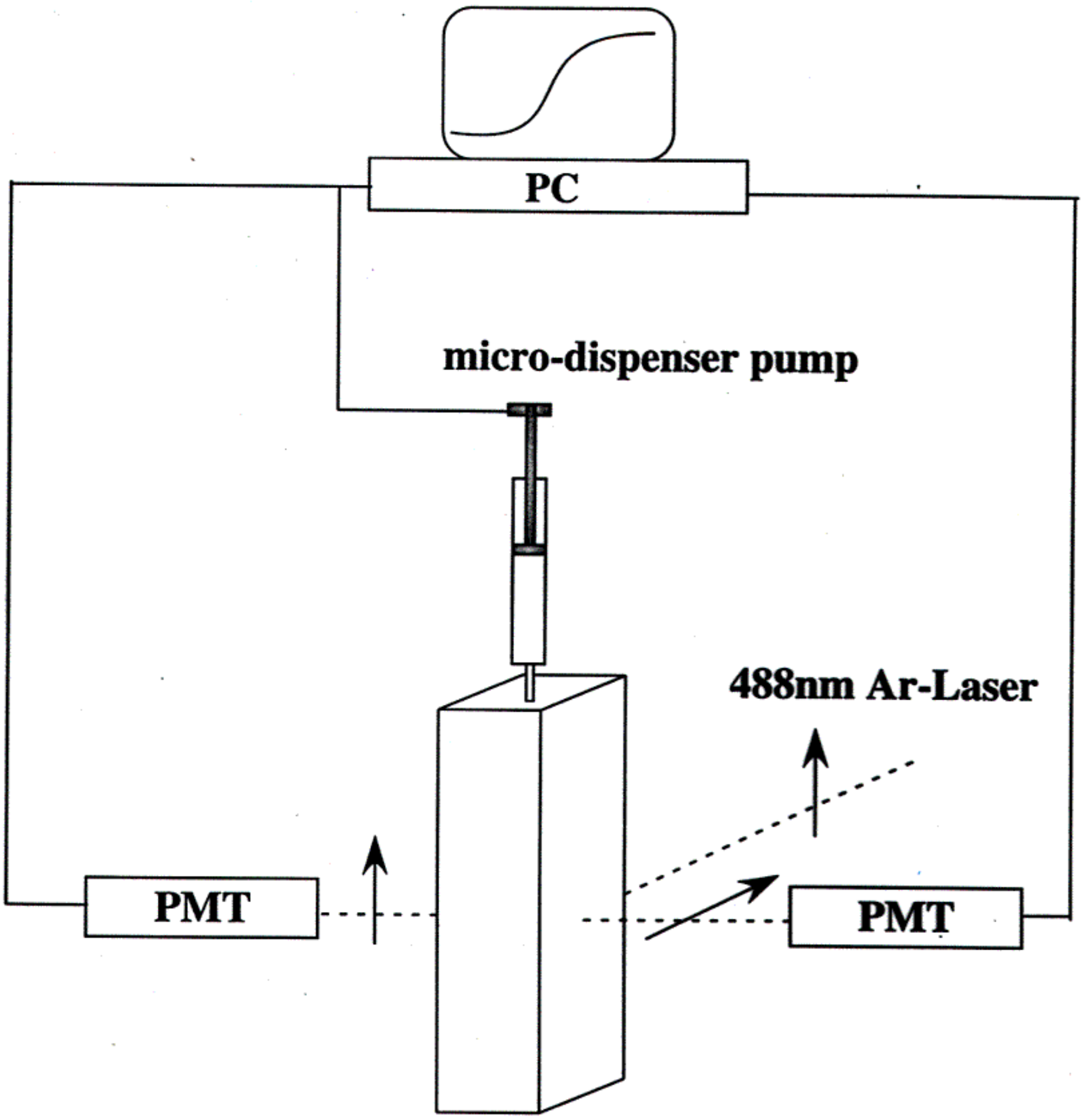


[12] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution



[13] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

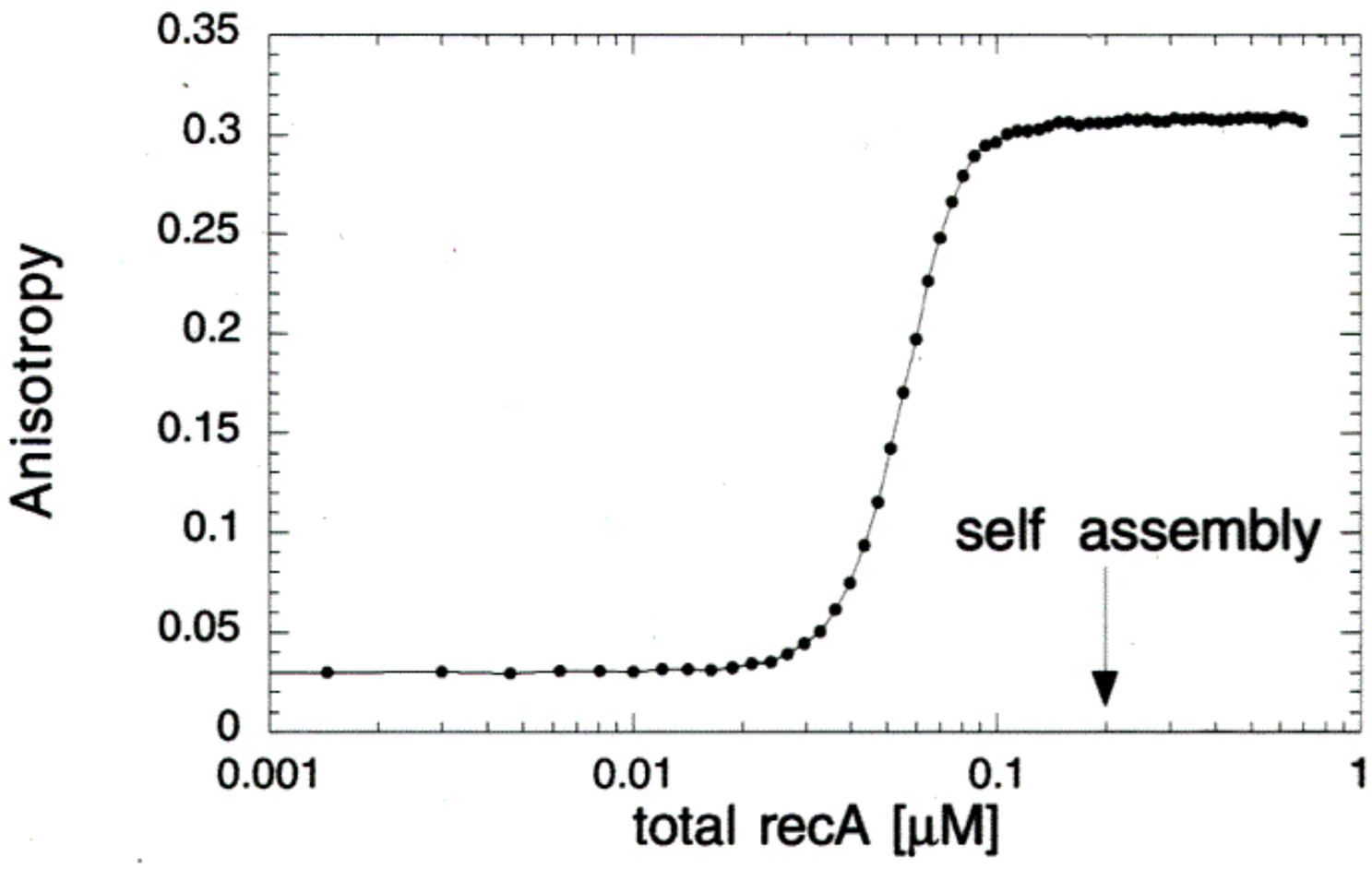
Measurement Setup

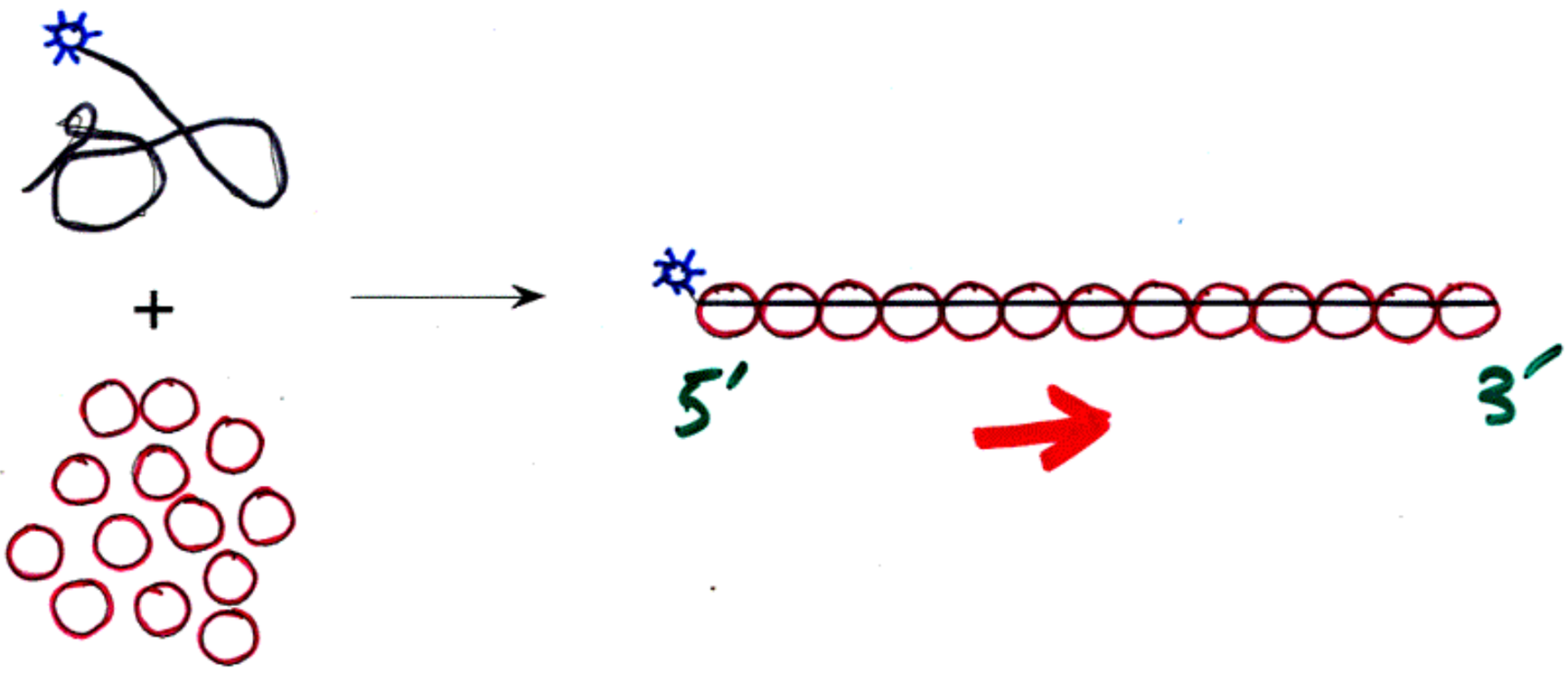


Condensation of recA-ATPγs on ssDNA

single stranded DNA 40 bases poly T, 20^{0}C

1nM T40-FL, 25mM Tris-HCl pH7.5, 150mM NaCl,
1mM $MgCl_2$, 1mM DTT, 1mM ATPγs

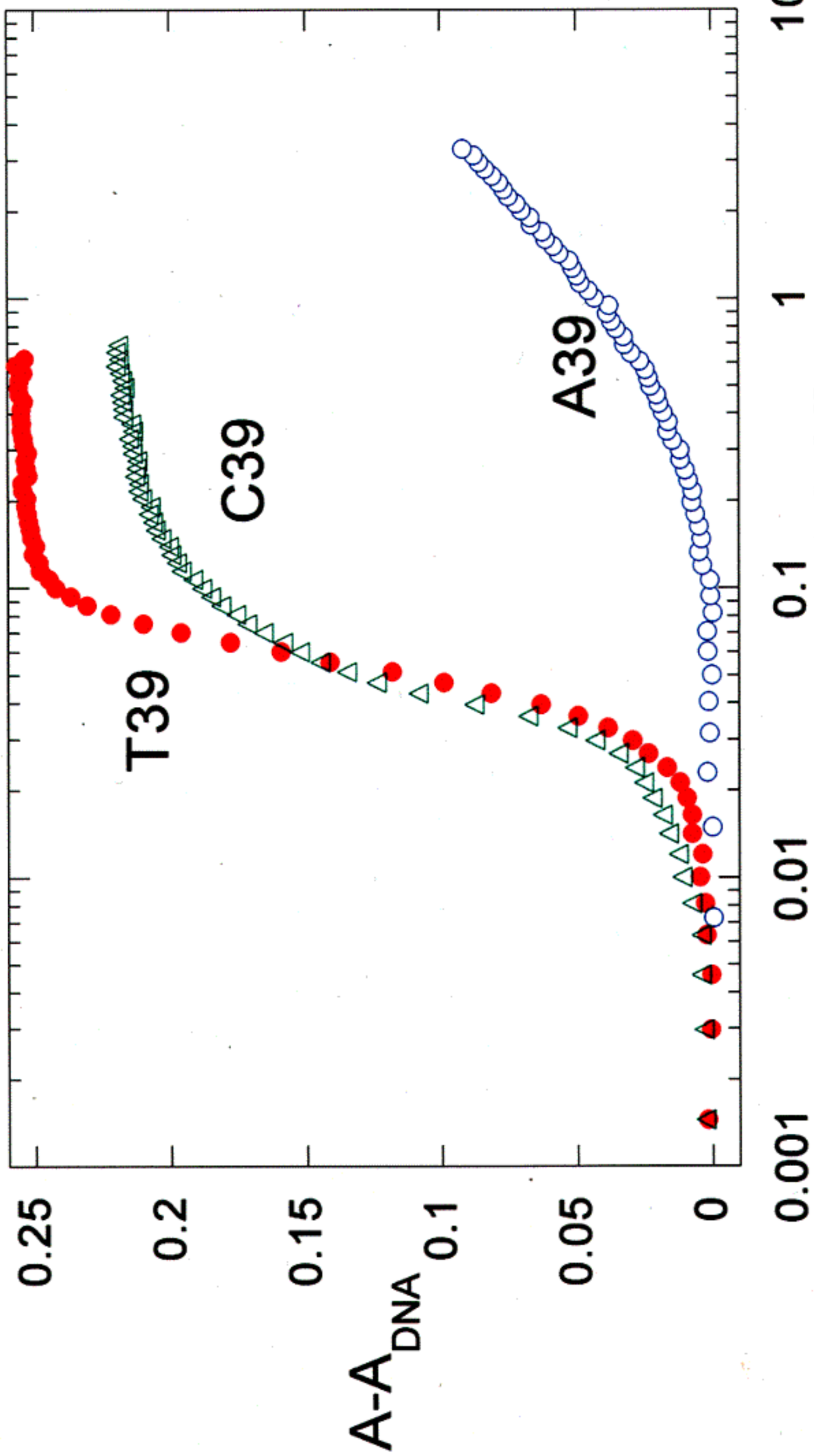




[15] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

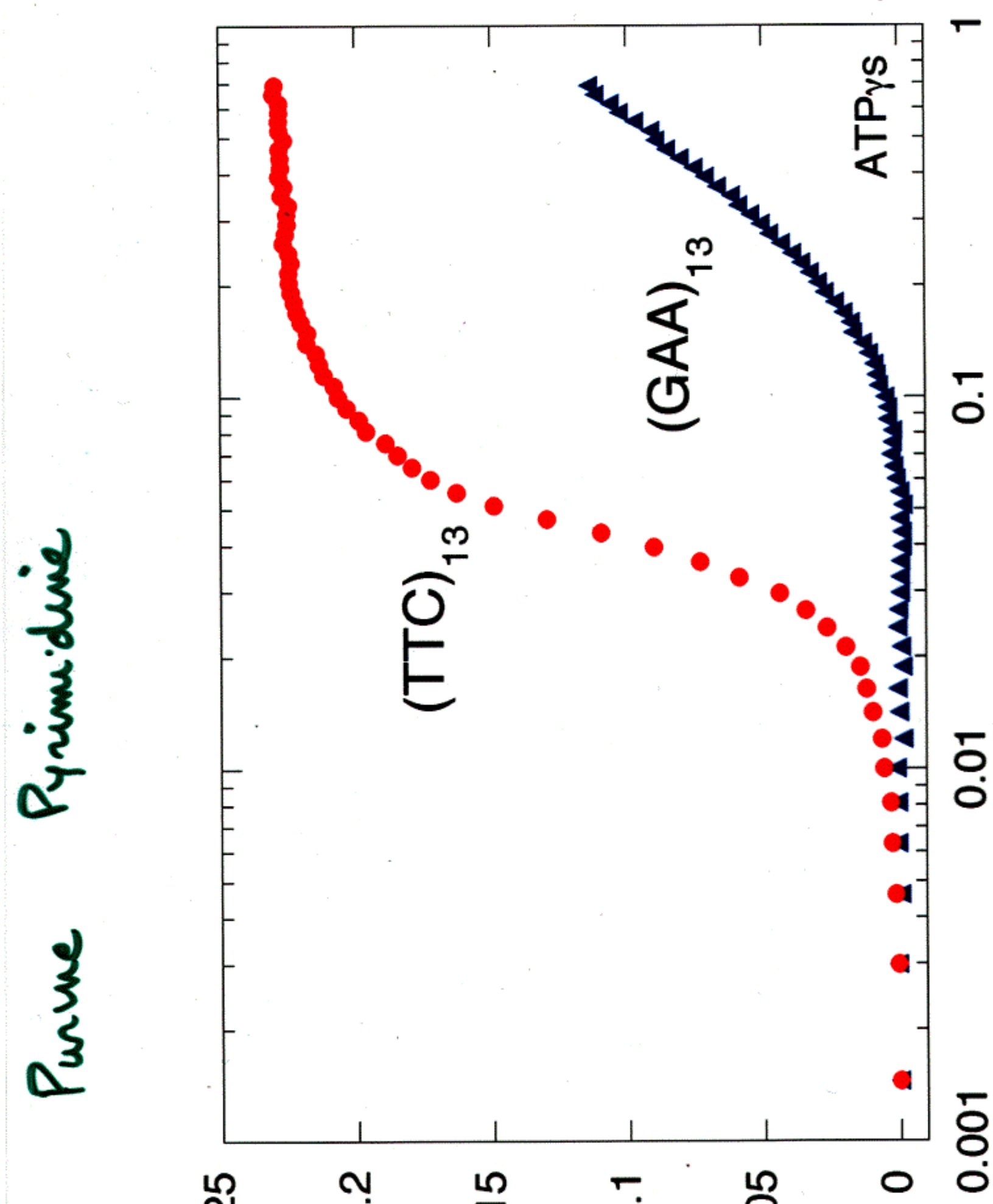
Stacking of Purines

Smaller



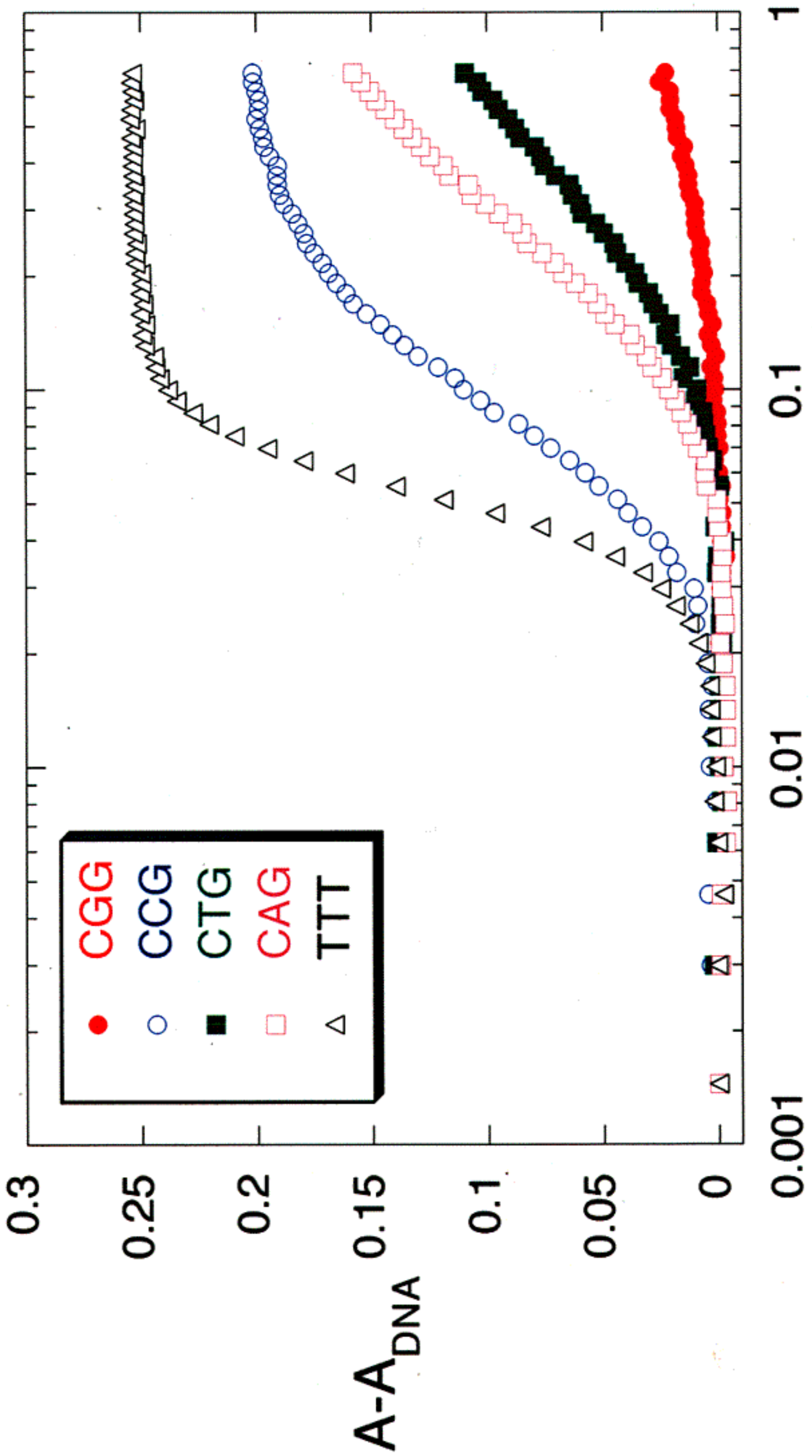
Biet, Sun, Dutreix
Nucl. Ac. Res. 33, 5981 (05)

[16] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution



[17] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

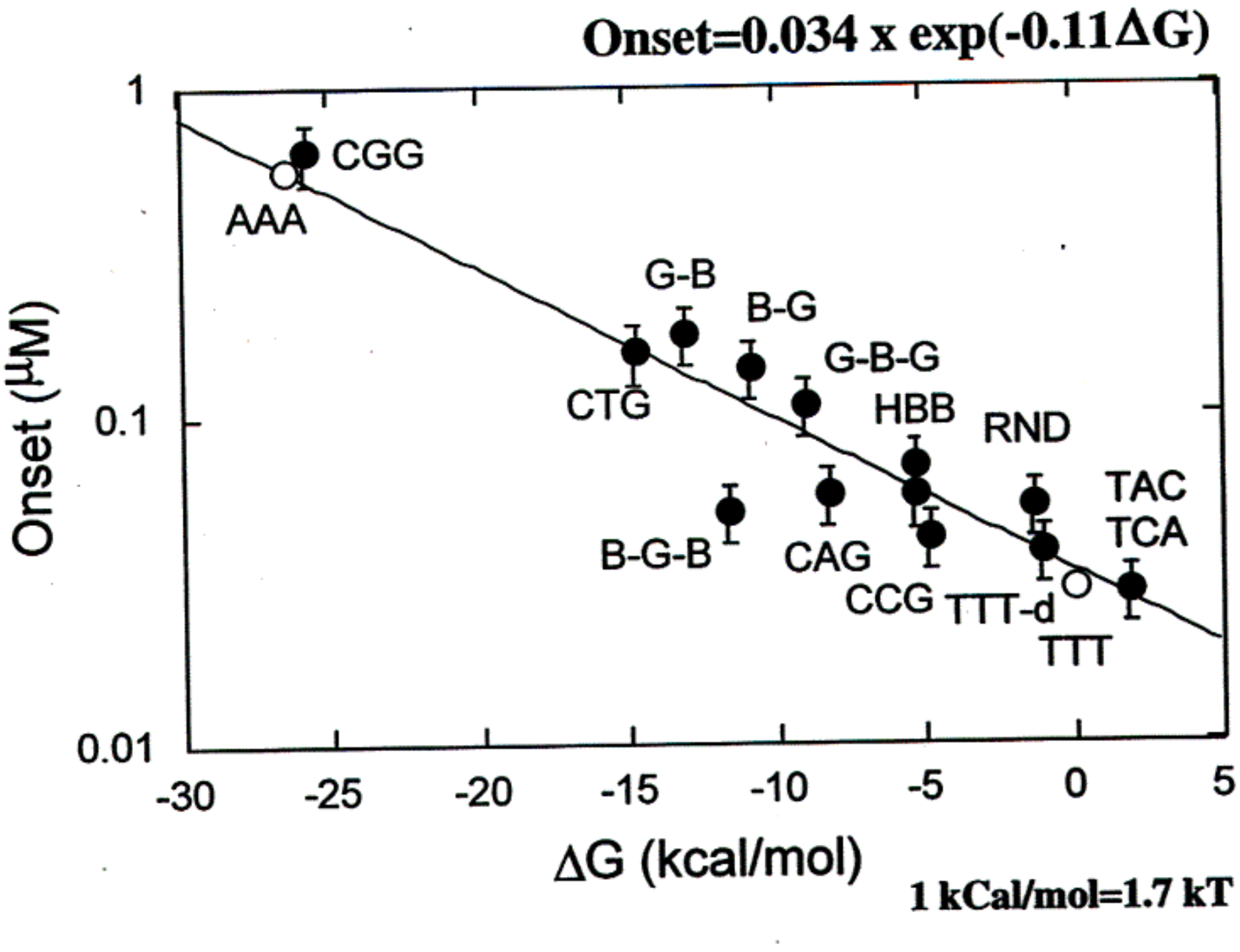
CNG secondary structure



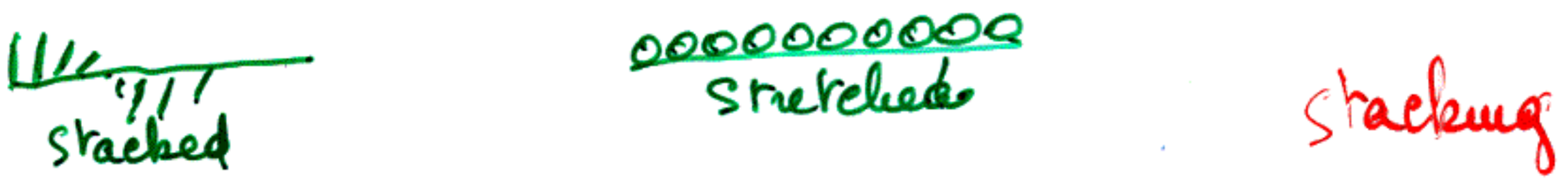
[18] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

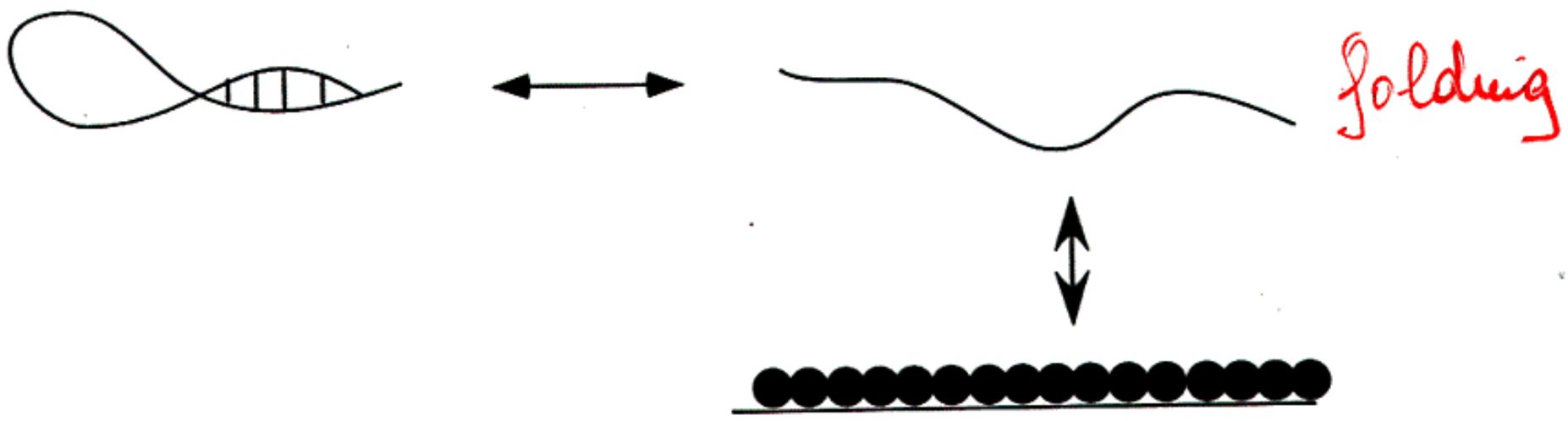
39 bases 13 x (codon)

DNA Folding and Stacking: A Nucleation Barrier for RecA Binding



1 kCal/mol=1.7 kT



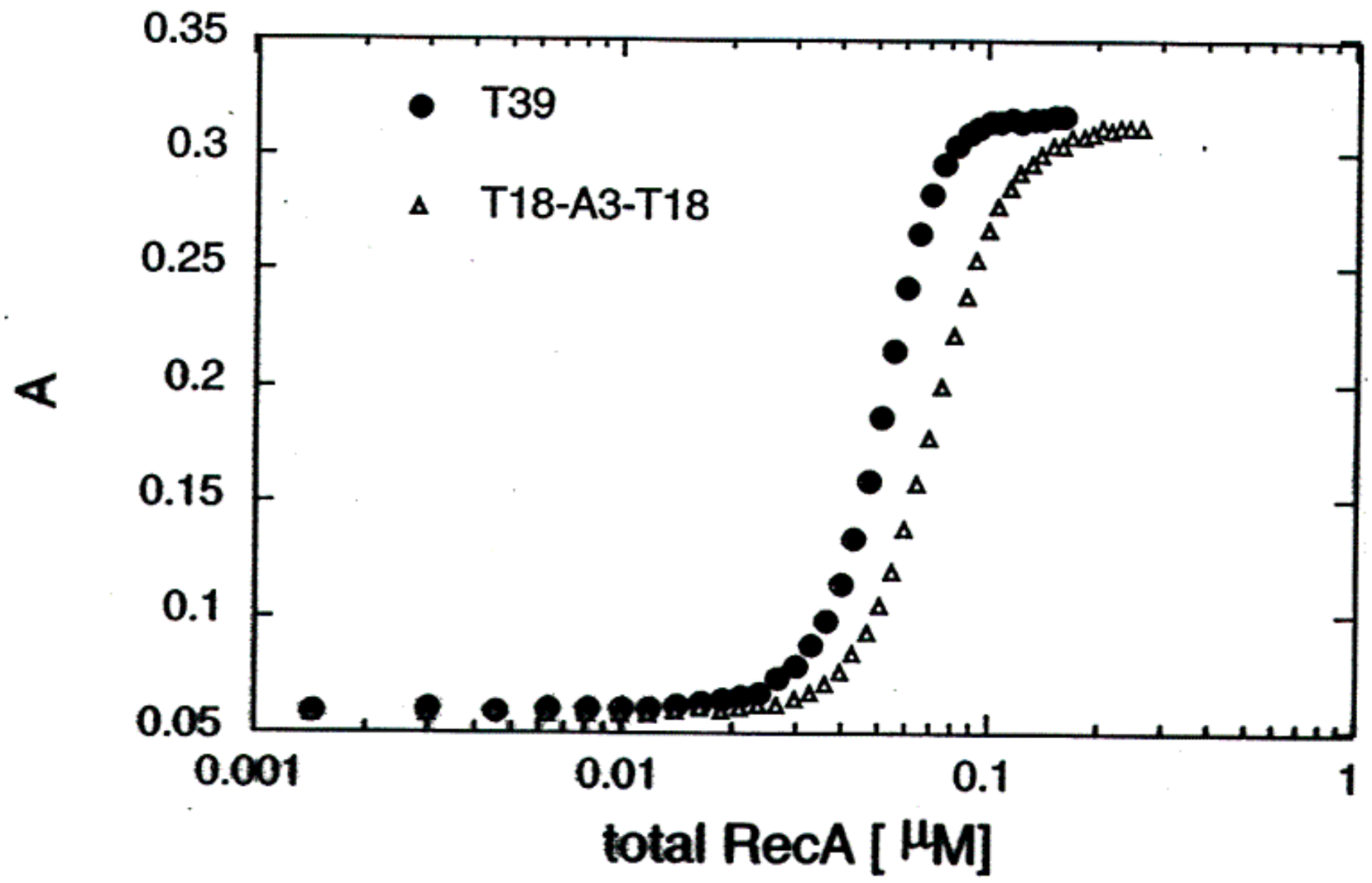


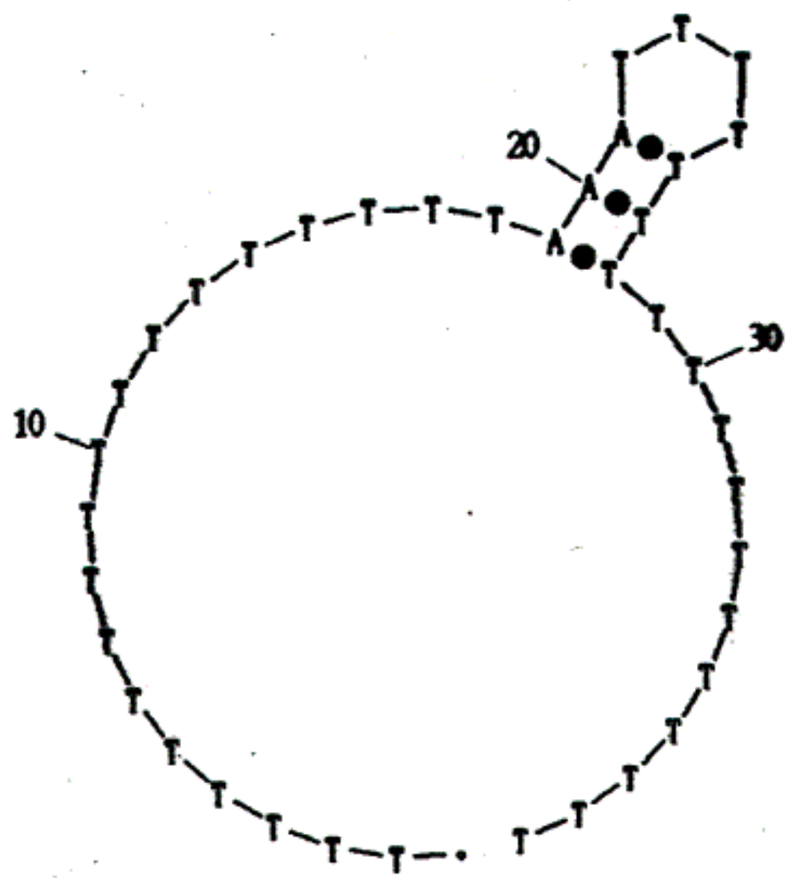
A G purine T C pyrimidine

A C amine G T keto.

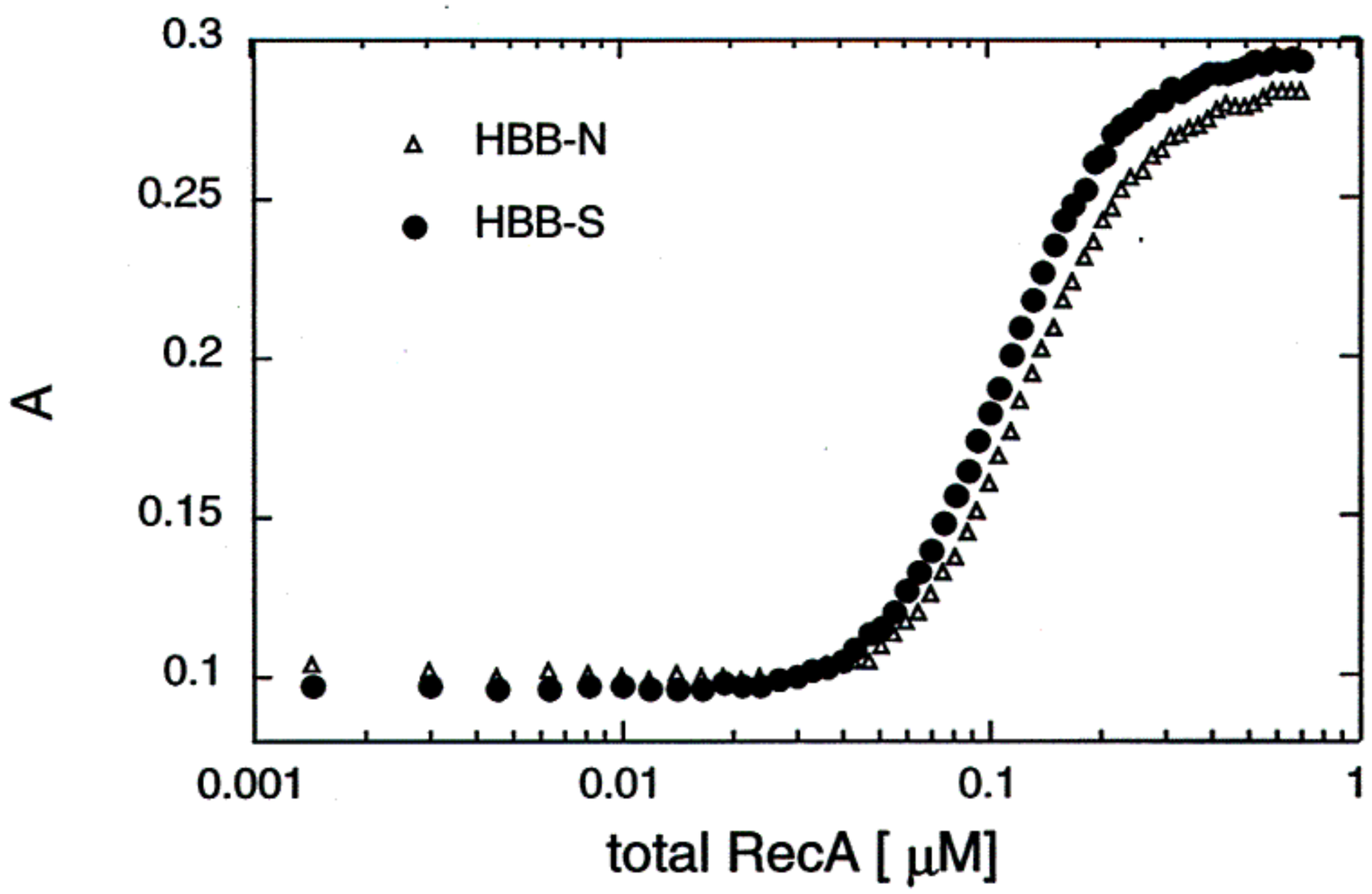
Effect of "Lattice Defects"

TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TTT TTT TTT TTT TTT TTT AAA TTT TTT TTT TTT TTT TTT





Single Mismatch Detection



Hemoglobin-Beta-Normal:

ATG GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT

Hemoglobin-Beta-Sickle:

ATG GTG CAC CTG ACT CCT GTG GAG AAG TCT GCC GTT ACT

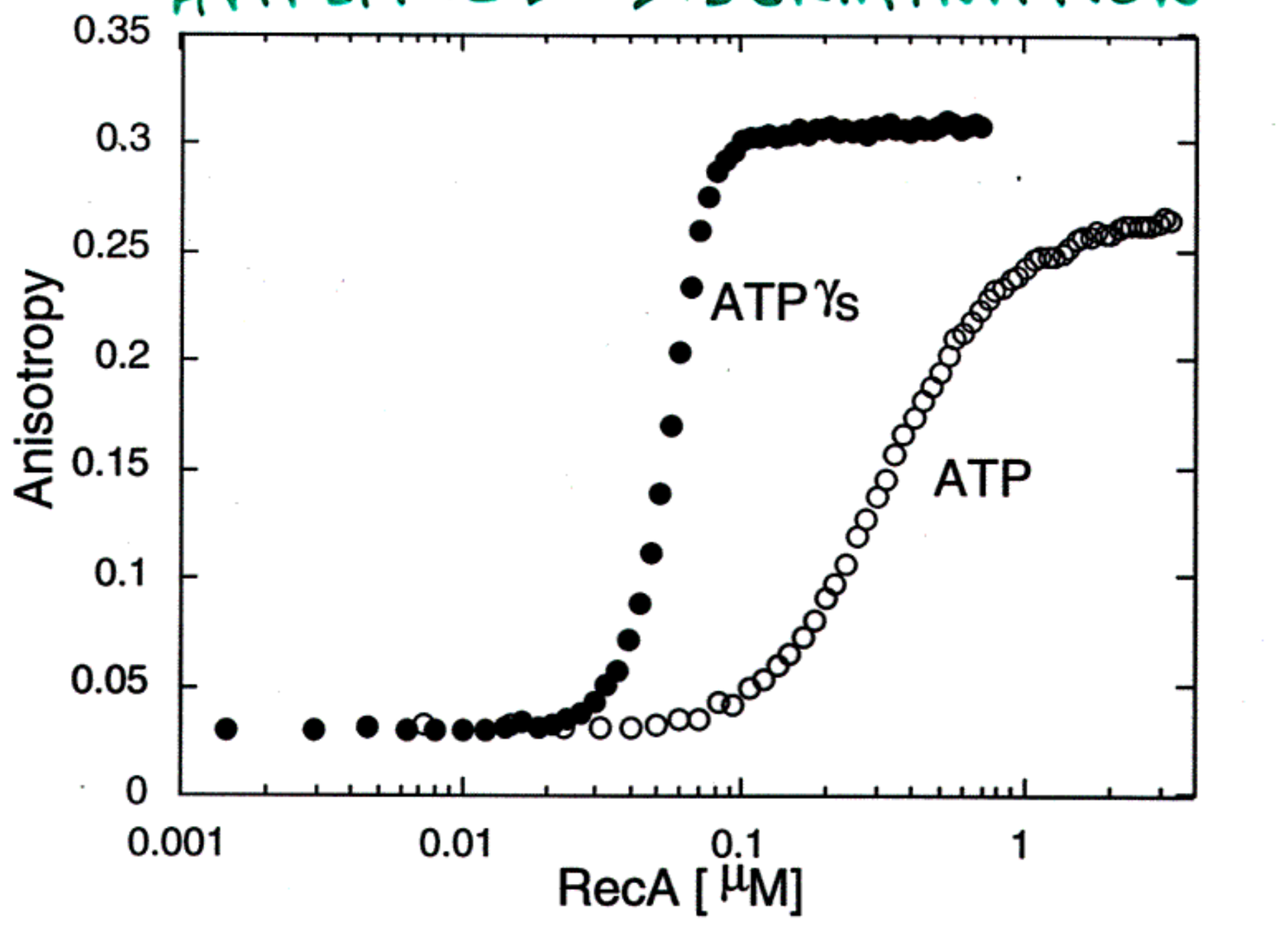
Glutamate --> Valine

can optimize discrimination signal to noise ratio with salt !

Condensation of RecA on ssDNA with ATP Hydrolysis: Assembly <--> Disassembly

Roy BAR ZIV Tsvi TLUSTY

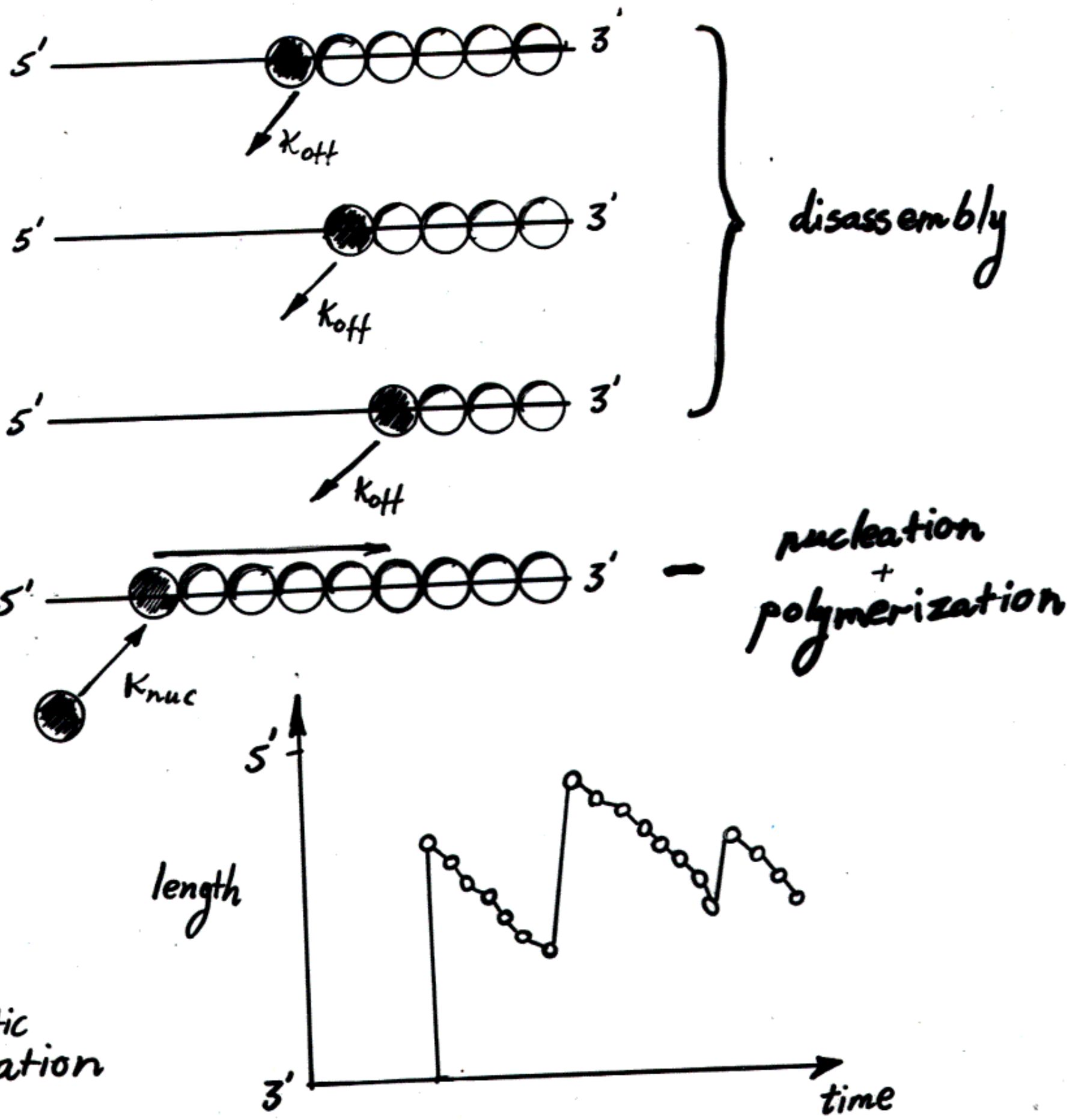
AMPLIFIED DISCRIMINATION



Arenson Tsodikov Cox JMB(99) 288 391-401

kinetics of end disassembly of RecA from ssDNA

ATP-driven assembly: Treadmill

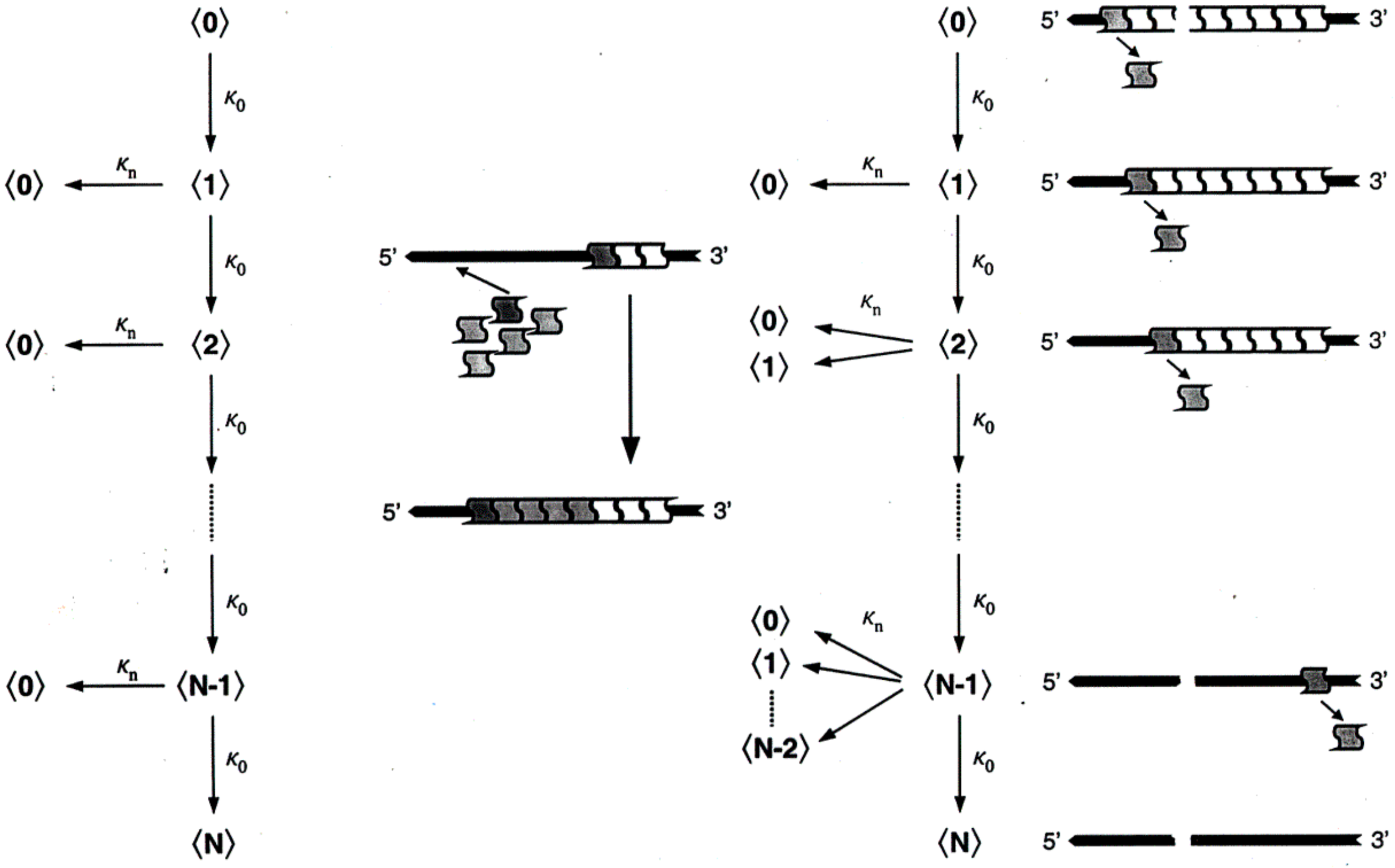


Stochastic simulation

irreversible cascade

J. Hopfield PNAS 71, 4135 (1974)
J. Nunio Biochimie 57, 587 (1975)

Gx JMB 288, 391 (1999)



THEORETICAL CONSIDERATIONS RESPECTING THE SEPARATION OF GASES BY DIFFUSION AND SIMILAR PROCESSES.

Lord Rayleigh

[*Philosophical Magazine*, XLII. pp. 493—498, 1896.]

THE larger part of the calculations which follow were made in connexion with experiments upon the concentration of argon from the atmosphere by the method of atmolysis*. When the supply of gas is limited, or when it is desired to concentrate the lighter ingredient, the conditions of the question are materially altered; but [illegible] problem which then presented itself of the simple diffusion of a gaseous mixture into a vacuum, with special regard to the composition of the residue. The diffusion tends to alter this composition in the first instance only in the neighbourhood of the porous walls; but it will be assumed that the forces promoting mixture are powerful enough to allow of our considering the composition to be uniform throughout the whole volume of the residue, and variable only with time, on account of the unequal escape of the constituent gases.

Let x, y denote the quantities of the two constituents of the residue at any time, so that $-dx, -dy$ are the quantities diffused out in time dt. The values of dx/dt, dy/dt will depend upon the character of the porous partition and upon the actual pressure; but for our present purpose it will suffice to express dy/dx, and this clearly involves only the ratios of the constituents and of their diffusion rates. Calling the diffusion rates μ, ν, we have

$$\frac{dy}{dx} = \frac{\nu y}{\mu x}. \quad \text{.....(1)}$$

In this equation x, y may be measured on any consistent system that may be convenient. The simplest case would be that in which the residue is maintained at a constant volume, when x, y might be taken to represent the partial pressures of the two gases. But the equation applies equally well when the volume changes, for example in such a way as to maintain the total pressure constant.

The integral of (1) is

$$y^{1/\nu} = Cx^{1/\mu}, \quad \text{.....} \quad (2)$$

where C is an arbitrary constant, or

$$y/x = Cx^{-1+\nu/\mu}. \quad \text{.....(3)}$$

If X, Y be simultaneous values of x, y, regarded as initial,

$$\frac{y/x}{Y/X} = \left(\frac{x}{X}\right)^{-1+\nu/\mu}, \quad \text{.....(4)}$$

so that

$$x = X\left(\frac{y/x}{Y/X}\right)^{\mu/(\nu-\mu)}. \quad \text{.....(5)}$$

In like manner

$$y = Y\left(\frac{x/y}{X/Y}\right)^{\nu/(\mu-\nu)}. \quad \text{.....(6)}$$

If we write

$$\frac{y/x}{Y/X} = r, \quad \text{.....(7)}$$

r represents the enrichment of the residue as regards the second constituent, and we have from (5), (6),

$$\frac{x+y}{X+Y} = \frac{X}{X+Y}r^{\mu/(\nu-\mu)} + \frac{Y}{X+Y}r^{\nu/(\nu-\mu)}, \quad \text{.....(8)}$$

an equation which exhibits the relation between the enrichment and the ratio of the initial and final total quantities of the mixture.

[25] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

distillation

Returning now to the separation of gases by diffusion into a vacuum, let us suppose that the difference between the gases is small, so that $(\nu-\mu)/\mu=\kappa$, a small quantity, and that at each operation one-half the total volume of the mixture is allowed to pass. In this case (8) becomes

$$\tfrac{1}{2}=\frac{X}{X+Y}r^{\frac{1}{\kappa}}+\frac{Y}{X+Y}r^{\frac{1+\kappa}{\kappa}}=r^{\frac{1}{\kappa}}\text{ nearly;}$$

so that

$$r=\frac{y/Y}{x/X}=(\tfrac{1}{2})^{\kappa}. \qquad \text{.....(15)}$$

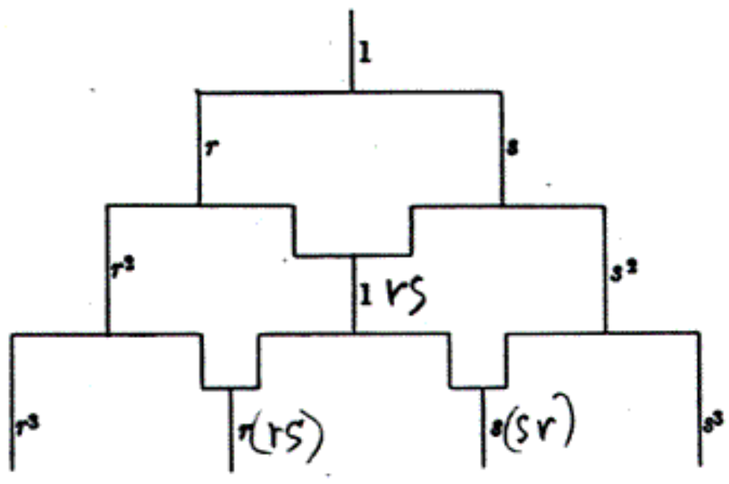
This gives the effect of the operation in question upon the composition of the residual gas. If s denote the corresponding symbol for the transmitted gas, we have

$$s=\frac{(Y-y)/Y}{(X-x)/X}=\frac{1-y/Y}{1-x/X}=\frac{1-rx/X}{1-x/X}=1+\frac{(1-r)x/X}{1-x/X}=2-r$$

approximately, since r is nearly equal to unity. Accordingly

$$\frac{1}{s}=\frac{1}{2-r}=r\text{ nearly,}$$

so that approximately s and r are reciprocal operations. For example, if starting with any proportions we collect the transmitted half, and submit it to another operation of the same sort, retaining the half not transmitted, the final composition corresponding to the operations sr is the same (approximately) as the composition with which we started, and the same also as would be obtained by operations taken in the reverse order, represented by rs. A complete scheme* on these lines is indicated in the diagram.



Representing the initial condition by unity, we may represent the result of the first operation by

$$\tfrac{1}{2}r+\tfrac{1}{2}s,\quad\text{or}\quad\tfrac{1}{2}(r+s),$$

in which the numerical coefficient gives the quantity of gas whose character is specified by the literal symbols. The second set of operations gives in the first instance

$$\tfrac{1}{4}r^2+\tfrac{1}{4}sr+\tfrac{1}{4}rs+\tfrac{1}{4}s^2,$$

or, after admixture of the second and third terms (which are of the same quality),

$$\tfrac{1}{4}(r^2+2rs+s^2)=\left(\frac{r+s}{2}\right)^2.$$

In like manner the result of the third set of operations may be represented by $\left(\frac{r+s}{2}\right)^3$, and (as may be formally proved by "induction") of n sets of operations by

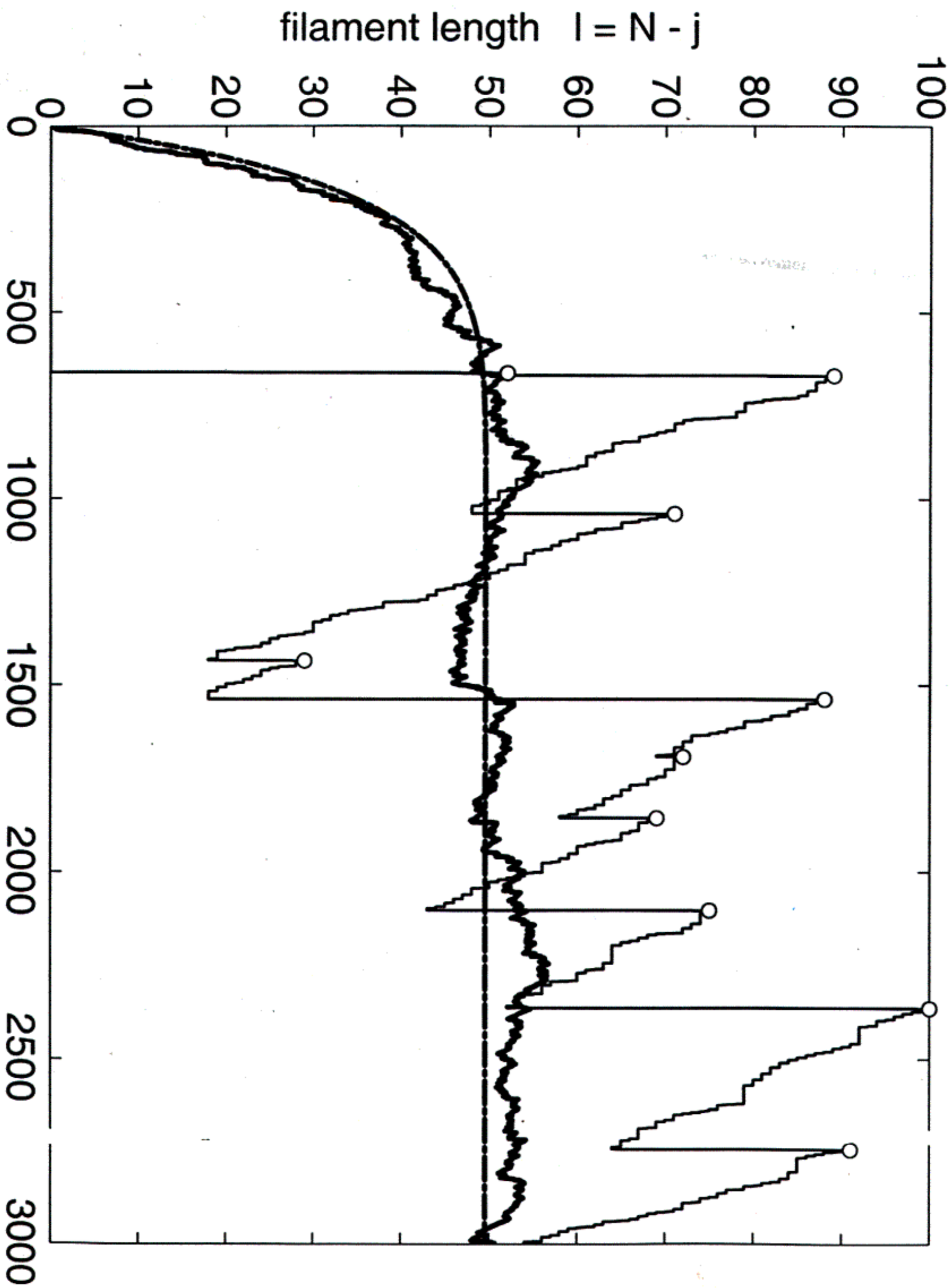
$$\left(\frac{r+s}{2}\right)^n. \qquad \text{.....(16)}$$

When we take account of the reciprocal character of r and s, this may be written

$$\frac{1}{2^n}\left\{r^n+nr^{n-2}+\frac{n(n-1)}{1.2}r^{n-4}+\ldots+nr^{-n+2}+r^{-n}\right\}, \qquad \text{.....(17)}$$

the number of parts into which the original quantity of gas is divided being

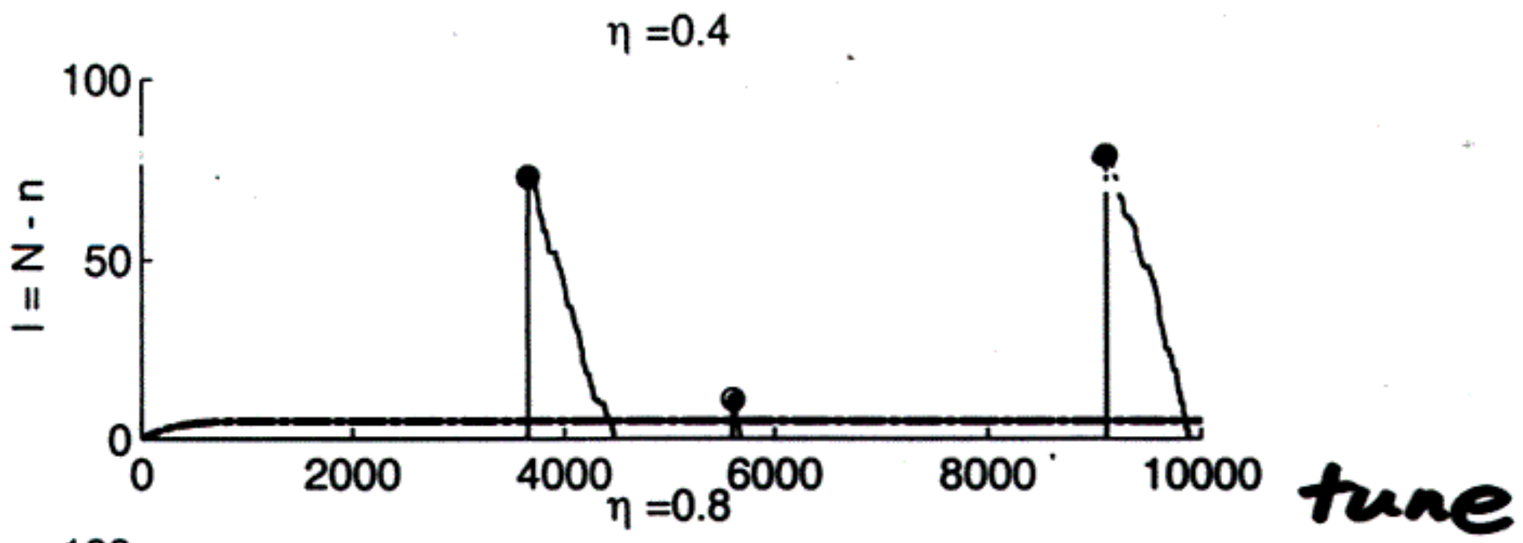
[26] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution



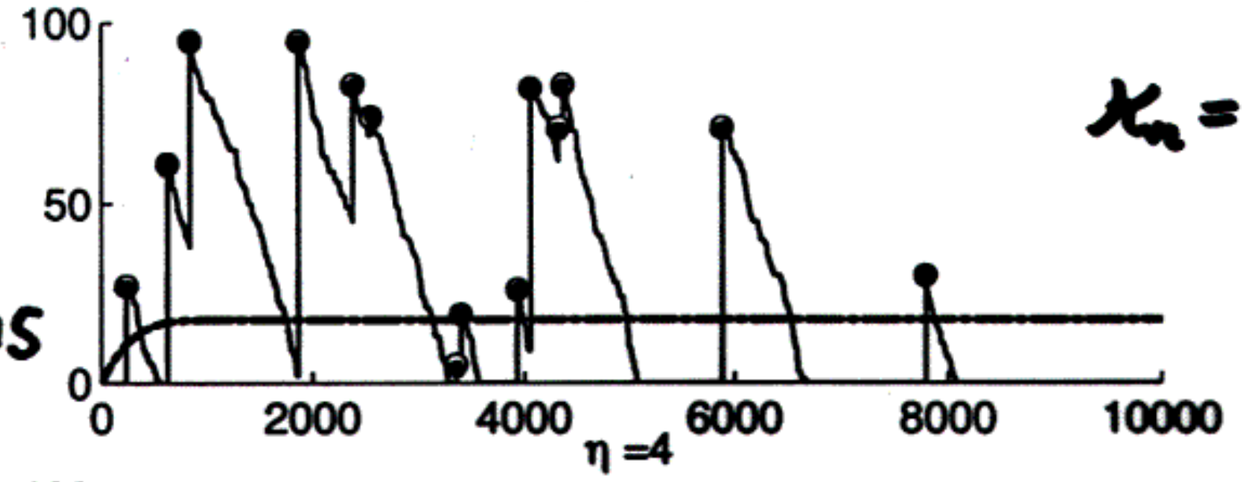
[27] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution

Assembly Fluctuations

onset

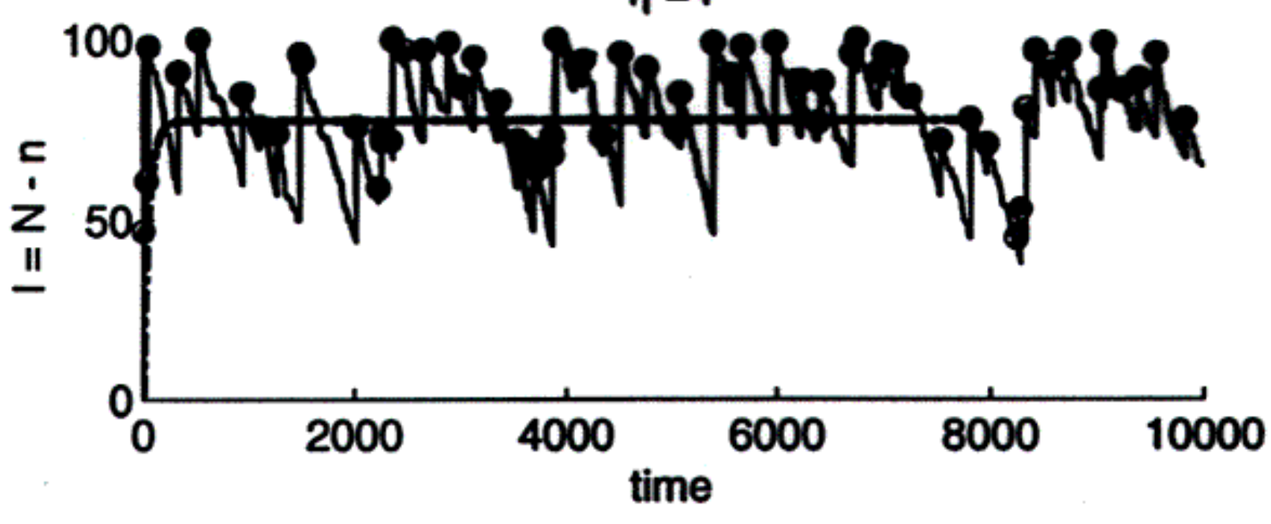


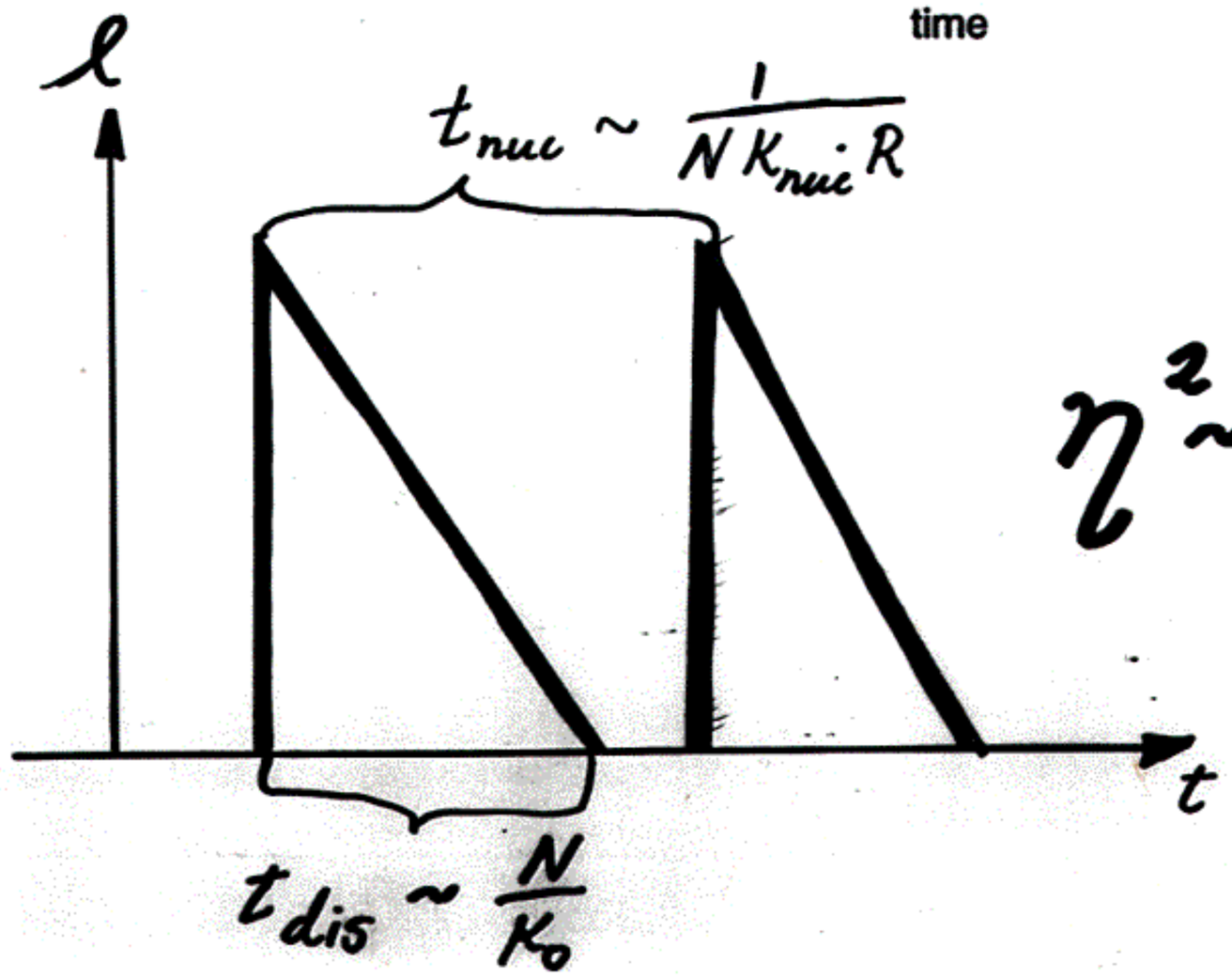
Strong fluctuations



$\kappa_n = \kappa_{nuc} R$

Saturation





$$\eta^2 \sim \frac{t_{dis}}{t_{nuc}} \sim \frac{K_{nuc} R}{K_0} N$$

$P(n,t)$ of RecA filament of length $N-n$ at time t.

$$\frac{\partial P}{\partial t} = -K_0 \frac{\partial P}{\partial n} - K_{NUC} R_n P$$

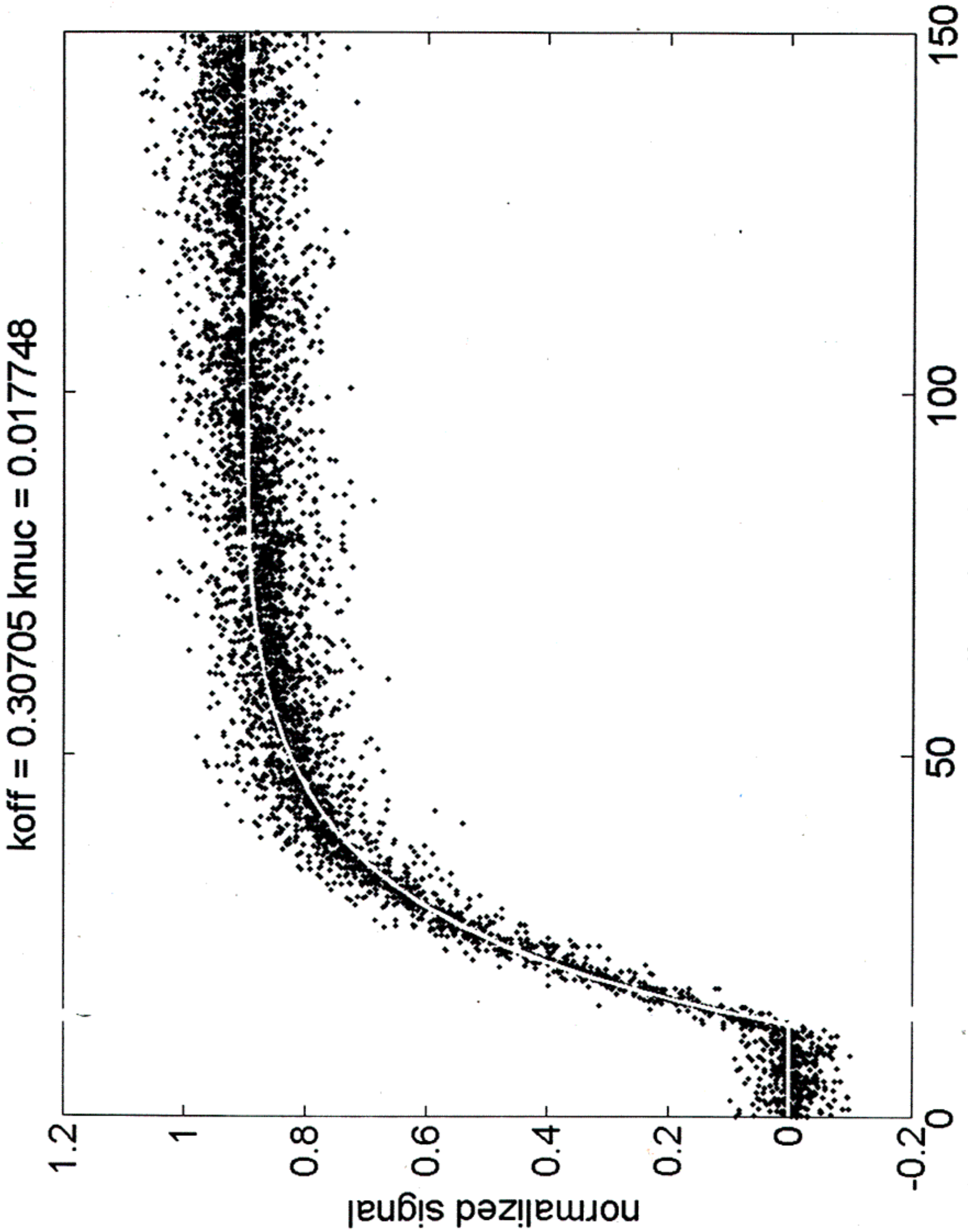
rates (sec^{-1}) $\frac{K_0}{n}$ $K_{nuc} R_n$

optimum when filament empties between nucleation time : maximum fluctuation

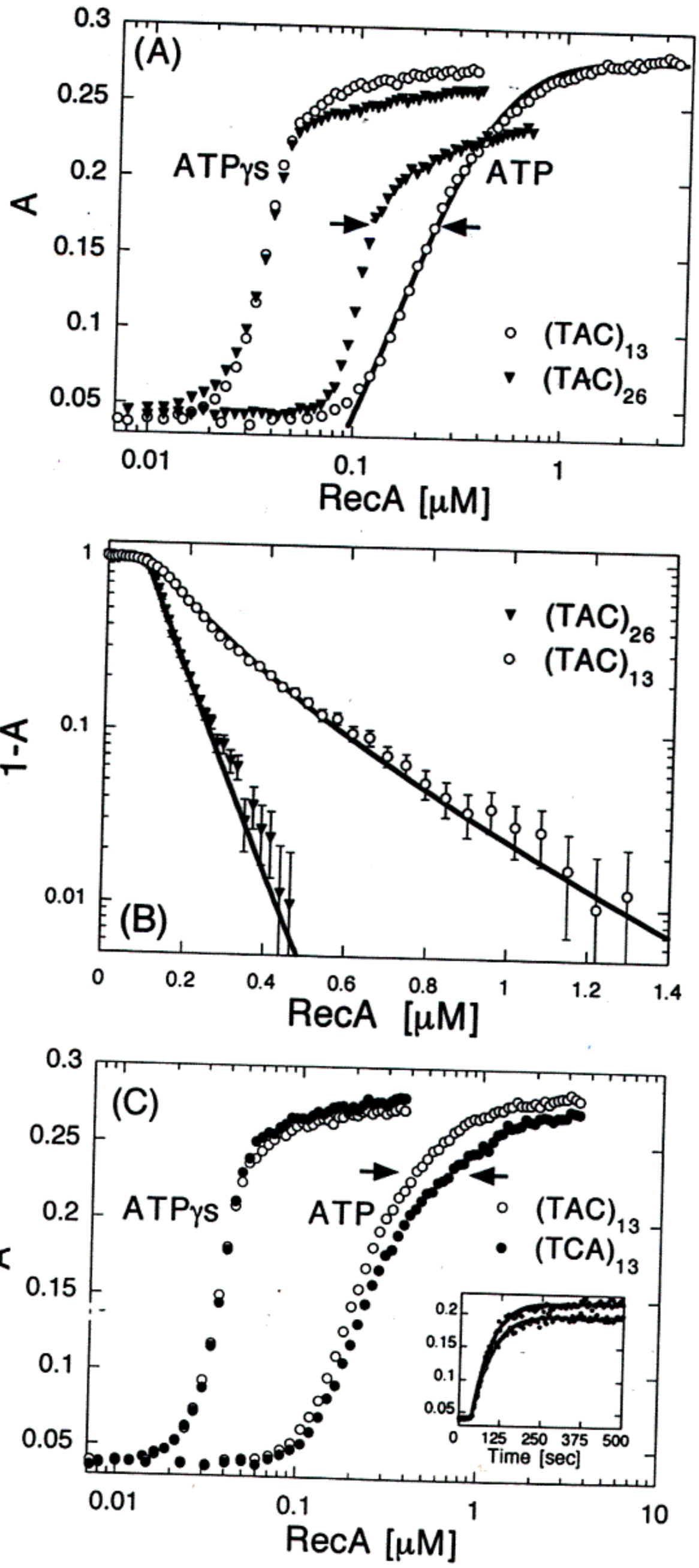
$$n^2 \left(\frac{K_{NUC}}{K_0}\right) R \sim 1$$

steady state $P(n) \simeq e^{-\frac{K_{NUC} R}{K_0} n^2}$

[29] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution



[30] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution



[31] Dr. A. Libchaber, Rockefeller U. (ITP 4/4/01) Molecular Evolution