

# Measuring the Formaldehyde Protein–DNA Cross-Link Reversal Rate

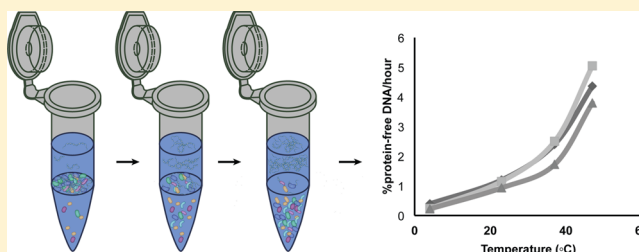
Julia Kennedy-Darling<sup>†</sup> and Lloyd M. Smith<sup>\*,†,‡</sup>

<sup>†</sup>Department of Chemistry, University of Wisconsin, 1101 University Avenue, Madison, Wisconsin 53706, United States

<sup>‡</sup>Genome Center of Wisconsin, Madison, Wisconsin 53706, United States

## S Supporting Information

**ABSTRACT:** Protein–DNA binding interactions play critical roles in important cellular processes such as gene expression, cell division, and chromosomal organization. Techniques to identify and characterize these interactions often utilize formaldehyde cross-linking for stabilization of the complexes. Advantages of formaldehyde as a cross-linking reagent include cell permeability, relatively fast cross-linking kinetics, and short cross-linker length. In addition, formaldehyde cross-links are reversible, which has the advantage of allowing complexes to be dissociated if desired but may also present a problem if undesired dissociation occurs in the course of an experiment. While the kinetics of formaldehyde cross-link formation have been well-established in numerous studies, there have been no reports of the rate of cross-link dissociation, even though it is clearly a critical variable when developing a biochemical protocol involving formaldehyde cross-linking. We present here a method for measurement of the rate of formaldehyde cross-link reversal based upon the Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) procedure and use it to determine the rate of cross-link reversal for cross-linked protein–DNA complexes from yeast cell lysate. The half-life of the protein–DNA cross-links varies from 179 h at 4 °C to 11.3 h at 47 °C, with a rate that increases exponentially with temperature and is independent of salt concentration.



Protein–DNA interactions are central to cellular function and vary widely with respect to parameters such as their binding affinity, binding specificity, and biological role. Histone proteins, for example, serve as structural DNA elements, binding ubiquitously throughout the genome with little sequence-based discrimination,<sup>1</sup> while transcription factors, another class of DNA-binding proteins, selectively bind to specific DNA sequences within the genome where they help to control the expression of target loci.<sup>2,3</sup> A variety of tools have been developed to characterize protein–DNA interactions in the cell including chromatin immunoprecipitation (ChIP),<sup>4</sup> formaldehyde-assisted isolation of regulatory elements (FAIRE),<sup>5,6</sup> proteomics of isolated chromatin segments (PiCH),<sup>7</sup> chromatin-affinity purification with mass spectrometry (ChAP-MS),<sup>8</sup> and hybridization capture of chromatin associated proteins for proteomics (HyCCAPP).<sup>9</sup> These technologies all utilize formaldehyde as a cross-linking reagent to stabilize native protein–DNA complexes in the cell.

While the kinetics of formaldehyde cross-link formation have been well-established in numerous studies,<sup>4,10,11</sup> there have been no reports of the rate of cross-link dissociation, even though it is clearly a critical variable when developing a biochemical protocol involving formaldehyde cross-linking. We present here a method for measurement of the rate of formaldehyde cross-link reversal based upon the FAIRE procedure and use it to determine the rate of cross-link reversal for cross-linked protein–DNA complexes from yeast cell lysate as a function of temperature and salt concentration. The methodology and measured rates of cross-link reversal are

useful in the design of studies involving formaldehyde cross-linking.

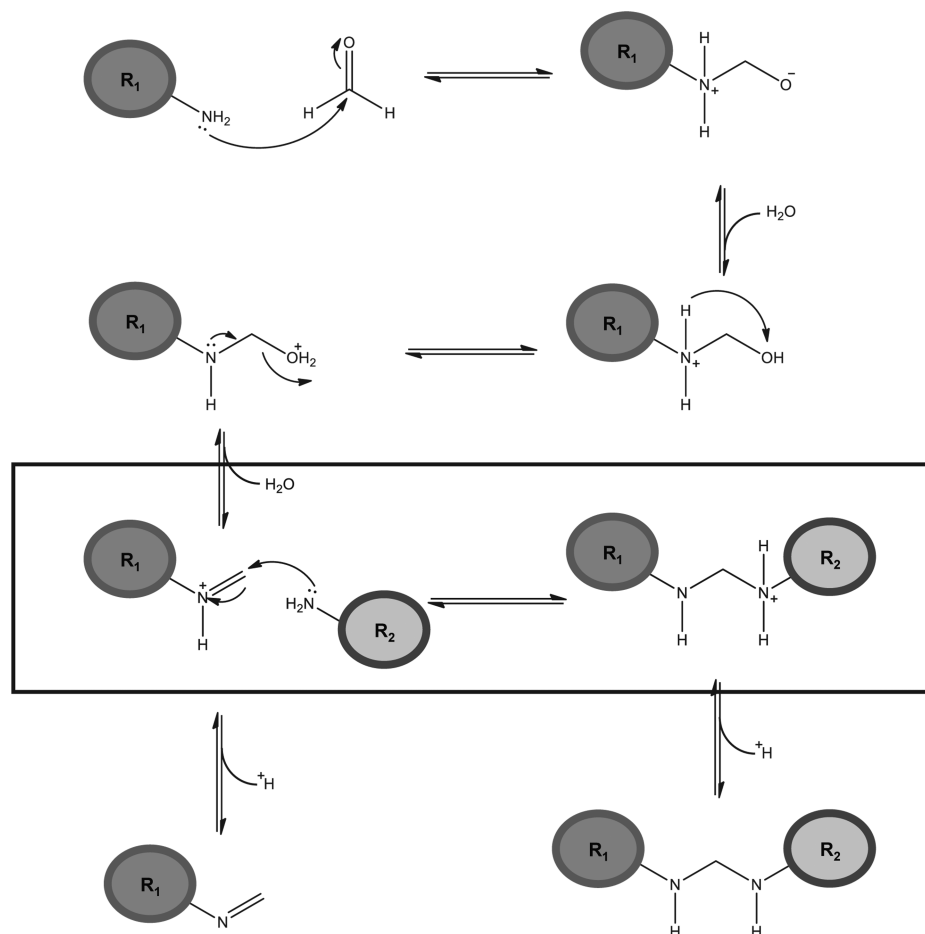
## EXPERIMENTAL SECTION

**Materials.** *Saccharomyces cerevisiae* strain Y1788 was obtained from Professor David Mitchell (University of Texas). Yeast extract peptone dextrose (Y1375, abbreviated YPD), 37% formaldehyde (F38775), phenol–chloroform–isoamyl solution 25:24:1 (77617, abbreviated phenol–chloroform), and protease inhibitors for fungal growth (P8215) were purchased from Sigma-Aldrich Co. (St. Louis, MO). The 20% sodium dodecyl sulfate (SDS) (161-0418) was purchased from Bio-Rad (Hercules, CA). The 10× phosphate buffered saline (PBS) (P0191), 5 M Tris pH = 8 (T5581), 1 M Tris pH = 7 (T1070), 5 M sodium chloride (NaCl) (S0250), and 500 mM tetraethylenediaminetetraacetic acid (EDTA) (E0307-06) were purchased from Teknova (Hollister, CA). RNaseA (12091-039) was purchased from Life Technologies (Carlsbad, CA). qPCR probes were ordered from Integrated DNA Technologies (Coralville, IA). The 96-well plates (04729692001) and Master Mix solutions (0470749001) for qPCR were purchased from Roche USA (Nutley, NJ). Proteinase K (P8107S) was purchased from New England Biolabs (Ipswich, MA).

Received: April 14, 2014

Accepted: May 21, 2014

Published: May 21, 2014



**Figure 1.** Formaldehyde protein–DNA cross-link/reversal mechanism between Lys and dG. All steps in the reaction mechanism are reversible. Lysine-containing proteins ( $R_1$ ) initially attack the electrophilic carbon of formaldehyde. After loss of water, the resultant Schiff base is susceptible to nucleophilic attack by the exocyclic amine of deoxyguanosine residues ( $R_2$ ) to form an amination linkage between protein ( $R_1$ ) and DNA ( $R_2$ ) components. Both the cross-linking and cross-link reversal steps are demarcated by the rectangle.

**Cell Lysate Preparation.** Yeast cells were grown to saturation in 5 mL of YPD overnight at 30 °C and shaken at 200 rpm in an Amerex 747 shaker/incubator. The cells were diluted into 1.5 L of YPD and grown to an  $OD_{600} \sim 2.0$  as measured using an Agilent 8453 UV–vis spectrophotometer. Formaldehyde was added to a final concentration just under 3% (122 mL) and incubated for 30 min at room temperature, after which unreacted formaldehyde was quenched with 250 mL of 5 M Tris. The cells were collected using an Avanti J-25I centrifuge at 5 000g for 20 min. The cell pellet was washed once with 1× PBS and either used right away or stored at –80 °C. Cells from 500 mL of culture ( $\sim 2$ –3 mL cell pellet) were resuspended in 50 mL of lysis buffer (20 mM EDTA, 200 mM NaCl, 50 mM Tris pH7, and protease inhibitors (1/200 from stock)). The cells were lysed at 30 kpsi using a Constant Systems TS Series Cell Disruptor. SDS was added to the lysate solution to a final concentration of 1% and the lysate was incubated at 65 °C for 5 min. The cross-linked chromatin was sonicated in 50 mL volumes using a MisoniX Ultrasonic Processor S4000 at 20 V for a total of 3.5 min with alternating intervals of 4 s on and 4 s off. The sample was then centrifuged at 8 000g for 12 min to separate the cellular debris from the soluble chromatin. The supernatant was removed from the pellet and diluted 5-fold into lysis buffer to decrease the total SDS concentration to 0.2%. RNaseA was added to a final

concentration of 60  $\mu\text{g/mL}$ , and the solution was shaken at 150 rpm at 37 °C for 60 min.

**Phenol–Chloroform Extraction to Purify Protein-Free DNA.** Samples of cell lysate (10 mL) were used for each formaldehyde cross-linking reversal measurement. Aliquots (700  $\mu\text{L}$ ) were removed from each sample prior to incubation to provide a reference sample for determination of the total amount of DNA present (referred to as input samples). Ten units of proteinase K were added to each of these aliquots followed by overnight incubation at 37 °C in order to degrade all proteins present and release the free DNA. The samples were incubated at 4, 23, 37, or 47 °C, and three additional 700  $\mu\text{L}$  aliquots were removed at each time point (0, 2.5, 5, 10, and 20 h). Phenol–chloroform (700  $\mu\text{L}$ ) was added to each aliquot followed by vigorous vortexing for 30 s. The samples were centrifuged at maximum speed (14 000g) for 5 min in a Eppendorf 5417R centrifuge, after which the organic and aqueous layers are completely separated. A volume of 500  $\mu\text{L}$  of the aqueous (upper) layer was removed and placed in a new eppendorf tube, followed by addition of a second 500  $\mu\text{L}$  portion of phenol–chloroform. Vortexing and centrifugation were repeated, 200  $\mu\text{L}$  of the aqueous phase was removed, and 200  $\mu\text{L}$  of the phenol–chloroform was added and the vortexing/centrifugation steps were repeated for a final time. A volume of 80  $\mu\text{L}$  of the aqueous phase was removed and diluted 10-fold into 1× TE (10 mM Tris pH = 7, 1 mM

EDTA). These samples were stored at  $-20\text{ }^{\circ}\text{C}$  prior to qPCR analysis. The input samples treated with proteinase K were subjected to the same phenol–chloroform extraction procedure.

**qPCR Analysis.** DNA isolated from each sample was measured using Taqman assays for each of the three genes studied (see the Supporting Information). Dilutions of purified yeast genomic DNA were used for a standard curve. Each DNA sample was analyzed in duplicate in a 96-well microtiter plate. Each well contained  $5\text{ }\mu\text{L}$  of sample,  $10\text{ }\mu\text{L}$  of LightCycler 480 probe master mix,  $4.5\text{ }\mu\text{L}$  of water, and  $0.5\text{ }\mu\text{L}$  of  $40\times$  primer probe mix. After pipetting, each plate was centrifuged for 2 min at  $2\text{ }000g$ . The samples were then analyzed using a Roche 480 LightCycler. The qPCR runs included a 5 min preincubation step at  $95\text{ }^{\circ}\text{C}$ , amplification cycles, and a 2 min cooling step at  $40\text{ }^{\circ}\text{C}$ . Each amplification cycle was composed of a 10 s  $95\text{ }^{\circ}\text{C}$  incubation with a temperature ramp of  $4.4\text{ }^{\circ}\text{C/s}$ , a 30 s incubation at  $60\text{ }^{\circ}\text{C}$  with a temperature ramp of  $2.2\text{ }^{\circ}\text{C/s}$ , and a third 1 s incubation at  $72\text{ }^{\circ}\text{C}$  with a temperature ramp of  $4.4\text{ }^{\circ}\text{C/s}$ . Detection of the FAM fluorophore was performed during the  $72\text{ }^{\circ}\text{C}$  incubation using a 483–533 filter set. Analysis of the resultant qPCR curves and calculation of  $C_p$  values were performed using the Roche 480 LightCycler software and the second quant/2nd derivative function. Absolute DNA amounts were calculated from the genomic DNA standard curve, and % protein-free DNA values are given by the ratio (protein-free DNA in aqueous phase/input DNA)  $\times 100\%$ .

## RESULTS AND DISCUSSION

**Measuring the Formaldehyde Cross-Link Reversal Rate.** Formaldehyde cross-link reversal involves dissociation of the methylene linkage between the protein and DNA moieties (Figure 1). The nature of these cross-links varies depending upon the amino acid and nucleotides involved and the chemical landscape within a cell subjected to formaldehyde is likely quite complex because of the various linkages present. Lu et al. recently reported the cross-linked species that result from addition of formaldehyde to mixtures of each amino acid and nucleotide.<sup>12</sup> They found that cross-linking occurs between the amino acids cysteine (Cys), histidine, tryptophan, and lysine (Lys) and the deoxynucleosides deoxyadenosine, deoxycytosine, and/or deoxyguanosine (dG), with the predominant reaction being that between Lys and dG to form an aminor linkage, and the second most prominent reaction being that between Cys and dG to form a hemiaminal thioether linkage.<sup>12</sup> A detailed reaction mechanism for the prototypic Lys–dG coupling reaction is shown in Figure 1. In the present work, we determine the aggregate cross-link reversal rate of formaldehyde protein–DNA complexes isolated from cross-linked yeast cells, which are likely to reflect the predominant Lys–dG and Cys–dG cross-links as well as small contributions from the other amino acid–deoxynucleoside pairs.

In general, to measure the rate of cross-link reversal, either the disappearance of reactants or the appearance of products must be measured. In the case of formaldehyde cross-links, this means the disappearance of protein–DNA complexes or the appearance of native protein and DNA moieties without the methylene linkages. Many methods exist to measure total DNA or protein levels including PicoGreen binding<sup>13</sup> for the former and BCA<sup>14</sup> and A260<sup>15</sup> assays for the latter. In the complex background of cell lysate, however, the disappearance of a single methylene bridge is not a substantial perturbation and these assays do not distinguish between cross-linked and non-

cross-linked species. Therefore, an additional purification step is useful to isolate one component of the reaction and allow for specific analysis.

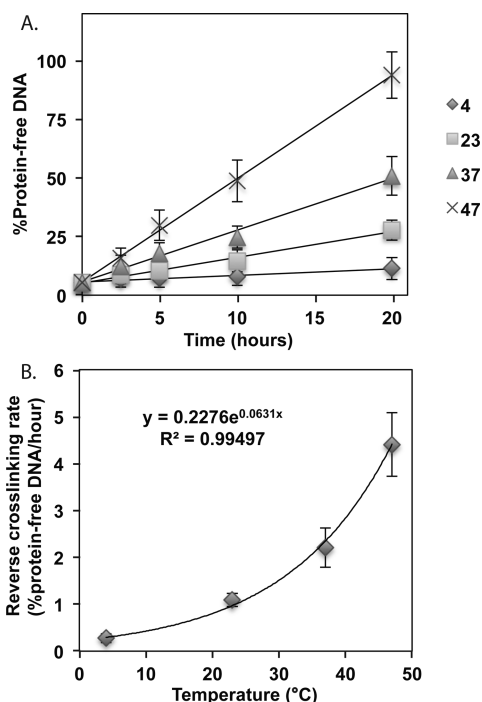
FAIRE was developed as a method to identify protein-free regions of the genome.<sup>6</sup> The assay uses phenol–chloroform extraction as a means to separate protein-free DNA fragments from cross-linked protein–DNA complexes and free protein. When subjected to phenol–chloroform extraction, proteins partition into the organic phase, while DNA remains in the aqueous phase. Protein–DNA complexes reside at the interface between the two phases. Identification of the DNA sequences found in the aqueous phase, the protein-free fraction, reveals those sequences in the genome that are not associated with protein, thereby providing a means of determining protein occupancy across the genome. We utilized FAIRE to determine the kinetics of cross-link reversal for formaldehyde-cross-linked protein–DNA complexes from yeast. Cross-linked cells were lysed, sonicated, and centrifuged, to shear the DNA and clear the lysate sample of debris. Aliquots of the lysate sample were then periodically subjected to phenol–chloroform extraction to isolate the protein-free DNA fraction of the lysate. The aqueous material was then interrogated using qPCR assays specific for three genes (*INO1*, X-element, and 25S rDNA) to measure the relative amounts of DNA corresponding to each gene present as a function of time. The DNA levels present in the aqueous phase of the phenol–chloroform extraction were plotted as a function of time and the cross-link reversal rate corresponds to the slope of the line (Figure 2). As very similar results were obtained for all three genes, the results for all three genes were averaged to provide an overall measure of cross-link reversal rate.

**Temperature Effects.** It is well-known that heat reverses formaldehyde cross-links;<sup>16</sup> however, the rates have not been measured nor has the dependence of the rate on temperature. We measured the cross-link reversal rate at four different temperatures: 4, 23, 37, and  $47\text{ }^{\circ}\text{C}$ . The first three temperatures were selected as they correspond to commonly employed temperatures in biochemical protocols, and the  $47\text{ }^{\circ}\text{C}$  temperature was chosen arbitrarily as a more extreme case (dissociation rates at other temperatures may be estimated if desired from the exponential fit to the data shown in Figure 2B). The percent of protein-free DNA relative to total input DNA is plotted versus time for each of the four temperatures examined (Figure 2A). The slope of each line is given in Table 1 as the reaction rate. As expected, the reaction rate increases with temperature, with the slowest reaction rate at the lowest temperature measured (0.3% protein-free DNA/h at  $4\text{ }^{\circ}\text{C}$ ) and the fastest reaction rate at the highest temperature measured (4.4% protein-free DNA/h at  $47\text{ }^{\circ}\text{C}$ ). Half-lives were also calculated for each temperature assuming a zero-order rate law for a unimolecular dissociation. The longest half-life was 179 h, at  $4\text{ }^{\circ}\text{C}$ , while the shortest half-life was 11.3 h, at  $47\text{ }^{\circ}\text{C}$ .

The temperature dependence of the cross-link reversal rate fits well to an exponential function, as shown in Figure 2B. On the basis of this dependence, we were able to fit the data to both the linear Arrhenius eq 1 and Eyring eq 2.

$$\ln(k) = \frac{-E_a}{R} \frac{1}{T} + \ln(A) \quad (1)$$

$$\ln\left(\frac{k}{T}\right) = \frac{-\Delta H^\ddagger}{R} \frac{1}{T} + \ln \frac{k_B}{h} + \frac{\Delta S^\ddagger}{R} \quad (2)$$



**Figure 2.** Temperature dependence of the cross-link reversal rate. The percentage of protein-free DNA measured by qPCR is plotted versus time for samples at four different temperatures: 4, 23, 37, and 47 °C (A). Each point is the average qPCR signal for the three different genes *INO1*, X-element, and 25S rDNA. The error bars are the standard deviation of qPCR signals. A best fit linear regression is shown for each temperature trace. The slope of each line was determined and plotted in panel B as a function of temperature (°C). These values represent the reverse cross-linking rate. The rate of reverse cross-linking as a function of temperature was fit to an exponential.

**Table 1. Reaction Rates and Half-Lives at Different Temperatures<sup>a</sup>**

temperature (°C)	reaction rate (% protein-free DNA/h)	$t_{1/2}$ (h)
4	0.3	179
23	1.1	45.9
37	2.2	22.7
47	4.4	11.3

<sup>a</sup>For each temperature studied, the reaction rate and half-life ( $t_{1/2}$ ) are listed.

From the Arrhenius plot, the activation energy ( $E_a$ ) was determined to be 47 kJ mol<sup>-1</sup>, while from the Eyring plot, the enthalpy of activation ( $\Delta H^\ddagger$ ) and entropy of activation ( $\Delta S^\ddagger$ ) were determined to be 44 kJ mol<sup>-1</sup> and -0.1 kJ mol<sup>-1</sup> K<sup>-1</sup>, respectively.

## CONCLUSION

We show here the use of FAIRE-qPCR to directly measure the rate of formaldehyde cross-link reversal for protein–DNA complexes. Formaldehyde cross-links are shown to depend exponentially on temperature, with half-lives varying from 11.3 h at 4 °C to 179 h at 47 °C. Characterization of the cross-link reversal rate is useful for experiments that utilize formaldehyde to stabilize protein–DNA interactions.

## ASSOCIATED CONTENT

### Supporting Information

qPCR primer/probe sequence information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [smith@chem.wisc.edu](mailto:smith@chem.wisc.edu). Phone: (608) 263-2594. Fax: (608) 265-6780.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the NIH Center of Excellence in Genomics Sciences Grant 1 P50 HG004952. We thank Professor Tehshik Yoon and Dr. Brian Frey for helpful discussions regarding the cross-link dissociation reactions and Professor Ron Raines for use of the cell disruptor.

## REFERENCES

- (1) Luger, K.; Richmond, T. J. *Curr. Opin. Struct. Biol.* **1998**, *8*, 33–40.
- (2) Ko, L. J.; Engel, J. D. *Mol. Cell. Biol.* **1993**, *13*, 4011–4022.
- (3) Osada, S.; Yamamoto, H.; Nishihara, T.; Imagawa, M. *J. Biol. Chem.* **1996**, *271*, 3891–3896.
- (4) Solomon, M. J.; Larsen, P. L.; Varshavsky, A. *Cell* **1988**, *53*, 937–947.
- (5) Simon, J. M.; Giresi, P. G.; Davis, I. J.; Lieb, J. D. *Nat. Protoc.* **2012**, *7*, 256–267.
- (6) Giresi, P. G.; Lieb, J. D. *Methods* **2009**, *48*, 233–239.
- (7) DeJardin, J.; Kingston, R. E. *Cell* **2009**, *136*, 175–186.
- (8) Byrum, S. D.; Raman, A.; Taverna, S. D.; Tackett, A. J. *Cell Rep.* **2012**, *2*, 198–205.
- (9) Kennedy-Darling, J.; Guillen-Ahlers, H.; Shortreed, M. R.; Scalf, M.; Frey, B. F.; Kendziora, C.; Olivier, M.; Gasch, A. P.; Smith, L. M. **2014**, submitted to *J. Proteome Res.*
- (10) Poorey, K.; Viswanathan, R.; Carver, M. N.; Karpova, T. S.; Cirimotich, S. M.; McNally, J. G.; Bekiranov, S.; Auble, D. T. *Science* **2013**, *342*, 369–372.
- (11) Solomon, M. J.; Varshavsky, A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 6470–6474.
- (12) Lu, K.; Ye, W. J.; Zhou, L.; Collins, L. B.; Chen, X.; Gold, A.; Ball, L. M.; Swenberg, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 3388–3399.
- (13) Dragan, A. I.; Bishop, E. S.; Casas-Finet, J. R.; Strouse, R. J.; Schenerman, M. A.; Geddes, C. D. *J. Immunol. Methods* **2010**, *362*, 95–100.
- (14) Brown, R. E.; Jarvis, K. L.; Hyland, K. J. *Anal. Biochem.* **1989**, *180*, 136–139.
- (15) Kalb, V. F.; Bernlohr, R. W. *Anal. Biochem.* **1977**, *82*, 362–371.
- (16) Jackson, V. *Cell* **1978**, *15*, 945–954.