Formaldehyde crosslinking: a tool for the study of chromatin complexes

Elizabeth A. Hoffman¹, Brian L. Frey², Lloyd M. Smith², and David T. Auble¹

¹From the Department of Biochemistry and Molecular Genetics, University of Virginia Health System, Charlottesville, VA 22908

²From the Department of Chemistry and Genome Center of Wisconsin, University of Wisconsin, Madison, WI 53706

*Running title: Formaldehyde crosslinking

To whom correspondence should be addressed: David T. Auble, Department of Biochemistry and Molecular Genetics, University of Virginia Health System, Charlottesville, VA 22908, USA. Tel.: (434) 243-2629; Fax: (434) 924-5062; E-mail: auble@virginia.edu.

Keywords: formaldehyde chemistry, chromatin immunoprecipitation (ChIP), transcription factor, protein cross-linking, chromatin structure, nucleic acid chemistry

ABSTRACT

Formaldehyde has been used for decades to probe macromolecular structure and function and to trap complexes, cells, and tissues for further analysis. Formaldehyde crosslinking is routinely employed for detection and quantification of protein-DNA interactions, interactions between chromatin proteins, and interactions between distal segments of the chromatin fiber. Despite widespread use and a rich biochemical literature, important aspects of formaldehyde behavior in cells have not been well described. Here we highlight features of formaldehyde chemistry relevant to its use in analyses of chromatin complexes, focusing on how its properties may influence studies of chromatin structure and function.

Prior to its use in the chromatin field, formaldehyde use had a long history in a number of fields, including vaccine production (1,2) and histology (3). In this review we focus on its use in chromatin immunoprecipitation approaches and protein-protein interaction studies applied to understand the location and abundance of transcription factor binding along DNA. A recent complementary perspective highlights gaps in knowledge with a particular focus on how formaldehyde crosslinking data have been used to interpret aspects of chromatin three-dimensional organization (4). Here, we briefly review prior work describing formaldehyde reactivity toward proteins, DNA, and their constituent monomers. This information provides a basis for understanding how formaldehyde functions in widely used assays in the chromatin field, and conversely, highlights less well-understood aspects of formaldehyde behavior in cells. These issues are of significance for designing crosslinking-based studies as well as for properly interpreting the resulting data. The analysis of formaldehyde-fixed chromatin has provided fundamental insights into where and when regulatory factors associate with the DNA template in vivo, but it in general does not provide unambiguous information about chromatin binding kinetics or an absolute measure of fractional occupancy. A major goal of ongoing work is to understand kinetic and thermodynamic aspects of chromatin complex assembly at single copy loci in vivo. Development of experimental strategies to achieve these goals will require a deeper and more comprehensive understanding of the effects mediated by formaldehyde in cells.

The following discussion provides a framework for understanding aspects of formaldehyde function when used to trap macromolecular complexes in cells, with the main features shown in Figure 1. Beginning with basic chemical
reactivity, this review will explore the ability of formaldehyde to crosslink with proteins and DNA to form protein-protein or protein-DNA complexes, common molecular quenchers, and the potential for crosslink reversal. Progress in capturing crosslinked complexes will also be discussed with an emphasis on the impact of a better understanding of formaldehyde chemistry in vivo.

**Basic chemistry**

Formaldehyde is the smallest aldehyde, an electrophilic molecule susceptible to chemical attack by a wide range of nucleophilic species of biological interest. The chemical complexity of formaldehyde-mediated reaction products was appreciated 70 years ago (5). Initially using amino acids, and subsequently proteins and other substrates, it was shown that formaldehyde reacts in vitro with a wide range of functional groups, forming a complex array of products (6,7). It has been known since the 1940s that such products can include intramolecular and intermolecular crosslinked species, and that the reaction conditions (e.g. pH, temperature) can strongly influence the nature, yield, and half-life of chemical modifications (8). The concentration of formaldehyde used, incubation times and other conditions can vary substantially among different applications employing formaldehyde fixation, yielding very different chemical products (reviewed in (9)).

Formaldehyde reacts with macromolecules in several steps (Figure 2). In the first step, a nucleophilic group on an amino acid or DNA base (for example) forms a covalent bond with formaldehyde, resulting in a methylol adduct which is then converted to a Schiff base. Methylols and Schiff bases can decompose rapidly (10-12) or may be stabilized in a second chemical step involving another functional group, often on another molecule, leading to formation of a methylene bridge (13). A methylene bridge might form between a solvent-exposed group on a macromolecule and a small molecule in solution such as glycine, which is frequently used as a formaldehyde quencher. Alternatively, and of most interest to biologists, is the formation of a covalent bond linking functional groups in two different macromolecules. The small size of formaldehyde dictates its linkage of groups that are ~2 angstroms apart, making it well suited for capture of interactions between macromolecules that are in close proximity (14,15). (Commercial preparations of formaldehyde may also contain formaldehyde aggregates (16) whose reactivities and distance spanning capabilities are unclear.)

**Formaldehyde reactivity with proteins**

As studies of formaldehyde reactivity became more sophisticated, it was found that conditions that more closely resemble those used for crosslinking components in cells yield a subset of the products identified in the earlier studies (17). Using model peptides, formaldehyde was found to react with N-terminal amino groups and side chains of cysteine, histidine, lysine, tryptophan and arginine (10). Reaction products were in some cases influenced by the peptide sequence, yielding intramolecular crosslinks as well as linkages of the N-terminus and histidine, asparagine, glutamine, tryptophan, tyrosine and arginine residues to glycine molecules added to the reaction (10). Despite the long incubation time (48 hours), adducts were not detected between glycine and peptide cysteine or lysine residues. Subsequent work employing model substrates along with formaldehyde concentrations and reaction times more in line with those used with cells identified a smaller subset of formaldehyde reaction products involving lysine, tryptophan and cysteine side chains as well as the peptide N-terminus (17). Such studies have often been motivated by interest in developing techniques for analysis of native protein complex subunit composition. As discussed more below, the rapid reactivity of formaldehyde with cellular constituents suggests that cells are highly permeable to formaldehyde, and the requirement for crosslinked groups to be closely apposed makes formaldehyde a good candidate for capturing macromolecular complexes in vivo containing specific but unstably bound subunits, which can then be analyzed by mass spectrometry (18).

In discussing the complexity of crosslinked complexes formed by incubation of cells with formaldehyde, it is important to distinguish
between two types of complexity. The first is the chemical complexity arising from the multiplicity of macromolecular functional groups that can potentially react with formaldehyde, and the second is complexity associated with the types and numbers of macromolecules crosslinked to each other. Although formaldehyde can potentially generate a great variety of chemically distinct products \textit{in vitro}, the biologically relevant chemical complexity is in all likelihood simpler under incubation conditions more typically used for analyses of macromolecular complexes \textit{in vivo}. This is due to several factors, including a lowered effective formaldehyde concentration in cells compared to most model experiments \textit{in vitro}, limiting the ability of formaldehyde to locate and interact with a functional group. While there is much greater macromolecular diversity in cells than in typical in vitro experiments, native macromolecules likely provide a smaller range of chemically reactive groups than model substrates used in vitro. As discussed below, N-terminal amino groups may be less available and side chains are less accessible to formaldehyde crosslinking due to protein tertiary structure in native proteins. These factors would decrease the proportion of potentially chemically reactive groups and allow for a smaller, less diverse set of chemical products \textit{in vivo}. For instance, reactivity with native proteins is limited to those nucleophilic groups that are accessible to formaldehyde, and indeed, studies exploring differential formaldehyde reactivity have been used to provide insight into enzyme structure and catalytic function (19). Solvent accessible lysine residues have been found to provide the most reactive functional groups in native proteins, and moreover, modification of native proteins by formaldehyde does not appear to perturb tertiary structure in native proteins. These factors would decrease the proportion of potentially chemically reactive groups and allow for a smaller, less diverse set of chemical products \textit{in vivo}. For instance, reactivity with native proteins is limited to those nucleophilic groups that are accessible to formaldehyde, and indeed, studies exploring differential formaldehyde reactivity have been used to provide insight into enzyme structure and catalytic function (19). Solvent accessible lysine residues have been found to provide the most reactive functional groups in native proteins, and moreover, modification of native proteins by formaldehyde does not appear to perturb tertiary structure very much (20). This is consistent with early work in the chromatin field that established that lysine residues are the predominant sites of formation of methylene bridges in histone complexes; such studies led to the suggestion as well that formaldehyde crosslinking does not in general perturb protein structure (21). The apparent preference of formaldehyde for accessible lysine residues may explain in part why formaldehyde has emerged as the crosslinker of choice for trapping protein-DNA complexes, as lysine residues are common mediators of interactions with DNA (22). The differential reactivity of accessible groups on protein surfaces has also been explored to understand how formaldehyde fixation impacts epitope recognition by antibodies (23). Of note, the potential for formaldehyde to affect antibody recognition could possibly impact a wide range of experiments that require quantification of recovered fixed material by immunoprecipitation. Conditions can often be worked out such that formaldehyde treatment does not adversely impact antibody recognition (24,25), but to our knowledge this has not been examined in great detail in the chromatin field. Importantly, the apparent predominance of a subset of reactive sites on macromolecules under typical experimental conditions is not to suggest that overall crosslinking complexity in cells is necessarily simple. Even though \textit{in vivo} crosslinking is probably predominated by a subset of the chemical products observed \textit{in vitro}, there is potential for macromolecules to become crosslinked together in multiple ways and in multiple combinations, forming larger daisy-chained structures that complicate \textit{in vivo} crosslinking results. Indeed, there is some evidence that formaldehyde treatment of cells can result in higher order chromatin or nuclear structures whose formation may yield misleading interpretations of chromatin association data by trapping factors within dense crosslinked networks (4,26).

\textbf{Formaldehyde reactivity with DNA}

Formaldehyde reacts with amino and imino groups of DNA bases, and extensive studies have been performed to document kinetic and thermodynamic aspects of such reactions (11,12,27-29). While formaldehyde reactivity with proteins does not appear to perturb protein tertiary structure, formaldehyde reactivity with DNA is notably different as covalent modification of DNA bases requires disruption of base pairing in duplex DNA, and in fact, formaldehyde was used in pioneering studies to probe DNA melting (30-34). Modified bases are thus precluded from base pairing and promote further DNA denaturation (30). This likely occurs to some extent in stretches of naked DNA in cells treated with formaldehyde, although under typical conditions employed for \textit{in vivo} studies, the
Formaldehyde crosslinking

recovered DNA is by and large suitable for enzymatic manipulation (35). Formaldehyde modification of naked DNA in vitro may be more extensive (36). Conformational changes in DNA that promote formaldehyde reactivity have been referred to as DNA “breathing” or base flipping. Measurement of the rates of such spontaneous conformational changes is an active area of investigation (37) and it is unclear what specific DNA conformational changes are required to allow reaction of DNA bases with formaldehyde (i.e., full extrahelical extrusion of a DNA base may not be required). The rates of formaldehyde reactivity with naked DNA in vitro were found to be orders of magnitude below diffusion-limited rates, although these studies make clear that reaction conditions can have large effects on reactivity and indeed, it was recognized early on that it would be difficult to extrapolate rates of reaction obtained in relatively simple in vitro systems to other more complex systems, let alone in vivo (11).

Capture of protein-DNA complexes

The early use of formaldehyde as a probe of macromolecular structure led to the discovery that formaldehyde can crosslink histones to DNA (38). Retrieval of the crosslinked complexes and analysis of the associated DNA then gave birth to the ChIP3 assay (14,39,40), which has become ubiquitous in the chromatin field in a multitude of variations (41-49). Although ChIP assays performed without crosslinking have proven valuable for analyses of stable chromatin complexes (50), crosslinking has made it possible to identify interactions that would not otherwise withstand the isolation procedure. Given the central utility of crosslinking and its critical role in establishing many of the principles underlying the current understanding of chromatin structure and function, a clear picture of formaldehyde chemistry is critical to ensure that any biases resulting from formaldehyde crosslinking are taken into account.

The ability of formaldehyde to crosslink amino acids to DNA bases has been examined systematically in vitro. In comparing the products of reactions containing lysine, cysteine, histidine, or tryptophan with each of the four DNA bases, the highest yield of crosslinked product was obtained with lysine and deoxyguanosine (51), consistent with lysine being most reactive among residues in native proteins described above. Similar results were obtained using short peptides and trinucleotides (51). In the context of protein-DNA interactions, the first chemical step could involve reaction with either an amino acid side chain in a protein, the protein N-terminus, or an amino or imino group on a DNA base; importantly, however, the ε-amino group on the lysine side chain is a better nucleophile than are the amino/imino groups on DNA bases whose lone pair electrons are delocalized in the aromatic ring. For this reason it seems reasonable to speculate that in most crosslinked protein-DNA complexes, a Schiff base is formed on a lysine residue first, followed by nucleophilic attack by the DNA base held in proximity to the side chain, resulting in a methylene bridge.

Interestingly, and in line with this idea, formaldehyde reactivity with DNA was stimulated substantially by adding amino acids or histones to an in vitro reaction, resulting in stable products that in some cases contained both DNA and the protein or amino acid (52). The ~20-30-fold stimulation in the reaction rates observed in these early experiments by addition of glycine or lysine (for example) was striking; furthermore, formaldehyde crosslinking of proximal functional groups on specific, stable macromolecular complexes presumably can occur even faster owing to the constrained physical proximity of the reacting species (53). In addition to the ubiquity of lysine side chains in DNA-binding proteins (for interaction with the phosphate backbone), the DNA bases provide a high density of amino and imino groups along the length of the nucleic acid. These two features may contribute to the relatively higher yield of protein-DNA crosslinks compared to protein-protein crosslinks as measured by conjugation of chromatin regulatory complexes that interact indirectly with DNA (54). It has been observed that for some transcriptional co-regulators, protein-protein crosslinks are not efficiently detected between factors that interact with chromatin indirectly when using 3C3 methods; this could be due to either inefficiencies in formaldehyde crosslinking between proteins (due to non-optimal reactive side chain
Several different crosslinking agents have been used in ChIP (55) but all of the features described above including cell permeability, short spacer length, rapid reactivity-as well as reversibility (discussed below)- have led to formaldehyde becoming the crosslinking agent of choice for ChIP. This utility has been borne out by many genome-wide studies that have shown how profiles of crosslinked complexes capture transcription factor binding to physiologically significant DNA sites (for example, see (56-58)). Since DNA site-specific transcription factors can also bind to nonspecific sites (59-62), crosslinking of non-specifically bound proteins to DNA would be expected to occur and may account in part for binding events detected in genome-wide studies that cannot be readily explained physiologically. Non-DNA binding proteins are not crosslinked to chromatin (14,63), and non-specifically bound factors are presumably bound to a multitude of disparate sites at low levels consistent with their relative occupancies (64,65). Analyses of chromatin binding by a series of mutants in the methyl CpG binding protein 2 gene led to the conclusion that there is a threshold interaction lifetime of about 5 seconds required for crosslinking (26). However, it has been possible to ChIP transcription factors whose interactions with chromatin are known from imaging studies to be highly transient (~few seconds time scale) (61,66-69). Importantly, in vivo ChIP signals have been found to correlate with DNA binding specificities and affinities measured in vitro (70,71), supporting the use of formaldehyde for measuring chromatin binding interactions in cells with quantitative rigor and over a broad thermodynamic range. A better understanding of formaldehyde’s effects in cells could potentially be obtained by biochemical studies of protein-DNA complex crosslinking in vitro. However, while the data are not extensive, it is noteworthy that there are examples in which in vitro and in vivo binding behaviors differ when assessed using formaldehyde in one system or another (14,55,70).

As mentioned above, formaldehyde crosslinking can be used to isolate physiologically relevant complexes for analysis by mass spectrometry (9,18). Formaldehyde crosslinking has been exploited to identify constituents of some cellular complexes, including the proteasome, for example, (72,73), and to identify protein-protein interactions in tissues (74). Unbiased efforts to identify proteins associated with specific DNA sequences in vivo have employed formaldehyde to stabilize such complexes during their isolation. It is technically challenging to obtain sufficient material for identification of the DNA-associated proteins by mass spectrometry, but this strategy has nonetheless been successfully employed for repeated DNA sequences such as telomeres (75-77) and in model in vitro systems (36), as well as for single copy and multicopy regions in yeast (75,76,78-80). The relatively mild formaldehyde reaction conditions used to maximize crosslinking of physiologically relevant associations and avoid spurious ones may yield a relatively low proportion of formaldehyde crosslinked material and thereby contribute to the technical challenges of this kind of approach (36).

Crosslinking kinetics, stability and reversal

Formaldehyde crosslinking in chromatin studies typically employ relatively low formaldehyde concentrations (1% or less). The facile detection of protein-DNA complexes following incubation times of 30 min or less suggests that macromolecular crosslinking occurs relatively rapidly, as suggested by the earliest ChIP experiments (14,40,81). Relatively rapid formaldehyde reactivity in cells is also consistent with the ability to distinguish ChIP signals over short time intervals (seconds to minutes) (63,82,83). Formaldehyde crosslinks are quite stable in vivo compared to the durations of most crosslinking experiments, with crosslink half-lives of ~10-20 hours depending on the cell type and conditions (15). In ChIP experiments, crosslinks are most often reversed by heat (21). The reversibility of formaldehyde crosslinking has been explored in some detail in an effort to recover proteins from fixed tissue and cell samples (84,85). The temperature and salt concentration dependence of the formaldehyde crosslink reversal rate has been established, revealing a crosslink half-life consistent with the estimate of crosslink half-life in cells (tens of hours at 37 C) (86). That study also quantitatively showed the extent to
which heat can increase the crosslink reversal rate. More such measurements on other aspects of crosslinking chemistry will be useful in developing a quantitative understanding of how the abundance of a particular crosslinked species obtained under some set of conditions (i.e. the effect of pH, quencher choice and concentration, and formaldehyde concentration) relates to dynamic aspects of complex assembly/disassembly and stability (83).

To limit formaldehyde reactivity to a particular time interval, unreacted formaldehyde is quenched with an excess of a small reactive molecule added to the reaction (Figure 3). Quenching is important but not well understood (9). Glycine has been typically used as a sink for unreacted formaldehyde in ChIP (44,55) as well as in approaches to map higher order chromatin structure (87-90). The efficacy of glycine is improved by reduced pH, but detailed studies of the quenching reaction have not been reported (9). In principle, formaldehyde crosslinking could be quenched by reaction of the quencher with formaldehyde molecules in solution or reaction with formaldehyde conjugates on other molecules in the cell, if the quencher is readily cell permeable. As discussed above, formaldehyde-mediated glycine conjugates have been detected or inferred in vitro, although there was no evidence for such conjugates seen in proteins analyzed from formaldehyde-treated cells (9,25,83). On the other hand, evidence suggests that glycine-DNA conjugates are formed in an in vitro reaction (36). Despite the fact that glycine has been used routinely to quench crosslinking, Tris is a more efficient quencher (9), which can be explained chemically by the ability of Tris to form a cyclic product upon reaction with formaldehyde (36) (Figure 3). However, at higher concentrations of Tris, which would likely be used for quenching, Tris can also facilitate crosslink reversal (91) thereby potentially impacting the yield of crosslinked material.

**Complex effects in the cellular milieu**

Given the pivotal role of this reagent in understanding chromatin biology and the continuing evolution of technologies exploiting its properties, it is critical that a deeper understanding is achieved of formaldehyde crosslinking as it occurs in cells. Although formaldehyde can mediate myriad chemical reactions in vitro, the conditions used for crosslinking in cells suggest that, with respect to chromatin, the chemical complexity of macromolecule-containing reaction products is more limited, with reactions occurring mainly with solvent-exposed lysine residues and endo- and exocyclic amino groups on bases. Formaldehyde has a number of other properties that make it well suited for trapping macromolecular complexes in cells, including cell permeability and the temperature-dependent stability of methylene bridge containing adducts. Macromolecules that do not interact are in general not crosslinked together efficiently, and methylol/Schiff base intermediates are reversibly formed and appear to be inefficiently trapped by reaction with quenchers in cells. This explains why proteins and DNA isolated from formaldehyde-treated cells appear unmodified in general (14,83,92). Within minutes of formaldehyde incubation, there is very little detectable free DNA (<10%) (36,86), and crosslinking appears to occur uniformly along DNA as well (14).

ChIP has been developed by empirical determination of seemingly optimal crosslinking conditions, with low recoveries occurring for either too little or too much crosslinking (93). If the formaldehyde concentration is too low or the incubation time is too short, not enough crosslinked material will be produced. On the other hand, too high of formaldehyde concentration or too long of incubation time also reduces recovery, presumably reflecting the formation of complexes that are insoluble or the masking of epitopes recognized by the antibody used for immunoprecipitation. We have observed little effect of formaldehyde incubation time on chromatin protein yield over a broad range of formaldehyde concentrations and incubation times (92), but formaldehyde concentrations above 3% do impact yield even after moderate incubation times, suggesting the formation of such complexes. Given the dense concentration of macromolecules in the nucleus, it is plausible that formaldehyde may cause the formation of higher order networks of crosslinked chromatin (4), as illustrated in Figure 4. This is an area worthy of
additional investigation, particularly since it may explain nonspecific DNA crosslinking that occurs in ChIP or ChIP-related methods (94). Conversely, other spurious enrichment phenomena, such as localization of unrelated proteins, can occur with ChIP at highly expressed genes; these, too, warrant deeper investigation (95,96) to ensure that apparent ChIP signals in fact represent true association with the loci of interest.

ChIP assays have been pivotal in establishing our current view of chromatin structure and function. As answers to deeper questions about chromatin binding dynamics and higher order structure are pursued, we feel it is imperative that we understand more fully how procedures employed to obtain snapshots of chromatin state may perturb the very properties being measured, particularly as variations in experimental design can yield different interpretations (4,97). In this regard, an understanding of the behavior and properties of formaldehyde in the cell are important for determining the best methods for measuring dynamic interactions using crosslinking. A better understanding of formaldehyde crosslinking may in turn lead to better quantitative models (98).
REFERENCES


Formaldehyde crosslinking


34. McGhee, J. D., and von Hippel, P. H. (1976) Fromaldehyde as a probe of DNA structure. 4. Mechanism of the initial reaction of formaldehyde with DNA. *Biochemistry* 16, 3276-3293


Formaldehyde crosslinking


Formaldehyde crosslinking


FOOTNOTES

*This work was supported by NIH grants R01 GM55763 (D.T.A.) and P50 HG004952 (L.M.S. and B.L.F.). E.A.H. was supported by NCI Cancer Training Grant T32 CA009109-38.

1 To whom correspondence may be addressed: Department of Biochemistry and Molecular Genetics, University of Virginia Health System, Charlottesville, VA 22908, USA. Tel.: (434) 243-2629; Fax: (434) 924-5062; E-mail: auble@virginia.edu.

2 From the Department of Chemistry and Genome Center of Wisconsin, University of Wisconsin, Madison, WI 53706, USA

3 The abbreviations used are: ChIP, chromatin immunoprecipitation; Tris, (Tris(hydroxymethyl)aminomethane); 3C, chromosome conformation capture.

FIGURE LEGENDS

FIGURE 1. This cartoon depicts the main aspects of formaldehyde reactivity in cells. The dashed arc represents cell or nuclear membranes, which are thought to be highly permeable to formaldehyde (red circles). The thick black curved line represents DNA, shown assembled as nucleosomes (light gray circles). A chromatin interacting factor is schematized in cyan, with other partner proteins shown in blue and purple. Small molecules such as glycine and Tris that react with formaldehyde and can therefore quench reactivity with cellular constituents are shown as green circles. Formaldehyde can crosslink macromolecules together as well as modify exposed groups on macromolecules, forming a product species potentially stabilized by reactivity with a quencher. Quenchers are ordinarily added to the extracellular milieu and may exert their main effects outside the cell.

FIGURE 2. Formaldehyde crosslinking of biomolecules occurs in two steps. First, formaldehyde reacts with a relatively strong nucleophile, most commonly a lysine ε-amino group from a protein. This reaction forms a methylol intermediate that can lose water to yield a Schiff base (an imine). Second, the Schiff base reacts with another nucleophile, possibly an amino group of a DNA base, to generate a crosslinked product. This second nucleophile might be from DNA, another protein, the same protein as the first nucleophile, a quencher molecule, or other endogenous small molecule, and therefore a protein-DNA crosslink is only one of many possible products. All of the reactions in this two-step process are reversible, which is a key feature of formaldehyde crosslinking for chromatin capture. A specific example of a protein-DNA crosslink is shown. The atoms are color coded to match those of Figure 1: protein (cyan), formaldehyde (red), DNA (black).

FIGURE 3. Formaldehyde quenching reactions with glycine and Tris, the two most common quenchers. The chemical reactions are analogous to those shown in Figure 2 with the amino group of glycine or Tris acting as the primary nucleophile. The Schiff base formed from glycine may or may not react with a second nucleophile, but regardless, the crosslinking between macromolecules has been quenched. The Tris molecule has readily available second nucleophiles (hydroxyl groups) that create stable intramolecular five-membered rings. It is also possible for Tris to react with two formaldehyde molecules, leading to the final product shown. The propensity for Tris to form these stable intramolecular products likely allows it to scavenge formaldehyde from other molecules and thereby facilitate crosslink reversal. The atoms are color coded: quencher (green), formaldehyde (red), miscellaneous nucleophile (brown).

FIGURE 4. Potential effects of formaldehyde in mediating formation of higher order chromatin structures. The black wavy lines denote chromatin fibers, which may become a crosslinked meshwork in the presence of formaldehyde (red circles). The formation of these potentially confounding structures may or may not be mediated by physiologically relevant higher order interactions captured by crosslinking (dashed gray rectangle). Such a meshwork may define localized neighborhoods in the
nucleus that trap proteins (cyan) that may or may not interact specifically with nearby DNA sequences in an unperturbed cell.
Figure 1

- Protein crosslinking
- Chromatin binding protein
- Formaldehyde quencher
- Permeability
- Covalent modification of solvent-exposed side chains
- Reaction with quencher
- Covalent modification of DNA
- Protein-DNA crosslinking
**Figure 2**

**Step 1**

Protein + Formaldehyde $\rightarrow$ Methylol $\rightarrow$ Schiff Base

\[
\begin{align*}
R_1\text{--NH}_2 + & \quad \text{Formaldehyde} \quad \text{Methylol} \quad \text{Schiff Base} \\
& \quad \text{R}_1\text{--NH} \quad \text{CH}_2\text{--OH} \quad \text{R}_1\text{--N}\text{--CH}_2
\end{align*}
\]

**Step 2**

Schiff Base + DNA or Protein $\rightarrow$ Crosslinked Product

\[
\begin{align*}
R_1\text{--N}\text{--CH}_2 + & \quad \text{Nuc}\text{--R}_2 \quad \text{Crosslinked Product} \\
& \quad \text{R}_1\text{--NH} \quad \text{CH}_2\text{--Nuc} \quad \text{R}_2
\end{align*}
\]

**Example Protein-DNA Crosslink**

Lysine Residue of a Protein $\rightarrow$ Guanine Base of DNA
Formaldehyde Schiff Base Quenching Products

Figure 3

Glycine or Tris + Formaldehyde → Schiff Base + Nuc—R → Quenching Products

Downloaded from http://www.jbc.org/ by guest on September 1, 2017
Figure 4