THE EMBO JOURNAL

Manuscript EMBO-2009-72853

Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry

Zhuo Chen, Anass Jawhari, Lutz Fischer, Claudia Buchen, Salman Tahir, Tomislav Kamenski, Morten Rasmussen, Laurent Lariviere, Jimi-Carlo Bukowski-Wills, Michael Nilges, Patrick Cramer and Juri Rappsilber

Corresponding author: Juri Rappsilber, University of Edinburgh

Review timeline:

Submission date:	13 October 2009
Editorial Decision:	19 November 2009
Revision received:	09 December 2009
Accepted:	10 December 2009

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 19 November 2009

I would like to thank you for submitting your manuscript for consideration by The EMBO Journal. I have received the final report of the three referees who evaluated your study and I enclose their reports below. As you will see from their comments the referees are positive regarding the study and require mainly modifications to the text and figures before publication in the EMBO Journal. Referee #3 would also like to see more information given regarding the Xaminatrix program and if the software is available on the web. Once these issues are addressed, we would be happy to publish a revised manuscript.

I should remind you that it is EMBO Journal policy to allow a single round of revision only and that, therefore, acceptance or rejection of the manuscript will depend on the completeness of your responses included in the next, final version of the manuscript. When you submit a revised version to the EMBO Journal, please make sure you upload a letter of response to the referees' comments. Please note that when preparing your letter of response to the referees' comments that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process initiative, please visit our website: http://www.nature.com/emboj/about/process.html

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Editor
The EMBO Journal

REFEREE COMMENTS

Referee #1 (Remarks to the Author):

This study shows that a combined crosslinking/mass spec approach can be used to make predictions about the structures of large, multisubunit complexes such as the TFIIF-pol II complex. The analysis is obviously very complicated, but it can help make connections in cases where partial structures are available. This paper nicely complements the accompanying paper from the Hahn lab. However, I found that parts of the discussion were perhaps overly speculative and that the main figure (4) was very hard to read. If these issues can be addressed I think the paper would be appropriate for the EMBO Journal.

p10 The discussion of the one crosslink of the dimerization domain that doesn't fit the model is confusing. First, it's never stated which residues are involved, so it's impossible to look at the figures and see just how far away they are. Are they shown in Fig 4A? Second, the text cites some data from Chen et al 2007 that may be consistent with the unsatisfied crosslink, but it's never explained what those results are. I took a quick look at the Chen paper and couldn't find anything obvious. If the crosslink here and in the Chen paper suggest an alternative position, it could be worth creating a supplementary figure to show it. Finally, it's never discussed whether this unsatisfied crosslink might actually be due to a mistake in the modeling of the dimerization domain (which is the yeast TFIIF modeled after a human TFIIF structure).

p13 I found the attempt to connect the Tfg1 winged helix linker with Tfg3, and then Tfg3 to the CTD linker, to be rather tenuous and not very convincing. This is especially true given a recent structure of the S. pombe pol II from Kornberg that extends the CTD linker out onto Rpb4/7.

p14/15 Similarly, most of the discussion concerning how TFIIF might function in open complex formation and early elongation was not really connected to the data here, since it involves unstructured regions of the protein. I found this part to be overly speculative and I recommend trimming it in length and level of detail.

p16 URA1 should be URA3

Fig 1 It would be helpful to denote the domains on the protein "bars" (as is done in Fig 3).

Fig 4 and p 10. It seems to me that the "alternative" location for the Tfg2 WH is not being taken very seriously, surprising since this may be in more agreement with the previously published EM studies. I think it would be reasonable to show both positions, especially if they can be more specifically docked. If not, a better explanation for what the functional implications of the alternative site are is needed. This could be easily expanded into a separate figure (see below).

FIg 4 This is the most important figure of the paper, yet I found it very difficult to read. There is a lot of information here, but the labeling is very inconsistent and the figure legend doesn't contain much in the way of explanation. What are the spheres (the entire side chain, the nitrogens?). On the top figures, what do the numbers in parentheses after the residues K137 and K149 mean? In the blowups, why is the format for polymerase residues different from those in TFIIF? Finally, if it's not too dense, it would be nice to see the crosslinks drawn in as done in the other figures. I would strongly recommend breaking this figure up into multiple figures: one dealing with the Tfg2 WH, one for dimerization, and a third with the integrated PIC model. With improved figure legends, the authors' points will be made much clearer.

Referee #2 (Remarks to the Author):

The manuscript submitted for publication by Chen et al. investigates the architecture of a complex

made of RNA polymerase II and TFIIF general transcription factor. Together these two proteins consist of 15 different polypeptides totalling 670 kDa. TFIIF plays a key role in preinitiation complex (PIC) formation since it associates with RNA polymerase II prior their incorporation in PIC. TFIIF is also important for transcription start site selection. How TFIIF and Pol II interact is an important because there are conflicting reports about it.

A very recent (2007) method that was devised and first applied to the structure of the Ndc80 complex (176 kDa) is used to investigate this question. It entails the cross-linking of lysines and protein N-termini with the BS3 cross-linker and the identification of the cross-linked peptides by mass-spectrometry. The distance between cross-linked C-alpha is around 28 angstroms. The method was first validated on Pol II, using the known structure, identifying around 100 contacts. The false positive error rate was estimated to be around 1%. The method is then applied to the analysis of the Pol II-TFIIF complex. Around 400 linkage sites were identified of which 182 connected Pol II to TFIIF. The structure of the dimerization domains of Tfg1 and Tfg2 and of the winged helices of the two subunits was modelled. These models and the structure of Pol II were used to model the Pol II-TFIIF complex using the linkages as constraints. The protein complex model indicates that the dimerization domain is positioned on Pol II lobe while Tfg2 winged helix crosslinks to the Pol II protrusion upstream face. Tfg2 linker domain extends along the Rpb2 protrusion. Tfg1 winged helix cannot be cross-linked and thus thought to be mobile. The model predicts that some of the interactions between Pol II and TFIIF prevent the non-specific interaction between the enzyme and DNA, a known property of TFIIF. These observations, using a totally different technology, are consistent with those made by Chen et al. (Genes Dev. 2007) but not with the model of the complex proposed by Chung et al. (Mol. Cell 2003) based on EM analysis.

This work brings a wealth of information on the Pol II-TFIIF complex and strongly argues that the EM model is incorrect. The method used is very interesting since it allows investigating the structure and organization of very large complexes like those involved in transcription. It is almost certain that this method will be very useful for the analysis of complexes hard to crystalize or as an independent tool to assess structure and models. This manuscript should be published in The EMBO journal.

Comments and suggestions:
-p7: The description of how the Pol II-TFIIF complex is prepared is somewhat confusing. I would eventually drop the problems in expressing TFIIF in E. coli and low yield of complexes obtained by TAP-tagging Tfg2. I would make clear that the complex was affinity-purified form a strain over-expressing the three TFIIF subunits.
-p7: The RNA extension assay should be briefly described.
-p28: The Figure 2C legends indicates "Pol II-TFIIF complex shows normal elongation activity (Methods)". There is nothing in the Methods section. My remark is also related to the previous one on the RNA extension assay.
-Figure 1E labels are difficult to read.
-Is it possible to estimate the number of contacts that should have been observed in Pol II and that were not (false negative rate)?

Referee #3 (Remarks to the Author):

General:
This manuscript describes the investigation of the RNA polymerase II (Pol II)/TFIIF complex structure using chemical cross-linking, charge-based enrichment of cross-linked species, and high-resolution (LTQ-Orbitrap) mass spectrometric analysis. The complex possesses a molecular weight of ~ 670 kDa and presumably is the largest assembly that has been stued by a chemical cross-linking/MS approach to date. This paper presents an impressive example of the potential of the cross-linking approach for studying large multiprotein assemblies.

General Comment:

-Supplementary table file 24249_0_supp_254411_krgf1l has not been correctly converted into a pdf file - at least in the version I got.

-In this report, the authors employ one single amine-reactive cross-linker (BS3) with a spaer length of ca. 11 Angstrom. Did the authors also consider using additional cross-linkers with different specificities and different specer lengths? Moreover, N-hydroxysuccinimide (NHS) esters, such as BS, are quite easily hydrolzyed at one site giving so-called "dead-end" cross-links. These cross-linking products do not yield any distance constraints within the protein complex; yet, they provide useful information on the solvent accessibility of the proteins in the complex. The authors should comment whether they did not search at all for these "dead-end" products or (in case they did) how many of these species were obtained. These products could be included into the model.

Specific Comments:

-Page 6, lines 6/7: N-hydroxysuccinimide esters, such as BS3, have been found to frequently react with hydroxyl groups of serines, threonines, and tyrosines (Kalkhof S, Sinz A, Analytical and Bioanalytical Chemistry 2008, 392, 305; Mädler S, Bich C, Touboul D, Zenobi R, Journal of Mass Spectrometry 2009, 44, 694). On page 20, the authors state that no linkage specificity was assumed for BS3, but that all fragments observed contained a lysine or N-terminus as most likely modification site. This is somewhat surprising and authors should comment on this.

-Page 6, line 9: It is not completely clear why the authors assume a CA-CA distance of 27.4 Angstrom for BS3.

-Page 6, last paragraph: It has been shown in MD simulations that distances larger than 19 Angstrom are highly unlikely for a reagent with a spacer length of ca. 11 Angstrom (Green NS, Reisler E, Houk KN, Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers, Protein Science, 10, 1293-1304 (2001)). Experimentally, it has been shown (Ye X, et al., Probabilistic cross-link analysis and experiment planning for high-throughput elucidation of protein structure. Protein Sci, 13, 3298-313 (2004)) that distances of larger than 19 Angstrom are indeed bridged with signifcantly lower frequency. These references should be included in this paragraph.

-Page 7, line 3: Although the authors explain on page 20 how decoy searches were performed, they should add one sentence here on how the error rate of 1% was determined.

-Page 8, line 4 from bottom: The authors should be more specific on how the modeling of the Pol II/TFIIF complex was performed. Unfortunately, in the Materials and Methods part nothing is stated on the modeling programs and the methods used. More information is clearly needed here.

-Page 10, first & second line: The statement that "enough spatial restraints were avaiable to position and orient the dimerization domain" is very ambiguous. The authors should clearly state how many cross-links were acutally employed to determine the orientation of the dimerizaton domain. From the Supplementary Table it is very difficult to get this information.

-Page 10, line 16: The authors claim that the Tfg1 WH domain is apparently mobile as no cross-links to Pol II were found. Yet, one of the greatest strengths of the chemical cross-linking approach consists in that information also on flexible regions within a protein is readily obtained. Admittedly, fewer cross-links are found in flexible protein regions compared to very rigid regions, but one would have expected to find at least some cross-links between the Tfg1 WH domain and Pol II. The authors should comment on this.

-Page 17, last line and page 18, line 6: potassium acetate.

-Page 19, 4th line from bottom: Is the Xi programm available on the Web?

-Page 20, line 9: The authors should give details on the Xaminatrix program. This software should be shared with the community as there is an urgent need to improve exisiting software tools for analyzing chemical cross-linking data.

Figures:

Figure 2D: The protein standard does not cover the high molecular weight range making it impossible to judge the apparent molecular weight of the cross-linked Pol II/TFIIF complex.

1st Revision - authors' response 09 December 2009

Response to reviewer's comments

Referee #1(Remarks to the Author):

This study shows that a combined crosslinking/mass spec approach can be used to make predictions about the structures of large, multisubunit complexes such as the TFIIF-pol II complex. The analysis is obviously very complicated, but it can help make connections in cases where partial structures are available. This paper nicely complements the accompanying paper from the Hahn lab. However, I found that parts of the discussion were perhaps overly speculative and that the main figure (4) was very hard to read. If these issues can be addressed I think the paper would be appropriate for the EMBO Journal.

Comments and suggestions:

p10 The discussion of the one crosslink of the dimerization domain that doesn't fit the model is confusing. First, it's never stated which residues are involved, so it's impossible to look at the figures and see just how far away they are. Are they shown in Fig 4A? Second, the text cites some data from Chen et al 2007 that may be consistent with the unsatisfied crosslink, but it's never explained what those results are. I took a quick look at the Chen paper and couldn't find anything obvious. If the crosslink here and in the Chen paper suggest an alternative position, it could be worth creating a supplementary figure to show it. Finally, it's never discussed whether this unsatisfied crosslink might actually be due to a mistake in the modeling of the dimerization domain (which is the yeast TFIIF modeled after a human TFIIF structure).

The residues involved in the conflicting cross-link are Tfg1 394 and Rpb2 228, this was added to the text. We have added a supplemental figure (Suppl. Fig. 5) that lists our data and the data of Chen et al 2007 in support of the lobe position of the dimerization domain and displays this data in our model. We have furthermore added a supplemental figure (Suppl. Fig. 6) that lists our data and the data of Chen et al 2007 in support of an alternative position of the dimerization domain (panel C) and displays this data in the alternative model. Tfg1 394 is near to the C-terminal boundary of the dimerization domain as predicted by us (residue 400 in S. cer) and Chen et al 2007 (residue 416 in S. mik, equivalent to 412 in S. cer). The residue is found to cross-link within the dimerization domain (Tfg2 127, Tfg1 400) and the homology model is consistent with these cross-links. Also, assuming Tfg1 394 being part of the charged region extending into the cleft from the lobe position of the dimerization domain moves the C-terminal boundary of the domain further upstream then just residue 394, to allow the Tfg1 chain to enter into the cleft for Tfg1 394 to reach a position proximal to Rpb2 228. This would affect likely two beta sheets with reasonable conservation to human and leave open the question how Tfg1 394 could cross-link from a position deep in the cleft to Tfg2 127 in the dimerization domain near the lobe. We feel that the best explanation for our data is to see Tfg1 394 as part of the dimerization domain and to propose two alternative positions. As also the Tfg2 winged helix domain is seen in more than a single position the binding of parts of TFIIF to Pol II appears to be possible in more than a single arrangement. These arrangements may reflect different phases the Pol II complex.

p13 I found the attempt to connect the Tfg1 winged helix linker with Tfg3, and then Tfg3 to the CTD linker, to be rather tenuous and not very convincing. This is especially true given a recent structure of the S. pombe pol II from Kornberg that extends the CTD linker out onto Rpb4/7.

The text was modified.

p14/15 Similarly, most of the discussion concerning how TFIIF might function in open complex formation and early elongation was not really connected to the data here, since it involves unstructured regions of the protein. I found this part to be overly speculative and I recommend trimming it in length and level of detail.

The text was modified.

p16 URA1 should be URA3

The text was modified.

Fig 1 It would be helpful to denote the domains on the protein "bars" (as is done in Fig 3).

We presume the comment is directed at Fig 2 and have added there the domain structure of TFIIF.

Fig 4 and p 10. It seems to me that the "alternative" location for the Tfg2 WH is not being taken very seriously, surprising since this may be in more agreement with the previously published EM studies. I think it would be reasonable to show both positions, especially if they can be more specifically docked. If not, a better explanation for what the functional implications of the alternative site are is needed. This could be easily expanded into a separate figure (see below).

It is not possible to dock the Tfg2 WH since the EM map was not deposited to the data bank. A mention regarding the agreement of the location of this domain with the previous EM work was added to the text. As suggested by the referee, we added supplementary figures 4-7 to illustrate the 2 possible locations of the Tfg2 WH and the data supporting them.

Fig 4 This is the most important figure of the paper, yet I found it very difficult to read. There is a lot of information here, but the labeling is very inconsistent and the figure legend doesn't contain much in the way of explanation. What are the spheres (the entire side chain, the nitrogens?). On the top figures, what do the numbers in parentheses after the residues K137 and K149 mean? In the blowups, why is the format for polymerase residues different from those in TFIIF? Finally, if it's not too dense, it would be nice to see the crosslinks drawn in as done in the other figures. I would strongly recommend breaking this figure up into multiple figures: one dealing with the Tfg2 WH, one for dimerization, and a third with the integrated PIC model. With improved figure legends, the authors' points will be made much clearer.

The figure was rearranged and broken into Fig. 4 and Fig 5 and additionally Supplemental Figs 4, 5, 6, and 7. In Fig 4, a zoom was added that displays the experimental data used to position the dimerization domain of the Pol II surface. This includes dashed lines between those residue pairs found to cross-link. The spheres show alpha carbon atoms of cross-linked residues which is now added to the figure legend. Labelling inconsistencies within the figure have been removed and the same way of labelling is now used throughout the paper. The Tfg2 winged helix positions are shown as a new supplemental figure (Suppl. Fig. 7).

Referee #2 (Remarks to Author):

The manuscript submitted for publication by Chen et al. investigates the architecture of a complex made of RNA polymerase II and TFIIF general transcription factor. Together these two proteins consist of 15 different polypeptides totalling 670 kDa. TFIIF plays a key role in preinitiation complex (PIC) formation since it associates with RNA polymerase II prior their incorporation in PIC. TFIIF is also important for transcription start site selection. How TFIIF and Pol II interact is an important because there are conflicting reports about it.
A very recent (2007) method that was devised and first applied to the structure of the Ndc80 complex (176 kDa) is used to investigate this question. It entails the cross-linking of lysines and protein N-termini with the BS3 cross-linker and the identification of the cross-linked peptides by mass-spectrometry. The distance between cross-linked C-alpha is around 28 angstro&#xms. The method was first validated on Pol II, using the known structure, identifying around 100 contacts. The false positive error rate was estimated to be around 1%. The method is then applied to the analysis of the Pol II-TFIIF complex. Around 400 linkage sites were identified of which 182

connected Pol II to TFIIF. The structure of the dimerization domains of Tfg1 and Tfg2 and of the winged helices of the two subunits was modelled. These models and the structure of Pol II were used to model the Pol II-TFIIF complex using the linkages as constraints. The protein complex model indicates that the dimerization domain is positioned on Pol II lobe while Tfg2 winged helix crosslinks to the Pol II protrusion upstream face. Tfg2 linker domain extends along the Rpb2 protrusion. Tfg1 winged helix cannot be cross-linked and thus thought to be mobile. The model predicts that some of the interactions between Pol II and TFIIF prevent the non-specific interaction between the enzyme and DNA, a known property of TFIIF. These observations, using a totally different technology, are consistent with those made by Chen et al. (Genes Dev. 2007) but not with the model of the complex proposed by Chung et al. (Mol. Cell 2003) based on EM analysis.
This work brings a wealth of information on the Pol II-TFIIF complex and strongly argues that the EM model is incorrect. The method used is very interesting since it allows investigating the structure and organization of very large complexes like those involved in transcription. It is almost certain that this method will be very useful for the analysis of complexes hard to crystalize or as an independent tool to assess structure and models. This manuscript should be published in The EMBO journal.

Comments and suggestions:

-p7: The description of how the Pol II-TFIIF complex is prepared is somewhat confusing. I would eventually drop the problems in expressing TFIIF in E. coli and low yield of complexes obtained by TAP-tagging Tfg2. I would make clear that the complex was affinity-purified form a strain over-expressing the three TFIIF subunits.

Due to the long history of Pol II-TFIIF preparation trials, in particular the toxicity of TFIIF subunits in E.coli and low yield of endogenous Pol II-TFIIF, we would like to leave this part. However the text was modified for clarity.

-p7: The RNA extension assay should be briefly described.

A description of the RNA extension assay was added to the method section

-p28: The Figure 2C legends indicates "Pol II-TFIIF complex shows normal elongation activity (Methods)". There is nothing in the Methods section. My remark is also related to the previous one on the RNA extension assay.

See before

-Figure 1E labels are difficult to read.

The labels in Fig 1E have been changed to increase readability.

-Is it possible to estimate the number of contacts that should have been observed in Pol II and that were not (false negative rate)?

This is a very interesting proposal and has also crossed our minds. The problems are in how to (1) predict which residues could actually cross-link and (2) which of the cross-linked peptides could be detected in our analysis. Predicting the reactivity of an individual residue requires knowledge of how the micro-environment affects the reaction. This is currently poorly understood in protein chemistry at large and would be highly desirable also in the context of predicting non-covalent interactions of small molecules with proteins. Once site-specific reactivities are established, one needs to determine the chance of two sites being linked together. This could presumably be addressed through molecular dynamics calculations in the context of the entire complex. Finally, we would need to estimate the site specific yield in our enzymatic proteolysis using trypsin, individual loss of peptides during the sample handling for example due to adsorption on tube surfaces, the ionization efficiency and matrix effect during mass spectrometric analysis and finally the fragmentation behaviour of the cross-linked peptide and how this affects our success in identifying it. The "detectability" of peptides in mass spectrometry is a current research area in itself, linked with large interests in the biomarker and other proteomics fields. Currently, there is not a clear solution for linear peptides available and we have thus not looked into this issue for

cross-linked peptides. We therefore refrained from predicting which cross-links we should see until such time that all these issues have been addressed.

Referee #3 (Remarks to Author):

General:
This manuscript describes the investigation of the RNA polymerase II (Pol II)/TFIIF complex structure using chemical cross-linking, charge-based enrichment of cross-linked species, and high-resolution (LTQ-Orbitrap) mass spectrometric analysis. The complex possesses a molecular weight of ~ 670 kDa and presumably is the largest assembly that has been stued by a chemical cross-linking/MS approach to date. This paper presents an impressive example of the potential of the cross-linking approach for studying large multiprotein assemblies.

General Comment:

-Supplementary table file 24249_0_supp_254411_krgf1l has not been correctly converted into a pdf file - at least in the version I got.

We checked all files after up-load for correct conversion and did not observe any problems. We are sorry for the difficulties experienced by the reviewer.

-In this report, the authors employ one single amine-reactive cross-linker (BS3) with a spaer length of ca. 11 Angstrom. Did the authors also consider using additional cross-linkers with different specificities and different specer lengths?

Using alternative cross-linkers will be an interesting approach to increase the resolution of our data once the entire analysis pipeline has been further optimized.

Moreover, N-hydroxysuccinimide (NHS) esters, such as BS, are quite easily hydrolzyed at one site giving so-called "dead-end" cross-links. These cross-linking products do not yield any distance constraints within the protein complex; yet, they provide useful information on the solvent accessibility of the proteins in the complex. The authors should comment whether they did not search at all for these "dead-end" products or (in case they did) how many of these species were obtained. These products could be included into the model.

Our sample preparation (using SCX fractionation) and mass spectrometric acquisition method (focussing at highly charge precursors) aim at enriching for cross-linked peptides. This is at the expense of linear peptides, including peptides modified by the hydrolyzed cross-linker. Testifying the success of our enrichment, we have identified a very small number of cross-linker modified peptides. We therefore did not further investigate their value for our models.

Specific Comments:

-Page 6, lines 6/7: N-hydroxysuccinimide esters, such as BS3, have been found to frequently react with hydroxyl groups of serines, threonines, and tyrosines (Kalkhof S, Sinz A, Analytical and Bioanalytical Chemistry 2008, 392, 305; Mädler S, Bich C, Touboul D, Zenobi R, Journal of Mass Spectrometry 2009, 44, 694). On page 20, the authors state that no linkage specificity was assumed for BS3, but that all fragments observed contained a lysine or N-terminus as most likely modification site. This is somewhat surprising and authors should comment on this.

We cannot exclude that other residues than Lysine or the protein N-termini were involved in cross-linking in our analysis of the Pol II complex. However, the abundance of such cross-link products was below the sensitivity of our analysis. The low yield of such alternative reaction sites can be explained by the reactivity of different amino acid side chains towards N-hydroxysuccinimide esters, such as BS3, being largely influenced by the pH of the buffer. It has been shown that at pH 6.0 the reactivity is greatly different from pH 8.4 where primarily N-termini and lysine side chains react (Leavell MD, Novak P, Behrens CR, Schoeniger JS, Kruppa GH. J Am Soc Mass Spectrom. 2004 Nov;15(11):1604-11.). Our work was conducted at pH 8.0 while work cited by the reviewer utilized pH 7.4 (Kalkhof S, Sinz A, Analytical and Bioanalytical Chemistry 2008, 392, 305). In a recent presentation of one of the authors, Andrea Sinz, reactivity under these conditions was given

as 15% towards Serine. How this would scale to pH 8.0 is unknown. The other work cited by the reviewer (Madler S, Bich C, Touboul D, Zenobi R, Journal of Mass Spectrometry 2009, 44, 694) investigates the influence of pH and neighbouring residues on the reactivity of Lysine, Arginine, Serine, Threonine and Tyrosine. The authors of that study did not find evidence of N-hydroxysuccinimide esters reacting with Arginine, Serine, Threonine and Tyrosine. However, this changed when a Histidine residue was in -2 position. A reaction mechanism was proposed that excludes the possibility that a Histidine-mediated reaction with N-hydroxysuccinimide esters takes part in a cross-link reaction with another residue. The Histidine catalysis bases on Histidine being part of a cross-link that is hydrolyzed readily in water.

-Page 6, line 9: It is not completely clear why the authors assume a CA-CA distance of 27.4 Angstrom for BS3.

The amines must be less than 11.4 Å apart, the maximal length of the BS3 spacer. Adding 16 Å to this, two times the length of a lysine side chain (6-6.5 Å for 1WCM) including an estimated coordinate error (1.5 Å for 1WCM), defines the maximal C-alpha distance of linkable lysine residues, 27.4 Å, when comparing our cross-link data with the available crystallographic data.

-Page 6, last paragraph: It has been shown in MD simulations that distances larger than 19 Angstrom are highly unlikely for a reagent with a spacer length of ca. 11 Angstrom (Green NS, Reisler E, Houk KN, Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers, Protein Science, 10, 1293-1304 (2001)). Experimentally, it has been shown (Ye X, et al., Probabilistic cross-link analysis and experiment planning for high-throughput elucidation of protein structure. Protein Sci, 13, 3298-313 (2004)) that distances of larger than 19 Angstrom are indeed bridged with signifcantly lower frequency. These references should be included in this paragraph.

Importantly, we are considering the maximal possible distance to identify a boundary at which our cross-link data deserves special questioning as it apparently disagrees with crystallographic data. Both references consider this to be 11.4 Å for the cross-linker used in our work, BS3. Green et al. determine in their MD calculations the median spacer length as 9.2 Å, with a range of 5.58-11.42 Å. Ye et al. - in full agreement with our estimations - consider the maximal possible alpha-carbon distance in cross-linked lysine pairs to be 24 Å (11.4 Å from the cross-linker + 2x 6.2 Å for the lysine side chains). We arrive at 27.4 Å using the actual observed maximal length of a lysine side chain in our reference structure (6.5 Å) and considering the coordinate error (estimated as 1.5 Å based on the crystallographic B-factor for Lysine alpha carbon atoms). Based on the work of Green et al. and own assumptions Ye et al. also define a second boundary of 19 Å. They feel this is justified as Young et al. (PNAS 97, 5802 (2000)) have observed a cross-link yield of 31% below and 24% above this value. This difference in yield appears in our eyes not to testify a "significantly lower frequency", especially keeping in mind the amount of data it bases on (18 cross-links in total). In our data, 79 (75%) links fell below 19 Å and 95 (90%) fell below 23 Å. The vast majority of our links therefore take place within the different boundaries defined by others. It will be interesting to discuss if our observations contradict the theoretical considerations or support them. However, we do feel that such a discussion does not fit the objective of our current manuscript and should be directed at a more technical audience.

-Page 7, line 3: Although the authors explain on page 20 how decoy searches were performed, they should add one sentence here on how the error rate of 1% was determined.

Our estimated error rate does in fact not rely on the decoy approach but uses the final results as validated by the crystal structure. We changed the sentence to clarify this into:" ..., with an experimentally determined error rate in the order of 1%, with one of the 106 observed cross-links being false."

-Page 8, line 4 from bottom: The authors should be more specific on how the modeling of the Pol II/TFIIF complex was performed. Unfortunately, in the Materials and Methods part nothing is stated on the modeling programs and the methods used. More information is clearly needed here.

The text was modified accordingly.

-Page 10, first & second line: The statement that "enough spatial restraints were avaiable to position and orient the dimerization domain" is very ambiguous. The authors should clearly state how many cross-links were acutally employed to determine the orientation of the dimerizaton domain. From the Supplementary Table it is very difficult to get this information.

We have added a new panel in Fig. 4 to display the experimental constraints in the model. This data is summarized in the new supplementary table 5.

-Page 10, line 16: The authors claim that the Tfg1 WH domain is apparently mobile as no cross-links to Pol II were found. Yet, one of the greatest strengths of the chemical cross-linking approach consists in that information also on flexible regions within a protein is readily obtained. Admittedly, fewer cross-links are found in flexible protein regions compared to very rigid regions, but one would have expected to find at least some cross-links between the Tfg1 WH domain and Pol II. The authors should comment on this.

The impact of mobility on cross-linking yields is a highly interesting and very important subject. Obviously there must be a limit to how short-lived a proximity is allowed to be in order to still be captured by a cross-linking reaction. In this context it is worth noting that in vivo work in connection with ChIP could establish recently a temporal threshold for formaldehyde crosslinking and fixation (Schmiedeberg L, Skene P, Deaton A, Bird A. PLoS One. 2009;4(2):e4636.) There is hence an upper limit to cross-linking speed. It remains to be seen into what time scale this translates for other chemistry such as N-hydroxysuccinimide esters used in our here presented study. Our results on the Tfg1 winged helix domain indicate, however, that a domain being held via a linker alone to a complex is not necessarily cross-linked at detectable levels to the complex. In contrast, the Tfg2 winged helix domain is excessively linked to the wall and clamp of Pol II. Equally, the Tfg2 linker and C-terminal regions are cross-inked to the same region of Pol II as the Tfg2 winged helix domain. In many instances two or even all three Tfg2 regions share the same linkage site on Pol II. This is consistent with all three regions of Tfg2 forming a flexible C-terminal tail of the protein that interacts dynamically with Pol II. There is an alternative explanation, however. As we are analyzing a population of complexes and have not isolated individual conformations it is possible that the Tfg2 C-terminal tail takes a series of distinct positions on Pol II. From a methodological point of view this is a formidable research question as the two winged helix domains of TFIIF may well serve as a prototypical example for dynamic versus random interactions, even strengthened by the random part of the equation (Tfg1 winged helix - Pol II) being forced into close proximity. The functional importance of the Tfg2 interaction patch on Pol II is unclear, however. Also worth adding is that the Tfg1 winged helix domain is not the only apparently flexible region in the complex that does not cross-link with the rest of the complex. Also the CTD is not observed with a single cross-link to the rest of Pol II or TFIIF. This is consistent with a disordered conformation of the CTD and a minimum of order being required for cross-linking to proceed.

-Page 17, last line and page 18, line 6: potassium acetate.

Corrected.

-Page 19, 4th line from bottom: Is the Xi programm available on the Web?

The Xi programme is in development and currently not available. It should be pointed out that one of the largest groups in proteomics, Ruedi Aebersold, has recently published a methods paper reporting on a programme, xQuest, to identify cross-linked peptides even in very complex mixtures like whole cell lysates (Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. Nat Methods. 2008 Apr;5(4):315-8. Epub 2008 Mar 9. Erratum in: Nat Methods. 2008 Aug;5(8):748.). Furthermore, GeneBio has launched a new product as part of the Phenyx protein identification suite that makes use of the approaches developed and published earlier by us (Maiolica A, Cittaro D, Borsotti D, Sennels L, Ciferri C, Tarricone C, Musacchio A, Rappsilber J. Mol Cell Proteomics. 2007 Dec;6(12):2200-11.). This means that large academic labs as well as commercial entities have provided solutions to this problem or are at least in the process of doing so.

-Page 20, line 9: The authors should give details on the Xaminatrix program. This software should be shared with the community as there is an urgent need to improve exisiting software tools for

analyzing chemical cross-linking data.

We hope that the Xaminatrix programme will be a useful addition to the community and work on making it transferable. In its current form it relies on our network architecture and will not run in any other environment. Fortunately, many labs involved in cross-link methods development are writing on some sort of software to address the needs of spectra annotation. Also, this problem is addressed in some form as part of the two programmes mentioned above, xQuest and Phenyx.

Figures:
Figure 2D: The protein standard does not cover the high molecular weight range making it impossible to judge the apparent molecular weight of the cross-linked Pol II/TFIIF complex.

We are aware of this shortcoming and regret that no molecular weight marker is currently available for denaturing gel electrophoresis that extents up to 670 kDa. For native gel electrophoresis markers extend to this and higher molecular weight. In Fig. 2E we present a native gel of the cross-linked Pol II-TFIIF complex together with a marker and the cross-linked Pol II complex. This allows some estimation of the molecular weight. Most importantly, the formation of aggregated complexes as an artefact of cross-linking can be excluded.

