THE EMBO JOURNAL

Manuscript EMBO-2014-88638

Architecture of the *S. cerevisiae* SAGA transcription coactivator complex

Yan Han, Jie Luo, Jeff Ranish and Steven Hahn

Corresponding author: Steven Hahn, Fred Hutchinson Cancer Research Center

Review timeline:

Submission date:	01 April 2014
Editorial Decision:	30 April 2014
Revision received:	14 July 2014
Editorial Decision:	31 July 2014

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.)

Editor: Alexander Kohlmaier

1st Editorial Decision 30 April 2014

Thank you for submitting your manuscript "Architecture of the S. cerevisiae SAGA transcription coactivator complex" for consideration to the EMBO Journal. We have now received comments of three expert referees that you will find pasted below.

I am pleased to inform you that all of the referees find your results of interest and importance. Pending adequate addressing of a limited number of specific concerns, we shall therefore be happy to consider a revised version of your manuscript further for publication in The EBMO Journal.

The referees' comments appear constructive and self-explanatory, and I will not repeat them in detail here. In terms of additional experimental work, referee #1 suggests to use a more conservative/minimal Sgf73 N-terminal mutation to more specifically assess the selective effect of the absence of the DUB module independently of interfering with the core of the SAGA complex. Moreover, referees #1 and #2 (and referee #3 in more general terms) concur that the description and quality assessment of crosslinks, including minimum requirements for cross-link identification, peptide length and maximum mass errors for MS1 and MS2 etc., should be very clearly listed. Referee #3 most clearly articulates which experimental and technical parameters should be mentioned or re-considered. This pertains also to the unexpected low number of crosslinks with the DUB module subunits, for which referee #1 requests quantifications of Ubp8, Sgf11 and Sus1 in the starting material to rule out more trivial explanations.
Finally, while we feel that trying to potentially reconcile differences between the previous and the current models, as suggested by referee #2's insightful comments, will likely make the manuscript

even more compelling, we will, however, not insist on that you would have to account for all differences in the form of a comparison figure, if you feel that this would be a digression.

Together, I am certain that modifications requested by the referees would result in an improved study, as already indicated by some of the referees' encouraging remarks. I would, therefore, be pleased if you invested the necessary time and efforts to address the reviewers' concerns.

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

Referee #1:

The authors provide a comprehensive view of the interconnectivity between the 19 subunits of the yeast SAGA coactivator complex. The conserved role of SAGA in transcription activation across eukaryotes and the importance of its role in activating mRNA transcription has made this a paradigm for understanding gene regulation and is thus of broad interest to the readership of EMBO J. While the proteins in this 1.9 MDa complex have previous been studied by mass spec analysis of SAGA with single subunit deletions, the authors take a fresh approach to studying SAGA connectivity and SAGA interactions with TBP by cross-linking SAGA subunits and then mapping the positions of peptide cross-links. This enables them to arrive at a far more detailed few of intersubunit interactions that provides a foundation for biochemical and structural studies. In addition, the authors analyze reciprocal effects of deletions in the two enzymatic modules (subcomplexes): the HAT (histone acetyltransferase) module, which contains GCN5, and the DUB (deubiquitinating) module, which contains Ubp8, and study effects on both subunit retention and HAT activity. Together, these results lead the authors to conclude that these enzymatic modules are near one another and that the DUB module modestly impacts HAT activity on nucleosomes. Finally, the authors attempt to pull together their results with the use of structures and homology models to propose a more detailed organization of the SAGA complex than previous studies.

While I very much liked the overall approach and found the writing and figures particularly lucid, I have a few significant issues with the data and interpretation.

1) There are remarkably few cross-links with any of the DUB module subunits except for Sgf73, which makes me wonder whether the remaining three subunits are present in stoichiometric amounts. In particular, both EDTA and EGTA (0.5 mM) were used in the TAP purifications, yet the DUB module is known to be sensitive to EDTA, which can strip out the bound zincs and cause Ubp8, Sgf11 and Sus1 to dissociate from Sgf73 (Samara 2010 Science). The gel in Fig. 1A was not blotted for any of the DUB module subunits and the complex in Fig. 5B was pulled down in the absence of EDTA (if I read the methods correctly). The iBAQ experiment Fig. 6B shows substoichiometric amounts of Ubp8 and Sgf11 (but curiously, much more Sus1) - does this reflect the amount in the complex in the cross-linking experiments? The authors should provide direct evidence that their starting material contained Ubp8, Sgf11 and Sus1 before the cross-linking experiments were done. They should also discuss the number of peptides that were recovered for these proteins, as it is possible that few cross-links were observed because of poor coverage. Coverage for Tra1 should also be commented upon, since this similarly has few cross-links with other subunits.

2) The "delta DUB' mutations are misinterpreted as being due to the loss of the four DUB module proteins, whereas they instead appear to be due to a disruption in the SAGA core complex. The authors create their delta DUB by studying internal deletions of Sgf73, whose N-terminus is part of the DUB module and whose C-terminus is integrated into the remainder of SAGA. The authors show that an Sgf73 internal deletion (Sgf73del2A, which lacks residues 350-400) disrupts the association of this subunit with the remainder of SAGA, without disrupting its association with Ubp8 (the catalytic DUB subunit). This makes sense, given that Ubp8 binds to the Sgf73 N-terminus whereas the deletion lies in a region that is extensively cross-linked to other SAGA subunits. The authors then purify SAGA from yeast strains expressing this version of Sgf73 (2A) in place of the

wild type and interpret their results in terms of the loss of the DUB module. While it is true that this Sgf73 deletion disrupts the association of the DUB module with the rest of SAGA, the residues they have deleted mediate extensive interactions with multiple SAGA subunits (see Fig. 2). Nothing is known about the structure of this part of Sgf73, but the fact that the C-terminal half of Sgf73 forms so many cross links with so many different SAGA subunits suggests that they may have removed a core subunit that either co-folds with other SAGA subunits (in the manner observed for the Sgf73 N-terminus, Ubp8, Sgf11, Sus1) or leaves a hole in the SAGA complex that impacts neighboring subunits including Ada2, which is part of the HAT complex. It is therefore inaccurate to call this version of SAGA a "delta DUB module" mutant, since the deletion is expected to have far broader impacts on SAGA overall.

The authors could test the effects of losing the DUB module, but nothing else, by using a more conservative Sgf73 deletion. Structural and biochemical studies showed that the DUB module itself includes roughly the first 100 residues of Sgf73 (Kohler 2010 Cell, Samara 2010 Science). There is also structural information on the Sca7 domain (Bonnet 2011 EMBO Rep), which is also N-terminal to the internal region deleted by the authors. To test the effect of simply removing the DUB module, the authors should use a minimal Sgf73 N-terminal deletion (see Kohler 2010 Cell). This would leave intact the remainder of Sgf73 and leave its association with other SAGA subunits unchanged.

3) While I appreciated the authors attempts to place their results in the context of their results and that of other labs (biochemical, mass spec, crystal structure, EM), I found Figure 7 somewhat confusing. I don't have any particularly helpful suggestions and do think some sort of figure is important to pull their observations together, but perhaps something a bit less detailed would be clearer.

Minor comment:

(1) The x-axis label of Figure 6C is confusing - perhaps the authors meant "time (min -1)"?

Referee #2:

In this study, Han et al. provide insight into the spatial relationship between the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex and the TATA binding protein (TBP). They utilize crosslinking to stabilize interactions between proteins and perform mass spec proteomic analysis to find regions that interact.

This report is a very interesting and informative resource. We do not feel that additional experiments are necessary, but we do have suggestions for clarifications. It would be useful if the authors clarified some aspects of the model and put them in the larger context of the two other models existing currently.

Significant issues
1. Regarding BS3 crosslinking.
a. There is a simultaneous argument that lack of crosslinking indicates very tight protein-protein interactions, and then an argument that lack of crosslinking indicates no interaction. Please clarify the interpretation of what BS3 crosslinking means. The nature of how BS3 crosslinks is important to the interpretation of the data, yet there are many open questions regarding how the crosslinking works. How drastic is the inhibitory effect of close protein interactions on crosslinking? It is suggested in the text that exposure to BS3 may be limited by close protein-protein interactions. Is the center of a complex cross-linked less than the outside of the complex? Supposedly this is used as argument for saying the Taf6-Taf9 interface is not readily accessible to BS3. Spt8, however, is peripheral, but is also one of the last to crosslink (Fig E1B). The DUBm components are intimately associated yet they fail to crosslink. At the same time we do not see the DUBm on the silver stain, resisting BS3 crosslinking (Fig E1B). In general, an argument is made that a lack of crosslinking indicates that two areas do not interact, but then it is also being considered as a sign that an interface is intimate and buried from access to BS3. In some cases - the DUBm, for example - there is no crosslinking but we know the interfaces are intimate and abundant.

b. Can BS3 crosslink if the distance is too short? A max range is given, but what about a min range? Can interactions be too close to see? Close interactions appear to be invisible to this technique considering the DUBm has been crystalized and the very intimate interactions seen in that crystal structure are not at all shown using the BS3 crosslinking method.
2. Interpretation of the data to create a model and compare it to previous models.
a. The statements "however, this model is at odds with some aspects of an earlier model for SAGA architecture based on EM and immune localization of several subunits" and "...mapping Spt7 and Spt20 within the SAGA EM structure led to an alternative view that suggested these two subunits are positioned at different locations within the complex" are too vague. Describe the differences for the reader.
b. It would be nice to have a comparison figure to go along with your discussion of the three current models. It is somewhat fuzzy to compare the two structural models with the module model which is not really a structure and provides no proposal for a 3D model. Perhaps some nuance is needed there. Phrases like "Tafs at the core" do not really apply to the modular model because the data is not presented in the shape of a complex. Instead, it is a cartoon to represent the probability matrix derived from the deletion analysis. It reveals that the complex is modular and provides some indication about which components in each module may be peripheral and which may be more internal. In all models the TAFs are close to each other, the HAT module is a module, the DUB module is a module, Tra1 appears to sit on the periphery and have limited contact with the rest of the complex, and so on... The way that the complex is presented here makes it hard to say that the TAFs alone are at the "core" while Spt 7 and Ada1 are not. They all seem to make the "core" with the HAT, Tra1, and DUB modules facing outward.
c. "...mapping Spt7 and Spt20 within the SAGA EM structure led to an alternative view that suggested these two subunits are positioned at different locations within the complex" "Consistent with this latter view, our crosslinking results suggest distinct roles for these two subunits in SAGA assembly as Spt7 and Spt20 crosslink to different sets of subunits but do not crosslink to each other." Subunits may be part of a module, as proposed by Lee et al, without every protein in that module directly interacting with the others. Other groups have used co-purification as a proxy for interaction/association. Is this a controversy that does not really exist? These are different types of data which are not necessarily conflicting. Instead, they seem to complement each other. In the case of the DUBm, the components of the module are known to intertwine, interacting intimately. But this does not have to be the only mechanism for formation of a module, especially a module containing many proteins. In that case, the module could be defined by co-purification analysis. This could define the module and even go so far as to say which components of the module are peripheral or central. The inner structure of the module would then be determined by cross-linking/mass-spec analysis. In fact, looking at Lee's model, one could fold it into a 3D orientation and get a structure very similar to what is being proposed in this paper (It is less intuitive to match up this model and Wu's model, but folding Wu's SAGA in half might be close.) Further, comparing the probabilistic protein network from Lee et al, to Figure 2 in this paper are also suggests they are closer than they are far apart. In other words, the models are derived using different techniques and they give different information. These data, combined with those from previous studies should, together, provide evidence that allows one to make a more accurate model than any of the techniques could make on their own.
d. Can you comment on the different shape of your model and Wu's model, even though it was created using an immunological technique at a time when the complex was less-well characterized (i.e. missing important components)?

Minor issues
3. Tafs and TFIID
a. A lot of emphasis is made on the similarity of the Tafs in SAGA and in TFIID. If this is a TFIID-like core, it would be interesting to compare it to the comparable TFIID section. In this work TBP is not associated with the TAFs. Instead, TBP is associated with Spt8 and Spt3 as shown in previous studies. It would be helpful to comment more explicitly on the relationship between TBP and the TAFs in TFIID and what functional consequences are implied by this relationship between SAGA and TBP. Do both complexes position TBP in the same relative position relative to the TAFs? What implication does this have for the potential for shared roles between SAGA and TFIID?

Referee #3:

In this manuscript the authors report the topological analysis of the transcriptional coactivator complex, SAGA, by combining chemical cross-linking and mass spectrometry. The budding yeast 1.8 MDa SAGA complex is composed of 19-20 subunits and previous analysis have revealed several modules involved in activator binding, TBP binding, histone acetylation and deubiquitination. These modules contribute activities for transcription initiation and early elongation events for a large set of inducible and developmentally regulated genes. The cross-linking analysis provided novel insights into the architecture of this complex indicating which subunits and domains form the different modules and how they are organized around a TFIID like core complex. The structural data were discussed with respect to previous biochemical and EM analyses and gave rise to mutational analysis revealing domains that mediate protein-protein interactions and suggesting a functional interaction between the DUB and HAT module which may be important for regulating its histone acetylation activity during transcription initiation.

This study provides novel insights into the subunit architecture of the SAGA complex, resolves conflicts arising from previous topological and interaction studies and also reveals functional implications, thus, I recommend this manuscript for publication in EMBO Journal.
As most of the follow-up experiments and conclusions are solely based on the structural interpretation of the cross-link derived distance restraints and at the same time, the description and quality assessment of the cross-link identifications are superficial I request major revision of this part of the manuscript. Judging the quality of the cross-linking analysis is not possible in the current form.

1. The authors state that they have identified 199 intermolecular and 240 intramolecular crosslinks. However, in Table S1 there are several redundant cross-link identifications indicating the linkage of the very same lysine residues. The actual number of unique cross-links, indicating the linkages of unique lysine-lysine pairs, is smaller. The reviewer recommends to distinguish between total number of cross-link identification, unique cross-links and unique distance restraints (unique peptide-peptide pairs) that have actually been used for the structural interpretation.

2. What are the minimum requirements for a positive cross-link identification? What is the minimum peptide length? What are the maximum mass errors for MS1 and MS2 yielding a valid/confident cross-link identification? Do all cross-link identifications display 3 consecutive ions even if there are more than 3 per peptide? MS/MS spectra of different cases for cross-link validation or of structurally/functionally important cross-links should be shown in the supplement.

3. What are the FPRs of the individual cross-link experiments?
Indicating which individual cross-links have been found more than once would significantly increase the quality assessment of the data?

4. In the methods section, the authors state that precursors with charge state +3 were rejected from MS/MS analysis. The few studies that actually report the charge state distribution for identified cross-links indicate that there is a significant number of cross-link identifications with a precursor charge state of +3. Researching the data and thereby including +3 precursors may increase the number of cross-links and distance restraints.

5. What is experiment "8" in Table S1. Does this indicate cross-linking at 8 mM BS3?
In addition to displaying some cross-link identification criteria (score, FPR, mass error) providing a legend for Table S1 will aid in understanding the data.

Minor comments:
page 4: ...molecular connections... "better" ...molecular interactions...
page 11: Our cross-linking... "better" ... Our crosslink derived distance restraints
page 15: Our cross-linking... "better" ... Our cross-link data "or" distance restraints
page 16: ...Spt3, has numerous crosslinks... "better" ... displays "or" gives rise to
page 17: One possibility to explain... "better" One explanation of this inhibitory effect...
page 17: Given the crosslinking results... "better" Given the domain topology derived from chemical cross-links

page 18: Our crosslinking shows that Spt20 crosslinks to either side of the Spt3 histone fold domain that has been implicated in TBP binding. "rephrase sentence"

1st Revision - authors' response 14 July 2014

First we thank the reviewers for their careful reading of the manuscript and for their constructive comments. Our response to each of the specific points raised is detailed below:

Referee #1:
1) There are remarkably few cross-links with any of the DUB module subunits except for Sgf73, which makes me wonder whether the remaining three subunits are present in stoichiometric amounts. In particular, both EDTA and EGTA (0.5 mM) were used in the TAP purifications, yet the DUB module is known to be sensitive to EDTA, which can strip out the bound zincs and cause Ubp8, Sgf11 and Sus1 to dissociate from Sgf73 (Samara 2010 Science). The gel in Fig. 1A was not blotted for any of the DUB module subunits and the complex in Fig. 5B was pulled down in the absence of EDTA (if I read the methods correctly). The iBAQ experiment Fig. 6B shows substoichiometric amounts of Ubp8 and Sgf11 (but curiously, much more Sus1) - does this reflect the amount in the complex in the cross-linking experiments? The authors should provide direct evidence that their starting material contained Ubp8, Sgf11 and Sus1 before the cross-linking experiments were done. They should also discuss the number of peptides that were recovered for these proteins, as it is possible that few cross-links were observed because of poor coverage. Coverage for Tra1 should also be commented upon, since this similarly has few cross-links with other subunits.

To address why few crosslinks were observed within the enzymatic portion of the DUB module, we analyzed MS data for peptide coverage in both the uncrosslinked and crosslinked samples. First, we consistently observed greater than 70% sequence coverage for the majority of SAGA subunits (including DUB subunits Ubp8 [>80%], Sgf11 [77.8%], and Sus1 [85.4%]; Sgf73 only shows 50.5% sequence coverage) upon MS analysis of several independent SAGA preparations that were not treated with crosslinker. These preparations were made using the same method as was the SAGA used in the crosslinking experiments. Second, we found 20%-30% sequence coverage for unmodified peptides from Sgf73, Ubp8 and Sus1 in the MS data from the crosslinked samples (2mM BS3 experiment), directly showing that these subunits were present in the SAGA preparation that was used in the crosslinking experiments. We did not observe any unmodified peptides originating from the small DUB subunit Sgf11 in the crosslinked samples, and we speculate that this may be due to the fact that we excluded +1, +2, +3 ions in our analyses.

It is true that iBAQ analysis of our uncrosslinked SAGA preparations indicates that Ubp8 and Sgf11, but not Sus1, are present at slightly lower levels (~2-fold) compared to most of the other SAGA subunits. This may explain in part why Ubp8 and Sgf11, but not Sus1, yielded few crosslinks. A further consideration for Sgf11 is that it only contains 2 crosslinkable lysine residues. To summarize, our SAGA prep used for crosslinking analysis contains all DUB subunits, but Ubp8 and Sgf11 may be present at slightly substoichiometric levels. This does not change our model as placement of the DUB module at a peripheral location is consistent with both our the crosslinking and deletion data, as well as the results of previous studies.

As for Tra1, we observed 77%-80% sequence coverage in the uncrosslinked samples, and 20%-30% sequence coverage in the crosslinked samples. Thus, the reason why Tra1 has few crosslinks with other subunits is not due to poor coverage in our analysis.

In the revised manuscript, we have listed the sequence coverage in **Table E1**.

2) The "delta DUB' mutations are misinterpreted as being due to the loss of the four DUB module proteins, whereas they instead appear to be due to a disruption in the SAGA core complex. The authors create their delta DUB by studying internal deletions of Sgf73, whose N-terminus is part of the DUB module and whose C-terminus is integrated into the remainder of SAGA. The authors show that an Sgf73 internal deletion (Sgf73del2A, which lacks residues 350-400) disrupts the association

of this subunit with the remainder of SAGA, without disrupting its association with Ubp8 (the catalytic DUB subunit). This makes sense, given that Ubp8 binds to the Sgf73 N-terminus whereas the deletion lies in a region that is extensively cross-linked to other SAGA subunits. The authors then purify SAGA from yeast strains expressing this version of Sgf73 (2A) in place of the wild type and interpret their results in terms of the loss of the DUB module. While it is true that this Sgf73 deletion disrupts the association of the DUB module with the rest of SAGA, the residues they have deleted mediate extensive interactions with multiple SAGA subunits (see Fig. 2). Nothing is known about the structure of this part of Sgf73, but the fact that the C-terminal half of Sgf73 forms so many cross links with so many different SAGA subunits suggests that they may have removed a core subunit that either co-folds with other SAGA subunits (in the manner observed for the Sgf73 N-terminus, Ubp8, Sgf11, Sus1) or leaves a hole in the SAGA complex that impacts neighboring subunits including Ada2, which is part of the HAT complex. It is therefore inaccurate to call this version of SAGA a "delta DUB module" mutant, since the deletion is expected to have far broader impacts on SAGA overall.

The authors could test the effects of losing the DUB module, but nothing else, by using a more conservative Sgf73 deletion. Structural and biochemical studies showed that the DUB module itself includes roughly the first 100 residues of Sgf73 (Kohler 2010 Cell, Samara 2010 Science). There is also structural information on the Sca7 domain (Bonnet 2011 EMBO Rep), which is also N-terminal to the internal region deleted by the authors. To test the effect of simply removing the DUB module, the authors should use a minimal Sgf73 N-terminal deletion (see Kohler 2010 Cell). This would leave intact the remainder of Sgf73 and leave its association with other SAGA subunits unchanged.

To directly address these comments, we made the Sgf73 mutation suggested by the Reviewer. This mutation, termed Sgf73 Δ4, removes residues 2-104 and causes release of subunits Ubp8, Sus1 and Sgf11 from SAGA while retaining the Sgf73 subunit (**Fig 5, 6, E5**). This mutant SAGA complex was characterized as before by MS and HAT assays. In contrast to the more severe Sgf73 Δ2A mutation, we found that simply disrupting the enzymatic portion of the DUB module with Sgf73 Δ4 had a modest inhibitory effect on HAT activity (**Fig 6, Fig E5**). Our combined results suggest that the Reviewer's interpretation of the Sgf73 Δ2A mutation was correct: release of Sgf73 from SAGA likely causes an indirect effect on HAT function due to changes in protein-protein interactions within SAGA. Our new result clearly shows that the enzymatic portion of the DUB module has a modest stimulatory effect on HAT activity. The text (pages 11-12, 14) and figures 5, 6, & E5 have been revised to incorporate these new results. Thank you for suggesting this experiment.

In addition, we have changed the name of the mutant SAGA complexes as suggested by the Reviewer to better reflect the mutations present in the complex, and to avoid any potential misunderstanding.

3) While I appreciated the authors attempts to place their results in the context of their results and that of other labs (biochemical, mass spec, crystal structure, EM), I found Figure 7 somewhat confusing. I don't have any particularly helpful suggestions and do think some sort of figure is important to pull their observations together, but perhaps something a bit less detailed would be clearer.

To address the Reviewer's concern, we have created a movie of the 3D SAGA model and this is now included as supplementary material (Movie E1). We think that the movie makes it easier to understand the model for SAGA architecture derived from our experiments. We think that it is valuable to have the model as detailed as possible (as long as it is supported by experimental data) as this will aid future structural and biochemical studies aiming to give a higher resolution picture of SAGA.

Minor comment:

(1) The x-axis label of Figure 6C is confusing - perhaps the authors meant "time (min -1)"?

This label has been corrected to: Time (min)

Referee #2:
This report is a very interesting and informative resource. We do not feel that additional experiments are necessary, but we do have suggestions for clarifications. It would be useful if the authors clarified some aspects of the model and put them in the larger context of the two other models existing currently.

Significant issues
1. Regarding BS3 crosslinking.
a. There is a simultaneous argument that lack of crosslinking indicates very tight protein-protein interactions, and then an argument that lack of crosslinking indicates no interaction. Please clarify the interpretation of what BS3 crosslinking means. The nature of how BS3 crosslinks is important to the interpretation of the data, yet there are many open questions regarding how the crosslinking works. How drastic is the inhibitory effect of close protein interactions on crosslinking? It is suggested in the text that exposure to BS3 may be limited by close protein-protein interactions. Is the center of a complex cross-linked less than the outside of the complex? Supposedly this is used as argument for saying the Taf6-Taf9 interface is not readily accessible to BS3. Spt8, however, is peripheral, but is also one of the last to crosslink (Fig E1B). The DUBm components are intimately associated yet they fail to crosslink. At the same time we do not see the DUBm on the silver stain, resisting BS3 crosslinking (Fig E1B). In general, an argument is made that a lack of crosslinking indicates that two areas do not interact, but then it is also being considered as a sign that an interface is intimate and buried from access to BS3. In some cases - the DUBm, for example - there is no crosslinking but we know the interfaces are intimate and abundant.
b. Can BS3 crosslink if the distance is too short? A max range is given, but what about a min range? Can interactions be too close to see? Close interactions appear to be invisible to this technique considering the DUBm has been crystalized and the very intimate interactions seen in that crystal structure are not at all shown using the BS3 crosslinking method.

The BS3 crosslinker has a linker arm of 11.4 Å when fully extended and can crosslink two lysine residues whose Cα atoms (alpha carbon of backbone peptide) are up to 30 Å apart (Merkley et al. (2014) Protein Sci. 23:747-59. doi: 10.1002/pro.2458). Merkley et.al. also report that the Cα atoms of crosslinked lysines can be as close as ~5 Å. Thus, the technology provides information about Cα atoms of lysines that are ~5-30 Å apart in native proteins/complexes. However, the crosslinking sites do not necessary correspond to interaction sites/interfaces. We and others have found that the crosslinking sites largely reflect domain-domain interactions (Knutson et al., submitted). We also found that lysines within flexible regions are often identified in crosslinks and may crosslink to many different sites. Accordingly, BS3 based crosslinking and MS provides information about the topology of proteins and protein complexes.

There are several scenarios that could result in the failure to detect topological features: 1) If a particular region lacks lysine residues or if the lysine residues are modified in a way that prevents reaction with BS3, this region will not crosslink; 2) The accessibility and the local chemical environment of lysine residues can affect crosslinking efficiency; 3) Crosslinked peptides may be poorly charged or may not transfer to the gas phase efficiently which would limit there detectability in the mass spec; 4) Crosslinked peptides that are too long or too short may not be identified because their m/z's are not in the range selected for MS analysis. Crosslinks containing short peptides can also be difficult to identify due to a lack of fragment ion information; 5) Certain protein domains or folds may be difficult to crosslink, e.g., we have found that it is difficult to identify BS3 crosslinks derived from cyclin-dependent kinase-cyclin complexes; 6) The complexity of the sample can limit detection of crosslinks due to duty cycle limitations during MS analysis. In most cases, it is difficult to determine the exact reason why a particular site of crosslinking was not identified. As always, extra caution should be taken in the interpretation of the negative results. In this study we have used the crosslinking-MS data, high resolution structural data (when available) and results of domain deletion studies to guide our modeling of SAGA architecture.

By far, the strongest conclusions we can make are based on positive crosslinking results. See the response to Reviewer 1 showing that the DUB module and Tra1 are present in our SAGA preparations. As to the comment why DUB subunits are not readily visible on silver stained gels, Sgf73 is visible, Ubp8 co migrates with several other subunits in the 50 kDa range, and Sus1 and Sgf11 are small and likely ran off the bottom of the gels used in Figs 1,and 6.

In response to this Reviewer comment, we have modified the text to more clearly outline the strengths and limitations of this approach and eliminated the statement regarding the accessibility of Tafs6-9 to BS3 (pgs 7, 8, 17).

2. Interpretation of the data to create a model and compare it to previous models.
a. The statements "however, this model is at odds with some aspects of an earlier model for SAGA architecture based on EM and immune localization of several subunits" and "...mapping Spt7 and Spt20 within the SAGA EM structure led to an alternative view that suggested these two subunits are positioned at different locations within the complex" are too vague. Describe the differences for the reader.

Text has been modified as suggested (pgs 4, 7).

b. It would be nice to have a comparison figure to go along with your discussion of the three current models. It is somewhat fuzzy to compare the two structural models with the module model which is not really a structure and provides no proposal for a 3D model. Perhaps some nuance is needed there. Phrases like "Tafs at the core" do not really apply to the modular model because the data is not presented in the shape of a complex. Instead, it is a cartoon to represent the probability matrix derived from the deletion analysis. It reveals that the complex is modular and provides some indication about which components in each module may be peripheral and which may be more internal. In all models the TAFs are close to each other, the HAT module is a module, the DUB module is a module, Tra1 appears to sit on the periphery and have limited contact with the rest of the complex, and so on... The way that the complex is presented here makes it hard to say that the TAFs alone are at the "core" while Spt 7 and Ada1 are not. They all seem to make the "core" with the HAT, Tra1, and DUB modules facing outward.

We disagree that our data do not indicate that the Taf and Taf-like subunits are at the core of the SAGA complex. Our data suggests that the Tafs occupy a unique position within SAGA where they make numerous crosslinks with all other SAGA modules. The text has been modified when discussing other models so as to make clear the similarities and differences between our work and previous studies. We don't think it is a good idea to have a comparison figure in the manuscript, but we have clearly referenced these other papers so readers can easily compare the different approaches and results. The previous EM, MS, and biochemistry were all important steps in understanding SAGA function and the Reviewer is correct that our new studies are consistent with much of this work. We have made this clear in the revised version.

c. "...mapping Spt7 and Spt20 within the SAGA EM structure led to an alternative view that suggested these two subunits are positioned at different locations within the complex" "Consistent with this latter view, our crosslinking results suggest distinct roles for these two subunits in SAGA assembly as Spt7 and Spt20 crosslink to different sets of subunits but do not crosslink to each other." Subunits may be part of a module, as proposed by Lee et al, without every protein in that module directly interacting with the others. Other groups have used co-purification as a proxy for interaction/association. Is this a controversy that does not really exist? These are different types of data which are not necessarily conflicting. Instead, they seem to complement each other. In the case of the DUBm, the components of the module are known to intertwine, interacting intimately. But this does not have to be the only mechanism for formation of a module, especially a module containing many proteins. In that case, the module could be defined by co-purification analysis. This could define the module and even go so far as to say which components of the module are peripheral or central. The inner structure of the module would then be determined by cross-linking/mass-spec analysis. In fact, looking at Lee's model, one could fold it into a 3D orientation and get a structure very similar to what is being proposed in this paper (It is less intuitive to match up this model and Wu's model, but folding Wu's SAGA in half might be close.) Further, comparing the probabilistic protein network from Lee et al, to Figure 2 in this paper are also suggests they are closer than they are far apart. In other words, the models are derived using different techniques and they give different information. These data, combined with those from previous studies should, together, provide evidence that allows one to make a more accurate model than any of the techniques could make on their own.

See response above for how we have revised the text to describe the contributions of previous work. Our definition of a module is a structural one, where groups of subunits associate independently

from the rest of the complex. In contrast, a functional module would be one where subunits are grouped by function only – for example if all Spt subunits were grouped because they gave the same phenotype when mutated. We have used the term structural module in the revised manuscript when first introducing the SAGA modules.

d. Can you comment on the different shape of your model and Wu's model, even though it was created using an immunological technique at a time when the complex was less-well characterized (i.e. missing important components)?

The shape of our model is very compatible with Schultz's original EM studies, which we think is clear in the revised manuscript (pgs 16-17).

Minor issues
3. Tafs and TFIID
a. A lot of emphasis is made on the similarity of the Tafs in SAGA and in TFIID. If this is a TFIID-like core, it would be interesting to compare it to the comparable TFIID section. In this work TBP is not associated with the TAFs. Instead, TBP is associated with Spt8 and Spt3 as shown in previous studies. It would be helpful to comment more explicitly on the relationship between TBP and the TAFs in TFIID and what functional consequences are implied by this relationship between SAGA and TBP. Do both complexes position TBP in the same relative position relative to the TAFs? What implication does this have for the potential for shared roles between SAGA and TFIID?

In the discussion, we briefly discuss how the TFIID-like core structure binds SAGA-specific subunits – at least partly because of the SAGA-specific Taf-like subunits Ada1 and Spt7. Considering the available data, we do not think the similarity between TFIID and SAGA extends outside the core, although this view may change when more structural and functional data is available.

Referee #3:
As most of the follow-up experiments and conclusions are solely based on the structural interpretation of the cross-link derived distance restraints and at the same time, the description and quality assessment of the cross-link identifications are superficial I request major revision of this part of the manuscript. Judging the quality of the cross-linking analysis is not possible in the current form.

1. The authors state that they have identified 199 intermolecular and 240 intramolecular crosslinks. However, in Table S1 there are several redundant cross-link identifications indicating the linkage of the very same lysine residues. The actual number of unique cross-links, indicating the linkages of unique lysine-lysine pairs, is smaller. The reviewer recommends to distinguish between total number of cross-link identification, unique cross-links and unique distance restraints (unique peptide-peptide pairs) that have actually been used for the structural interpretation.

We had duplicated the crosslinks identified in both experiments in the original Table E1. That's the reason for the redundancy, and we have corrected it. Thank you for pointing this out. We have identified 1005 total crosslinks, of which 439 are unique, including 199 inter- and 240 intra-molecular crosslinks. Table E1 reports the unique crosslinking pairs. To better illustrate the crosslinking positions, we have sorted the crosslinked peptides by the sequences of peptide 1. It should be easily seen that there are no redundant entries for the crosslinked peptides.

2. What are the minimum requirements for a positive cross-link identification? What is the minimum peptide length? What are the maximum mass errors for MS1 and MS2 yielding a valid/confident cross-link identification? Do all cross-link identifications display 3 consecutive ions even if there are more than 3 per peptide? MS/MS spectra of different cases for cross-link validation or of structurally/functionally important cross-links should be shown in the supplement.

We have added the requested information in the Methods section, including minimum requirements for positive crosslink identification, minimum peptide length, maximum mass errors for MS1 and MS2, and we have included MS2 spectra for a number of crosslinks that were used for modeling purposes in Fig E6.

3. What are the FPRs of the individual cross-link experiments?
Indicating which individual cross-links have been found more than once would significantly increase the quality assessment of the data?

Each experiment has been searched by pLink with FPR of 5%. After manual inspection of each spectrum, about 14% of the spectra were discarded due to their poor quality. Since there is no FPR estimation for manual inspection, the final FPR estimation will be difficult. Conservatively, it should be less than 5%. We have added the sentence that "14% of spectra were removed after manual inspection" in the Methods section. We also added the number of spectra identified for each crosslink in Table E1.

4. In the methods section, the authors state that precursors with charge state +3 were rejected from MS/MS analysis. The few studies that actually report the charge state distribution for identified cross-links indicate that there is a significant number of cross-link identifications with a precursor charge state of +3. Researching the data and thereby including +3 precursors may increase the number of cross-links and distance restraints.

We have designed the MS method to get as much information as possible for higher charge states. We have increased the FT AGC target to 50,000 and AGC maximum 250ms, so that each duty cycle can be as long as 0.5s. There are many +3 ions for the unmodified peptides. It could be too costly to include the +3 ions for the MS analysis. While we agree with the reviewer that there were studies reporting identification of +3 crosslinked peptides and we might miss some crosslinks with +3 charge states, we think there are many more benefits in our experimental design: A. For such a complex sample, selection of +3 ions will spend more time analyzing the unmodified peptides at the expense of missing the higher charge state crosslinked peptides; B. We can have better quality spectra if more ions are let into the trap so that once we have selected the higher charge states, we have a better chance to confidently identify them; C. Many crosslinks and peptides in general have multiple charge states. If one crosslinked peptide has a +3 charge state, it is likely it will also have +4 or higher charge states. We may miss the crosslinked peptides at +3 charge states, but we may have a better chance to select them out in the higher charge states. Of course, we agree with the reviewer that if we have more material and more machine time, we can definitely explore more options, such as more fractionation, gas phase separation, more charge state selection etc.

5. What is experiment "8" in Table S1. Does this indicate cross-linking at 8 mM BS3?
In addition to displaying some cross-link identification criteria (score, FPR, mass error) providing a legend for Table S1 will aid in understanding the data.

That was a letter "B" in the experiment column, and it meant the crosslink was identified in both of the experiments. We have updated Table E2, and added a legend as suggested.

Minor comments:
page 4: ...molecular connections... "better" ...molecular interactions...
page 11: Our cross-linking... "better" ... Our crosslink derived distance restraints
page 15: Our cross-linking... "better" ... Our cross-link data "or" distance restraints
page 16: ...Spt3, has numerous crosslinks... "better" ... displays "or" gives rise to
page 17: One possibility to explain... "better" One explanation of this inhibitory effect...
page 17: Given the crosslinking results... "better" Given the domain topology derived from chemical cross-links
page 18: Our crosslinking shows that Spt20 crosslinks to either side of the Spt3 histone fold domain that has been implicated in TBP binding. "rephrase sentence"

Text revised as suggested.

2nd Editorial Decision 31 July 2014

Thank you for submitting your revised manuscript for our consideration. I am pleased to inform you that we have now accepted it for publication in The EMBO Journal.

Please find two referees' responses attached below. Referee #3 was not available to assess your revised manuscript version. We have, therefore, consulted an external editorial board advisor and expert in this field, who regularly ensures consistency in our editorial handling. The advisor assessed whether you have successfully addressed the technical advices of referee #3 and the advisor was satisfied with your revision.

Congratulations to your work!

I am looking forward to your response.

Referee #1:

The authors have addressed all of my points satisfactorily. This is a very nice paper that represents an important advance in elucidating the role that SAGA architecture plays in its activity.

Referee #2:

In this revision, Han et al. provide additional insights into the interpretation of their model, which looks at the spatial relationship between the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex and the TATA binding protein (TBP). They also Analyze a new deletion of Sgf73 and arrive at a new conclusion about the effect of losing the DUBm on acetylation - that loss of the DUBm only slightly lowers acetylation levels.

This revised report has largely addressed our concerns.

Remaining point:

1. The authors have put the minimum BS3 crosslinking distance in the rebuttal, but they should also put it in the text.

