1. Initial referee #2 comment: The manuscript presented in the current form is more like a Measurement Report, as its atmospheric implication is not fully discussed or addressed. The only three nighttime samples also limits a broader atmospheric implication. If I understood correctly, the samples were combined before the freezing and determination of retention coefficient. Why not perform the experiments for individual samples, which one can exam the variation of retention for the same compounds?

Author Response: There are a variety of reasons why we chose to combine filter samples for our experiment. Using combined samples instead of individual samples eliminates variability to allow for experimental reproducibility and valid statistical evaluation of the retention coefficients. Specifically, some of these confounding variables that are controlled by this method could be large concentration differences between samples, background matrix influences, pH, biasing of trace species, etc. The individual samples on their own also do not produce adequate mass loadings for experimental viability. Combining sample extracts allows for enough liquid volume to be produced for the experiment without unduly diluting the sample. Further, comparative analysis of the individual filter samples is outside objectives of this publication and would also require examining the synoptic conditions during filter sample collection.

Referee #2 Comment: The response doesn't answer my questions. My concerns remain unsolved.
First, as the authors stated: "Specifically, some of these confounding variables that are controlled by this method could be large concentration differences between samples, background matrix influences, pH, biasing of trace species, etc.". Does this mean that the conclusion from this study may not be applied to aerosol particles from different regions where background matrix can be different from Beijing?
Second, If I understand it correctly, the Research articles must include substantial advances and general implications for the scientific understanding of atmospheric chemistry and physics. The authors response: "Further, comparative analysis of the individual filter samples is outside objectives of this publication and would also require examining the synoptic conditions during filter sample collection.". This is also not a good justification. I strongly suggest the authors either present it as a Measurement Report or at least incorporate these above reasons in their response in the manuscript.

Editor comment: I agree with the referee.
I acknowledge that you do explain nicely potential atmospheric implications based on your observed trends in the conclusion section. However, the potential caveats and limitations of your study due to the unavoidable mixing of samples should be also discussed. Please address it in the Results and Discussion Section and also in the Conclusions.

2. Initial Referee #2 Comment: In the Abstract, the statement in Line 20-22 is overstated as the S- and N-containing compounds are not solely from the NOx and Sox chemistry. The statement in Line 24-26 is not supported by the results of non-sigmoidal relationship or the related discussion. The lower surface-to volume ratio of the large drop size investigated may be one of the reasons. The differences in the experimental techniques, as I understand, may be the main reason, the wind tunnel and levitation, which have totally different air dynamic environments and surrounding corresponding gas-phase organic concentrations that will affect the gas-liquid diffusion and partitioning, and thus the retention.

Author Response: We have amended lines 20-22 to more accurately state what compounds are being referenced. We have amended lines 24-26 to more precisely state the confidence of correlation. We agree that the differing physical parameters of this experiment is a significant barrier to its comparison with wind tunnel studies; we present discussion on this at the end of section 3.3 and in the conclusions in lines 469-474. One of the objectives of this study is to establish untargeted complex mixture analysis as a method to measure retention before we make more direct comparisons to the wind tunnel. We have future experiments planned as a separate publication involving wind tunnel experiments. With interest for this discussion, our preliminary data from the tunnel shows similar results to this experiment in the levitator. This gives us reason to believe that the physical differences of these experiments is not the cause for the absence of sigmoidal behavior here.

Referee #2 Comment: As the authors agree the significant barrier to its comparison with wind tunnel studies, then it is not fair to state that the sigmoidal behavior is an overfitting. In the revision, the authors didn't make any adjustment or include the reasoning in the related discussion. If the authors plan to present in a separate publication involving wind tunnel experiments. I would suggest remove the discussion in the comparison with data from the previous wind tunnel studies.

Editor comment: I am still confused how you conclude that N- and S-containing compounds are mostly anthropogenic SOA. There is a huge body of literature that discuss the formation of such compounds also from biogenic sources (isoprene, terpenes etc). If you are sure that your samples contain organics with N and/or S as heteroatoms, just state it like that in the abstract, e.g. "The findings here also indicate that N- and S- containing organics have enhanced retention."
I disagree with the referee to remove the wind tunnel data from Figure 10. However, I would like to see a better explanation for the disagreement of low H* species. Why would different physical parameters lead to bad agreement for low H* species but good agreement for high H* species? Were there any other differences in terms of species properties, e.g. water-solubility (which does not necessarily correlate with H*), polarity, functional groups, molecular weight – just to name a few.

The term 'overfit' has a negative connotation, as obviously also expressed by the referee who calls it 'unfair'. Fact is that low retention of low H* species were measured in the previous study. The red line in Figure 10 seems a good fit to this data. However, whether this data – and therefore the fit - is relevant for atmospheric conditions is a different question. I suggest toning down the reference to 'overfitting' in your text and phrase it more cautiously.

Comments by Referee #3
General editor comment: Although it is clear that your paper has a companion paper with related experiments, each paper should stand on its own. It cannot be expected that the referee or any reader has to read both papers to understand fundamental findings in the present study. Therefore, please address some of the comments by referee #3 in more detail as suggested below.

Initial referee #3 comment: 444-459. The discussion of the relationship with effective Henry's law constant does not adequately address the potential influence of the conditions of freezing and freezing kinetics that are expected to have important impacts on retention coefficients for species with lower effective Henry's constant, as discussed in previous literature. Although the formation of an ice shell as inhibiting expulsion is mentioned at the end here, it appears to be largely an afterthought. The data are consistent with freezing conditions that enhance trapping, increasing retention even for the lowest H* compounds (which are expected to have high variability in retention dependent on freezing conditions). The assumption that a sigmoidal shape is expected, irrespective of freezing conditions does not do the previous literature on the retention phenomena adequate justice. (See for example Stuart and Jacobson 2003, 2004 (cited in this manuscript) and 2005, doi: 10.1007/s10874-006-0948-0).

Author response: We have discussed the influence of physical parameters on freezing retention more thoroughly in the Part I publication. https://doi.org/10.5194/egusphere-2024-3917 We agree that the differing physical parameters of this experiment is a significant barrier to its comparison with wind tunnel studies; we present discussion on this at the end of section 3.3 and in the conclusions in lines 469-474. One of the objectives of this study is to establish untargeted complex mixture analysis as a method to measure retention before we make more direct comparisons to the wind tunnel. We have future experiments planned as a separate publication involving wind tunnel experiments. With interest for this discussion, our preliminary data from the tunnel shows similar results to this experiment in the levitator. This gives us reason to believe that the physical differences of these experiments is not the cause for the absence of sigmoidal behavior here.

Editor comment: Given that you highlight this large discrepancy between your current study and those that found a sigmoidal R-H* relationship as one of the main findings, it would seem reasonable to include more discussion on previous literature data such as Stuart and Jacobson 2005 and possibly others.

Initial referee #3 comment: 453-455 (and 469-473). Although this is an appropriate limitation to discuss, it is too narrow. Why focus only on the potential effect of the surface to volume ratio rather than other conditions of freezing, when there are other factors that have been suggested previously as likely important based on theory and modeling?

Author response: We have discussed the influence of physical parameters on freezing retention more thoroughly in the Part I publication. https://doi.org/10.5194/egusphere-2024-3917

Editor comment: A brief discussion should be added also in this paper. It does not have to be as detailed in Part I.

Initial referee #3 comment: 482- 490. Solubility is expected to influence retention but was not studied here and is only discussed as a rationale for why AA don’t follow polarity and VP trends. The expected effects of solubility on retention (and its relationship with Henry’s law constant) should be discussed more broadly, along with lack or solubility information limiting the findings.

Author response: Aqueous solubility is largely dependent on structural properties that this method of untargeted MS is not able to adequately resolve for analysis. The limitation on discussion of solubility stems from the inability to assign high confidence structural information. This is an issue commonly discussed in untargeted complex MS analysis. A comprehensive discussion on this issue would only distract readers from the supported conclusions with speculation that is outside of the current measurement capabilities of this method. Further, solubility is also already well described in the literature as a piece of the calculation to estimate H*. We were able to avoid some of the issues with calculating H* without high confidence structural assignment by accepting the average difference between the H* for structural isomers as found by Isaacman-Vanwertz and Aumont, 2021 (https://doi.org/10.5194/acp-21-6541-2021). This is discussed in lines 176-180.

Editor comment: I appreciate that you explain here that you explain here why solubility is not taken into account as a parameter. However, this information should be also mentioned in the manuscript text in a couple of sentences. I don’t think this would be distractive.
Water solubility and Henry’s law constants do not necessarily correlate. The former describes how much of a substance can be dissolved in water and above which limit it forms a solid or supersaturated

solution ; the latter describes how a compound partitions between the gas and aqueous phases. Thus it describes two different thermodynamic concepts.
There are very distinct differences. For example, small alcohols are fully miscible in water but have very low H*; in contrary, some salts are highly water soluble, others not – but all of them basically have infinite Henry's law constants.
==

Additional editor comments
(line numbers refer to your manuscript version without track-change)

l. 13/14: The structure of this sentence does not seem right; 'incurs' might not be the right verb here. Better use 'causes' or similar.

l. 15: The number and variety of compounds that can form new particles is very limited. Why are so specific here? Do you have evidence that indeed new particle formation occurs in the upper troposphere by compounds released from freezing droplets?
Just the fact that compounds can be released upon vertical transport seems a sufficient motivation to study retention coefficient. These compounds may then take part in any chemical reactions.

l. 25: differs from… (not 'with')

l. 26: I am missing in the abstract a concluding sentence that summarizes why the current study is more reliable/refined/better/… than the previous literature studies.

l. 34: SOA formation is not necessarily part of new particle formation. Compounds with sufficiently low vapor pressure may condense on any particle, whether primary or secondary. Also the initial particle may be inorganic (e.g. by new particle formation due to sulfuric acid) or organic.

l. 49 – 51: I am not convinced that uncertainties in gas transport are the main factor of uncertainty of NPF in the upper troposphere. Isn't it much more the uncertainties in NPF precursors, i.e. which compounds nucleate new particles at what rate?

l. 56: 'incurred' seems wrong here. Do you mean 'occurred'?

l. 74: What is the difference between 'liquid water content' and 'droplet size' here? Isn't the liquid water content simply the water volume and thus proportional to size?
Unless you mean something different, I suggest removing liquid water content here.

l. 82: 'cloud size range' seems odd here. Do you mean 'droplet size range'?

l. 251: Do nitrophenols from other sources (e.g. industrial emissions) have different H/C ratios?

l. 253/254: Are the compounds specific to automobile and shipping emission or could it be also other organosulfates?

l. 256: Sulfonic acid (oxidation state of S = +4) is not sulfate (oxidation state +6)

l. 257: The study by Iinuma et al was conducted at a forested, not urban, site. Given that terpenes are biogenic compounds, I do not think that camphorsulfonic acid is only formed under urban conditions.

l. 266/7 and following: Just providing the sum formula of the compounds is rather confusing since these formula could also represent compounds with functional groups other than amines (cf e.g., https://en.wikipedia.org/wiki/C9H11NO2#) Please either write the formulae such that the molecular structure is clear and/or add the compound names.

l. 277: One of the main oxidation products of isoprene include methyl vinyl ketone (among others), with an H/C ratio of 1.5. Thus, I don't think that the H/C range is a valid indicator of biogenic vs anthropogenic. If I am wrong, please add more recent references demonstrating it.

Tables 2 and 3: These tables could be moved to the supplement as they are not discussed in the text, but oly referred to in a very brief sentence (cf also comment by Referee 3). Their main information is included in Figure 3 . Readers who are interested in the data can find them in supplement.

l. 333/34: CHNOS represents more of the heaviest species in the sample set while CHN is entirely AA. This sentence is quite cryptic. Please reword.

l. 345: The paper by Daito et al does not refer to the atmosphere. The experiments were performed under very different conditions, not relevant to atmospheric ones. Unless you can add a relevant reference here that demonstrates that aqueous phase reactions in the atmosphere lead to NOx removal, I suggest removing this part of the sentence.

l. 419: The paper by Weschler and Nazaroff is about semivolatile organics only and in indoor environments. During the last decade there have been many papers on the volatility ranges of SOA species in the atmosphere. It would seem more appropriate to cite one of them.

Conclusions: Define AA and VP here once more for the readers who only read the conclusion section.

l. 493: 'nitrogen and sulfur inclusion' is not a commonly used term. I think what you mean is 'functional groups containing sulfur or nitrogen'

l. 494/5: cf my comment on the abstract

l. 497: cf my comment about this reference above

l. 501 – 503: Atmospheric chemical processing generally tends to oxidatively degrade large nonpolar species into more water soluble, less volatile species (Iavorivska et al., 2016)

This sentence is not quite correct. Chemical processing does not necessarily lead to degradation of organics but can also lead to functionalization. For this you could cite any atmospheric chemistry textbook rather than the paper by Iavorivska et al that is about deposition.

l. 508/9: Why do you limit your discussion here to SOA precursors? By far not all organics form SOA. Just the fact that less oxidized (or 'fresher') organics are less likely to be retained and vertically transported than more aged organics is an interesting result. Whether they eventually form SOA is not relevant.

l. 515 – 520: I got confused by this text (and obviously also referee 2). What do you want to say here? Would you expect that during summer the WSOC composition is completely different and therefore your results are irrelevant? What is known about WSOC differences between summer and winter?

l. 531ff: Can you explain better what you mean here? What should be included in a transport model? Usually retention coefficients are included for individual species or species groups, independent of their mass distribution. Retention is usually just described as a mass fraction that remains in the ice phase. Why should this depend on species abundancy?
Are you saying that the retention coefficients measured in your study are only valid for this particular composition? If so, why? It would imply that the presence of all compounds in a sample affects the retention of an individual compound.