

1 We wish to thank the reviewers for their insightful and constructive reviews. We will attempt to address some of the
2 raised questions and points below, and indicate several places where we will update both the camera-ready manuscript
3 and the supplemental material.

4 First, reviewer three is absolutely correct that inverse problems are hard and require rigorous and careful experimentation
5 to validate new approaches. The literature often refers to “inverse crimes”, where you show that your method can invert
6 data from your own forward model. Being able to invert simulated data from your own forward model is a necessary,
7 but far from sufficient, criterion to having a useful working method. Thus, while figure 5 and the associated table are
8 useful in understanding some of the robustness of our approach, they are not sufficient. This is why we (on the fourth
9 line of table 1) specifically evaluate our method on raw experimental data, not simulated data. This is data that our
10 approximated forward model has never seen, and our inverse model has never seen. This reflects the most realistic
11 real-world application of our approach to unseen data.

12 In the camera-ready manuscript, we will attempt to make the exact evaluation role of each row in Table 1 more clear.
13 In particular, the first row is an “inverse crime” – we are evaluating our ability to invert spectra that both our forward
14 model and our inverse model have seen during training. We do this to show how well our model could work in the best
15 case scenario, and to provide some context for understanding the relative accuracy of the other rows.

16 Reviewer three also pointed out that we do not define AUC in our paper, an omission which we will correct in the
17 camera ready. Our method ultimately calculates the difference between the (predicted) spectrum of a candidate structure
18 with the experimentally-observed spectrum. The difference (in ℓ_2) between these spectra allows us to set an error
19 threshold, below which we will consider the structure to be “correct”. We can compute the area under the curve that
20 arises from varying this threshold, thus giving rise to the AUC. To reviewer one’s comment that the AUCs seem low:
21 we find that most users of our method will want a high degree of confidence that the recovered structure is correct, and
22 will likely set this threshold very conservatively, thus making AUC a less useful assessment of overall performance. For
23 predicted structures above the threshold, the user can then discard the predictions – our algorithm simply could not find
24 a sufficiently good structure. This is why we have the other columns in table 1, which should be interpreted as follows:
25 if you set the threshold such that you predict a candidate structure N% of the time, how correct are those structures? We
26 see that in the real-world use case (row 4), if you set the threshold so conservatively that our method only predicts a
27 structure 20% of the time, 97.1% of the those returned structures are correct.

28 Finally, our fast forward model is described in another publication that is currently in press, and thus we glossed over it
29 in this paper. It is a fairly straight-forward application of a graph convolutional network to predict per-vertex properties
30 (in this case, the chemical shift values) with some minor extensions for incorporating uncertainty. In our camera ready
31 we will add several paragraphs to the supplemental section more fully describing this model, as well as linking to the
32 paper describing this approach.

33 All reviewers asked if the resulting learned inverse model was interpretable, or if the learned features captured the
34 intuition that humans use when solving this problem. Preliminary experiments looked at which bonds and bond types
35 were placed with which nuclei, and suggest the answer is yes. Initially higher-order bonds (double and triple) are often
36 placed first in sequence by our approach, as they are often associated with the most unique (highest entropy) observed
37 spectral values. We will include the results of these analysis in the supplemental section of the camera-ready.

38 Reviewer one asked about network architecture; we made a reasonable (although not exhaustive) effort to identify
39 a good architecture, considering many architectures derived in the graph prediction literature. In this context, the
40 referenced Battaglia paper was of considerable help, as their “relational network” formulation unified these approaches
41 and allowed us to zero-in on a reasonable set of hyperparameters, allowing us to incorporate per-vertex measurements,
42 sparse per-edge observations (the edges of the current state), and yet output a dense probability distribution over
43 subsequent edges. We have not considered a transformer-like architecture thus far but may in the future, given their
44 success in other structured prediction (NLP) problems.

45 Reviewer asked for clarification of “Beta” on line 3; when we are generating training data we delete edges in the source
46 molecule with a probability p where p is sampled from a mixture of a uniform and a beta distribution,

$$p \sim \frac{3}{10}\text{Unif}(0, 1) + \frac{7}{10}\text{Beta}(3, 3)$$

47 The beta component guarantees we sample a lot of “partially completed” molecules for training. For these sorts of
48 sequential structured prediction problems, “distributional shift” is often a challenge, but we found empirically that
49 by over-sampling partially completed molecules according to this distribution we achieved superior reconstruction
50 performance compared to a mere uniform sample. The investigation of active-learning approaches (like DAGGAR and
51 SEARN) is another way and represents an avenue for exploration in the future.

52 We thank the reviewers again for their helpful and constructive feedback.