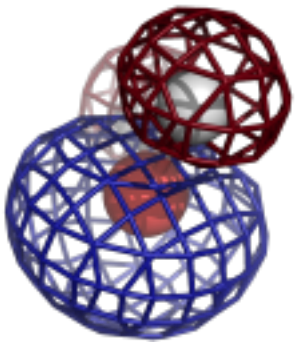
VOLUME ELEVEN



COMPUTATIONAL CRYSTALLOGRAPHY NEWSLETTER

JANUARY MMXX

ENSEMBLE REFINEMENT

Table of Contents

Editor

Nigel W. Moriarty, NWMoriarty@LBL.Gov

Phenix News

Announcements

New Phenix Release

Highlights for the 1.17 version of Phenix include:

- Improved handling of SHELX data in *phenix.reflection_file_converter*
- *eLBOW* can output files for Amber and supports the Orca QM package
- *dials.image_viewer* is used for viewing diffraction images
- Updated map smoothing
- Fix inconsistency in clashscore values in *phenix.validation_cryoem* when hydrogen atoms are in the model

Please note that this new publication should be used to cite the use of Phenix:

Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ, Adams PD: Acta Cryst. (2019). D75, 861-877.

A new tool, *phenix.homology*, is available in the nightly and discussed on page 5 of this newsletter.

Downloads, documentation and changes are available at phenix-online.org

Ensemble refinement produces consistent R-free values but smaller ensemble sizes than previously reported

Stephanie A. Wankowicz[a,b] and James S. Fraser[a,b,c,d]

a - Biophysics Graduate Program, University of California San Francisco, San Francisco, CA, USA.
b - Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA.
c - Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA, USA.
d - Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA
Correspondence email: jfraser@fraserlab.com

Introduction

Ensemble refinement combines molecular dynamics (MD) simulations with crystallographic data to provide a model of atomic fluctuations that are present in the crystal lattice. As implemented in *phenix.ensemble_refinement*, MD simulations are performed where the model is restrained by a time-averaged X-ray restraint (Burnley et al. 2012). Because the agreement with observed structure factors is calculated by averaging of several recent snapshots of the MD simulation, ensemble refinement differs significantly from traditional refinement where a single structure is used to calculate the agreement. To attempt to control for crystalline disorder, a Translation/Libration/Screw (TLS) model model is fitted prior to the simulation, leaving the simulation to fit the residual difference density. After the simulation is run, a procedure reduces the ensemble size down from all snapshots acquired during the period to a minimal set that will reproduce the R-free within a tolerated value. In the original paper describing *phenix.ensemble_refinement*, this yielded 39-600 ensemble members in the 20 PDB depositions that were subjected to refinement. The structural diversity across these ensemble members is a representation of the residual conformational heterogeneity after accounting for the disorder modeled by the TLS model.

We set out to run ensemble refinement on a large number of publicly available X-ray crystallography structures. Although some parameter names and default values had apparently changed since the original paper, the online documentation provided a guide to reasonable values (Phenix documentation: ensemble_refinement.html). For our analysis, all structures had a resolution between 1-2.5 Angstroms. Using Phenix version 1.15, we pursued the following workflow (code is available on github[1]).

1. Download existing model and structure factor files
2. Run *phenix.ready_set*
3. Re-refinement of model using *phenix.refine*
4. Ensemble refinement over a grid search of parameters
5. Selection of best model based on R_{free}

All input parameters for our analysis are available (https://ucsf.app.box.com/folder/95195345802).

Non-default inputs

- `wxray_coupled_tbath_offset: grid search of 2.5, 5, 10`

Errors

About 10% of the structures failed during refinement. There were numerous reasons for these failures including, a lack of appropriate

[1]https://github.com/stephaniewanko/Fraser_Lab/tree/master/phenix_pipeline

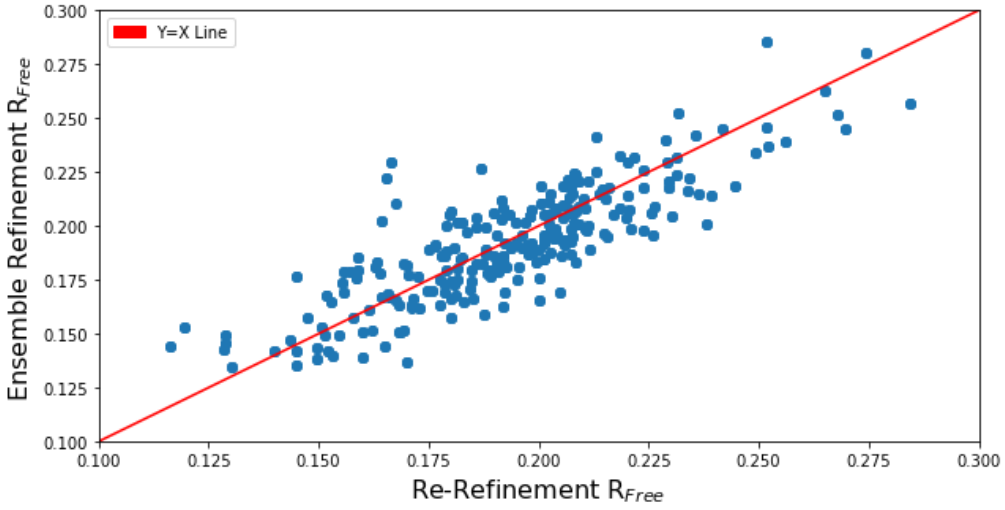


Figure 1. *R_{free} values from re-refinement and ensemble refinement are correlated (R^2=0.91). In 418 (57.6%) structures, the ensemble refinement R_{free} value was lower than the refinement R_{free} value. In 307 (42.3%) structures, the ensemble refinement R_{free} value was higher than the refinement R_{free} value.*

intensities or amplitude information, poor maps, and issues with ligands.

Conclusion

There were two major differences between our analysis and the original Burnley 2012 paper (Burnley et al. 2012). First, in the Burnley paper all 20 structures had reduced R_{free} values when subjected to ensemble refinement. In our study, overall, ensemble refinement R_{free} was comparable to re-refinement R_{free}, with 57.6% of structures having an improved R_{free} with ensemble refinement compared to traditional refinement, as shown in figure 1. This may be due to non-optimal parameter selection or insufficient model preparation. Second, it was unclear why we were getting such smaller ensemble size compared to the 2012 paper. We were expecting many ensemble sizes to be greater than 100; however, all of our structures returned ensembles <100, as demonstrated in figure 2. Although the ensembles obviously contain more diversity than single structures, we were curious as to the underlying cause of the greatly reduced ensemble size. To further investigate, we tested our ensemble refinement pipeline on the 20 PDB models originally analyzed in Burnley 2012 paper.

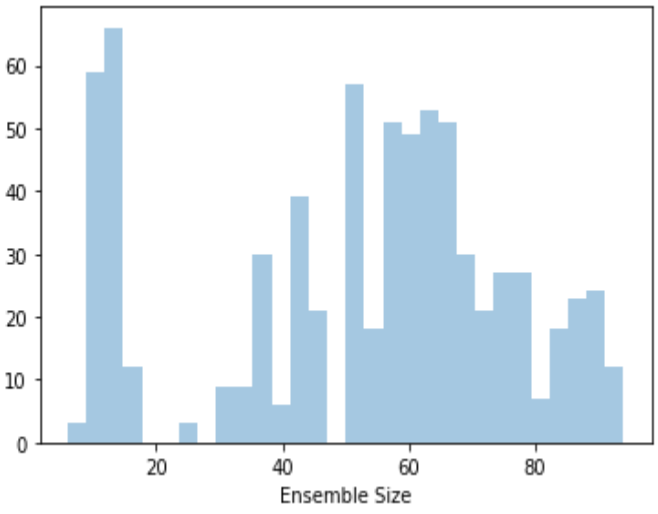


Figure 2. *Most structures have smaller ensemble sizes (the number of models in the ensemble output) than we expected based on the results in Burnley 2012.*

Recreating Burnley 2012 Paper

To recreate the results from the Burnley 2012 paper, we followed the same pipeline outlined above. Of note, while we automatically re-refined the models coming from the PDB, we did not perform any manual refinement, which

Table 1. *Input R values from Burnley 2012 compared to our input to our recreation.*

PDB	Resol.	Original Ensemble Size	R_{work} Burnley 2012	R_{free} Burnley 2012	R_{work} Recreation	R_{free} Recreation	Recreation Lowest R_{free} Ensemble Size
1kzk	1.1	600	0.125	0.153	0.155	0.179	100
3k0m	1.3	250	0.104	0.129	0.127	0.144	167
3k0n	1.4	209	0.115	0.133	0.117	0.143	167
2pc0	1.4	250	0.145	0.188	0.231	0.252	125
1uoy	1.5	167	0.104	0.137	0.136	0.165	125
3ca7	1.5	40	0.149	0.184	0.237	0.292	56
2r8q	1.5	200	0.132	0.162	0.164	0.188	125
3ql0	1.6	70	0.204	0.254	0.217	0.256	50
1x6p	1.6	400	0.121	0.149	0.141	0.163	134
1f2f	1.7	143	0.128	0.168	0.170	0.210	84
3ql3	1.8	80	0.160	0.208	0.171	0.207	56
1ytt	1.8	84	0.139	0.174	0.179	0.206	63
3gwh	2.0	39	0.160	0.200	0.198	0.230	67
1bv1	2.0	78	0.149	0.182	0.188	0.240	84
1iep	2.1	200	0.183	0.238	0.207	0.256	63
2xfa	2.1	100	0.171	0.217	0.226	0.261	60
3odu	2.5	50	0.208	0.269	0.247	0.297	32
1m52	2.6	50	0.161	0.211	0.198	0.240	32
3cm8	2.9	67	0.194	0.235	0.231	0.264	39
3rze	3.1	72	0.210	0.280	0.250	0.289	32

left us with input structures with slightly higher R_{free}/R_{work} compared to the Burnley 2012 paper (table 1). We extended our grid search to include three parameters suggested by the Phenix documentation (pTLS, wxray_coupled_tbath_offset, tx).

- pTLS defines the fraction of atoms included in the TLS fitting procedure. This is intended to model static crystalline lattice disorder and varying this parameter results in movement being absorbed by the TLS B-factors rather than by atomic fluctuations.
- wxray_coupled_tbath_offset controls the X-ray weight. This helps ensures that the simulation runs at the target temperature.
- tx dictates the structure factor memory relaxation time. This governs the time period for which a particular conformation retains its influence. The higher the number, the more a particular conformation affects the average.

Additionally, we added harmonic restraints for all ligands in each structure. Of note, while Burnley 2012 paper reported only one ensemble structure per PDB, we had 36 ensemble structures (corresponding to a 3 x 3 x 4 grid search of the parameters pTLS, tx, wxray_coupled_tbath_offset) and choose one select ensemble structure based on the criteria of lowest R_{free}. This test was run on Phenix version dev-3584 (a mid 2019 version).

Non-default inputs

- `wxray_coupled_tbath_offset: grid search of 2.5, 5, 10`
- `pTLS: grid search of 0.6,0.8,1.0`
- `tx: 0.5, 0.8, 1, 1.5`

Outputs

As shown in table 1, in almost all cases, the ensemble sizes were lower than what was found in the Burnley 2012 paper. Figure 3 illustrates that we found that R_{free} correlated with ensemble size (R^2 =-0.61). Similarly, resolution was slightly correlated with ensemble size (R^2=-0.48). Overall, the recreated R_{free} were highly correlated with the R_{free} from the Burnley 2012 paper (R^2=0.892). We could not identify any pattern between the parameter values correlated with R_{free} and the optimal parameter value as judged by R_{free} was idiosyncratic for each structure.

As we wanted to use ensemble refinement to assess dynamics, we want to see if different parameter values (pTLS, wxray_coupled_tbath_offset, tx) change the RMSF. We examined all structures, but focus our analysis below on C-ABL kinase domain in complex with STI-571 (PDB: 1IEP).

While the RMSF values of C-ABL kinase domain in complex with STI-571 (PDB: 1IEP) were highly correlated (>0.8) across all parameter values, highlighted in figure 4, there were only some notable deviations in magnitude for the pTLS parameter values, demonstrated in figure 5.

Because of the lack of correlation between the parameters and R_{free} values, and the relative consistency of the RMSF calculations, we evaluated each PDB independently and chose parameters that yielded the lowest R_{free}. At least one of the 20 PDBs had an optimal ensemble using each of the wxray_coupled_tbath_offset and pTLS parameter values. For the tx parameter only 3 out of the 5 values were used in optimal ensembles (0.8, 1.0, 1.5).

Conclusions

Overall, we were still getting much smaller ensemble sizes compared to the Burnley 2012 paper. However, our R_{free} values correlated very well with the R_{free} values from the paper giving

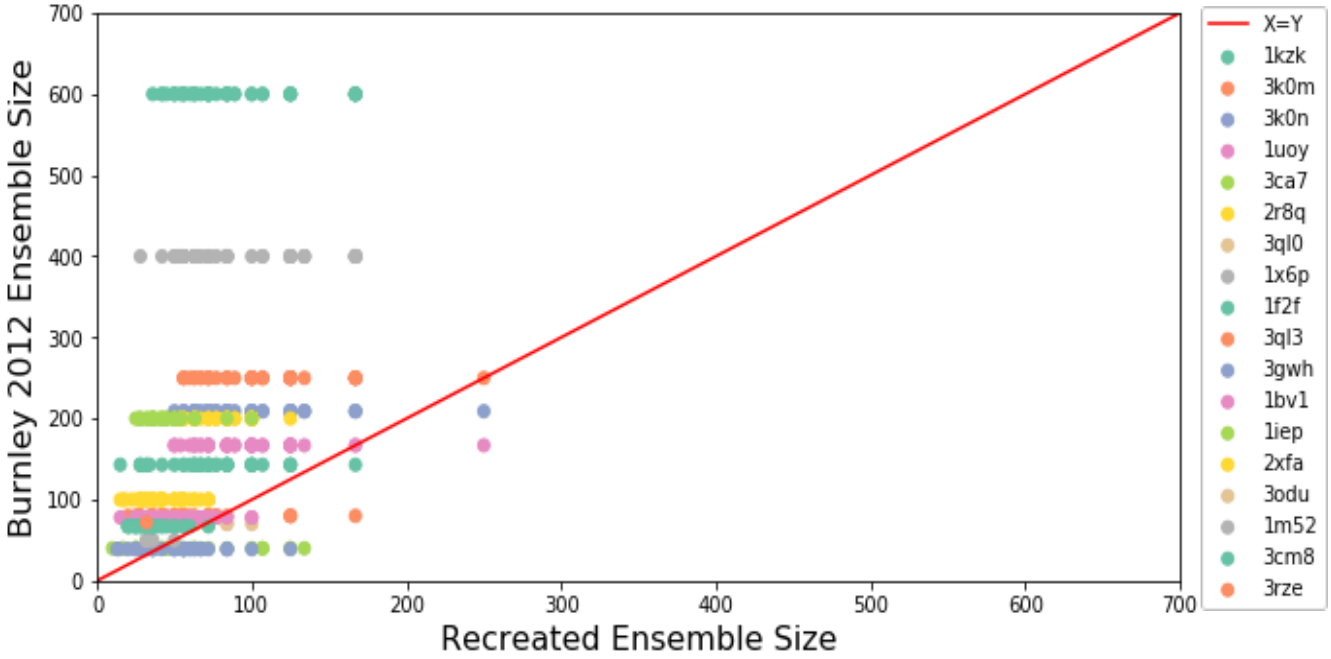


Figure 3. *Burnley 2012 paper ensemble size compared to our recreated ensemble size. In almost all cases, the ensemble sizes were lower than what was found in the Burnley 2012 paper.*

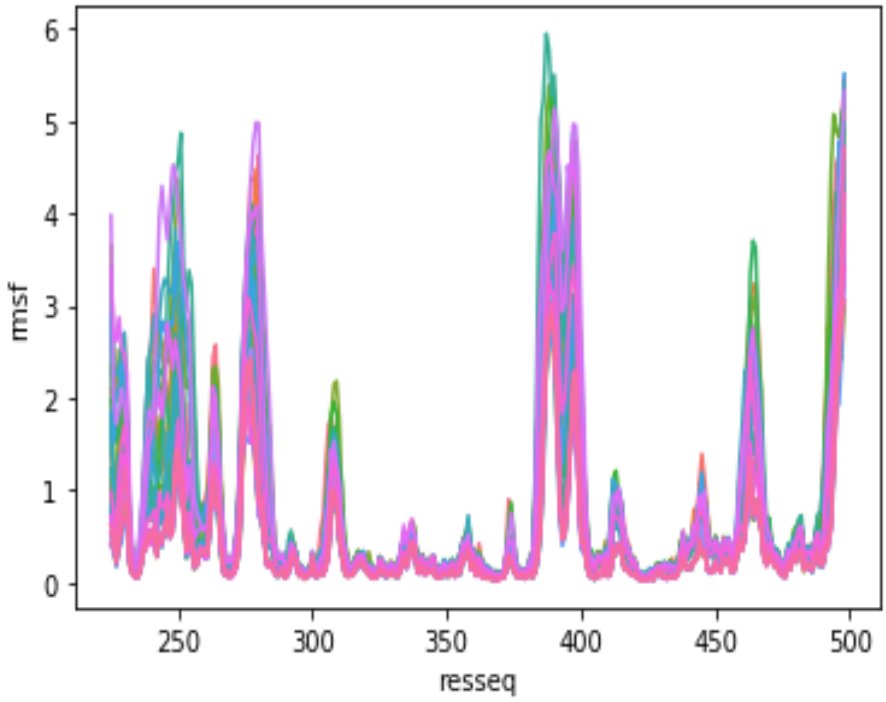


Figure 4. *RMSF of C-ABL kinase domain in complex with STI-571(PDB: 1IEP) across all 45 parameter values.*

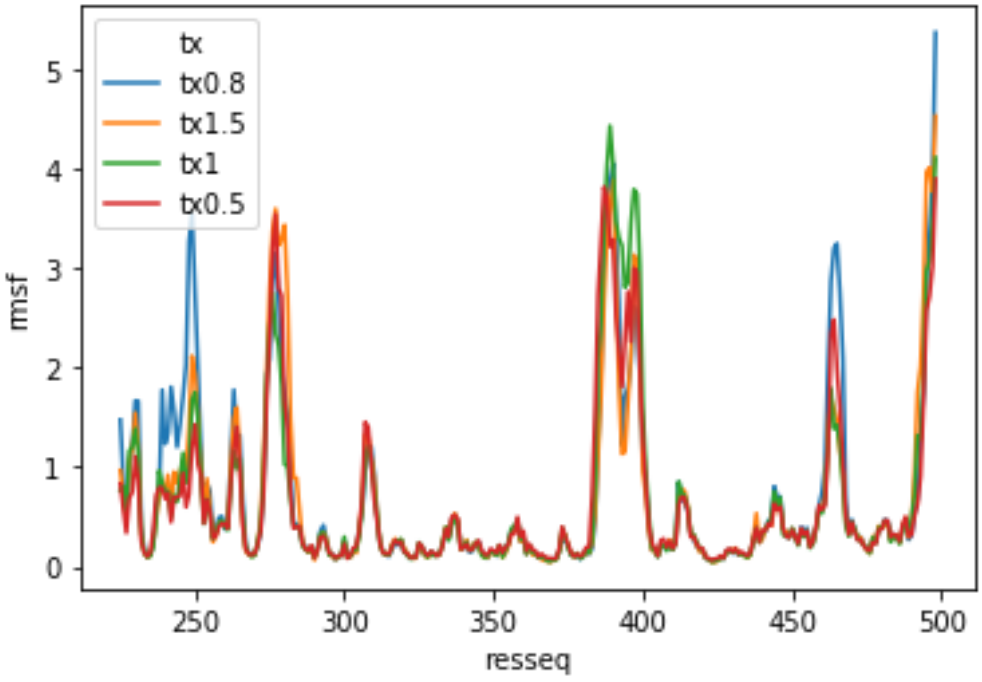


Figure 5. *RMSF of C-ABL kinase domain in complex with STI-571(PDB: 1IEP) across all tx parameter values.*

us confidence in the underlying procedure. The wxray_coupled_tbath_offset, pTLS, or tx parameter values were not correlated with R_{free}, R_{work}, or the ensemble size. In terms of RMSF changes, the only parameter that produced a major difference was pTLS, as expected. pTLS determines the percentage of atoms included in

the TLS model, which predicts the local positional displacement of atoms in a crystal structure with the underlying assumption that the atoms included are members of a rigid body. In our results, lower pTLS values (fewer atoms included in the pTLS model) have higher RMSF on average. It is unclear to us if choosing a model based on the best R_{free} will result in accurate results for protein conformational heterogeneity, especially when comparing two protein structures with different pTLS values.

Investigating the ensemble size difference

To try to resolve the discrepancy in the ensemble sizes from the original 2012 paper to our recreation of their results, we used the optimal parameter values from the test above for each PDB and tested four other keyword changes that we predicted might give us results closer to the Burnley 2012 paper.

1) Using the Phenix version released most closely to the Burnley 2012 paper (version 1.8.2, the first release to contain the *phenix.ensemble_refinement* command).
2) Removing the use of the conformation dependent restraint library.
3) Re-setting the ensemble R_{free} tolerance parameter to 0.001
4) Re-setting the ensemble reduction feature to false

Testing Phenix version 1.8.2

The Burnley 2012 paper was run using a different Phenix version than we used with our recreation. We ran ensemble refinement on Phenix version 1.8.2, which corresponds to the public release of the method after the Burnley 2012 paper.

Non-default inputs (Phenix version 1.8.2)

- `wxray_coupled_tbath_offset, pTLS, tx parameter values corresponding to the optimal` R_{free} `for each individual PDB from the previous tests.`

Errors

Only five out of the 20 structures ran ensemble refinement successfully. There were multiple reasons for failures. These included a pTLS error with chain breaks and errors reading in parameters fed into ensemble refinement.

Conclusions

Many structures failed to run ensemble refinement. However, for the five structures that finished, the ensemble sizes were still smaller than expected based on the Burnley 2012 paper and were highly correlated with our previous recreation, as seen in figure 6 and 7. Figure 8 demonstrated that we continue to observe a good correlation between the original and updated R_{free} values (R^2=0.95).

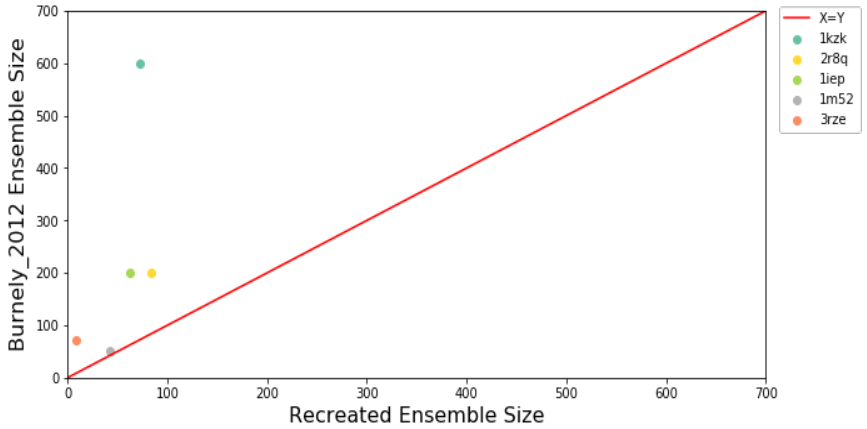


Figure 6. *Recreated ensemble sizes are smaller compared to the ensemble sizes in Burnley 2012 (R^2=0.57).*

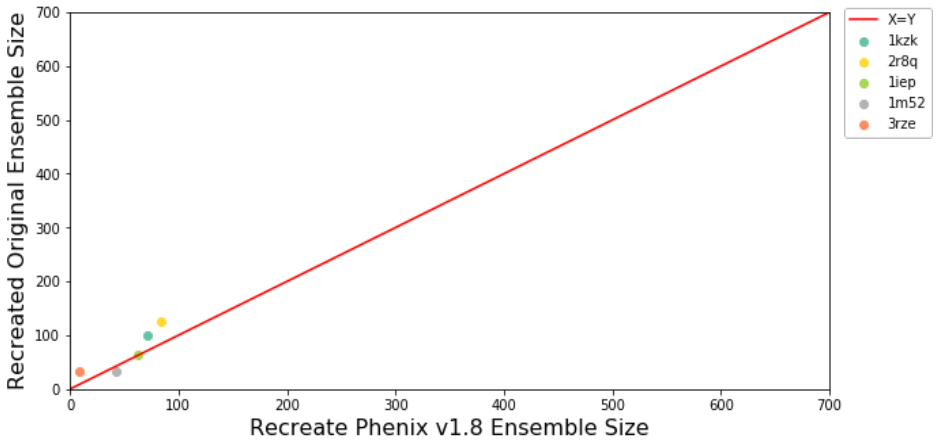


Figure 7. *Recreated ensemble sizes with Phenix version are similar to the initially recreated ensemble sizes (R^2=0.88).*

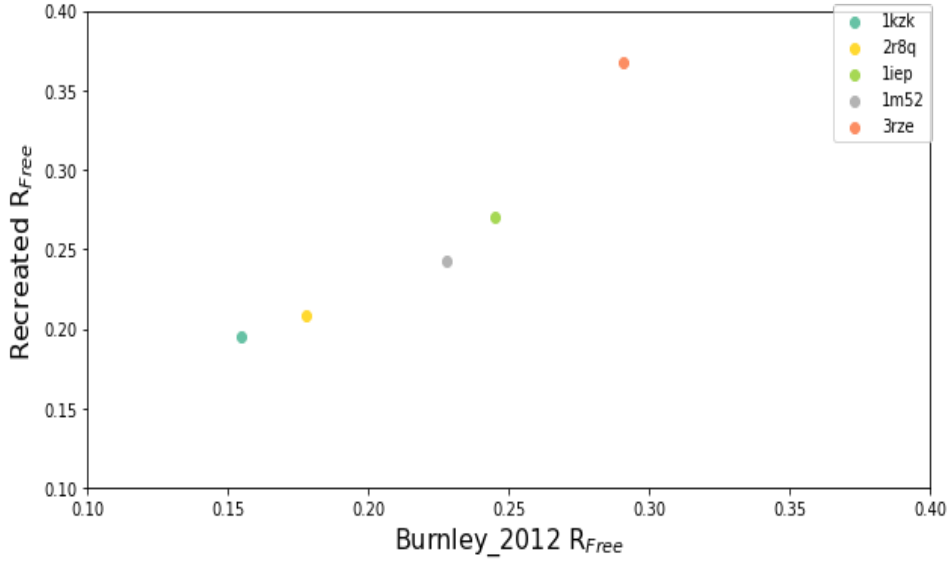


Figure 8. *Recreated R_{free} were highly correlated with the* Burnley 2012 *paper (R^2=0.95).*

Reverting R_{free} Tolerance to 0.001

The last step of ensemble refinement takes all of the snapshots saved from the MD simulation and selects the lowest number of models that together have an R_{free} within the percentage of the full ensemble R_{free}. This percentage is defined by the R_{free} tolerance parameter. The current version of Phenix (1.16), defaults this parameter to 0.0025 but in the Burnely 2012 paper, it was set to 0.001. Therefore, we tested if we could increase the ensemble size by changing this parameter back to what was used in the paper using Phenix version 1.16.

Non-default inputs (Phenix version 1.16)

- `Wxray_coupled_tbath_offset, pTLS, tx parameter values corresponding to the best` R_{free} `for each individual structure.`
- `ensemble_reduction_rfree_tolerance = 0.001`

Conclusions

Reducing the R_{free} tolerance parameter back to where it was initially set did increase our recreated ensemble size (median increase: 26 models), see figure 9. However, for many of these structures the

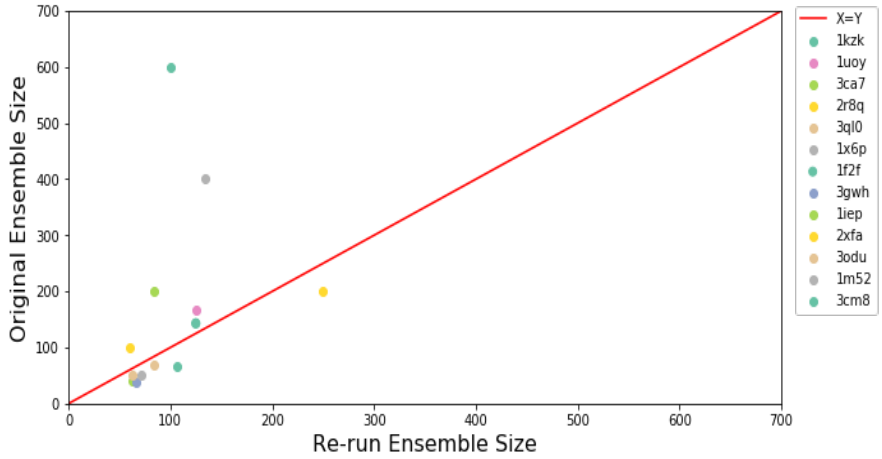


Figure 9. Burnley 2012 paper *ensemble size compared to our recreated ensemble size with an R_{free} Tolerance of 0.001 (R^2=0.307).*

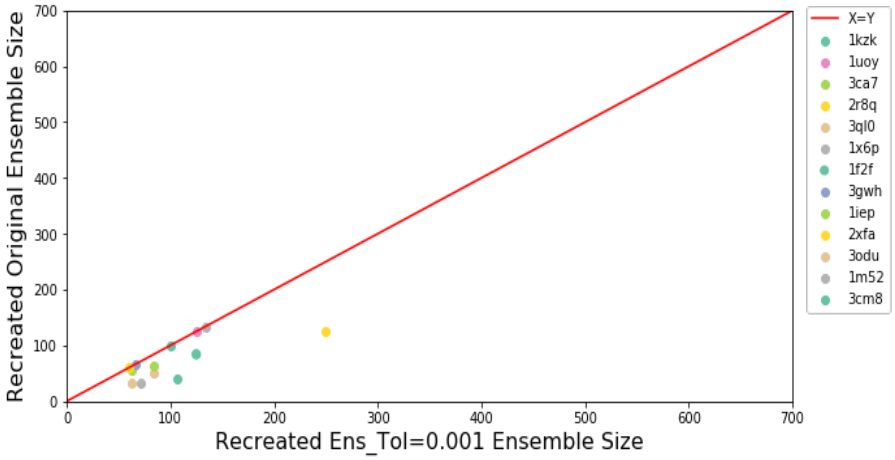


Figure 10. *Recreated ensemble sizes with ensemble_rfree_tolerence parameter =0.001 are mostly larger than the initially recreated ensemble sizes(R^2=0.71).*

number of models were still far below what was observed in the Burnley 2012 paper, as seen in figure 10. Figure 11 shows that the R_{free} correlation was still observed between the recreation on the Burnley 2012 paper.

Removing Conformational Dependent Library (CDL)

In the Burnley 2012 paper, the default restraints were Engh and Huber, but more recent versions of phenix use the conformation dependent library (CDL). One hypothesis is that the older restraints would bias ensembles to have energetically reasonable angles and bond lengths compared to the modern CDL restraints, leading the ensemble sizes to decrease under CDL restraints. Therefore, we set the cdl restraint library to false. Of note there are three other library (omega_cdl, rdl, and hpdl) that are also available as parameters but are set as false as the default.

Non-default inputs (Phenix version 1.16)

- `Wxray_coupled_tbath_offset, pTLS, tx parameter values corresponding to the best` R_{free} `for each individual structure.`
- `restraints_library_cdl = False`

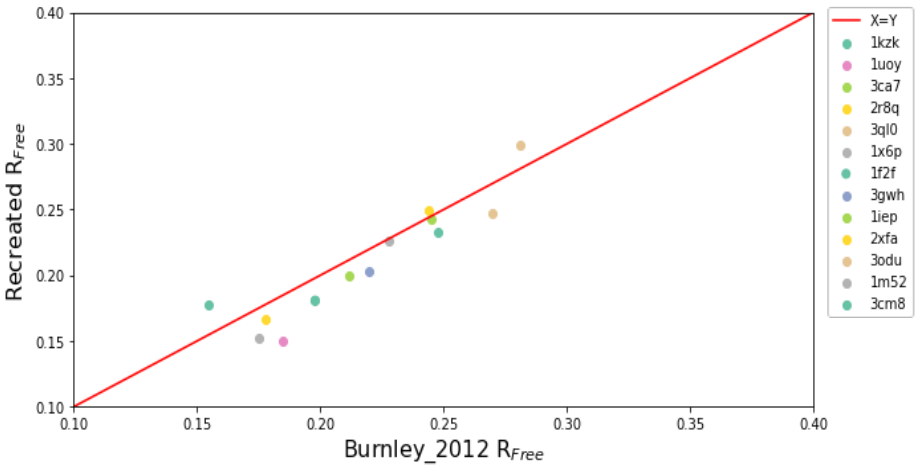


***Figure 11**. Recreated R_{free} with an R_{free} tolerance of 0.001 were highly correlated (R^2 =0.9) with the R_{free} Burnley 2012 paper*

Conclusions

By turning the CDL restraints off, we did not observe an increase in ensemble sizes, see figure 12. For one structure (PDB:3GWH), the size of the ensemble did increase, see figure 13. We suspect this is due to the input model having a high number of geometry outliers. By turning off CDL, we may have further increased the geometry problem, resulting in a larger ensemble size. There was still a high correlation between the original and recreated R_{free} (R^2=0.96) as shown in figure 14.

Testing Ensemble Reduction

When the ensemble reduction parameter is turned to false, ensemble refinement outputs all of the models in the ensemble rather than selecting down a smaller number of models to match the R_{free} tolerance value. By turning this value off, we hypothesized that the size of the ensembles would all be 500, since that is the number of models created in ensemble refinement (based on default parameters).

Non-default inputs

```
wxray_coupled_tbath_offset, pTLS, tx
parameter values corresponding to
the best Rfree for each individual
structure.
ensemble_reduction = False
```

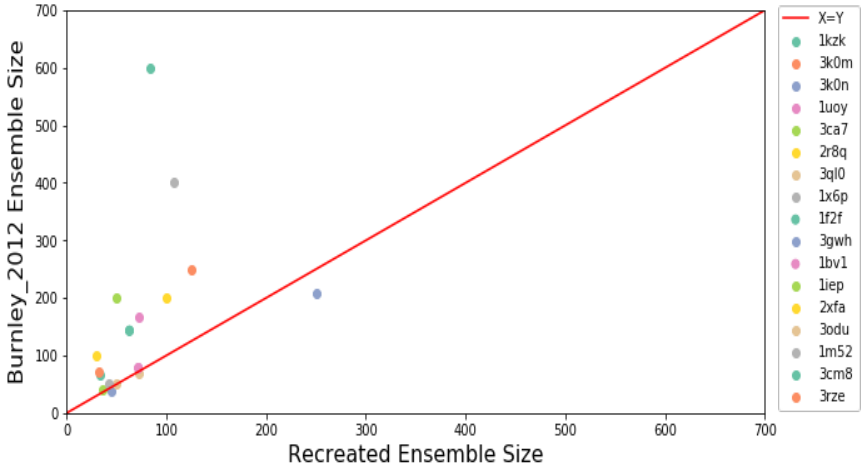


***Figure 12.** Recreated ensemble sizes are smaller compared to the ensemble sizes in Burnley 2012 (R^2=0.4).*

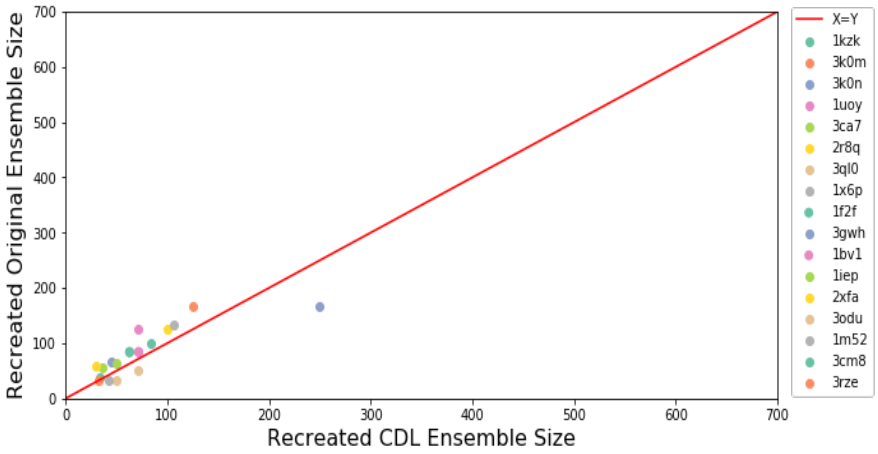


Figure 13. *Recreated ensemble sizes with CDL parameter=False are similar to the initially recreated ensemble sizes(R^2=0.82).*

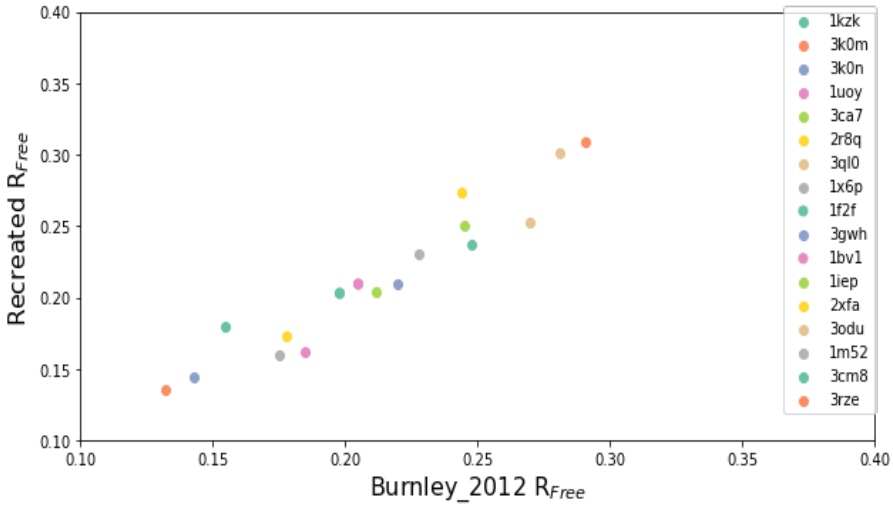


Figure 14. *Recreated R_{free} were highly correlated with the* Burnley 2012 *paper (R^2=0.96).*

Conclusion

While turning off the ensemble reduction parameter did increase the ensemble size, we only observed two ensemble size. There was not a trend of the ensemble size observed in Burnley 2012 and our recreated ensemble size (R^2=-0.03), see figure 15. There was still a high correlation between the original and recreated R_{free} (R^2=0.95) as shown in figure 16.

Using specific TLS selections from the 2012 paper

After discussing our results with the original authors, we realized that the authors used specific TLS selections and other values for the pTLS, tx, and wxray_coupled_tbath_offset parameters. We then set the following parameters from their log files in ensemble refinement Phenix version 1.16. Of note, there were additional parameters that were different between the two versions that we were not able to change.

Parameters changed:

- `pTLS`
- `tx`
- `wxray_coupled_tbath_offset`

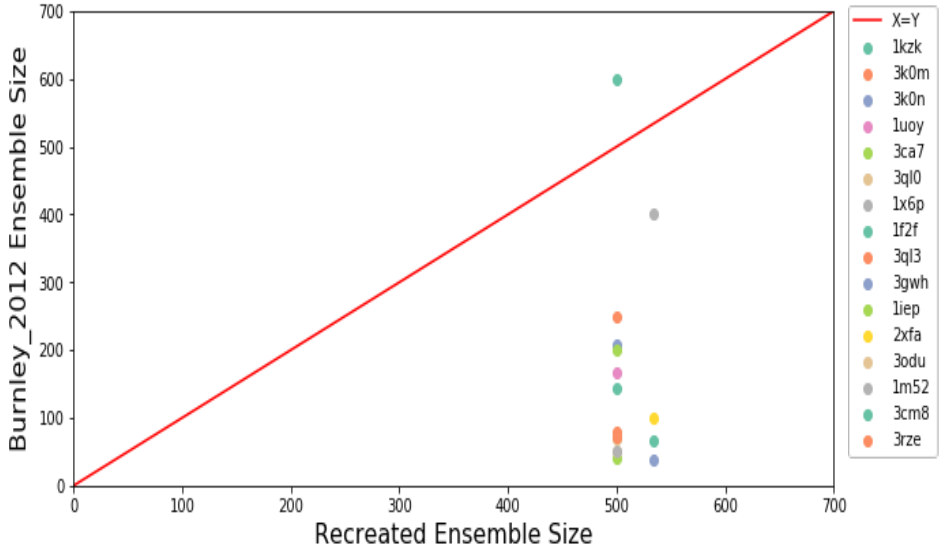


Figure 15. *Recreated ensemble sizes are larger compared to the ensemble sizes in Burnley 2012 (R^2=-0.03).*

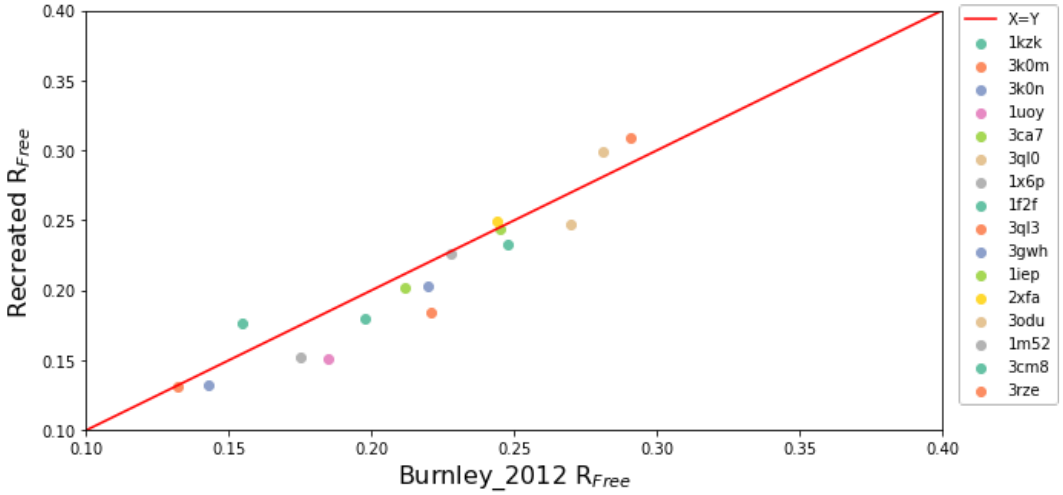


Figure 16. *Recreated R_{free} were highly correlated with the* Burnley 2012 *paper (R^2=0.95).*

- `pTLS selections`
- `harmonic restraints`

Conclusion

Altering these parameters more closely resembled our original recreation both in ensemble size and Rfree and does not underly the larger ensembles in the 2012 paper, as shown in figure 17.

Overall conclusions

While we were not fully able to recreate the ensembles in the Burnley 2012 analysis, we are confident that ensemble refinement is stable and outputs interesting representations of conformational heterogeneity. Metrics that can be used to assess those representations, such as R values or RMSF are not greatly affected by the changes to the method. We would advise future users of the ensemble refinement methods that you may observe lower number of models in each ensemble compared to the Burnley 2012 paper. The only parameter change that seemed to increase the ensemble size was changing

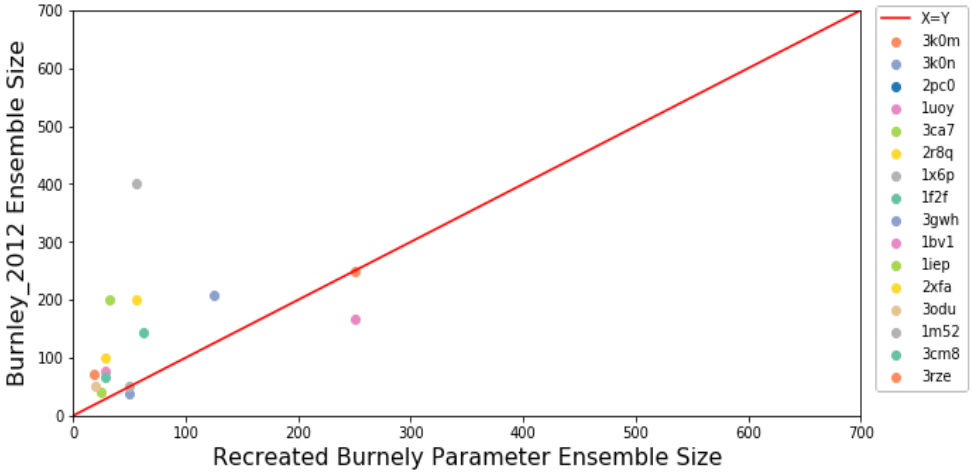


Figure 17. *Recreated ensemble sizes with the Burnley 2012 parameters not correlated to the Burnley 2012 parameters (R^2=0.41).*

ensemble_reduction to false but these values did not correlate with the ensemble sizes observed in Burnley 2012 paper. As advised on the phenix website, we suggest that users perform a grid search over the parameters of tx, wxray_coupled_tbath_offset, and pTLS. Additionally, adding harmonic restraints on all non-water HETATMS is important. All input parameters for our analysis are available (https://ucsf.app.box.com/folder/95195345802) . Moving forward, we would like to encourage publishing the exact parameters used in any refinement procedure for reproducibility.

References

Burnley, B. Tom, Pavel V. Afonine, Paul D. Adams, and Piet Gros. 2012. "Modelling Dynamics in Protein Crystal Structures by Ensemble Refinement." *eLife*.