

Glycan structure determination using CheSweet and Bayesian inference

Pablo G. Garay, Agustina Arroyuelo, Jorge A. Vila, and Osvaldo A. Martin

IMASL-CONICET, Universidad Nacional de San Luis, Italia 1556, 5700-San Luis, Argentina.

*garaypablo01@gmail.com

Introduction

Glycans are the most abundant and structurally diverse biomolecules in nature. The knowledge of their tridimensional structure is necessary to understand in detail, at atomic level, the molecular processes in which they are involved. Chemical Shifts (CS) are observables obtained from Nuclear Magnetic Resonance experiments that are highly sensitive probes to sense conformational changes. CS can be calculated accurately using quantum chemical tools, although these computations are very demanding for routine computations of more than a few conformations. For that reason we have developed CheSweet⁽¹⁾, a Python module for accurate and fast computation of CS. Here we discuss some details of its implementation, and a result of using CheSweet as part of a Bayesian model implemented in PyMC3⁽²⁾ to predict glycan structures.

Results and Discussion

2LIQ – Energy profiles

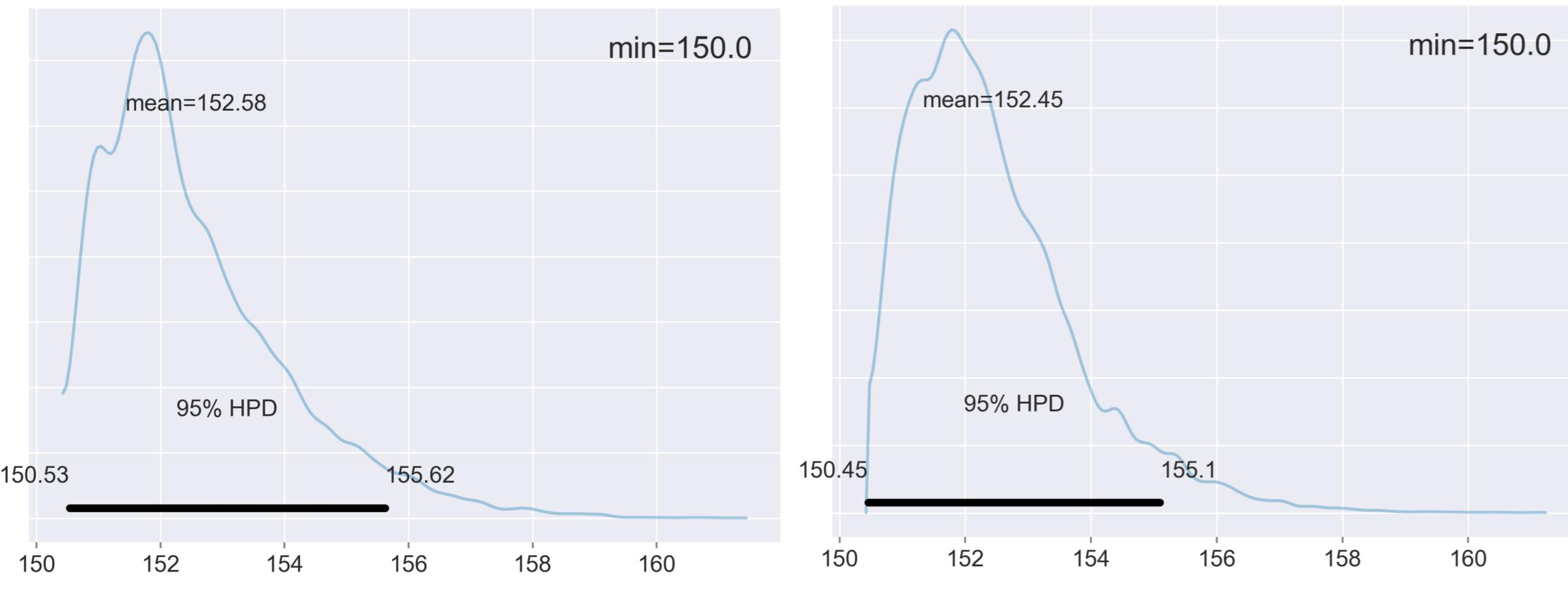


Figure 4a. Energy profile for 2LIQ without CS.

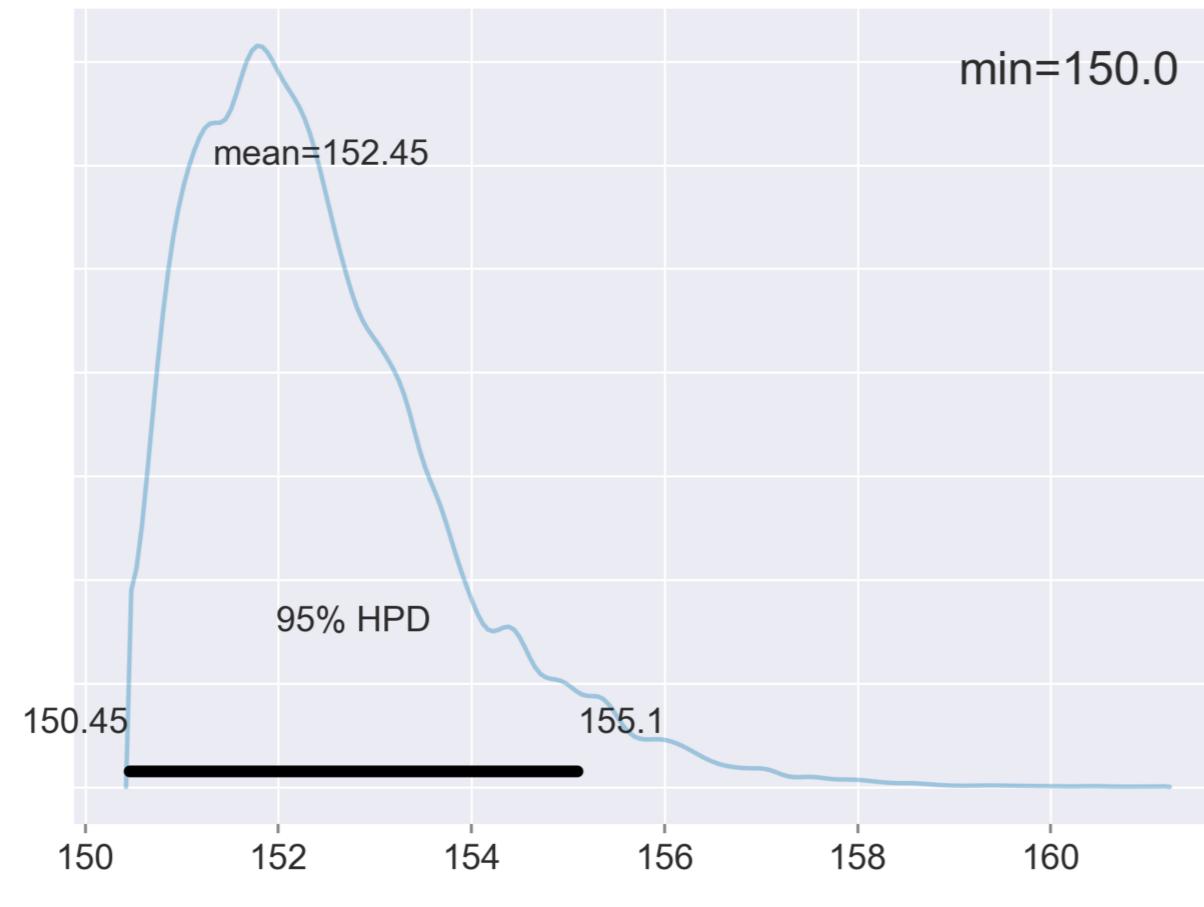


Figure 4b. Energy profile for 2LIQ with CS.

2LIQ – Ramachandran plots

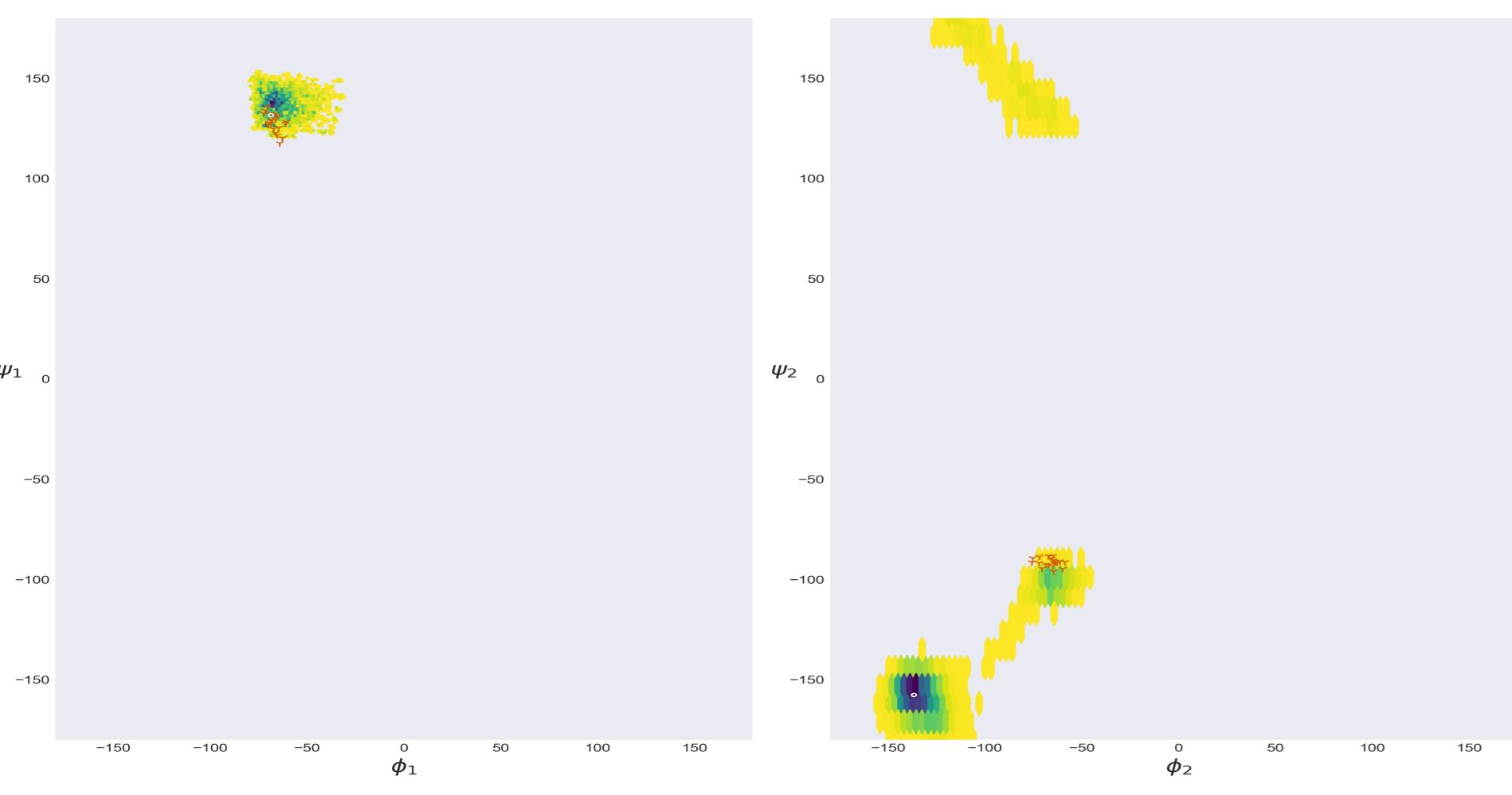


Figure 5a. Ramachandran plot for 2LIQ without CS.

2KQO – Energy profiles

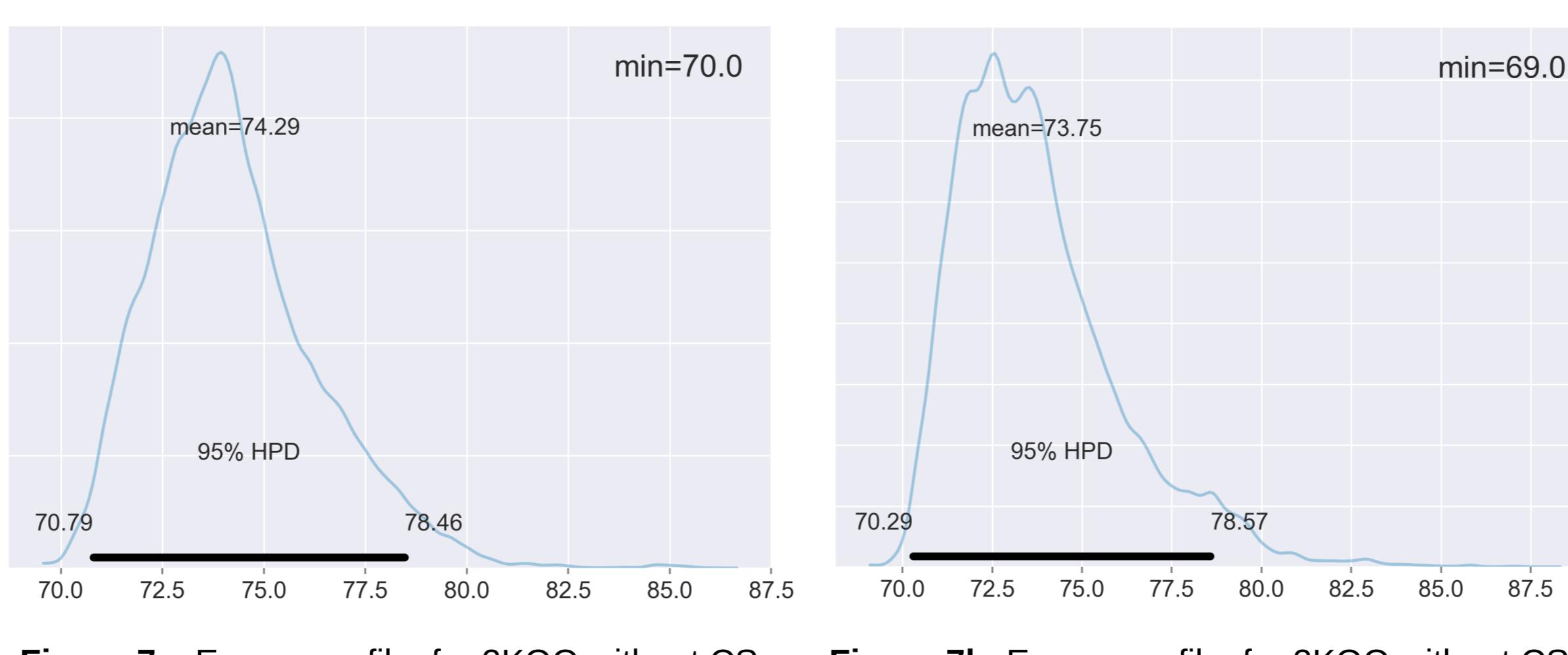


Figure 7a. Energy profile for 2KQO without CS.

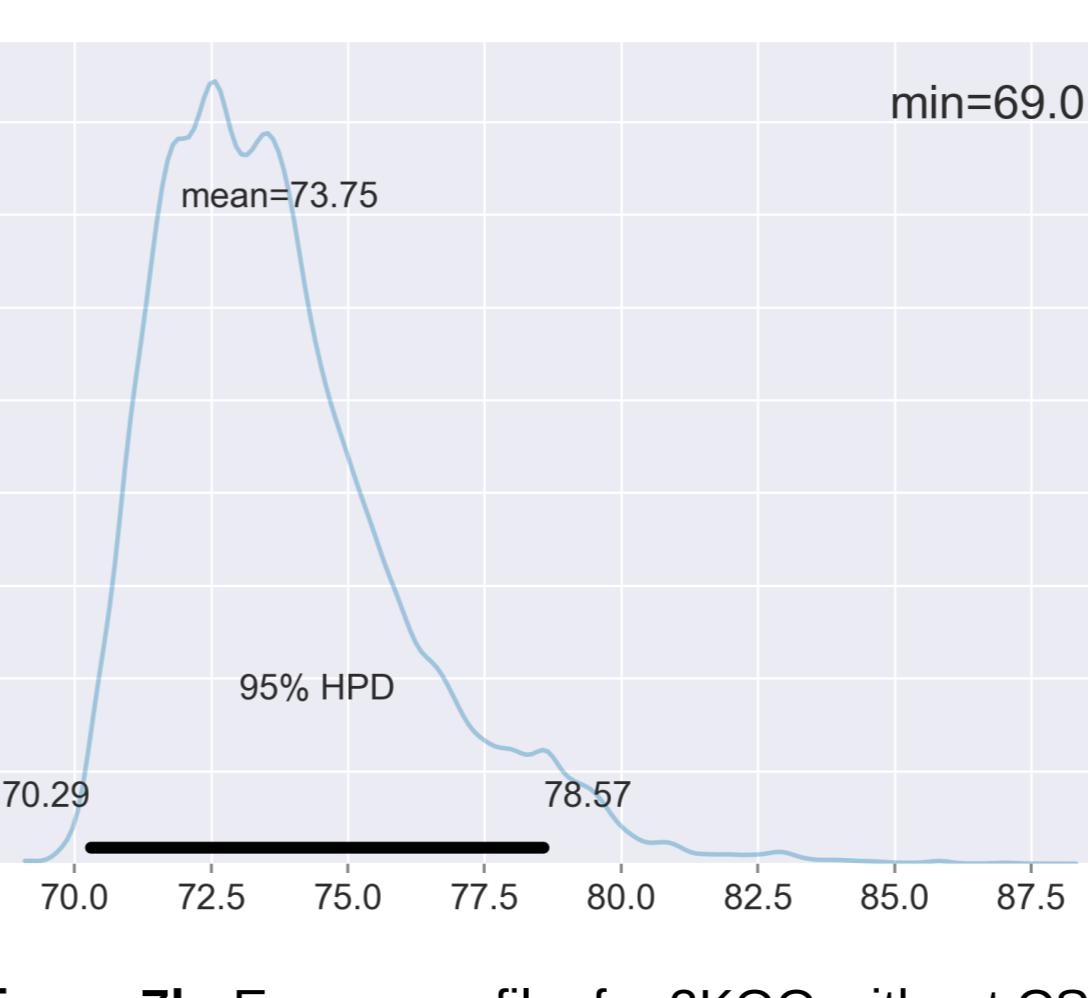
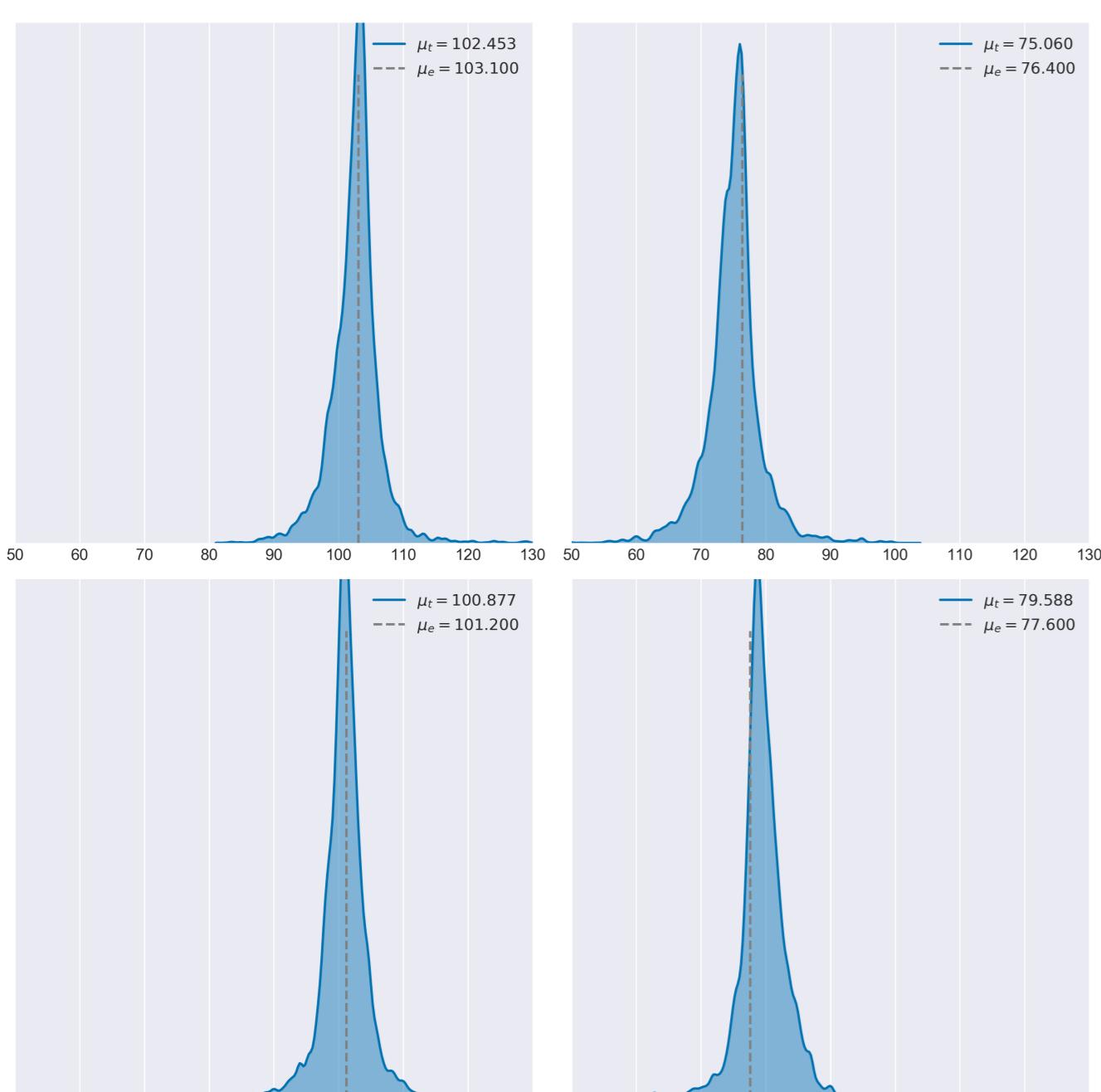
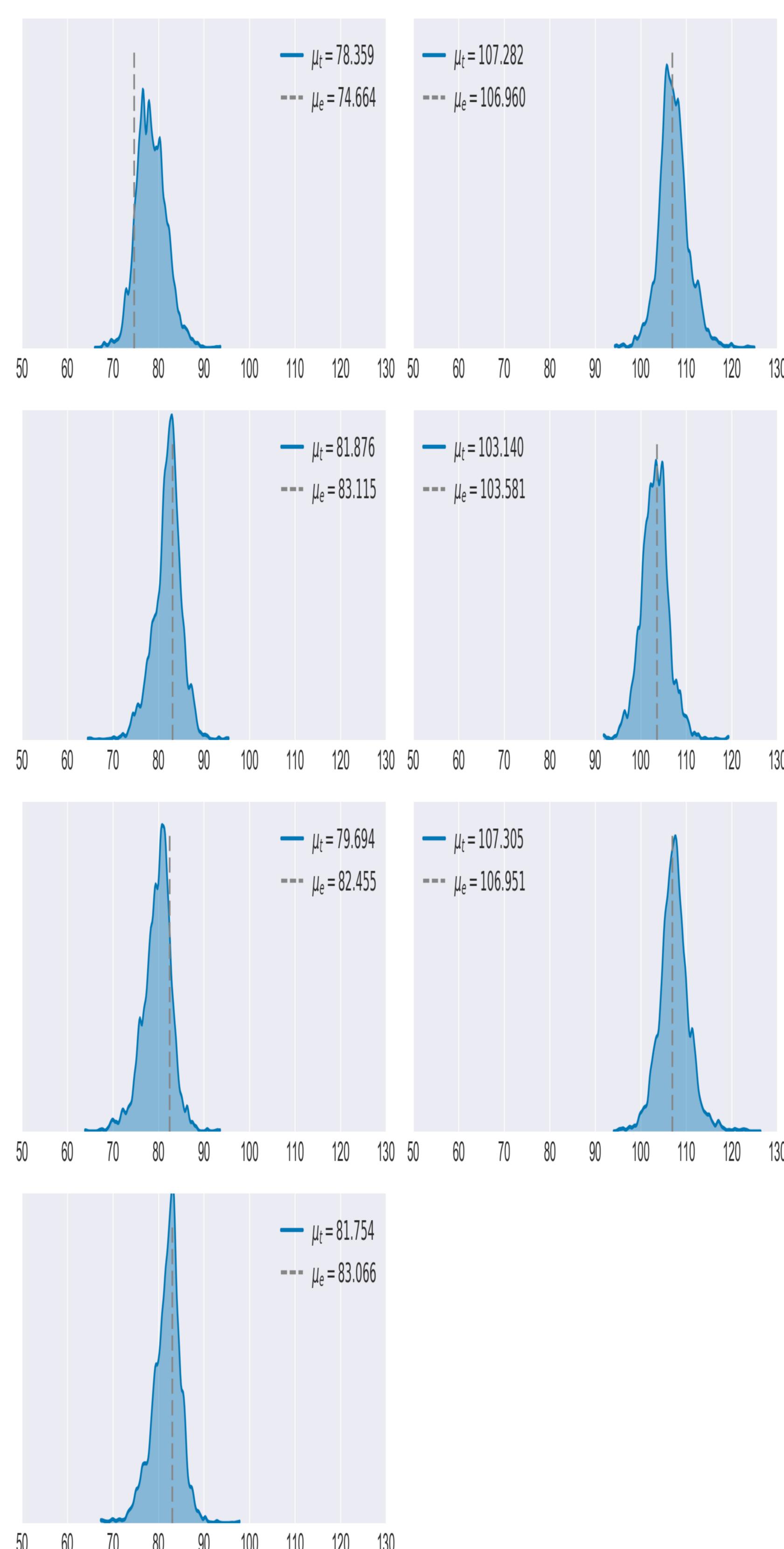


Figure 7b. Energy profile for 2KQO with CS.

2LIQ – mu distributions


 Figure 6. μ distributions and experimental CS values of 2LIQ.

2KQO – mu distributions


 Figure 9. μ distributions and experimental CS values of 2KQO.

Conclusions

Even when there are no differences in the energy profiles or in the Ramachandran plots between the calculation with or without CS, the use of CS helps to avoid sampling a zone far from the experimental torsional angles. In an earlier version of this work we used another sampler (with a worse performance than the used here) and in that case we observed that CS help find the correct minima.

The good agreement between means and the experimental CS is evidence of the accuracy of CheSweet computations.

References

- (1) CheSweet <https://github.com/BIOS-IMASL/chesweet>
- (2) Salvatier J, Wiecki TV, Fonnesbeck C (2016) Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2:e55. doi: 10.7717/peerj-cs.55
- (3) Garay PG, Martin OA, Scheraga HA, Vila JA. 2014. Factors affecting the computation of the ¹³C shieldings in disaccharides. *Journal of Computational Chemistry* 35:1854-1864.
- (4) BoMeBa <https://github.com/BIOS-IMASL/bomeba0>

ACKNOWLEDGEMENTS: This work was supported by ANPCyT-Argentina PICT-2014-0556 and PICT-joven 0218 Probabilistic Programming for Structural Bioinformatic.

Figure 1. Glycosidic portion of 2LIQ with the sequence: GlcNAc β 1,4[Fuc α 1,3]GlcNAc β

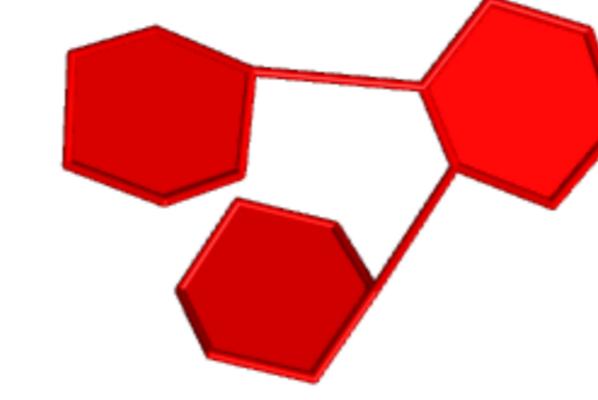

 Figure 1. Glycosidic portion of 2LIQ with the sequence: GlcNAc β 1,4[Fuc α 1,3]GlcNAc β

Figure 2. Glycosidic portion of 2KQO with the sequence: GlcpA- β 1,3-GalpNAc- β 1,4-GlcpA- β 1,3-GalpNAc- β 1,4-GlcpA- β 1,3-GalpNAc β

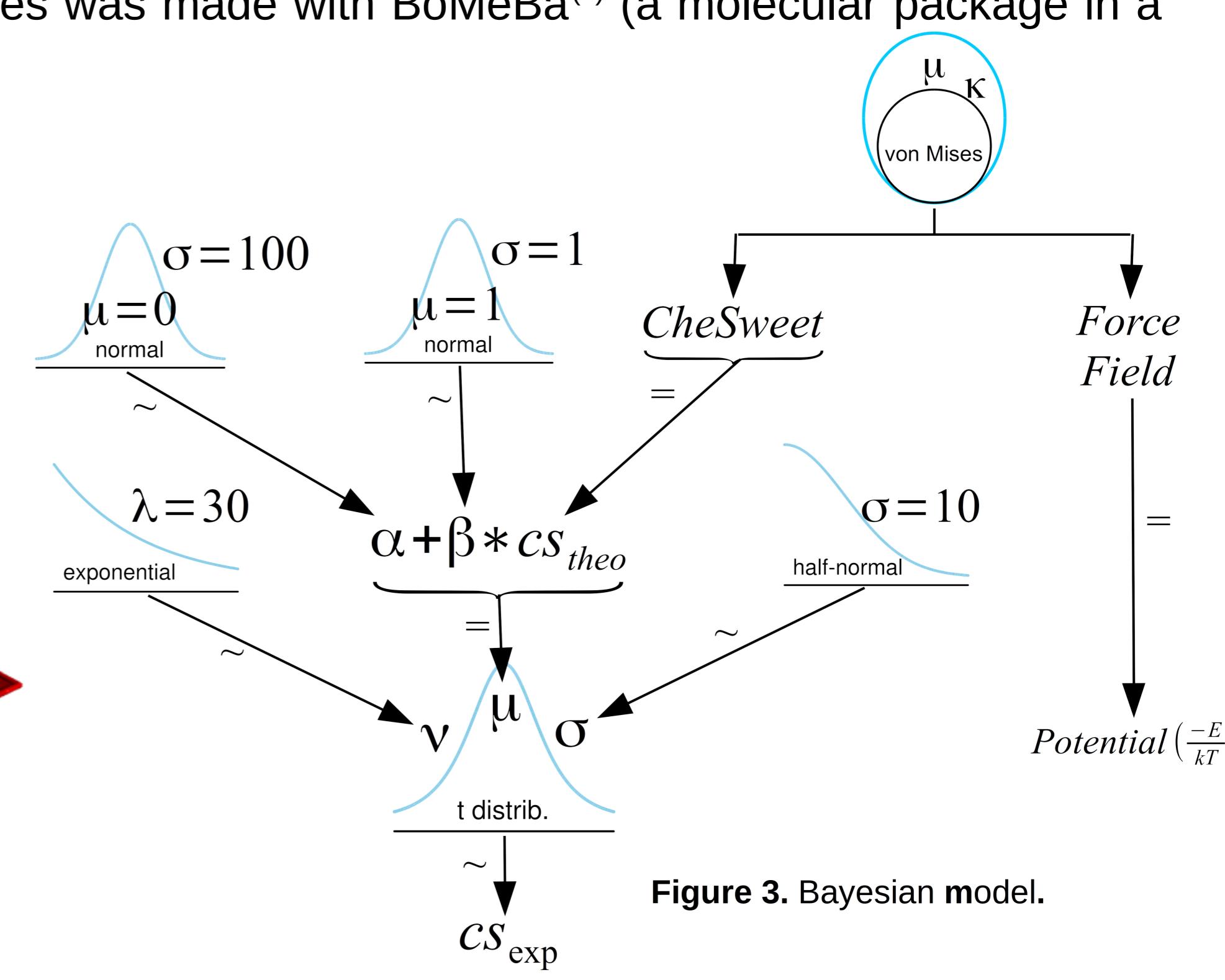

 Figure 2. Glycosidic portion of 2KQO with the sequence: GlcpA- β 1,3-GalpNAc- β 1,4-GlcpA- β 1,3-GalpNAc- β 1,4-GlcpA- β 1,3-GalpNAc β


Figure 3. Bayesian model.

2KQO – Ramachandran plots

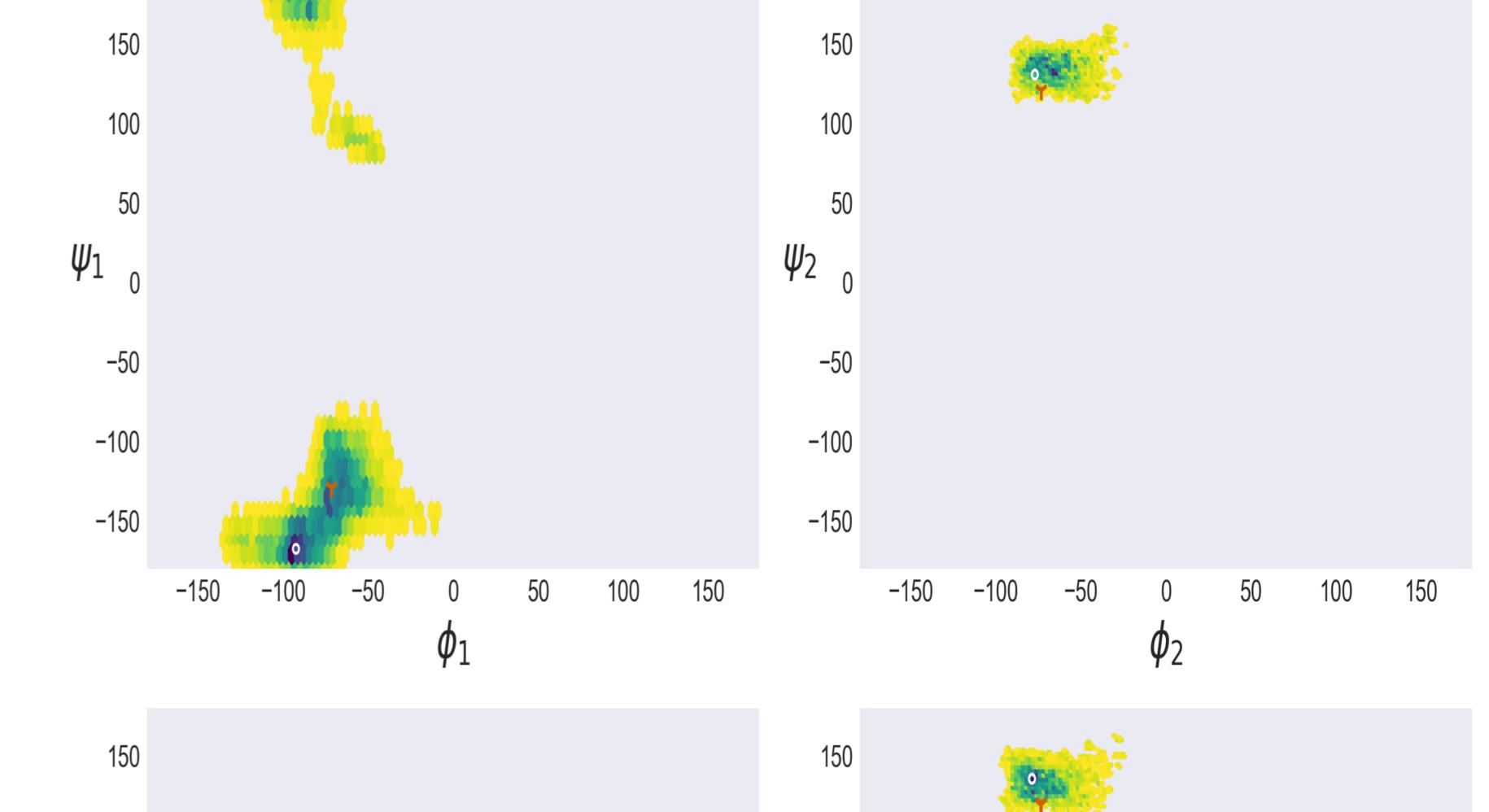


Figure 8a. Ramachandran plots for 2KQO without CS.

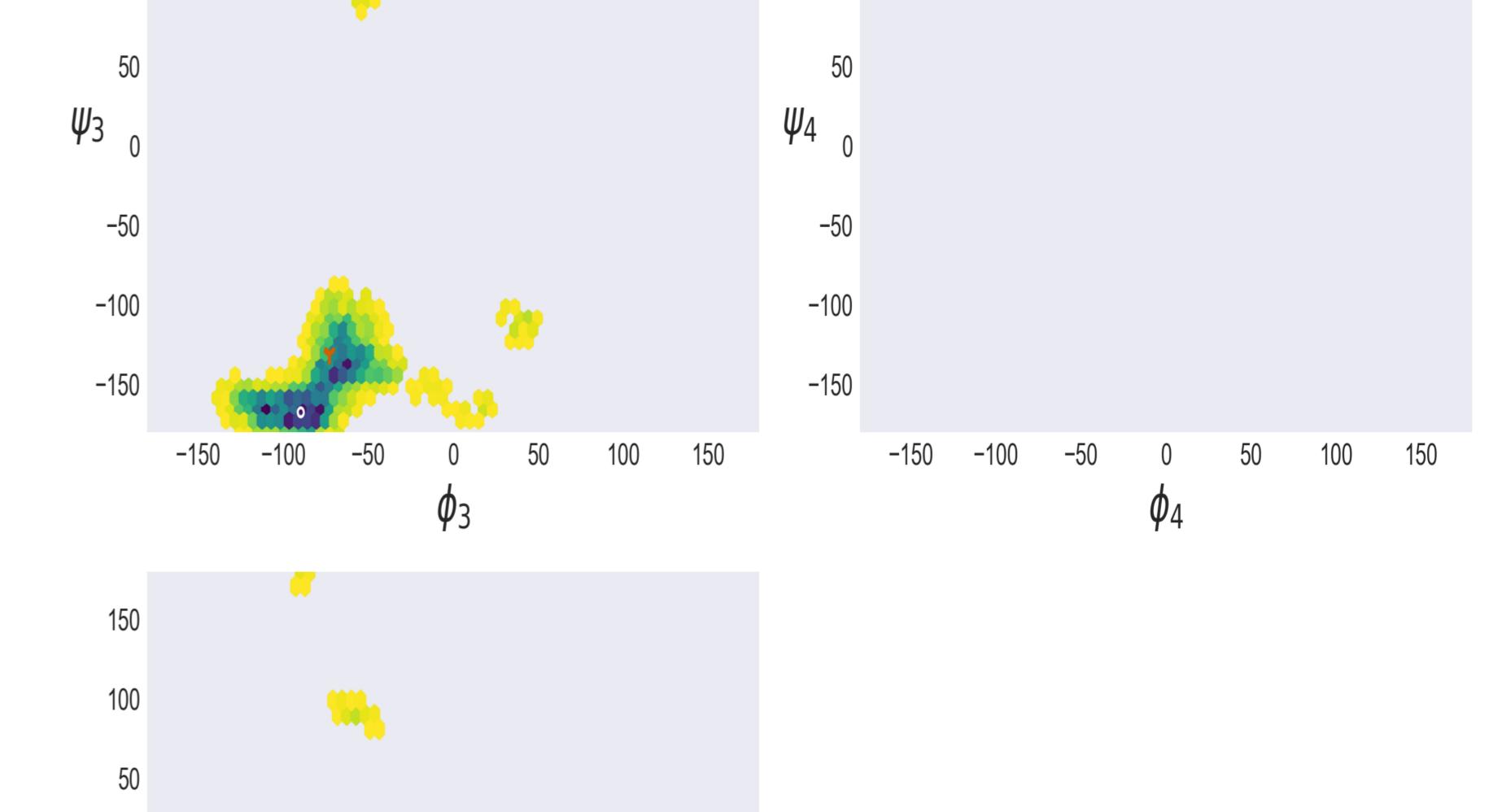


Figure 8b. Ramachandran plots for 2KQO with CS.

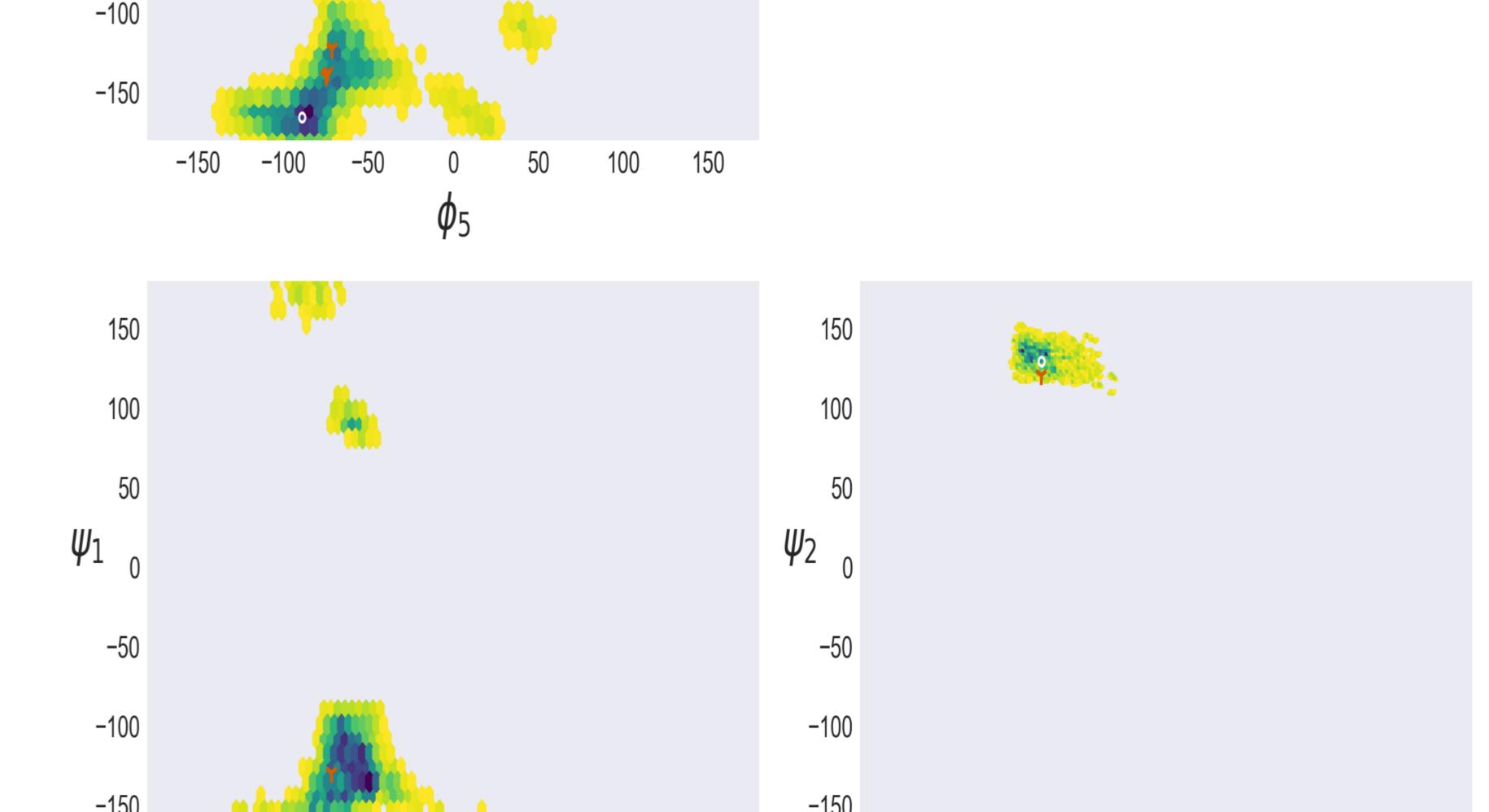


Figure 8c. Ramachandran plots for 2KQO with CS.

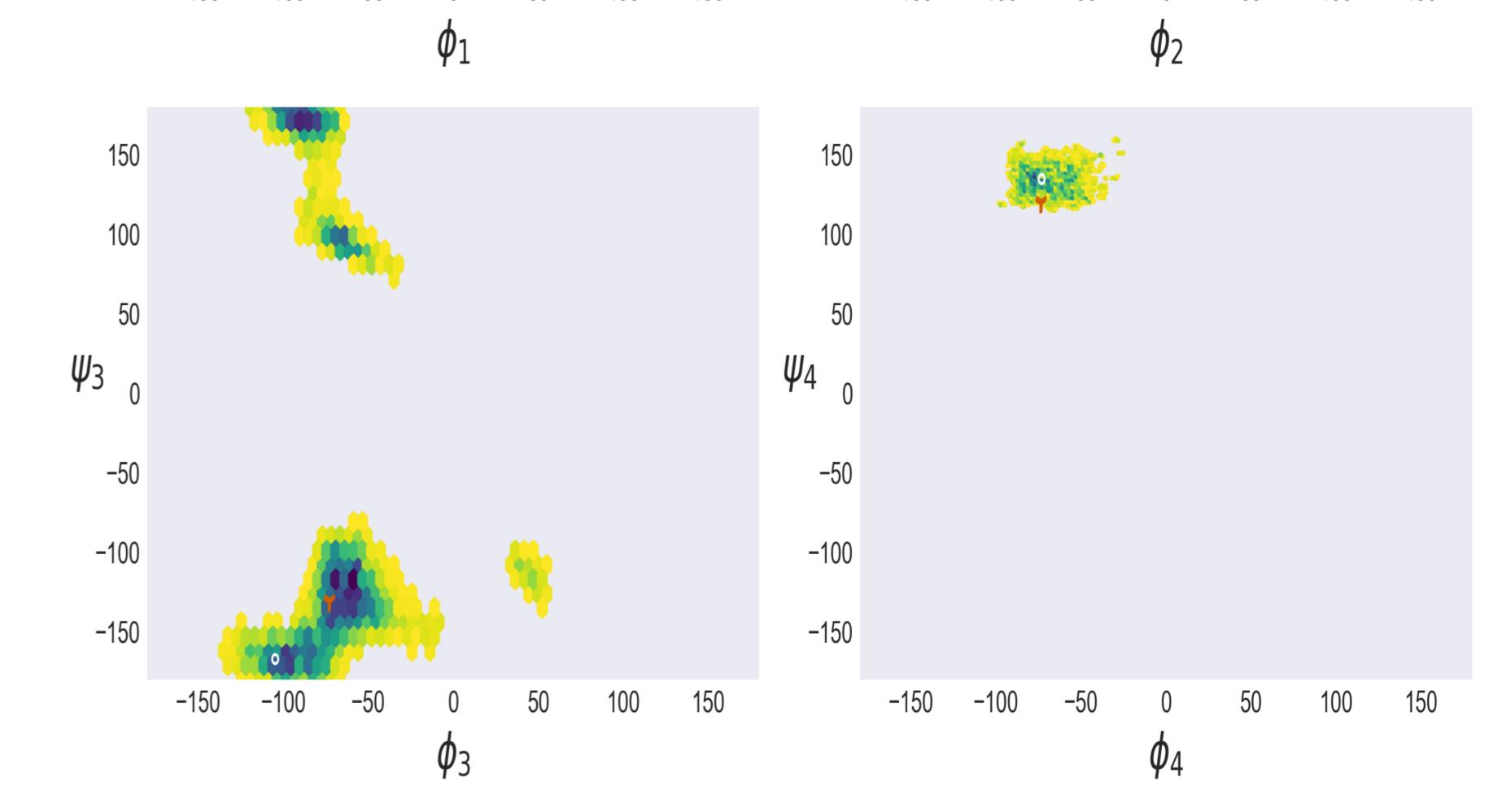


Figure 8d. Ramachandran plots for 2KQO with CS.

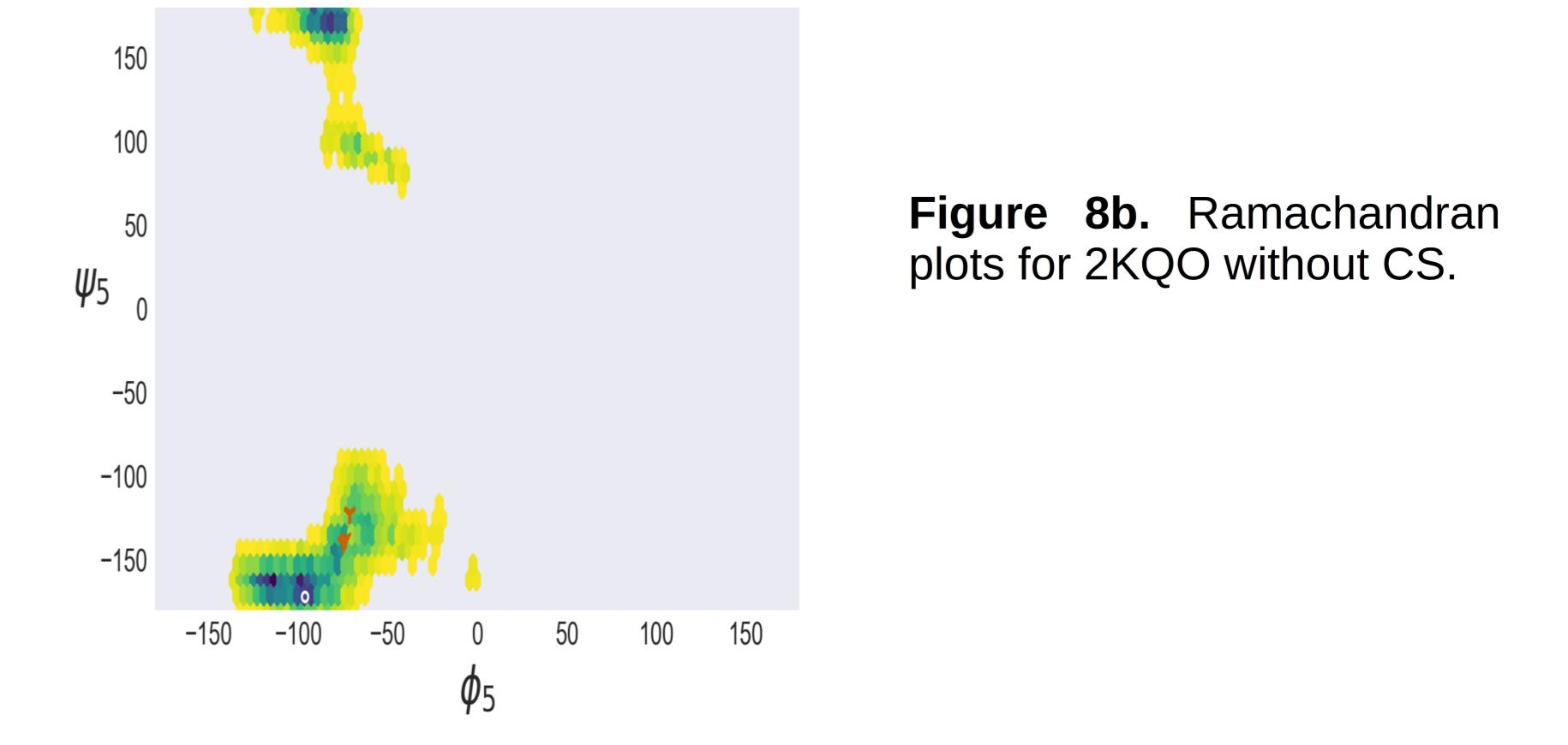


Figure 8e. Ramachandran plots for 2KQO with CS.