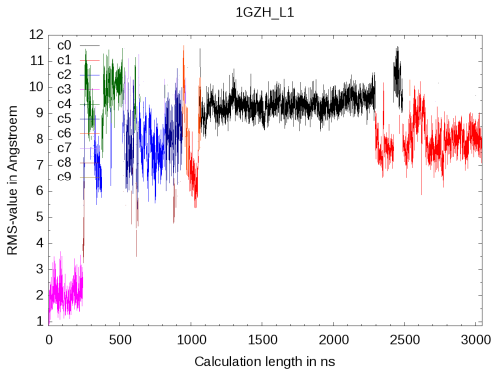
Do we need to analyze µs MD simulations differently?

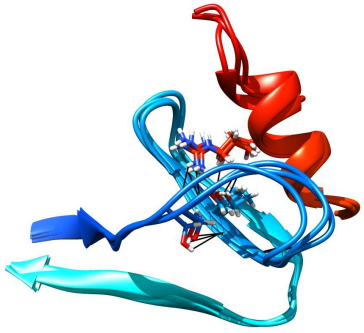
Achim Sandmann, Harald Lanig, Tim Clark

Department of Chemistry and Pharmacy, Computer Chemistry Center,

Friedrich Alexander University, Naegelsbachstr. 25, 91052 Erlangen, Germany

Biologically relevant timescales of 1 µs and more are becoming accessible in MD simulations because of the continuing increase in computational performance. Among the available methods for sampling the conformational space associated with such timescales, the brute force method has the advantage of being highly analogous to the actual processes and using little *a priori* knowledge [1]. We now present some methods that are useful for interpreting the vast amount of data associated with these kinds of simulations.





The system under investigation, the p53 core domain, has several flexible regions, as can be seen directly from B-factor plots. After cluster analysis of six 2 – 4 µs simulations, and comparison of the representative structures via the RMSD matrix, it is apparent that still no structural convergence is reached.

Subsequently, individual flexible regions were investigated. A set of cluster analyses was performed, in each of which only residues of one individual region were regarded (example shown in the left picture). Hydrogen-bond analyses can show structural homogeneity within clusters of those individual regions (example shown in right picture), as well as among clusters with similar representative structures. They can also show structural differences between clusters with high RMSD differences and thermodynamic explanations for sudden changes in the RMSD trajectory of the respective region.

For flexible loop regions, clustering by means of a DASH analysis, which is based on internal coordinates [2], also shows a good match to the results of Cartesian clustering based on atomic positions.

These analyses show that the ensembles obtained with cluster analysis indeed represent discrete conformations. The vastly simplified data provided by cluster analysis can be used to investigate individual regions systematically and identify reoccurring structures. In the p53 core-domain system, frequently reappearing structures are found in some of the flexible regions. While still not converged, the conformational space of these regions is thought to be sampled to a larger extent, compared to other regions or the whole protein domain.

[1] M. C. Zwier, L. T. Chong, *Current opinion in pharmacology* **2010**, *10*, 745-752.

[2] D. W. Salt, B. D. Hudson, L. Banting, M. J. Ellis, M. G. Ford, *Journal of Medicinal Chemistry* **2005**, *48*, 3214-3220.