

## Microsymposium

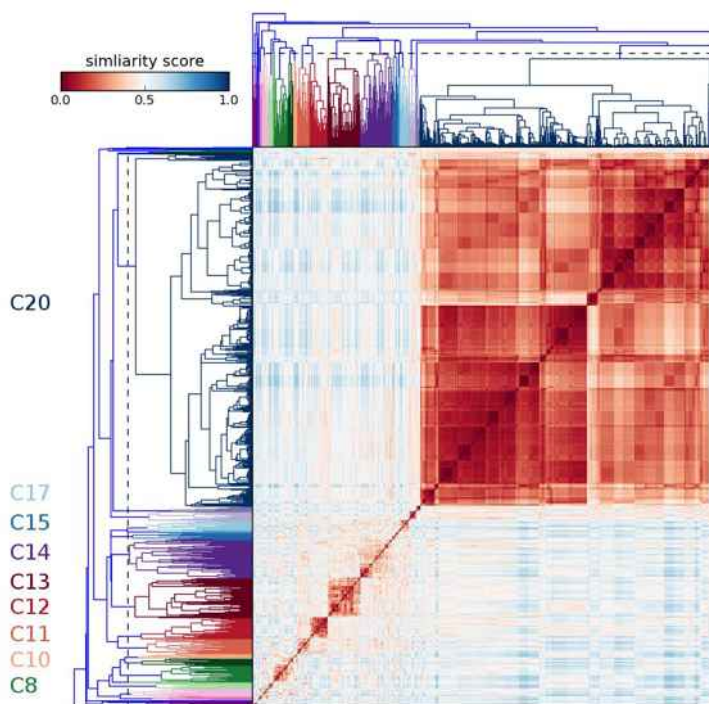
MS77.O06

### *Chemical clustering and visualization applied to macromolecular crystallography*

A. Bruno<sup>1</sup>, A. Ruby<sup>1</sup>, J. Luft<sup>2</sup>, T. Grant<sup>3</sup>, J. Seetharaman<sup>4</sup>, J. Hunt<sup>4</sup>, G. Montelione<sup>5,6</sup>, E. Snell<sup>2,3</sup>

<sup>1</sup>State University of New York (SUNY), Center for Computational Research, Buffalo, NY, USA, <sup>2</sup>Hauptman-Woodward Medical Research Institute, Buffalo NY, USA, <sup>3</sup>SUNY Buffalo Dept. of Structural Biology, Buffalo NY, USA, <sup>4</sup>Columbia University, Department of Biological Sciences, The Northeast Structural Genomics Consortium, New York, NY, USA, <sup>5</sup>Rutgers, The State University of New Jersey, Northeast Structural Genomics Consortium, Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, Piscataway, NJ, USA, <sup>6</sup>University of Medicine and Dentistry of New Jersey, Department of Biochemistry, Robert Wood Johnson Medical School, Piscataway, NJ, USA.

Many bioscience fields employ high-throughput methods to screen multiple biochemical conditions. The analysis of these becomes tedious without a degree of automation. Crystallization, a rate limiting step in biological X-ray crystallography, is one of these fields. Screening of multiple potential crystallization conditions (cocktails) is the most effective method of probing a proteins phase diagram and guiding crystallization but the interpretation of results can be consuming. To aid this empirical approach a cocktail distance coefficient was developed to quantitatively compare macromolecule crystallization conditions and outcome. These coefficients were evaluated against an existing similarity metric developed for crystallization, the C6 metric, using both virtual crystallization screens and by comparison of two related 1,536-cocktail high-throughput crystallization screens. Hierarchical clustering was employed to visualize one of these screens and the crystallization results from an exopolyphosphatase-related protein from *Bacteroides fragilis*, (BfR192) overlaid on this clustering. This demonstrated a strong correlation between certain chemically related clusters and crystal lead conditions. While this analysis was not used to guide the initial crystallization optimization, it led to the re-evaluation of unexplained peaks in the electron density map of the protein and the insertion and correct placement of a sodium, potassium and phosphate atoms in the structure. With these in place, the resulting structure of the putative active site demonstrated features consistent with active sites of other phosphatases which are involved in binding the phosphoryl moieties of nucleotide triphosphates. The new distance coefficient appears to be robust in this application and coupled with hierarchical clustering and the overlay of crystallization outcome reveals information of biological relevance. While tested with a single example the potential applications appear promising.



**Keywords:** Crystallization, Hierarchical clustering, Chemical similarity