



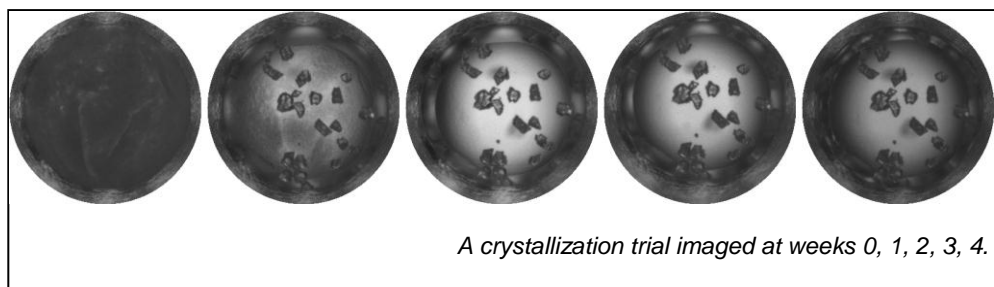
October 2010 Update

Milestones: 50% of work units completed, work units increased

In June The Help Conquer Cancer project briefly reached the 50%-complete milestone the week of June 2. Due to the eagerness of HCC volunteers and recent performance improvements to HCC code, and anticipating future performance improvements, HCC scientists added 20,000,000 additional extra work units (images) to the project in late June. These additional images were obtained from proteins screened by Hauptman-Woodward Medical Research Institute since the start of the HCC project. With the added work units, the total completed units regained the 50%-complete mark the week of August 26.

Modeling protein crystallization over time

Using HCC results, we have built a model of protein crystallization over time. A single crystallization trial (one protein + one chemical cocktail) generates multiple images as the experiment is photographed at multiple time points. The images relate to one another, as the crystallization outcome usually evolves over the time course. In the example below, a crystallization trial proceeds from precipitate (left), through mixed precipitate and crystal, to crystal (right).

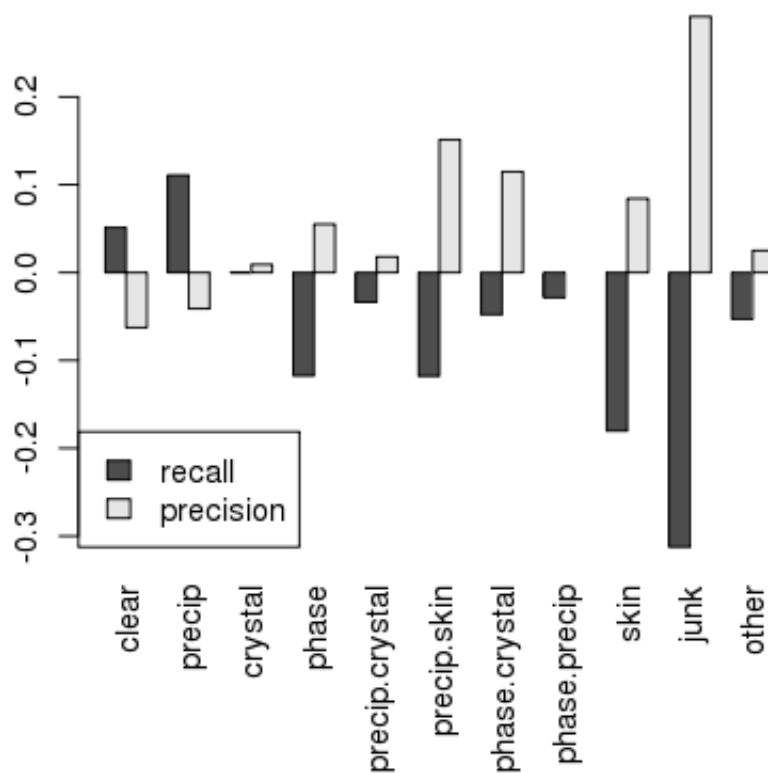


We are using HCC results from complete time courses to improve automatic image classification. We consider 11 different crystallization states, and model the changing crystallization state over time using a *Hidden Markov Model*, a tool frequently used in biological sequence analysis.

The accuracy of the classifier is measured in terms of precision and recall (i.e., specificity and sensitivity). Preliminary results show a mixed effect: adding the time dimension increases the recall of the most common crystallization states, *clear*, and *precipitate*, and decreases recall of all others. The effect on precision is the opposite: decreased for *clear*, and *precipitate*, and increased for all others.



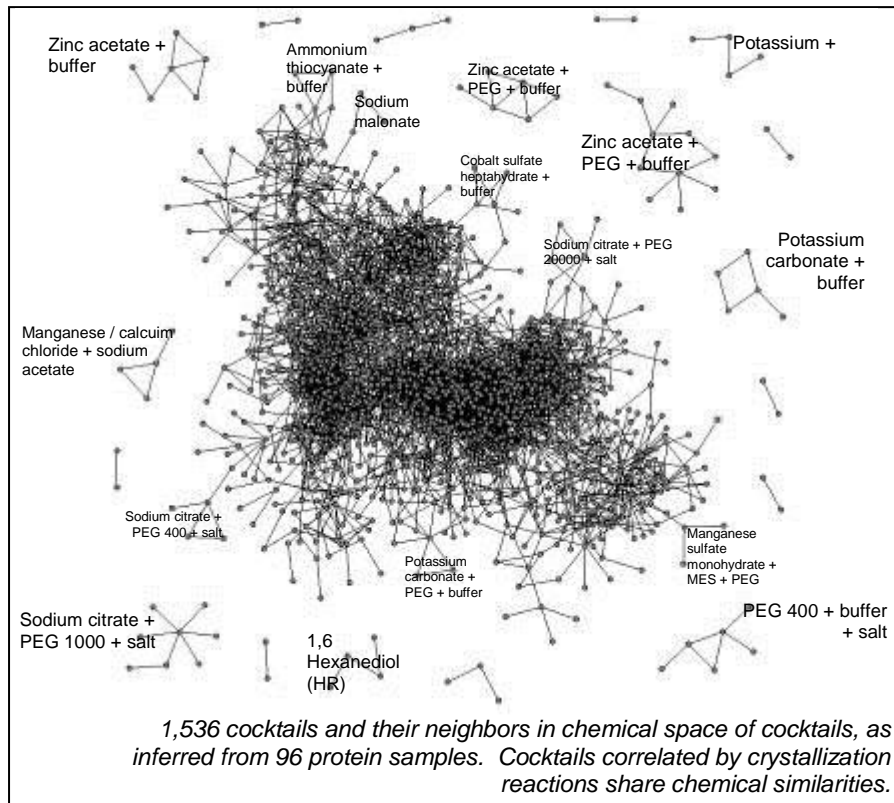
Net effect of time-series modeling

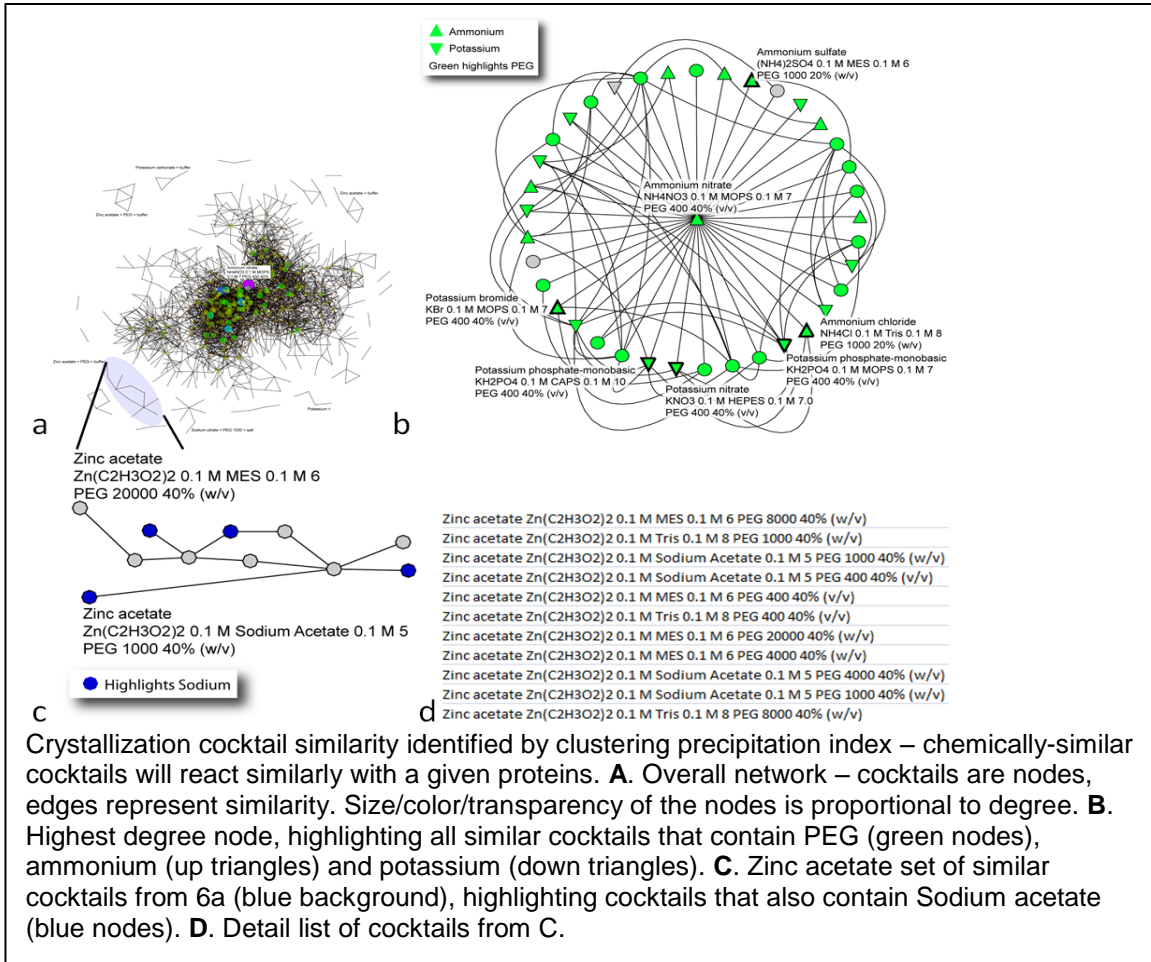


The model can be enriched further by adding other relationships between images. Just as multiple images of the same experiment over time are related, so are multiple trials on the same protein with similar chemical cocktails. Chemically-related cocktails form a connected network of correlated crystallization outcomes. Work in this area is ongoing.



Help Conquer Cancer





Thank you,

C. A. Cumbaa, Yulia Kotseruba and I. Jurisica

