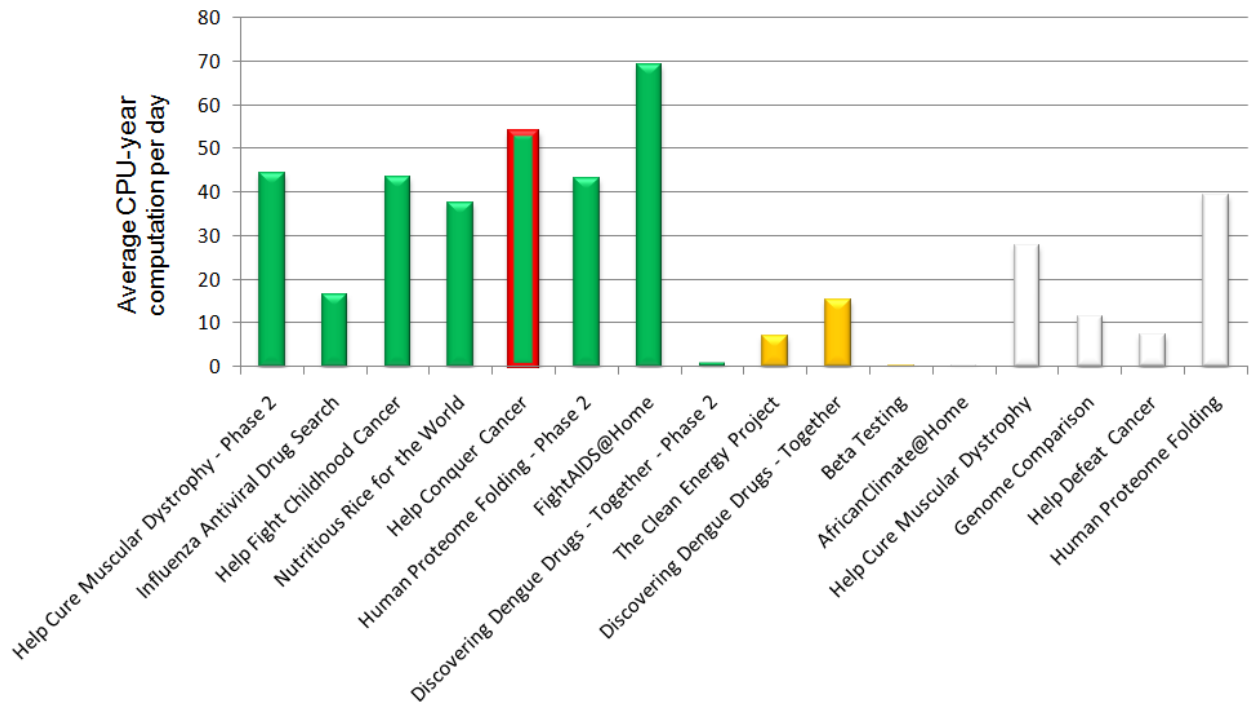




# Update March 2010

Your continued support of the Help Conquer Cancer project (HCC) is highly appreciated.

Since the launch of HCC project in November 2007, WCG members contributed over 46,906 years of run time, averaging over 53 years of computation per day. To date 64,724,807 results were returned (*Statistics Last Updated: 3/26/10 00:05:59 (UTC)*). We are thrilled and thankful to continue to rank as the second “most CPU/day computation” project on WCG (in the graph below, green bars show active projects, orange bars represent intermittent projects, and white bars correspond to completed projects).



## Summary

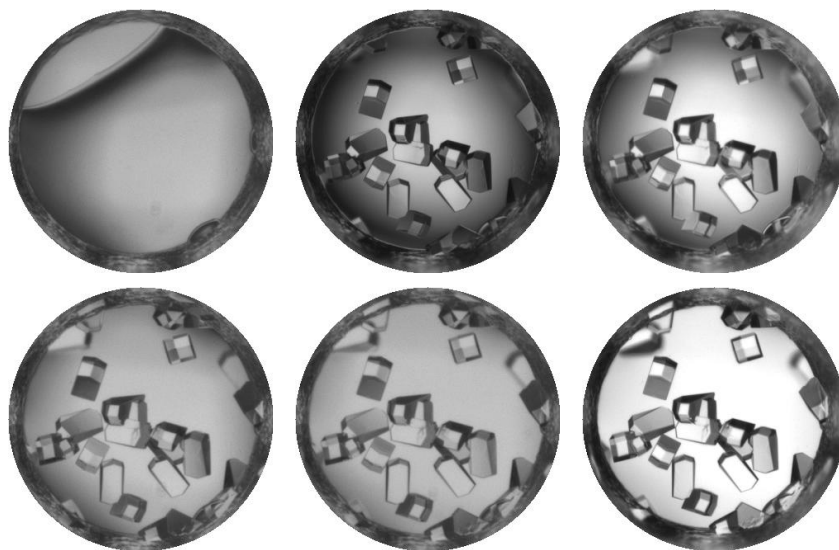
12,375 primary features computed by the HCC project are augmented by a set of 2,533 features computed from the primary, creating a final set of 14,908 features. These image features and a massive training set of 165,351 hand-scored images were used to train multiple Random Forest classifiers that accurately recognize multiple crystallization outcomes, including crystals, clear drops, precipitate, and others. The classifier achieves good performance, as highlighted by the results from testing on 8,528 classified images:



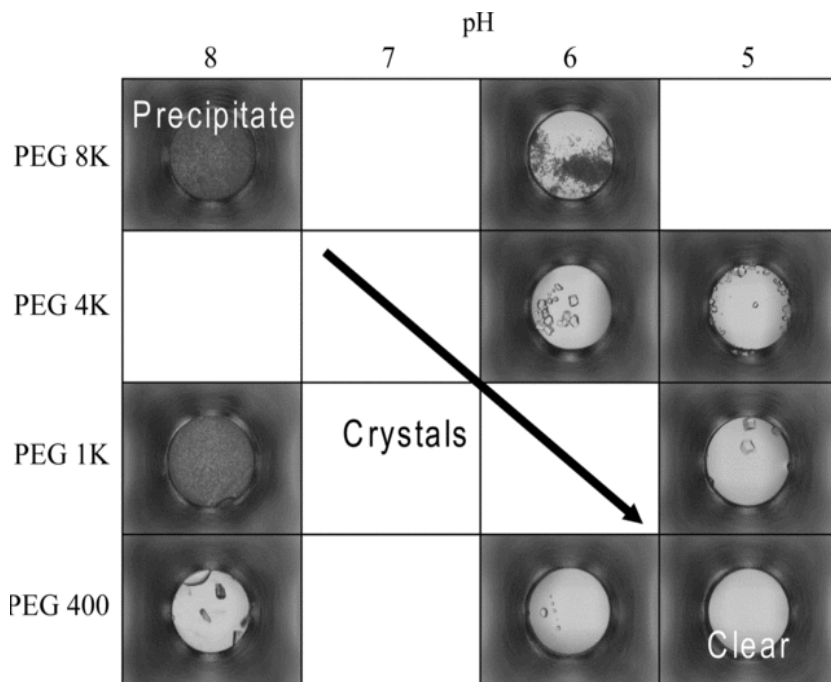
		Machine classification										Total	Recall
		clear	precip	crystal	phase	precip & crystal	precip & skin	phase & crystal	phase & precip	skin	junk		
Truth	clear	2746	3	60	53	0	0	0	0	1	24	2887	<b>0.951</b>
	precip	29	2213	45	77	343	185	1	0	2	2	2897	<b>0.764</b>
	crystal	66	9	531	22	96	6	11	0	5	2	748	<b>0.710</b>
	phase	48	5	21	469	8	8	3	0	0	2	564	<b>0.832</b>
	precip & crystal	0	55	84	19	425	33	3	0	0	0	619	<b>0.687</b>
	precip & skin	1	46	13	6	16	378	0	0	6	0	466	<b>0.811</b>
	phase & crystal	3	4	67	48	33	2	42	0	1	1	201	<b>0.209</b>
	phase & precip	0	13	3	3	20	6	0	0	0	0	45	<b>0.000</b>
	skin	16	0	4	8	0	0	0	0	32	2	62	<b>0.516</b>
	junk	9	0	3	1	0	1	0	0	0	25	39	<b>0.641</b>
	Total	2918	2348	831	706	941	619	60	0	47	58	8528	
	Precision	<b>0.941</b>	<b>0.943</b>	<b>0.639</b>	<b>0.664</b>	<b>0.452</b>	<b>0.611</b>	<b>0.700</b>		<b>0.681</b>	<b>0.431</b>		

The system successfully recognizes 80% of crystal-bearing images and 95% of clear drops. Some of these results have recently been summarized and published in the *Journal of Structural and Functional Genomics*.

Some of these results identify beautiful crystal leads.



However, we are working toward a system that will be able to take advantage of the sampled crystallizability and chemical space to determine how to optimize crystallization conditions even for proteins that do not lead to crystals. This work is enabled by the consistent and accurate classification of images on WCG.



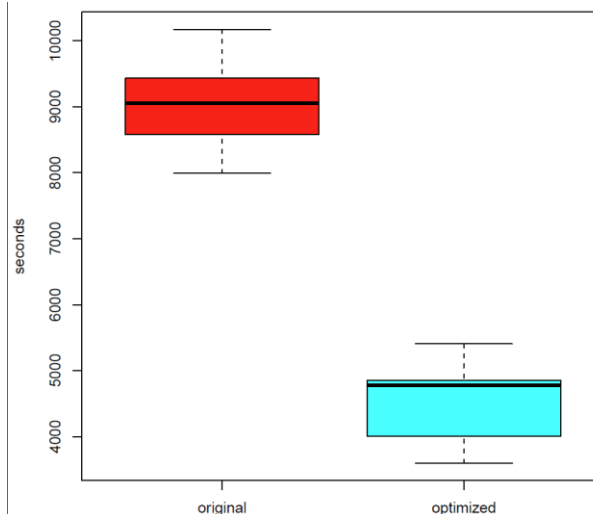
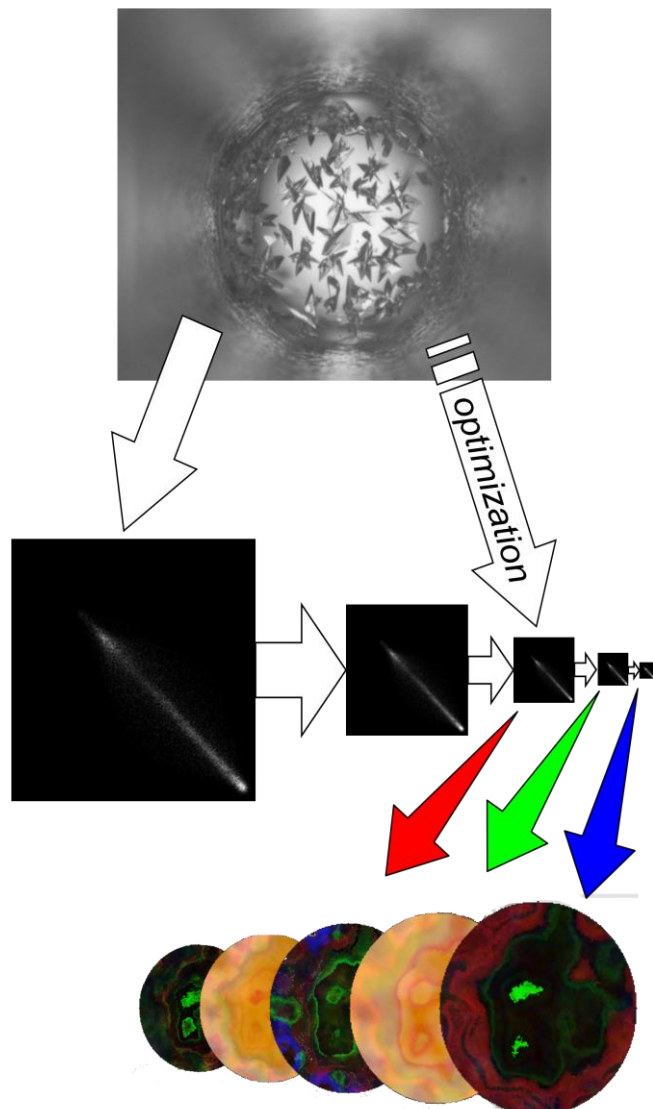
(Source: Snell EH, Nagel RM, Wojtaszczyk A, O'Neill H, Wolfley JL, Luft JR. (2008). The application and use of chemical space mapping to interpret crystallization screening results. *Acta Cryst. D64*, 1240-1249.)

## Performance Improvements

We are happy to announce significant performance improvements to the *Help Conquer Cancer* project. Resulting from a collaboration with the IBM Toronto Software Lab, the optimized HCC software will process work units in less than half the time.

Working with Miso Cilimdžić of IBM DB2 Performance and Solutions Development, we were able to identify a CPU-intensive section of the project code responsible for creating, destroying, and manipulating millions of matrices of  $256^2$  elements. By re-ordering certain calculations in the image analysis and modifying the critical section, we were able to re-cast the computation in terms of  $64^2$ -element matrices, and drastically reduce the number of memory allocations, de-allocations, and accesses. Altogether, the optimizations generate results that are bit-identical to the original HCC software, with an average 54% reduction in run-time.

The optimized HCC project passed Beta testing and has been released to production on Thursday, March 25, 2010.



We wish to thank Miso Cilimdžić for his hard work and expertise, and Roy Bryant of the Department of Computer Science, University of Toronto, for suggesting other code optimizations. We would also like to thank the World Community Grid staff, and the World Community Grid's enthusiastic Beta-testing community.

## Recently published work

### Crystallography

1. Cumbaa, C. A. and Jurisica, I. Protein crystallization analysis on the World Community Grid, *J Struct Funct Genomics*, 11(2): R13, 2010.

### Cancer research

1. Radulovich, N., Pham, N. A., Strumpf, D., Leung, L., Xie, W., Jurisica, I., Tsao, M. S. Differential roles of cyclin D1 and D3 in pancreatic ductal adenocarcinoma. *Mol Cancer*, 9(1): 24, 2010.



- Hui, A.B.Y., M. Lenarduzzi, T. Krushel, L. Waldron, M. Pintilie, W. Shi, B. P.-Ordonez, I. Jurisica, B. O'Sullivan, J. Waldron, P. Gullane, B. Cummings, F-F. Liu, Comprehensive microRNA profiling for head & neck squamous cell carcinomas, *Clin Cancer Res*, 16(4): 1129-39, 2010.
- Deribe, Y.L., P. Wild, A. Chandrashaker, J. Curak, M. H.H. Schmidt, Y. Kalaidzidis, N. Milutinovic, I. Kratchmarova, L. Buerkle, M. J.Fetchko, P. Schmidt, S. Kittanakom, K. R. Brown, I. Jurisica, B. Blagoev, M. Zerial, I. Stagljar, and I. Dikic, Regulation of Epidermal Growth Factor Receptor Trafficking by Lysine Deacetylase HDAC6, *Sci Signal*, 2(102): RA84, 2009. Epub: 2009/12/24.
- Cervigne, N. K., Reis, P. P., Machado, J., Sadikovic, B., Bradley, G., Galloni, N. N., Pintilie, M., Jurisica, I., Gilbert, R., Gullane, P., Irish, J., and Kamel-Reid, S. Identification of a microRNA signature associated with progression of leukoplakia to oral carcinoma, *Hum Mol Genet*, 18(24): 4818-29, 2009. Epub 2009/09/25.

### Recent (selected) presentations

- Waldron, L, Jurisica, I., Pintilie, M. Assessment of strategies for the application of Lasso penalized likelihood regression to survival analysis of simulated high-dimensional gene expression data, *Statistics in the Life Sciences, High-dimensional Inference Workshop*, Groningen, the Netherlands, 23-25 November, 2009. [Best Poster Award.]
- Luft, J.R., Cohen, A.E., Dumont, M.E., Grayhack, E.J., Gruner, S.M., Hodgson, K., Jurisica, I., McPherson, A., Phizicky, E.M., Snell, E.H., Soltis, S.M., Weeks, C.M., Malkowski, M.G., DeTitta, G.T. Cloning through diffraction technologies developed at the Center for High-throughput Structural Biology. *Keystone Symposia. Structural Biology*, 2010.
- Craddock, K.J., Buys, T.P.H., Zhu, C.Q., Strumpf, D., Pintilie, M., Ding, K., Seymour, L., Jurisica, I., Shepherd, F.A., Lam, W.L., Tsao, M.S. High resolution genomic analysis of NSCLC reveals regions of DNA copy number gain that may be predictive of benefit from adjuvant chemotherapy, *IASLC*, 2009.
- Elschenbroich, S., Ignatchenko, V., Shaw, P. A., Jurisica, I. Kislinger, T. Identification of putative biomarkers for ovarian cancer from human ascitic fluid. *HUPO*, Toronto, October, 2009.

Thank you,

C. A. Cumbaa and I. Jurisica

