

Aaron Smalter Hall

Senior Data Scientist

Federal Reserve Bank of Kansas City
Kansas City, Missouri

aaron.smalterhall@kc.frb.org

www.kansascityfed.org

Office: (816) 283-7883

Research Interests

- Machine learning and data mining techniques for large, heterogeneous, and structured data sets
- Objective function regularization and optimization
- Kernel methods and structured similarity analysis

Education

Ph. D. Computer Science, *University of Kansas*

August 2011

M. S. Computer Science, with Honors, *University of Kansas*

December 2008

B. S. Computer Science, *University of Kansas*

May 2006

Positions

- 07/2015 – Present **Senior Data Scientist**, Center for the Advancement of Data and Research in Economics, Federal Reserve Bank of Kansas City
- 03/2013 – 07/2015 **Director**, Molecular Graphics and Modeling Laboratory, University of Kansas
- 05/2012 – 03/2013 **Bioinformatics Specialist**, K-INBRE Bioinformatics Core Facility, University of Kansas
- 09/2011 – 05/2012 **Post-doctoral Researcher**, K-INBRE Bioinformatics Core Facility, University of Kansas
- 01/2007 – 09/2011 **Graduate Research Assistant**, Molecular Graphics and Modeling Laboratory, University of Kansas
- 12/2004 – 01/2007 **Undergrad. Research Asst.**, Macromolecule and Vaccine Stabilization Center, University of Kansas

Honors & Awards

- **Student Travel Award** 2009
IEEE International Conference on Bioinformatics & Biomedicine
- **Student Travel Award** 2009
IEEE International Conference on Data Mining
- **School of Engineering Scholarship** 2009
Dept. of Electrical Engineering and Computer Science
- **Paul and Virginia B. Miller Scholarship** 2005–2006
Dept. of Electrical Engineering and Computer Science
- **Finalist, Linkugel Speech Competition** 2005
Dept. of Communication Studies
- **National Merit Scholarship** 2001–2005
University of Kansas

Research Supports

1. NIH/NIA P01AG029531 (PI: George Bousfield) 04/01/09 – 03/31/14
The Aging Pituitary-Gonadal Axis Total: \$6,236,045

This proposal seeks to structurally and functionally characterize post-translational modifications to specific targets within the pituitary-gonadal axis according to their contribution to aging-related degradation of biochemical function. The informatics core of this effort supports protein purification and mass spectrometric analyses within this project.

Role: **Core Leader**

2. FDA/CDER U01FD005285 (PI: Volkin) 9/01/14 – 8/31/16
Mathematical Model for Characterization of Complex Molecules Total: \$800,000

This proposal will develop, implement and validate integrated mathematical algorithms to assess the similarity of multiple batches of two different complex molecules (a biopolymer mixture and IgG-based glycoproteins) using analytical comparability data sets from characterization (biological, chemical and physical) methods.

Role: **Co-PI**

3. NIH/NINDS R01NS088059 (PI: Nancy Muma) 4/01/15 – 3/31/19
HTS Identification of Novel Drugs for Huntington's Disease Total: \$2,020,800

The goal of this project is to identify biological probes that disrupt the binding of mhht to CaM and do so selectively without inhibiting the normal function of CaM. The top compounds will further protect against both the deleterious effects of mhht in neuronal cells and the transamidation of mhht in neurons. The goal is to identify lead scaffolds for future optimization and development by medicinal chemistry, and pharmacokinetic and toxicology testing.

Role: **Key Personnel**

Professional Activities

Invited Reviewer	2013
International Journal of Data Mining and Bioinformatics, Inderscience	
Invited Reviewer	2013
PHYSICA-A: Statistical Mechanics and its Applications, Elsevier	
Invited Reviewer	2013
Combinatorial Chemistry and High Throughput Screening, Bentham Science	
Associate Member	2012 – Present
KU Cancer Center, program in Drug Discovery, Delivery and Experimental Therapeutics	

Publications

Journal Articles

1. Viktor Y. Butnev, Vladimir Y. Butnev, Jeffrey V. May, Bin Shuai, Patrick Tran, William K. White, Alan Brown, **Aaron Smalter Hall**, David J. Harvey, and George R. Bousfield. *Production, purification, and characterization of recombinant hFSH glycoforms for functional studies*. *Molecular and cellular endocrinology* 405, 2015: 42-51.
2. George R. Bousfield, Vladimir Y. Butnev, William K. White, **Aaron Smalter Hall**, and David J. Harvey. *Comparison of Follicle-Stimulating Hormone Glycosylation Microheterogeneity by Quantitative Negative Mode Nano-Electrospray Mass Spectrometry of Peptide-N Glycanase-Released Oligosaccharides*. *Journal of glycomics & lipidomics* 5, no. 1, 2015.

3. Hieu Sy Vu, Sunitha Shiva, **Aaron Smalter Hall**, and Ruth Welti. *A Lipidomic Approach to Identify Cold-Induced Changes in Arabidopsis Membrane Lipid Composition*. Plant Cold Acclimation, pp. 199-215. Springer New York, 2014.
4. George R. Bousfield, Vladimir Y. Butnev, Monica A. Rueda-Santos, Alan Brown, **Aaron Smalter Hall**, and David J. Harvey. *Macro-and Micro-heterogeneity in Pituitary and Urinary Follicle-Stimulating Hormone Glycosylation*. Journal of glycomics & lipidomics 4, 2014.
5. **Aaron Smalter Hall**, Yunfeng Shan, Gerald Lushington, Mahesh Visvanathan. *An Overview of Computational Life Science Databases & Exchange Formats of Relevance to Chemical Biology Research*, Combinatorial Chemistry and High Throughput Screening, Volume 16, Number 3, March 2013, pp. 189-198(10)
6. The MicroArray Quality Control (MAQC) Consortium, *The MAQC-II Project: A Comprehensive Study of Common Practices for the Development and Validation of Microarray-based Predictive Models*, Nature Biotechnology, 2010.
7. Xiaohong Wang, Jun Huan, **Aaron Smalter**, Gerald Lushington, *Application of Kernel Functions for Accurate Similarity Search in Large Chemical Databases*, BMC Bioinformatics, Vol. 11 (Suppl 3):S8, 2010.
8. **Aaron Smalter**, Jun Huan, Gerald Lushington. *GPD: A Graph Pattern Diffusion Kernel for Accurate Graph Classification with Applications in Cheminformatics*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 7, No.2, pp. 197-207, 2010.
9. **Aaron Smalter**, Jun Huan, Gerald Lushington. *Graph Wavelet Alignment Kernels for Drug Virtual Screening*. Journal of Bioinformatics and Computational Biology, Vol. 7 (3), pp. 473-497, 2009.
10. D.T. Brandau, S.B. Joshi, **A. Smalter**, S. Kim, B. Steadman, and C.R. Middaugh. *Stability of the Clostridium botulinum type A neurotoxin Complex; An Empirical Phase Diagram Based Approach*. Journal of Molecular Pharmaceutics, 4(4):571-582, 2007.
11. N. Salamat-Miller, J. Fang, C. Seidel, **A. Smalter**, Y. Assenov, M. Albrecht, and C.R. Middaugh. *A network-based analysis of polyanions binding proteins utilizing yeast protein arrays*. Journal of Molecular & Cellular Proteomics, 5:2263-2278, 2006.

Refereed Conference Proceedings and Workshops

1. **Aaron Smalter Hall**, Jun Huan. *KUChemBio: A repository of computational chemical biology data sets*, in Proceedings of the 1st International Workshop on BigData in Bioinformatics and Health Care Informatics (BBH), 2013.
2. **Aaron Smalter**, Jun Huan, Gerry Lushington. *Similarity Boosting for Label Noise Tolerance in Protein-Chemical Interaction Prediction*, in Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM BCB), 2011.
3. Said Bleik, Min Song, **Aaron Smalter**, Jun Huan. *CGM: A biomedical text categorization approach using concept graph mining*. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshop (BIBM), 2009.
4. **Aaron Smalter**, Jun Huan, Gerald Lushington. *Feature Selection in the Feature Tensor Product Space*. In Proceedings of the 9th IEEE International Conference on Data Mining (ICDM), 2009.
5. Xiaohong Wang, Jun Huan, **Aaron Smalter**, Gerald Lushington. *Application of Kernel Functions for Accurate Similarity Search in Large Chemical Databases*. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedecine (BIBM), 2009.
6. Xiaohong Wang, **Aaron Smalter**, Jun Huan, and Gerald Lushington. *G-Hash: Towards Fast Kernel-based Similarity Search in Large Graph Databases*. In Proceedings of the 12th International Conference on Extending Database Technology (EDBT), 2009.

7. **Aaron Smalter**, Jun Huan, and Gerald Lushington. *GPM: A Graph Pattern Matching Kernel with Diffusion for Chemical Compound Classification*. In Proceedings of the 8th IEEE International Conference on Bioinformatics and BioEngineering (BIBE), 2008.
8. **Aaron Smalter**, Jun Huan, Jia Yi, and Gerald Lushington. *GPD: A Graph Pattern Diffusion Kernel for Accurate Graph Classification*. In Proceedings of the 8th International Workshop of Data Mining in Bioinformatics (BIOKDD), 2008.
9. **Aaron Smalter**, Jun Huan, Gerald Lushington. *Graph Wavelet Alignment Kernels for Drug Virtual Screening*. In Proceedings of the 7th Annual International Conference on Computational Systems Bioinformatics (CSB), 2008.
10. **Aaron Smalter**, Jun Huan, Gerald Lushington. *Structure-based Pattern Mining for Chemical Compound Classification*. In Proceedings of the 6th Asia Pacific Bioinformatics Conference (APBC), 2008.
11. **Aaron Smalter** and Seak Fei Lei and Xue-wen Chen. *Human Disease-Gene Classification with Integrative Sequence-Based and Topological Features of Protein-Protein Interaction Networks*. In Proc. of IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2007.

Book Chapters

1. Vu, H.S., Shiva, S., **Hall, A.S.**, and Welti, R. 2013. *A lipidomic approach to identify cold-induced changes in Arabidopsis membrane lipid composition*. In *Plant Cold Acclimation, Methods in Molecular Biology*. Ed., D. Hinch. Humana Press, New York, NY. In press.
2. Xiaohong Wang, Jun Huan, **Aaron Smalter**, and Gerald Lushington, *G-hash: Towards Fast Kernel-based Similarity Search in Large Graph Databases*, Graph Data Management: Techniques and Applications, Sherif Sakr and Eric Pardede ed, IGI Global, ISBN 161350053X, 2011.

Thesis

1. **Aaron Smalter**. *Genome-wide Protein-Chemical Interaction Prediction*. University of Kansas, 2011.
2. **Aaron Smalter**. *Kernel Functions for Graph Classification*. University of Kansas, 2008.