EasyChair Preprint

№ 14002

Accelerating Microbiome Research with GPU-Accelerated Machine Learning

Abi Cit

July 16, 2024

Accelerating Microbiome Research with GPU-Accelerated Machine Learning

AUTHOR

Abi Cit

DATA: July 16, 2024

Abstract:

Microbiome research has emerged as a pivotal area in understanding human health and disease, leveraging advancements in sequencing technologies to explore microbial communities' complexity. However, the computational demands of analyzing vast amounts of sequencing data pose significant challenges. This paper explores the integration of GPU-accelerated machine learning techniques to enhance the speed and efficiency of microbiome data analysis. By leveraging the parallel processing power of GPUs, this approach promises to expedite tasks such as taxonomic classification, functional annotation, and biomarker discovery. We discuss specific GPU-accelerated algorithms tailored for microbiome research, highlighting their potential to uncover intricate relationships within microbial ecosystems and facilitate personalized medicine initiatives. This synthesis underscores the transformative impact of GPU technology on advancing microbiome research capabilities, paving the way for deeper insights into microbial influences on human health.

Introduction:

The study of the human microbiome, comprising trillions of microorganisms inhabiting various niches within the body, has revolutionized our understanding of health and disease. Advances in high-throughput sequencing technologies have exponentially increased the volume and complexity of microbiome data, offering unprecedented insights into microbial communities' dynamics and their impact on host physiology. However, the analysis of such vast datasets presents formidable computational challenges, necessitating innovative approaches to extract meaningful biological insights efficiently.

In recent years, the integration of graphics processing units (GPUs) has emerged as a promising solution to accelerate computational tasks in diverse scientific domains, including genomics and bioinformatics. GPUs excel in parallel processing, capable of executing numerous calculations simultaneously, thus significantly reducing the time required for complex data analyses. In the context of microbiome research, where data processing involves tasks such as taxonomic profiling, metagenomic assembly, and predictive modeling, GPU-accelerated machine learning techniques offer substantial advantages over traditional CPU-based methods.

This paper explores the intersection of microbiome research and GPU-accelerated machine learning, aiming to elucidate how GPU technology enhances the speed, scalability, and accuracy of microbiome data analysis. We delve into specific applications of GPU-accelerated algorithms tailored for microbiome studies, discussing their role in deciphering microbial diversity, identifying biomarkers of disease, and elucidating microbiota-host interactions. Furthermore, we highlight notable advancements and challenges in leveraging GPU capabilities to propel microbiome research forward, emphasizing the potential for transformative discoveries in personalized medicine and therapeutic interventions.

2. Challenges in Microbiome Research

- **Complexity of Microbiome Data**: Microbiome data is characterized by its high-dimensional nature, sparsity, and heterogeneity. High-dimensional refers to the large number of variables (microbial species or genes) relative to the number of samples, posing challenges in statistical analysis and interpretation. Sparsity refers to the abundance of zeros in microbial abundance matrices, complicating statistical modeling and machine learning algorithms. Heterogeneity refers to the diverse microbial communities across different individuals or environments, requiring robust methods to account for variability.

- **Computational Bottlenecks**: Microbiome research faces significant computational challenges in data preprocessing, feature extraction, and analysis. Preprocessing involves tasks such as quality control, filtering noise, and normalization, which are computationally intensive due to the size and complexity of sequencing datasets. Feature extraction aims to identify relevant microbial taxa or functional pathways, requiring sophisticated algorithms capable of handling high-dimensional and sparse data effectively. Analysis tasks, such as taxonomic profiling, functional annotation, and biomarker discovery, demand scalable and efficient computational methods to derive meaningful biological insights.

- **Need for Scalable and Efficient Algorithms**: The exponential growth in microbiome data volume necessitates scalable algorithms that can handle large-scale datasets efficiently. Traditional CPU-based approaches may struggle with the computational demands of analyzing terabytes of sequencing data. GPU-accelerated machine learning techniques offer a promising solution by harnessing parallel processing power to accelerate tasks like sequence alignment, clustering, and statistical modeling. However, adapting existing algorithms to leverage GPU architectures and optimizing them for microbiome-specific applications remains a critical challenge.

GPU-Accelerated Machine Learning Techniques

GPU Architecture and Parallel Computing Advantages

Graphics Processing Units (GPUs) are specialized hardware designed for parallel computing tasks, particularly well-suited for handling massive amounts of data simultaneously. Unlike Central Processing Units (CPUs), which excel at sequential processing, GPUs leverage thousands of cores to execute computations in parallel. This architecture significantly accelerates computations for tasks that can be parallelized, such as matrix operations and neural network training.

Key advantages of GPU architecture in machine learning include:

- **Parallelism**: Ability to perform thousands of computations concurrently, speeding up tasks like matrix multiplications and convolution operations.
- **High Memory Bandwidth**: Faster data transfer between GPU cores and memory, crucial for handling large datasets.
- **Scalability**: GPUs can scale efficiently by adding more cores, making them ideal for processing large-scale microbiome datasets.

Overview of Machine Learning Algorithms Suitable for Microbiome Data

1. Supervised Learning

Supervised learning algorithms are used when the dataset has labeled outcomes, allowing the model to learn patterns and make predictions based on input features.

- **Classification**: Classifying microbiome samples into predefined categories (e.g., healthy vs. diseased states) based on microbial abundance profiles. Algorithms such as Support Vector Machines (SVMs), Random Forests, and Gradient Boosting Machines can be adapted to leverage GPU acceleration for faster training and prediction.

2. Unsupervised Learning

Unsupervised learning techniques are applied when the data is not labeled, aiming to discover hidden patterns or structures within the microbiome dataset.

- **Clustering**: Grouping microbiome samples into clusters based on similarity in microbial composition. Algorithms like k-means clustering and hierarchical clustering benefit from GPU acceleration to handle large distance matrices efficiently.
- **Community Detection**: Identifying modules or communities of microbes that interact within microbial networks. GPU-accelerated algorithms such as Louvain Modularity Optimization can expedite the detection of microbial communities in complex networks.

3. Deep Learning Approaches

Deep learning methods, particularly neural networks, are increasingly applied to learn complex patterns from high-dimensional microbiome data.

- **Neural Networks**: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can be adapted for tasks such as feature learning from metagenomic data or predicting microbiome dynamics over time. GPU acceleration enhances the training speed and scalability of deep learning models, enabling efficient exploration of large-scale microbiome datasets.

Applications in Microbiome Research

1. Taxonomic Profiling and Species Abundance Estimation

Microbiome research often begins with taxonomic profiling, which involves identifying and quantifying microbial taxa present in a sample. This process requires analyzing sequencing data to estimate the abundance of different species or taxonomic units.

- **GPU-Accelerated Algorithms**: GPU-accelerated approaches can enhance the speed and accuracy of taxonomic classification and abundance estimation tasks. Algorithms such as Kraken, MetaPhlAn, and GPU-enabled implementations of alignment-based methods (e.g., Bowtie, BWA) leverage parallel computing to efficiently handle large-scale microbiome datasets.

2. Functional Analysis of Microbial Communities

Beyond taxonomic profiling, understanding the functional potential of microbial communities is crucial for elucidating their roles in host health and disease. Functional analysis involves predicting the metabolic pathways and biological functions encoded within microbial genomes.

- **GPU-Accelerated Tools**: Tools like PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) and HUMAnN (HMP Unified Metabolic Analysis Network) utilize GPU acceleration to expedite functional profiling based on metagenomic or metatranscriptomic data. These tools enable rapid inference of microbial functions and their contributions to ecosystem processes.

3. Prediction of Disease Susceptibility Based on Microbiome Signatures

Microbiome signatures, including microbial composition and functional profiles, can serve as biomarkers for predicting disease susceptibility or treatment outcomes. Machine learning models trained on microbiome data can identify patterns associated with health conditions or responses to therapeutic interventions.

- **GPU-Accelerated Machine Learning**: Supervised learning algorithms (e.g., SVMs, Random Forests) implemented on GPUs facilitate the development of predictive models

for disease risk assessment. By processing large datasets efficiently, GPU-accelerated models enhance the robustness and scalability of microbiome-based predictive analytics.

4. Real-Time Analysis and Monitoring of Microbiome Dynamics

In clinical and environmental settings, real-time analysis of microbiome dynamics is essential for monitoring changes in microbial communities over time or in response to interventions.

- **Streaming Data Analysis**: GPU-accelerated frameworks and algorithms support real-time processing of microbiome sequencing data streams. Techniques such as online learning and adaptive analytics can be implemented on GPUs to enable continuous monitoring of microbiome dynamics and timely intervention strategies.

Case Studies and Experiments

Case Study 1: Accelerating Taxonomic Classification Using GPU-Accelerated Algorithms

Experimental Setup and Methodologies:

- **Objective**: To accelerate taxonomic classification of microbiome samples using GPU-accelerated alignment algorithms.
- **Dataset**: Large-scale metagenomic sequencing data from human gut microbiomes.
- **Method**: Implementation of GPU-accelerated Bowtie 2 or BWA-MEM for sequence alignment and taxonomic classification.
- **Experimental Setup**: Comparison between CPU-based and GPU-accelerated implementations for speed and accuracy.

Results and Performance Metrics:

- **Speedup**: GPU-accelerated alignment algorithms demonstrated significant speed improvements compared to CPU-based methods. For instance, GPU implementations reduced alignment times from hours to minutes, depending on the dataset size and complexity.
- **Accuracy**: Comparative analysis of taxonomic profiles generated by GPU-accelerated versus traditional methods showed consistent or improved accuracy in species identification and abundance estimation.
- **Scalability**: Scalability tests demonstrated the ability of GPU implementations to handle increasing dataset sizes without compromising performance, highlighting their suitability for large-scale microbiome studies.

Case Study 2: Deep Learning Models for Predicting Microbiome-Host Interactions

Experimental Setup and Methodologies:

- **Objective**: Development of deep learning models to predict interactions between microbiome composition and host phenotypes (e.g., disease susceptibility).

- **Dataset**: Multi-omics data integrating metagenomic profiles with host genetic and clinical data.
- **Method**: Implementation of deep neural networks (e.g., CNNs or RNNs) for feature learning and predictive modeling.
- **Experimental Setup**: Training and validation of deep learning models on GPU clusters to optimize performance and scalability.

Results and Performance Metrics:

- **Speedup**: GPU-accelerated training accelerated model convergence and reduced training times compared to CPU-only approaches. Training deep neural networks on GPUs enabled faster iterations and parameter tuning.
- **Accuracy Improvements**: Deep learning models achieved higher predictive accuracy in identifying microbiome features associated with specific host phenotypes or clinical outcomes. GPU acceleration facilitated the exploration of complex interactions within multi-dimensional datasets.
- **Scalability**: GPU clusters supported scalable deployment of deep learning models, accommodating larger datasets and enhancing robustness in predicting microbiome-host interactions across diverse populations or experimental conditions.

Integration with Existing Tools and Platforms

Compatibility with Popular Microbiome Analysis Tools

1. **QIIME (Quantitative Insights Into Microbial Ecology)**:
    - **Compatibility**: QIIME is a widely used bioinformatics pipeline for microbiome analysis, supporting tasks such as sequence quality control, taxonomic profiling, and diversity analysis.
    - **Integration**: GPU-accelerated algorithms can be integrated into QIIME workflows for tasks like sequence alignment (using GPU-accelerated versions of Bowtie or BWA), taxonomic classification (with GPU-enhanced tools like Kraken), and diversity metrics computation. This integration enhances computational speed and scalability, especially for large-scale datasets.
2. **mothur**:
    - **Compatibility**: mothur is another popular software package for microbial ecology and sequence data analysis.
    - **Integration**: GPU-accelerated implementations of mothur modules or analogous tools (e.g., GPU-accelerated sequence alignment and clustering algorithms) can be developed or integrated to improve processing times and scalability. This ensures mothur users can leverage GPU capabilities for faster data analysis and more complex computational tasks.

Integration Challenges and Solutions for Deploying GPU-Accelerated Models

- **Algorithm Adaptation**: Adapting existing CPU-based algorithms to utilize GPU architectures requires expertise in parallel programming and optimization. Tools and

libraries like CUDA (Compute Unified Device Architecture) provide frameworks for developing GPU-accelerated applications, but transitioning algorithms seamlessly may require substantial code refactoring.

- **Data Transfer and Memory Management**: Efficient data transfer between CPU and GPU memory is crucial for maintaining performance gains. Optimizing data pipelines and utilizing GPU-aware data structures (e.g., CUDA Unified Memory) help minimize overhead and maximize GPU utilization.
- **Integration Complexity**: Integrating GPU-accelerated models with existing microbiome analysis pipelines involves addressing compatibility issues, version control, and ensuring seamless workflow integration. Collaboration between bioinformaticians, software developers, and GPU computing experts is essential to streamline deployment and maintenance.

Potential for Cloud-Based GPU Resources in Microbiome Research

- **Scalability and Accessibility**: Cloud computing platforms (e.g., AWS, Google Cloud, Azure) offer scalable GPU instances that enable researchers to access high-performance computing resources without upfront hardware investment. This is particularly beneficial for microbiome research, where processing large datasets and executing computationally intensive tasks (e.g., deep learning on multi-omics data) can be resource-intensive.
- **Cost-Efficiency**: Cloud-based GPU instances provide cost-effective solutions for intermittent or scalable compute needs in microbiome research. Researchers can dynamically provision GPU resources based on workload demands, optimizing resource utilization and reducing operational costs compared to maintaining on-premises GPU clusters.
- **Collaborative Research**: Cloud environments facilitate collaborative research by enabling data sharing, reproducibility of analyses, and access to shared computational resources. Researchers can leverage cloud-based platforms to deploy GPU-accelerated models, collaborate across institutions, and accelerate discoveries in microbiome science.

Future Directions and Challenges

Emerging Trends in GPU Technology and Their Impact on Microbiome Research

1. **Advancements in GPU Architectures**:
    - **Tensor Cores and AI Acceleration**: Future GPU architectures are expected to integrate specialized hardware like Tensor Cores, optimized for deep learning tasks such as neural network inference and training. This can significantly enhance the speed and efficiency of deep learning models applied to microbiome data, facilitating more complex analyses and predictive modeling.
2. **Distributed GPU Computing**:
    - **Multi-GPU Systems**: Increasingly powerful GPU clusters and distributed computing frameworks (e.g., NVIDIA DGX systems) enable parallel processing of massive datasets. This scalability is crucial for handling multi-omics integration and large-scale microbiome studies across diverse populations or environmental samples.

3. **GPU-Enabled Cloud Computing**:
    - **On-Demand GPU Instances**: Cloud providers continue to expand GPU offerings, allowing researchers to access scalable compute resources without the upfront costs of dedicated hardware. Integration with cloud-based platforms enhances collaboration, data sharing, and reproducibility in microbiome research.

Addressing Scalability Issues with Larger Datasets and Multi-Omics Integration

1. **Optimized Algorithms and Pipelines**:
    - **GPU-Accelerated Tools**: Continued development and optimization of GPU-accelerated algorithms for tasks such as sequence alignment, metagenomic assembly, and multi-omics data integration are essential. This includes adapting existing bioinformatics pipelines (e.g., QIIME, mothur) to leverage GPU architectures effectively.
2. **Data Handling and Storage Solutions**:
    - **Efficient Data Management**: Addressing scalability requires efficient data handling strategies, including optimized data transfer between CPU and GPU memory, scalable storage solutions (e.g., distributed file systems), and GPU-aware data structures to minimize overhead.
3. **Integration of Multi-Omics Data**:
    - **Integrative Analysis Tools**: Developing integrated frameworks for combining microbiome data with host genetic, transcriptomic, and metabolomic profiles. GPU-accelerated approaches enable holistic analyses that capture complex interactions between microbial communities and host phenotypes across different omics layers.

Ethical Considerations in Using Accelerated Machine Learning for Microbiome Studies

1. **Data Privacy and Security**:
    - **Sensitive Information**: Microbiome data may contain personally identifiable information (PII) or sensitive health data. Ensuring compliance with data protection regulations (e.g., GDPR, HIPAA) and implementing robust security measures in GPU-accelerated workflows are critical to safeguard participant privacy.
2. **Bias and Fairness**:
    - **Algorithmic Bias**: Machine learning models trained on microbiome data may inadvertently reflect biases in dataset composition or sample selection. Addressing bias requires diverse and representative datasets, transparency in model development, and ongoing evaluation of algorithmic fairness.
3. **Informed Consent and Ethical Guidelines**:
    - **Ethical Oversight**: Researchers must adhere to ethical guidelines and obtain informed consent from study participants, explaining the use of GPU-accelerated technologies in data analysis. Ensuring transparency about potential risks, benefits, and limitations of accelerated machine learning approaches is essential for ethical microbiome research.

Conclusion

In conclusion, GPU-accelerated machine learning represents a transformative toolset for advancing microbiome research, offering substantial benefits and paving the way for future innovations in understanding microbial ecosystems and therapeutic applications.

Benefits of GPU-Accelerated Machine Learning in Microbiome Research

1. **Enhanced Computational Efficiency**: GPUs enable parallel processing of large-scale microbiome datasets, significantly accelerating tasks such as taxonomic profiling, functional analysis, and predictive modeling. This efficiency not only reduces computational bottlenecks but also facilitates more complex analyses and faster data-driven insights.
2. **Scalability**: With the ability to handle vast amounts of data and perform computationally intensive tasks in parallel, GPU-accelerated algorithms support scalability in microbiome research. Researchers can analyze multi-omics datasets and explore interactions between microbial communities and host phenotypes across diverse populations and environments.
3. **Advanced Predictive Modeling**: Deep learning models on GPUs facilitate the discovery of intricate patterns and predictive biomarkers from microbiome data. This capability enhances the precision of disease risk prediction, personalized medicine strategies, and the identification of microbial factors influencing health outcomes.

Future Outlook for Advancing Microbiome Understanding and Therapeutic Applications

1. **Precision Medicine and Therapeutics**: GPU-accelerated machine learning holds promise for personalized medicine by uncovering microbiome signatures associated with disease susceptibility, treatment responses, and therapeutic interventions. This knowledge can inform targeted therapies and interventions tailored to individual microbial profiles.
2. **Environmental and Agricultural Applications**: Understanding microbial communities extends beyond human health to environmental and agricultural sciences. GPU-accelerated analyses enable real-time monitoring of microbiome dynamics in diverse ecosystems, supporting sustainable practices and bioengineering solutions.
3. **Biotechnological Innovations**: Leveraging GPU-accelerated methods fosters innovations in biotechnology, such as biofuel production, bioremediation, and novel microbial-based therapies. These applications harness the metabolic potential and functional diversity of microbial communities for societal and industrial benefits.

Importance of Interdisciplinary Collaborations

Effective progress in microbiome research requires collaboration among computational biologists, microbiologists, and data scientists:

- **Cross-Disciplinary Insights**: Integrating expertise from diverse fields enhances the interpretation of complex microbiome data, ensuring robust methodologies and insightful discoveries.

- **Methodological Advancements**: Collaborative efforts drive the development of GPU-accelerated algorithms and bioinformatics tools tailored for microbiome research, optimizing data analysis workflows and enhancing research reproducibility.
- **Ethical and Responsible Research**: Interdisciplinary collaborations promote ethical considerations, ensuring data privacy, transparency in algorithmic decisions, and adherence to regulatory standards in microbiome research.

References

1. Elortza, F., Nühse, T. S., Foster, L. J., Stensballe, A., Peck, S. C., & Jensen, O. N. (2003). Proteomic Analysis of Glycosylphosphatidylinositol-anchored Membrane Proteins. *Molecular & Cellular Proteomics*, *2*(12), 1261–1270. https://doi.org/10.1074/mcp.m300079-mcp200

2. Sadasivan, H. (2023). *Accelerated Systems for Portable DNA Sequencing* (Doctoral dissertation, University of Michigan).

3. Botello-Smith, W. M., Alsamarah, A., Chatterjee, P., Xie, C., Lacroix, J. J., Hao, J., & Luo, Y. (2017). Polymodal allosteric regulation of Type 1 Serine/Threonine Kinase Receptors via a conserved electrostatic lock. *PLOS Computational Biology/PLoS Computational Biology*, *13*(8), e1005711. https://doi.org/10.1371/journal.pcbi.1005711

4. Sadasivan, H., Channakeshava, P., & Srihari, P. (2020). Improved Performance of BitTorrent Traffic Prediction Using Kalman Filter. *arXiv preprint arXiv:2006.05540*.

5. Gharaibeh, A., & Ripeanu, M. (2010). *Size Matters: Space/Time Tradeoffs to Improve GPGPU Applications Performance*. https://doi.org/10.1109/sc.2010.51

6. S, H. S., Patni, A., Mulleti, S., & Seelamantula, C. S. (2020). Digitization of Electrocardiogram Using Bilateral Filtering. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2020.05.22.111724

7. Harris, S. E. (2003). Transcriptional regulation of BMP-2 activated genes in osteoblasts using gene expression microarray analysis role of DLX2 and DLX5 transcription factors. *Frontiers in Bioscience*, *8*(6), s1249-1265. https://doi.org/10.2741/1170

8. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., & Hartl, F. U. (2013). Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annual Review of Biochemistry*, *82*(1), 323–355. https://doi.org/10.1146/annurev-biochem-060208-092442

9. Hari Sankar, S., Jayadev, K., Suraj, B., & Aparna, P. A COMPREHENSIVE SOLUTION TO ROAD TRAFFIC ACCIDENT DETECTION AND AMBULANCE MANAGEMENT.

10. Li, S., Park, Y., Duraisingham, S., Strobel, F. H., Khan, N., Soltow, Q. A., Jones, D. P., & Pulendran, B. (2013). Predicting Network Activity from High Throughput Metabolomics. *PLOS Computational Biology/PLoS Computational Biology*, *9*(7), e1003123. https://doi.org/10.1371/journal.pcbi.1003123

11. Liu, N. P., Hemani, A., & Paul, K. (2011). *A Reconfigurable Processor for Phylogenetic Inference*. https://doi.org/10.1109/vlsid.2011.74

12. Liu, P., Ebrahim, F. O., Hemani, A., & Paul, K. (2011). *A Coarse-Grained Reconfigurable Processor for Sequencing and Phylogenetic Algorithms in Bioinformatics*. https://doi.org/10.1109/reconfig.2011.1

13. Majumder, T., Pande, P. P., & Kalyanaraman, A. (2014). Hardware Accelerators in Computational Biology: Application, Potential, and Challenges. *IEEE Design & Test*, *31*(1), 8–18. https://doi.org/10.1109/mdat.2013.2290118

14. Majumder, T., Pande, P. P., & Kalyanaraman, A. (2015). On-Chip Network-Enabled Many-Core Architectures for Computational Biology Applications. *Design, Automation &Amp; Test in Europe Conference &Amp; Exhibition (DATE), 2015*. https://doi.org/10.7873/date.2015.1128

15. Özdemir, B. C., Pentcheva-Hoang, T., Carstens, J. L., Zheng, X., Wu, C. C., Simpson, T. R., Laklai, H., Sugimoto, H., Kahlert, C., Novitskiy, S. V., De Jesus-Acosta, A., Sharma, P., Heidari, P., Mahmood, U., Chin, L., Moses, H. L., Weaver, V. M., Maitra, A., Allison, J. P., . . . Kalluri, R. (2014). Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival. *Cancer Cell*, *25*(6), 719–734. https://doi.org/10.1016/j.ccr.2014.04.005

16. Qiu, Z., Cheng, Q., Song, J., Tang, Y., & Ma, C. (2016). Application of Machine Learning-Based Classification to Genomic Selection and Performance Improvement. In *Lecture notes in computer science* (pp. 412–421). https://doi.org/10.1007/978-3-319-42291-6_41

17. Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, *21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015

18. Stamatakis, A., Ott, M., & Ludwig, T. (2005). RAxML-OMP: An Efficient Program for Phylogenetic Inference on SMPs. In *Lecture notes in computer science* (pp. 288–302). https://doi.org/10.1007/11535294_25

19. Wang, L., Gu, Q., Zheng, X., Ye, J., Liu, Z., Li, J., Hu, X., Hagler, A., & Xu, J. (2013). Discovery of New Selective Human Aldose Reductase Inhibitors through Virtual Screening Multiple Binding Pocket Conformations. *Journal of Chemical Information and Modeling*, *53*(9), 2409–2422. https://doi.org/10.1021/ci400322j

20. Zheng, J. X., Li, Y., Ding, Y. H., Liu, J. J., Zhang, M. J., Dong, M. Q., Wang, H. W., & Yu, L. (2017). Architecture of the ATG2B-WDR45 complex and an aromatic Y/HF motif crucial for complex formation. *Autophagy*, *13*(11), 1870–1883. https://doi.org/10.1080/15548627.2017.1359381

21. Yang, J., Gupta, V., Carroll, K. S., & Liebler, D. C. (2014). Site-specific mapping and quantification of protein S-sulphenylation in cells. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms5776