PROTEOFORM RESEARCH:

INTELLECTUAL PROPERTY THEFT AND EXPORT DIVERSION RISK ASSESSMENT

OCTOBER 2024

ARTEMIS RESEARCH

EXECUTIVE SUMMARY

The field of proteoform research is an emerging area of study with the potential to revolutionize the biotechnology industry, while simultaneously leading to intellectual property protection and compliance challenges. Proteoform research involves study of proteins in the form in which they function in the body, and knowledge of their presence, nature, and abundance of their different forms is critical to understanding the function of a biological system.

Companies which produce or sell the high-end chromatographs, mass spectrometers, and proteoform analysis software are likely at greatest risk of intellectual property theft or expert diversion over the near term by prohibited actors—such as sanctioned or export-restricted entities—wishing to illicitly obtain technology for top-down proteomics research. As the field of proteoform research progresses, these actors will likely increase focus on methods to overcome existing challenges in sample preparation, cataloging proteins in low abundance or in membranes, and the ability to obtain highly-specific molecular information while avoiding protein interference.

Acquisition of this technology and intellectual property by prohibited actors puts producers and distributors at risk of fines, reputational damage, and loss of market share from inadvertent dealings with these restricted entities. Countering this will require a thorough understanding of the value chain for this equipment to enable targeted mitigation of problematic acquisitions without stifling the collaboration so critical for scientific research.

TABLE OF CONTENTS

INTRODUCTION

PROTEOFORM STUDY AND ASSOCIATED TECHNOLOGY DEVELOPMENT

The field of proteoform research is an emerging area of study with the potential to revolutionize the biotechnology industry, while simultaneously leading to intellectual property protection and compliance challenges. Proteoform research involves study of proteins in the form in which they function in the body. Proteins act as the intermediaries between the genome and the physical characteristic encodes, and knowledge of their presence, nature, and abundance of their different forms is critical to understanding the function of a biological system.

Much like the Human Genome Project of the 1990s, the study of human proteoforms has the potential to completely transform the understanding of human health. Therefore, the scientific community launched the Human Proteoform Project in 2021, aiming to characterize the entire set of proteins encoded by the human genome. Their strategy involves a two-pronged approach to develop insights through deep analysis of medically relevant proteins, while investing heavily in research, technology development, and deployment of this technology for large-scale proteoform analysis.[1]

To accomplish the goals of the project, the scientific community will need to develop and apply new technologies and processes to overcome previous limitations in sample preparation and instrumentation. [2] A major development has been the technique of "top-down proteomics," in which a protein is analyzed in its entirety, in the folded structure in which it is used in the body. This is in contrast to previous techniques which first broke the protein into subcomponents before analysis.[3] In the top-down proteomics process, researchers 1) first obtain a sample and preserve the cells, before 2) breaking the cells open cells to release the proteins. They then 3) use a coupled set of instruments for analysis, including a 4) liquid chromatograph to separate the proteins, and then 5) a mass spectrometer to characterize them. Finally, researchers 6) employ specialized software to process and analyze the data.[4]

Just as the Human Genome Project converted a nearly $4 Billion public investment into over $700 Billion of economic activity and new industries, the economic implications of the Human Proteoform Project could be similarly transformative.[5] Therefore, we conclude this field is particularly ripe for intellectual property theft and diversion of resulting technology to prohibited entities.[i] This study aims to identify technology areas most vulnerable to export diversion and IP theft, enabling biotechnology companies to prioritize resources and develop strategies to better protect this technology while maintaining the open dialogue so critical to scientific progress.

[i] In this paper, the term "prohibited entities" include those subject to export restrictions, sanctions, or other prohibited business transactions based on the laws and regulations of the U.S. Government and other nations with similar export laws.

METHODOLOGY

This study is intended to help the leadership of biotechnology companies identify risk and allocate resources to best protect their company's investment from theft, misuse, and diversion. The equipment identified in this study is drawn from several top-down proteomics protocols compiled by the non-profit organization, Consortium for Top-Down Proteomics.[6] The future projections detailed in the Outlook section were based on existing challenges and technology needs obtained from an article published in a 2021 edition of *Science* magazine officially announcing the Human Proteoform Project. This information is current as of August 25th, 2024.

RESULTS

We determined that proteoform research is particularly reliant on high-end chromatographs, mass spectrometers, and proteoform analysis software, based on analysis of the types of technology appearing in the top-down proteomics protocols compiled by the Consortium. Therefore, we judge companies which produce or sell this technology are likely at risk of intellectual property theft or expert diversion over the near term by those wishing to illicitly obtain technology to support top-down proteomics research.

IDENTIFIED TOP-DOWN PROTEOMICS TECHNOLOGY

We identified these key technologies by comparison of six different protocols developed for different brands of laboratory instrumentation. We compiled a list of the different instruments, assays, and software used in these protocols, and identified the manufacturer for each. This resulted in 24 different technology items in eight categories manufactured by eight different companies, which sell their products via a network of authorized distributors around the world.[7, 8, 9, 10, 11, 12]

FIGURE 1: KEY LABORATORY EQUIPMENT AND MANUFACTURERS

Type	Instrument	Manufacturer
Assay	Pierce Fluorescent Assay Kit	Pierce (part of Thermo Fisher)
Assay	Proteograph Assay Plate	Seer
Automation	SP100 automation instrument	Seer
Column	Acquity UPLC Column	Waters
Column	Agilent Bio SEC-5 Column	Agilent
HPLC	1260 Infinity II HPLC	Agilent
HPLC	nanoAdvance HPLC system	Bruker
Ion Source	CE-MS Nanospray Interface	CMP Scientific Corp.
Ion Source	CESI 8000 Plus CE System	SCIEX
LC-MS/MS	maXis II ETD QTOF	Bruker Daltonics
MS	Fusion Lumos	Pierce (part of Thermo Fisher)
MS	impact II QTOF	Bruker
MS	Orbitrap Eclipse	Thermo Fisher
MS	Orbitrap Elite	Thermo Fisher

MS	Orbitrap Exploris 480	Thermo Fisher
MS	Orbitrap Fusion Lumos	Thermo Fisher
MS	Orbitrap QE-HF	Thermo Fisher
MS	Q-Exactive Plus	Thermo Fisher
MS	solariX 12T FTICR	Bruker Daltonics
MS	SolariX FT-ICR	Bruker
MS	timsTOF Pro mass spec	Bruker
MS	Xevo G2-S QTOF	Waters Corporation
UPLC	nanoAcquity UPLC System	Waters Corporation
UPLC	Dionex UltiMate 3000	Thermo Fisher

Acronyms:

CE	Capillary Electrophoresis
ETD	Electron Transfer Dissociation
FTICR	Fourier Transform Ion Cyclotron Resonance
HPLC	High-Performance Liquid Chromatograph
LC	Liquid Chromatograph
MS	Mass Spectrometer
UPLC	Ultra-Performance Liquid Chromatograph
QE	Q Exactive
QTOF	Quadrupole Time-of-Flight

This study also identified 14 different software suites, listed below in Figure 2, many of which were developed by academic researchers and available for free access or download from the internet.[13]

FIGURE 2: TOP-DOWN PROTEOMICS SOFTWARE

Type	Developer	Notes
Genome Analysis Toolkit	Broad Institute	
MASH Suite Pro	Ge Lab, University of Wisconsin	Available for free at http://crb.wisc.edu/ginglab/software.html; Supported by NIH grants
Metascape	Metascape Team	Available for free download at https://metascape.org; Supported by NIH grants
NRTDP TDPortal	National Resource for Translational and Developmental Proteomics (NRTDP), Northwestern University	User accounts can be requested at http://nrtdp.northwestern.edu/tdportal-request
OpenLAB Software	Agilent Technologies	
Perseus	Max Planck Institute of Biochemistry	Available for free download at https://maxquant.net/maxquant/; used for statistical analysis of data from the MaxQuant proteomics software package

ProSight Lite	NRTDP, Northwestern University	Available for free download at http://prosightlite.northwestern.edu
ProSightPC 4.0		Available for free download at http://proteinaceous.net/database-warehouse
Proteograph Analysis Suite	Seer	
QIAGEN IPA	QIAGEN	Software used to generate figures
Swiss-Prot Human Genome Database	Uniprot/SwissProt	Freely accessible database of protein sequences available at https://www.uniprot.org/
TDValidator	NRTDP, Northwestern University	Available for purchase through https://www.proteinaceous.net/tdvalidator
TopFD	Xiauwen Liu, Tulane University	Available for download at https://www.toppic.org/software/toppic/register.html
TopPIC Suite	Xiauwen Liu, Tulane University	Available for download at https://www.toppic.org/software/toppic/register.html

OUTLOOK

As the field of proteoform research progresses, we judge those wishing to illicitly obtain the most cutting-edge technology and intellectual property will likely focus on methods to overcome existing challenges in sample preparation, catalogue proteins in low abundance or in membranes, and obtain highly-specific molecular information while avoiding protein interference. This includes recent developments in proteoform pre-fractionation and high-resolution benchtop mass spectrometry, though in the future will likely include more sensitive MS-based technologies, single-molecule proteoform sequencing, more diverse assays, affinity reagents to capture the proteins of each gene, and ability for increased throughput, based on the journal article announcing the Human Proteoform Project launch.[14]

Additionally, desired areas of enabling technology will likely include nanopore sequencing, cryoelectron microscopy and visual proteomics, single-cell proteomics, single-molecule protein arrays, and development of proteoform reference sets and databases of molecular fragmentation spectra. [15] Among this group, we judge nanopore sequencing and cryoelectron microscopy technologies are the most likely targets for purchase, as they involve research instruments applicable to a variety of scientific questions. Compared with other sequencing techniques, nanopore sequencing is faster, more cost effective, and more portable, which we judge makes it attractive from a commercial perspective as well as a scientific one.[16] Cryoelectron microscopy can quickly elucidate protein structures difficult to identify with other techniques, and its application to protein research has expanded drastically over the past decade.[17, 18] Alternatively, we assess illicit acquisition related to assays, reference sets, and spectra is more likely to focus on obtaining data and intellectual property than purchasing an instrument itself, because we assume advancements in these areas are more likely to result from concerted applications of existing technologies rather than revolutionary technology design.

IMPLICATIONS AND RECOMMENDATIONS

Illicit acquisition of proteoform-related intellectual property or technology carries the potential for significant profit loss and regulatory violations. Companies can mitigate this risk by allocating additional due diligence resources to the technology identified in this article, with a particular focus on entities involved in the product's value chain to enable targeted mitigation of problematic acquisitions without stifling the collaboration so critical for scientific research.

Conducting thorough due diligence on new and existing distributors would reduce the chance of inadvertent sale to a restricted entity. This should include paying particular attention to distributors' prior trade activity, including the identity of their customers and any contact with export-restricted entities. Additionally, manufacturers should ensure distributors have a thorough understanding of the requirements surrounding resale of the items and export control obligations.

Similarly, conducting due diligence on potential partners and collaborators would mitigate risk of intellectual property theft and regulatory violations. Prominent researchers have faced fines and jail time for intellectual property theft, undeclared research relationships and prohibited sources of funding, including unauthorized grants from the Chinese Government.[19, 20] Additionally, several biotechnology start-ups have been fined for inadvertently accepting venture capital funding from sanctioned entities.[21] Therefore, it is critical for both researchers and biotechnology companies to be fully informed regarding their partners affiliations, relationships, and funding sources.

ENDNOTES

[1] Lloyd M. Smith et al. The Human Proteoform Project: Defining the human proteome. *ScienceAdvances.* **7**, eabk0734(2021). DOI:10.1126/sciadv.abk0734
[2] Toby, T.K., Fornelli, L., Srzentić, K. et al. A comprehensive pipeline for translational top-down proteomics from a single blood draw. Nat Protoc 14, 119–152 (2019). https://doi.org/10.1038/s41596-018-0085-7
[3] Toby, T.K., Fornelli, L., Srzentić, K. et al. A comprehensive pipeline for translational top-down proteomics from a single blood draw. Nat Protoc 14, 119–152 (2019). https://doi.org/10.1038/s41596-018-0085-7
[4] Toby, T.K., Fornelli, L., Srzentić, K. et al. A comprehensive pipeline for translational top-down proteomics from a single blood draw. Nat Protoc 14, 119–152 (2019). https://doi.org/10.1038/s41596-018-0085-7
[5] Drake, N. What is the human genome worth? Nature (2011). https://doi.org/10.1038/news.2011.281
[6] https://ctdp.org/resources/methods/
[7] Donnelly, D.P., Rawlins, C.M., DeHart, C.J. *et al.* Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods* **16**, 587–594 (2019). https://doi.org/10.1038/s41592-019-0457-0
[8] Ge Lab. LC/MS Top-Down Proteomics Standard Operation Protocol. University of Wisconsin Madison, http://www.topdownproteomics.org/wp-content/uploads/2017/09/Top-Down-Methods-for-Bruker-instruments.pdf; Accessed 25 August 2024.
[9] DeHart, C.J., Fornelli, L. NRTDP Top-Down Standard SOP: Orbitrap Elite. Northwestern University (2017). https://nrtdp.northwestern.edu/wp-content/uploads/2017/10/NRTDP-Top-Down-Standard-SOP_001.1.pdf. Accessed 25 August 2024.
[10] Human Plasma Sample Preparation Workflow Using the Seer Proteograph Product Suite. Seer (2022). http://www.topdownproteomics.org/wp-content/uploads/2023/02/Seer-PlasmaPrepWorkflow.pdf. Accessed 25 August 2024.
[11] McCool, Elijah N. et al. Deep top-down proteomics revealed significant proteoform-level differences between metastatic and non-metastatic colorectal cancer cells. *ScienceAdvances.* **8**, 51 (2022) doi:10.1126/sciadv.abq6348
[12] Su, Pei et al; Highly multiplexed, label-free proteoform imaging of tissues by individual ion mass spectrometry; *ScienceAdvances.* **8,** 32 (2022) DOI: 10.1126/sciadv.abp9929
[13] Ge Lab. LC/MS Top-Down Proteomics Standard Operation Protocol. University of Wisconsin Madison, http://www.topdownproteomics.org/wp-content/uploads/2017/09/Top-Down-Methods-for-Bruker-instruments.pdf. Accessed 25 August 2024.
[14] Lloyd M. Smith et al. The Human Proteoform Project: Defining the human proteome. *ScienceAdvances* **7**, eabk0734(2021).DOI:10.1126/sciadv.abk0734
[15] Lloyd M. Smith et al. The Human Proteoform Project: Defining the human proteome. *ScienceAdvances* **7**, eabk0734(2021).DOI:10.1126/sciadv.abk0734
[16] Tianyuan Zhang, Hanzhou Li, Mian Jiang, Huiyu Hou, Yunyun Gao, Yali Li, Fuhao Wang, Jun Wang, Kai Peng, Yong-Xin Liu. Nanopore sequencing: flourishing in its teenage years. *Journal of Genetics and Genomics (*2024) ISSN 1673-8527, https://doi.org/10.1016/j.jgg.2024.09.007.
[17] Callaway, Ewen. Revolutionary cryo-EM is taking over structural biology. *Nature* **578**, 201 (2020). https://doi.org/10.1038/d41586-020-00341-9
[18] Callaway, E. The revolution will not be crystallized: a new method sweeps through structural biology. *Nature* **525**, 172–174 (2015). https://doi.org/10.1038/525172a
[19] https://www.justice.gov/usao-ndca/pr/former-ceo-and-coo-jhl-biotech-sentenced-conspiring-steal-trade-secrets-and-commit-wire
[20] https://www.justice.gov/opa/pr/harvard-university-professor-convicted-making-false-statements-and-tax-offenses
[21] https://www.theguardian.com/world/2023/jan/06/roman-abramovich-trusts-transfer-leak-russia-sanctions