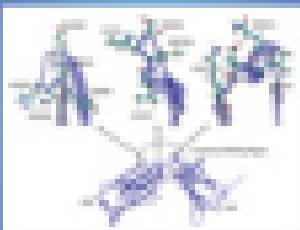


## Developments in Biotechnology and Bioscience

# Development in Biotechnology and Bioscience



Volume 10  
Anne Koenigsdorf, Pauline Arai,  
Jonathan L. Kellman, and Kavita Dangal

# **Developments in Biotechnology and Bioprocessing**



ACS SYMPOSIUM SERIES 1125

# Developments in Biotechnology and Bioprocessing

**Anne Kantardjieff**, Editor

*Alexion Pharmaceuticals  
Cheshire, Connecticut*

**Prashanth Asuri**, Editor

*Santa Clara University  
Santa Clara, California*

**Jonathan L. Coffman**, Editor

*Pfizer, Inc.  
Andover, Massachusetts*

**Karthik Jayapal**, Editor

*Bayer HealthCare  
Berkeley, California*

**Sponsored by the  
ACS Division of Biochemical Technology**



American Chemical Society, Washington, DC

Distributed in print by Oxford University Press



## Library of Congress Cataloging-in-Publication Data

Developments in biotechnology and bioprocessing / Anne Kantardjieff, editor, Alexion Pharmaceuticals Cheshire, Connecticut, Prashanth Asuri, editor, Santa Clara University Santa Clara, California, Jonathan L. Coffman, editor, Pfizer, Inc. Andover, Massachusetts, Karthik Jayapal, editor, Bayer HealthCare Berkeley, California ; sponsored by the ACS

Division of Biochemical Technology.

pages cm. -- (ACS symposium series ; 1125)

Includes bibliographical references and index.

ISBN 978-0-8412-2910-5

1. Biomass energy. 2. Biomass energy--Technological innovations. 3. Biochemical engineering--Research. 4. Biotechnology--Research. I. Kantardjieff, Anne., editor of compilation. II. Asuri, Prashanth., editor of compilation. III. Coffman, Jonathan L., editor of compilation. IV. Jayapal, Karthik., editor of compilation. V. American Chemical Society. Division of Biochemical Technology, sponsoring body.

TP339.D48 2013

333.95'39--dc23

2012050112

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48n1984.

Copyright © 2013 American Chemical Society

Distributed in print by Oxford University Press

All Rights Reserved. Reprographic copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Act is allowed for internal use only, provided that a per-chapter fee of \$40.25 plus \$0.75 per page is paid to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. Republication or reproduction for sale of pages in this book is permitted only under license from ACS. Direct these and other permission requests to ACS Copyright Office, Publications Division, 1155 16th Street, N.W., Washington, DC 20036.

The citation of trade names and/or names of manufacturers in this publication is not to be construed as an endorsement or as approval by ACS of the commercial products or services referenced herein; nor should the mere reference herein to any drawing, specification, chemical process, or other data be regarded as a license or as a conveyance of any right or permission to the holder, reader, or any other person or corporation, to manufacture, reproduce, use, or sell any patented invention or copyrighted work that may in any way be related thereto. Registered names, trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

PRINTED IN THE UNITED STATES OF AMERICA

# Foreword

The ACS Symposium Series was first published in 1974 to provide a mechanism for publishing symposia quickly in book form. The purpose of the series is to publish timely, comprehensive books developed from the ACS sponsored symposia based on current scientific research. Occasionally, books are developed from symposia sponsored by other organizations when the topic is of keen interest to the chemistry audience.

Before agreeing to publish a book, the proposed table of contents is reviewed for appropriate and comprehensive coverage and for interest to the audience. Some papers may be excluded to better focus the book; others may be added to provide comprehensiveness. When appropriate, overview or introductory chapters are added. Drafts of chapters are peer-reviewed prior to final acceptance or rejection, and manuscripts are prepared in camera-ready format.

As a rule, only original research papers and original review papers are included in the volumes. Verbatim reproductions of previous published papers are not accepted.

## ACS Books Department

# Preface

The chapters in this book were selected as the best presentations from the 2011 and 2012 annual meetings of the Biochemical Technology division of the American Chemical Society. A panel of industrial and academic members helped select the content of this volume, along with nominations from individual session chairs. The Biochemical Technology (BIOT) division brings together multiple disciplines, including biotechnology, biochemistry and microbiology. The division has always sought to provide a balanced representation of current research in industry and academia. The chapters in this volume are a reflection of this.

This volume brings together cutting-edge research in the areas of upstream and downstream process development, biofuels and biosynthesis. With an emphasis on current challenges facing the bioprocess community, this book contains innovative research by both academics and members of industry. A number of case studies from the biopharmaceutical industry are presented, including practical examples of methodologies for enhanced process development. In addition, the application of new and emerging technologies in the fields of biotechnology and bioprocessing are discussed.

This symposium book was intended to bring the best of the annual BIOT meetings to a greater audience. We extend special thanks to our technical writer, Christine Crane, who was instrumental in helping realize this objective.

**Anne Kantardjieff**  
Alexion Pharmaceuticals  
Cheshire, Connecticut

**Prashanth Asuri**  
Santa Clara University  
Santa Clara, California

**Jonathan L. Coffman**  
Pfizer Research and Development  
Andover, Massachusetts

**Karthik Jayapal**  
Bayer HealthCare  
Berkeley, California

## Chapter 1

# Toward a Biological Replacement of Petroleum

James C. Liao\*

Department of Chemical and Biomolecular Engineering and Institute for  
Genomics and Proteomics, University of California Los Angeles,  
5531 Boelter Hall, 420 Westwood Plaza, Los Angeles, California 90095

\*E-mail: [liaoj@seas.ucla.edu](mailto:liaoj@seas.ucla.edu). Phone: +1 (310) 825-1656.

Fax: +1 (310) 206-1642

Petroleum is the source of most of the world's raw materials and it also provides the majority of the world's energy. As the price of petroleum fluctuates and its combustion has significant negative environmental impact, there is growing interest in research into alternative sources of energy. Current biorefining schemes use plant biomass or algae to produce ethanol, biodiesel, and some chemicals used as raw materials. While these approaches are the leading processes, next-generation biorefinery should consider other options for starting materials, schemes for manufacturing, and the biofuel itself. For example, direct photosynthetic production of advanced biofuels (such as *n*-butanol or isobutanol) or raw materials from CO<sub>2</sub> would be an attractive and more environmentally friendly alternative; this strategy bypasses the need for lignocellulose deconstruction and algal lipid processing. Non-photosynthetic, electricity-powered CO<sub>2</sub> reduction to generate biofuels further by-passes the need for large light exposed surface areas used in photo-bioreactors and represents an option for biochemical and biofuel production as well as electricity storage.

### 1. Introduction

The world's energy needs have grown a staggering 30% (4 terawatts) in the past decade (2000-2011) (1). This astonishing increase in energy intake includes the consumption of natural fossil fuels (coal, natural gas, petroleum) as well as relatively small amounts of renewables and nuclear energy. Fossil fuels power

85% of the world's energy (1). Although fossil fuels provide the bulk of the world's energy, their use is not without significant negative side effects that include an ever-fluctuating supply and price. Further, the world's supply is dwindling and the debate is ongoing as to the environmental impact of fossil fuel combustion, whether it may result in an unsustainable carbon imbalance.

Research into alternative fuels is therefore essential to meet the world's increasing energy needs. To combat these issues and meet the demand, alternative energy sources are needed. Ultimately, energy supplies should derive from the most abundant resource. These include energy from the sun and the carbon from atmospheric CO<sub>2</sub>. Current approaches for harvesting sunlight using man-made device include: photovoltaic cells (also known as solar panels), which convert solar energy to electricity directly, and windmills, which harvest wind power indirectly generated by solar energy. In addition, plants and algae can harvest sunlight and carbon dioxide (CO<sub>2</sub>) to generate biomass that can be further converted into biofuels.

**Table 1. Carbon flux and energy yield from various plants to biofuels (4).**  
**Data obtained from ARPA-E (4)**

Source	Carbon Captured (t <sub>c</sub> /ha/yr)	Carbon Harvested (t <sub>c</sub> /ha/yr)	Carbon processed (t <sub>c</sub> /ha/yr)	Overall fuel yield (GJ /ha/yr)	% Energy Yield*
Corn (ethanol)	17.7	3.9	1.5	79	0.13
Soybean (FAME)	3.1	1.3	0.34	17	0.03
Sugar Cane (ethanol)	24	16	4.0	207	0.33

t<sub>c</sub> = carbon tonne. FAME= fatty acid methyl ester. \* Energy yield is calculated based on an average solar energy input of 200 W/m<sup>2</sup>.

The theoretical efficiency to convert sunlight to electricity using photovoltaic cells is 33.7%, with a practical efficiency of 10-20% (2). The theoretical efficiency to harness sunlight to biomass in plants is 4.6% for C<sub>3</sub>-plants and 6% for C<sub>4</sub>-plants, with a practical efficiency of 1% (3). Processing of plant biomass to generate sugars that are subsequently converted into a biofuel, such as ethanol, further reduces the above theoretical efficiency from 4-6% to 0.1-0.3% (according to calculations) (4). As demonstrated in Table 1, the energy efficiency of several plants is compared from the stage of CO<sub>2</sub> capture to the production of a biofuel. Comparatively, sugar cane to ethanol has the highest overall efficiency (0.3%)

followed by corn to ethanol (0.12%) and then soybean to biodiesel (0.03%) (4). Although these processes have been commercialized, they are in need of improvement and innovation at every stage from photosynthesis, CO<sub>2</sub> fixation, even to the final steps of making the fuels. Thus the present research focused on ways to genetically engineer the biochemical process of CO<sub>2</sub> fixation and fuel synthesis in microbes capable of harnessing sunlight or electricity-derived reducing power to power biosynthesis.

It is well-known that advanced biofuels, such as *n*-butanol, can be produced in microorganisms. In addition to its use as a solvent, plasticizer, and a raw material, *n*-butanol has another potentially important use, as a gasoline substitute (4). Similarly, isobutanol can also act as a gasoline substitute as it has an energy density of more than 100 mega Joules per gallon compared to gasoline that has an energy density of 128 mega Joules/gallon, and as such is considered a next generation biofuel (4). To develop isobutanol and *n*-butanol as a next-generation biofuels, we began to investigate the biosynthesis of *n*-butanol and isobutanol in a biological host system and aimed to generate biofuels such as *n*-butanol and isobutanol from sugar or CO<sub>2</sub>.

## 2. Results and Discussion

### 2.1. Genetic Manipulation of *E. coli* To Produce Biofuels

In *C. acetobutylicum*, six genes *thl*, *hbd*, *crt*, *bcd*, *etf*, and *adhE2* are required for the biosynthesis *n*-butanol from acetyl-CoA (Figure 1) (5). To characterize the biosynthetic process, the *n*-butanol biosynthetic machinery was cloned and expressed in *Escherichia coli*, which allowed us to study the production of *n*-butanol in this easily controllable model organism (5). As all the genes responsible for this transformation have been identified, it was hypothesized that as long the genes could be expressed in *E. coli* the biosynthesis of *n*-butanol from glucose should proceed rapidly. Although the process was successful, the highest yield that could be obtained, even under optimized conditions, was no greater than 0.55 g/liter of *n*-butanol (Figure 2) (5). Although this was an interesting result, the production rate and titer were much lower than that required for commercial feasibility.

Similar to the production pathway used for the biosynthesis of *n*-butanol in *C. acetobutylicum*, the biochemical pathway for the synthesis of isobutanol consists of 5 genes: *alsS*, *ilvC*, *ilvD*, *kdc*, and *adh* (Figure 3) (6). These isobutanol biosynthesis genes were cloned from various organisms, combined, and expressed in a single host system, *E. coli*. Surprisingly, using this engineered *E. coli* system isobutanol was produced from glucose to a titer of more than 20 g/L with a yield of 86% of theoretical maximum (Figure 4) (6). This level of production titer and yield rival the bioproduction of *n*-butanol used in its native organism *C. acetobutylicum*, which produces *n*-butanol at 19.6 g/L in 83% yield (7). Thus by means of genetic engineering tools a new strain of *E. coli* was engineered to produce isobutanol at a level comparable to the native organism (6).

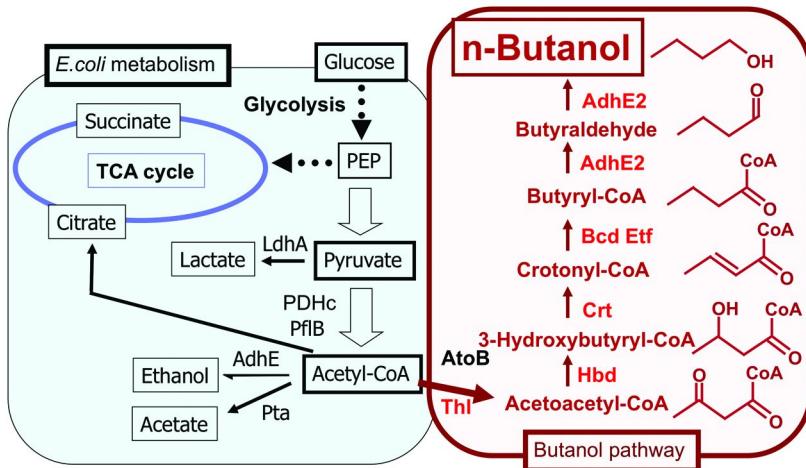


Figure 1. Biosynthesis of *n*-butanol from glucose in *C. acetobutylicum*. All the genes required for this biosynthesis were cloned into *E. coli* and the biosynthesis of *n*-butanol was studied (5).

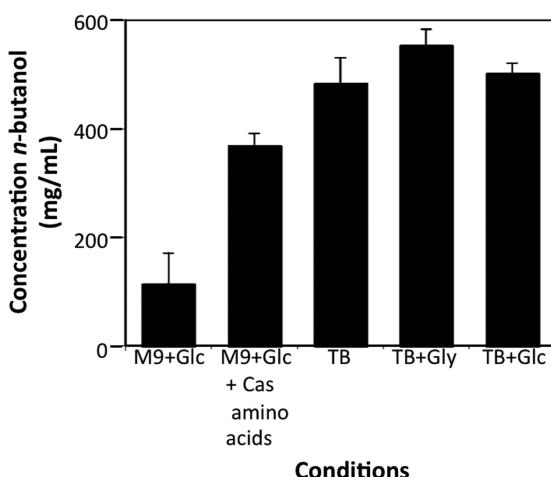


Figure 2. Optimization of *n*-butanol production from genetically modified *E. coli* expressing the *n*-butanol biosynthesis pathway from *C. acetobutylicum*. Maximum *n*-butanol produced was 0.55 g/L. Legend: M9, minimal salt medium; Glc, glucose; Cas, Cas amino acids; TB, “terrific broth” rich medium; Gly, glycerol. Modified from reference (5).

From a graph of optical density (OD) versus time, it was found that the majority of the isobutanol was produced after *E. coli* cell's growth stopped (Figure 4). It is interesting and as yet unclear, how the cells continued to produce isobutanol to this concentration when in fact isobutanol and *n*-butanol are toxic to *E. coli* cells at concentration of more than 6-7 g/L (8, 9).

Isobutanol was also produced in *E. coli* using an evolutionary strategy that selectively pressured the bacteria to produce isobutanol, as opposed to engineering a strain that produced this biofuel (10). Evolutionary methods were employed to identify a strain of *E. coli* that could produce isobutanol by norvaline selection to a final titer of more than 20 g/L (10). Similar to the engineered isobutanol producing strain, the evolutionary derived strain also stopped growing after 24 hours, but continued to produce isobutanol (10).

Scaling the production of isobutanol in a bioreactor achieved an effective total production of 50 g/L by simply removing or gas-stripping off the product as it was produced in the bioreactor (11).

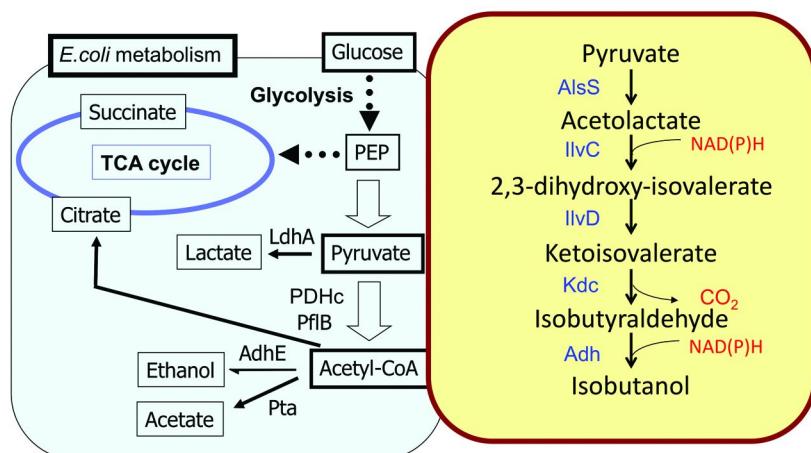


Figure 3. The genes required for the biosynthesis of isobutanol starting from glucose were cloned from different organisms, combined and expressed in *E. coli*. The biosynthesis of isobutanol was then studied as the cells grew. Modified from reference (6).

## 2.2. Driving the Bioproduction of *n*-Butanol

Genetic engineering strategies provided markedly different final titers for the biosynthesis of *n*-butanol versus isobutanol. Despite the fact that both used identical shake flask fermentation techniques, *n*-butanol production reached a plateau at 0.5 g/L whereas isobutanol production reached greater than 20 g/L. To solve the discrepancy the biochemical pathways used to generate ethanol, *n*-butanol, isobutanol, and isopropanol on large scale were carefully analyzed (Scheme 1). Employing these engineering techniques ethanol (12), isobutanol

(11), and isopropanol (13, 14) could all be produced in good yields (not all data presented), thus we sought a rational explanation for why in non-native producers *n*-butanol production lagged behind.

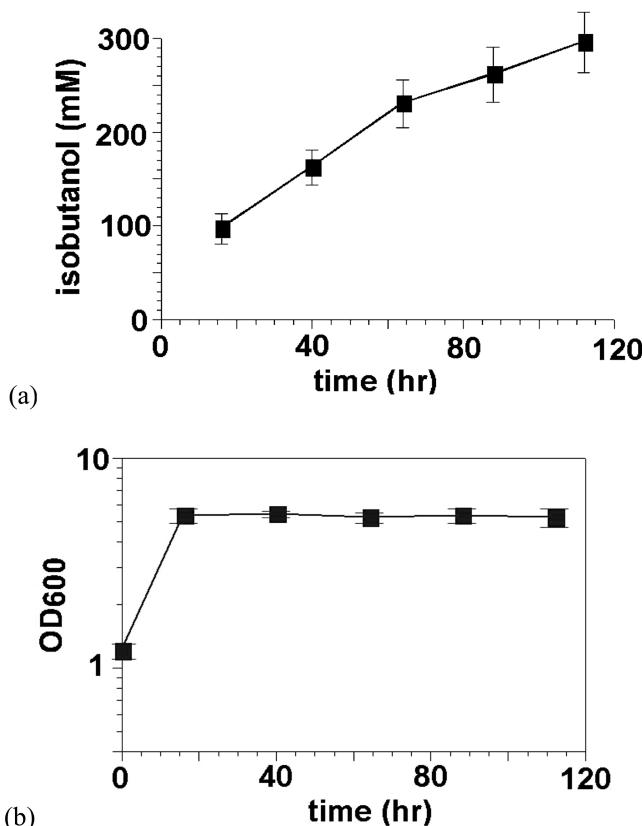
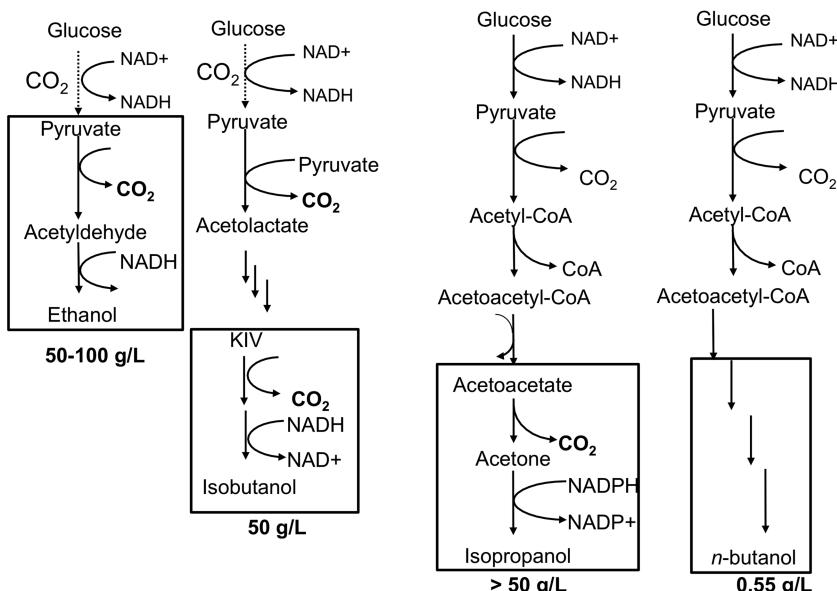


Figure 4. Isobutanol production and cell growth over time in *E. coli* engineered to express an isobutanol biosynthesis pathway. (a) Concentration of isobutanol produced in mM, 300 mM is ca. 21 g/L. (b) Optical density (OD) of *E. coli* cells producing isobutanol over 120 hours. Taken from reference (6).

According to the biochemical pathways outlined in Scheme 1 it was apparent that the higher yielding processes all had one thing in common, a late stage CO<sub>2</sub> generating step. A CO<sub>2</sub> evolution step serves as an irreversible kinetic trap that drives a metabolic process toward the desired products. In ethanol production during the conversion of pyruvate into acetaldehyde, CO<sub>2</sub> is released (Scheme 1). Similarly, in the biochemical pathway used to generate isopropanol, CO<sub>2</sub> is released during the conversion of acetoacetate into acetone (Scheme 1). Notice however that in the biochemical pathway used to generate *n*-butanol no such

driving force exists (Scheme 1). Thus it was hypothesized that this was the cause of the discrepancy in *n*-butanol production in non-native producers, despite the fact that it is the same production pathway used *in vivo* by native organisms used to generate *n*-butanol.

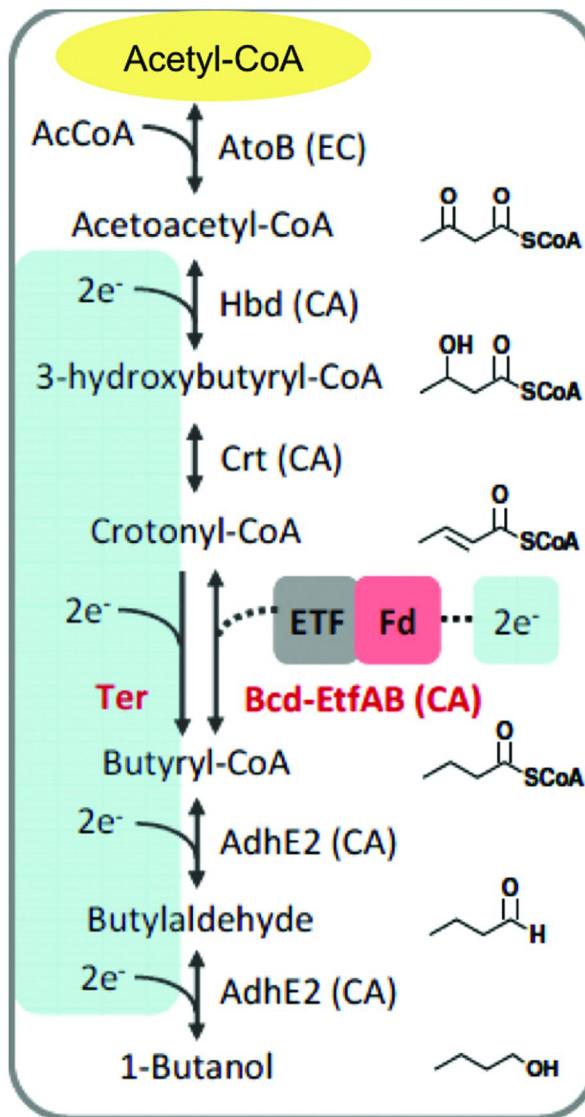


*Scheme 1. Outline of the biosynthetic pathways used to generate biofuels in engineered *E. coli* strains from glucose. Ethanol was produced on largest scale from 20-100 g/L followed by isopropanol at more than 50 g/L, and isobutanol at 50 g/L. Notice each pathway includes a late stage irreversible step, where CO<sub>2</sub> is released (boxed in black), except in the biosynthesis of *n*-butanol, which provided a reason why its biosynthesis lagged so far behind other similar alcohols at 0.55 g/L (5)*

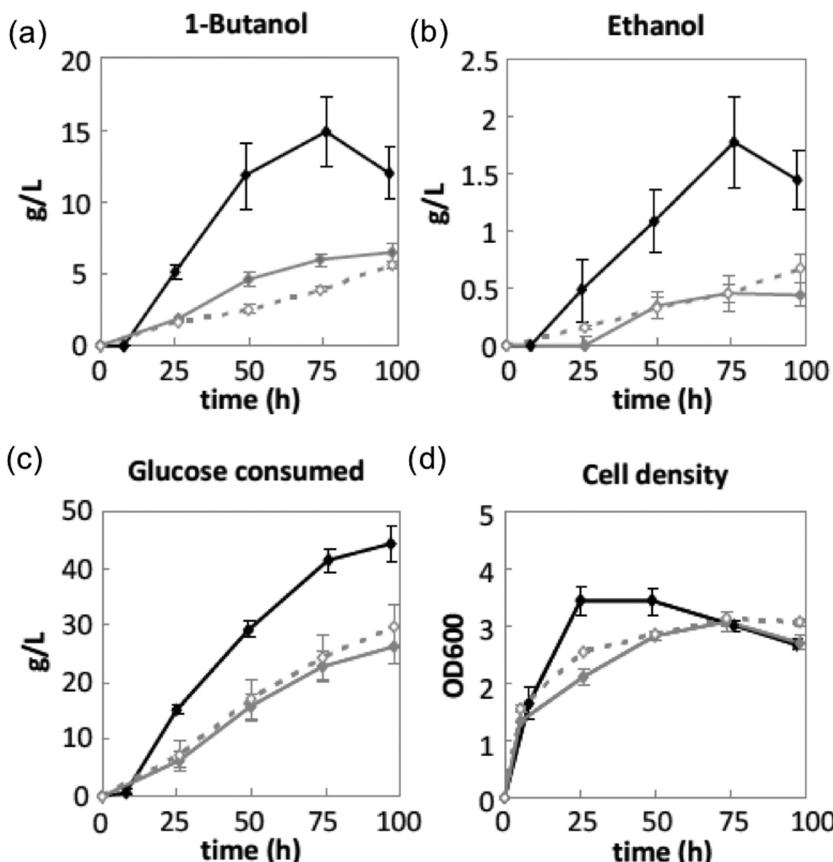
To drive the pathway toward *n*-butanol synthesis several genetic modifications of the pathway in *E. coli* were investigated. These modifications included the use of a common electron donor molecule, enzyme manipulation, and substrate saturation.

The biochemical reactions that generate *n*-butanol involve several reduction processes (Scheme 2). In particular, in step 4 where crotonyl-CoA is converted to the intermediate butyryl-CoA by the enzyme Bcd-EtfAB (CA, Scheme 2), where Etf derives the electrons for the reduction reaction possibly from ferredoxin (15). It was hypothesized that shuttling the reducing power by using a single reducing agent, such as the reduced form of the coenzyme nicotinamide adenine dinucleotide (NADH), would help to drive the entire process forward. As NADH is not the natural electron source for the Bcd-EtfAB complex, the gene for the complex was replaced in an *E. coli* host expressing the entire pathway with another gene called *ter* (16). *Ter* expresses a protein *trans*-enoyl-CoA reductase that also catalyzes the conversion of crotonyl-CoA to butyryl-CoA, but it uses

NADH as the electron donor (16). This genetic engineering strategy accomplished two objectives. The first was to forcibly drive all reduction steps to use a single reducing agent, NADH. The second was to add a late stage, irreversible step that could provide an additional driving force to propel this pathway in the forward direction.



*Scheme 2. Creating artificial driving forces for n-butanol biosynthesis in genetically engineering *E. coli* expressing an n-butanol biosynthesis pathway. Notice *bcd-EtfAB* (CA) was replaced with *ter*. Figure modified from reference (15)*



*Figure 5. Concentration of n-butanol (g/L), glucose (g/L), glucose consumed (g/L), and cell density as a function of time for *E. coli* engineered to express modified n-butanol biosynthesis pathway (15).*

The first step in the *n*-butanol biosynthesis pathway, the conversion of acetyl-CoA to acetoacetyl-CoA, is not thermodynamically favorable. To overcome this thermodynamic barrier and make this step more favorable the intracellular concentration of acetyl-CoA was increased, by knocking out a gene called *pta*, which codes for phosphate acetyltransferase (PTA) an enzyme that consumes acetyl-CoA. By knocking the *pta* gene out the intracellular acetyl-CoA concentration was effectively increased (15).

By using this three pronged approach (1) replacement of the *bcd-Etf* gene with *ter*, (2) use of a single electron donor, NADH, and (3) increasing the intracellular acetyl-CoA concentration the production of *n*-butanol was increased from 0.5 g/L to 15 g/L in 75 hours (Figure 5) (15). The results indicate that with a proper driving force a pathway can be forced into the desired direction.

### 2.3. Engineering the Biosynthesis of Fuels from CO<sub>2</sub>

Until this point *n*-butanol and isobutanol (as well as other alcohols not reported) were biosynthesized from glucose. However, our ultimate goal is not to prepare large amounts of fuels from sugar, rather from CO<sub>2</sub>. To do this we looked for ways to fix CO<sub>2</sub> to generate biofuels. The desire is to achieve this biochemically using a prokaryotic organism that can be easily manipulated. Photosynthetic cyanobacteria harvest sunlight using photosystems and produce energy in the form of adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide phosphate (NADPH), which are used to fix CO<sub>2</sub> *via* the Calvin cycle to synthesize sugars. One of the intermediate metabolites downstream of the Calvin cycle is pyruvate, which as we have seen previously, can be diverted into isobutanol or *n*-butanol biosynthesis by genetic manipulation of the cyanobacteria. To derive these products from sunlight and CO<sub>2</sub>, we genetically modified cyanobacteria to express biofuel production pathways.

As discussed previously, the essential genes for isobutanol biosynthesis are well characterized (6). These genes were cloned and expressed in the cyanobacteria *Synechococcus elongatus* (strain PCC7942) (17). After the cells were grown under a particular set of photosynthetic conditions, isobutanol was detected in the culture medium (~500 mg/L) (17). Encouraged by this result, the reaction was pushed even further forward by gas-stripping off an intermediate, isobutyraldehyde to a final titer of 1.2 g/L in 8 days (17). Sheehan compared the current production levels achieved employing our biosynthetic isobutanol and isobutyraldehyde production method from CO<sub>2</sub> to current corn and cellulosic ethanol production methods (18). Although it is important to note the production scales were vastly different, Sheehan found that our biosynthetic method was two to three fold higher than other production methods (18).

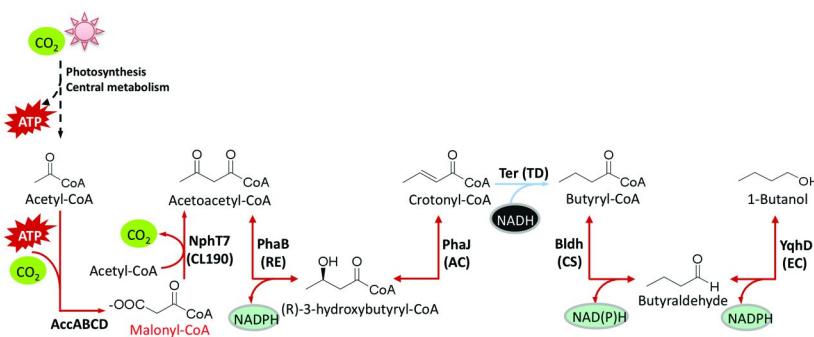
As isobutanol could be successfully produced from CO<sub>2</sub> in the cyanobacterium *S. elongatus*, the strategy was extended to the production of *n*-butanol in *S. elongatus*. Initially, we investigated the biosynthesis of *n*-butanol using the same set of genes used in the production of *n*-butanol in *E. coli* (See Scheme 2 for an outline of these enzymes) (19, 20). As the genes required for the biosynthesis of *n*-butanol from acetyl-CoA are well characterized (5) they were cloned and expressed in the cyanobacterium *S. elongatus* (see Scheme 2 for an overview of the biosynthesis of *n*-butanol) (19, 20). However, as seen previously in *E. coli*, *n*-butanol was not detected in *S. elongatus*.

To overcome this problem, the lessons learned in the biosynthesis of *n*-butanol in *E. coli* were applied in *S. elongatus*. Namely the pathway was engineered to proceed in the forward direction by manipulating driving forces. Instead of forcing the cyanobacteria to use NADH, or increasing the intracellular concentration of acetyl-CoA - as accumulation of these compounds in cyanobacteria proved to be difficult - other driving forces were identified. Specifically, an ATP-driven irreversible step was introduced into the biosynthetic pathway, and the cyanobacterium was forced to utilize exclusively NADPH as the reducing agent (20). Cyanobacteria favor the use of NADPH as a reducing agent over NADH, thus the enzymes used in the *E. coli* *n*-butanol biosynthetic pathway were exchanged with enzymes that would perform the desired transformation

with NADPH (Scheme 3). Second, a gene coding for the enzyme AccABCD that catalyzes the conversion of acetyl-CoA to malonyl-CoA was recruited to the pathway. This step produces malonyl-CoA, which is subsequently consumed by the enzyme NphT7 (CL190) (21) to generate acetoacetyl-CoA, in an irreversible and energetically favorable reaction that releases CO<sub>2</sub> (Scheme 3). The first step requires ATP, which is usually not favorable, however in this case ATP is used as a driving force to push the reaction forward (Scheme 3). As a result of this engineering strategy the pathway produced *n*-butanol in cyanobacteria from CO<sub>2</sub>, using ATP as a driving force and NADPH as reducing power to a titer of 30 mg/L (20).

## 2.4. Using Photovoltaic Cells To Generate Biofuels from Atmospheric CO<sub>2</sub>

To date this CO<sub>2</sub> fixation technology has been utilized to generate isobutanol, *n*-butanol, 2-methyl-1-butanol, and 3-methyl-1-butanol in cyanobacteria. Although this was a very exciting accomplishment, the scale-up of this process to reach production goals would require a great deal of sunlight exposed surface area, which could be very costly. In an attempt to make the process more cost effective we devised alternative way to harness energy from sunlight and couple that to CO<sub>2</sub> fixation and production of biofuels, employing electrochemical photovoltaic cells.



*Scheme 3. Outline of the optimized biosynthetic pathway used for the biosynthesis of *n*-butanol in *S. elongates* from CO<sub>2</sub>. Note the additional two step conversion of acetyl-CoA into malonyl-CoA and then acetoacetyl-CoA that is catalyzed by AccABCD and NphT7, respectively. Further, note the use of alternate NADPH using enzymes outlined in Scheme 2. Figure modified from reference (20)*

Photosynthesis is a two-stage reaction process comprised of light reactions (light is used to generate ATP and NADPH) and dark reactions or the Calvin cycle (assimilation of CO<sub>2</sub>). To facilitate scale-up the CO<sub>2</sub>-derived biofuels, it was thought that separation of the light and dark reactions would reduce the amount

of light exposed surface area needed. To achieve this separation, photovoltaic cells were envisioned to harvest sunlight that would then be used to generate electricity. The electricity would in turn be used to drive an electrochemical cell that would be used to generate hydrogen or formic acid that could be fed into the dark reaction. Two strategies were put forward, the first involving hydrogen production and second using formic acid. The former is problematic because it has a low solubility in water and it is also explosive.

Formic acid is a hydrogen alternative, it is water-soluble, and can be generated using an electrochemical reaction (Figure 6a). In this strategy the photovoltaic cells would be used to harvest sunlight and the energy generated coupled to formic acid production from water and CO<sub>2</sub>. The formic acid could then be used to produce NADPH (or NADH) to generate biofuels *via* the Calvin cycle.

*Ralstonia eutropha* is an organism that is capable of utilizing both hydrogen and formic acid as an electron donor source to fix CO<sub>2</sub> (22). *R. eutropha* was genetically engineered to divert the metabolite pyruvate, from the Calvin cycle, into isobutanol or 3-methyl-1-butanol biosynthesis, by cloning the genes necessary for biofuel production into *R. eutropha* (22). To first confirm that the engineered *R. eutropha* strain, expressing the biofuel biosynthesis pathways, would produce the biofuels, the cells were allowed to grow in the presence of hydrogen or formate as energy source (22). Gratifyingly, isobutanol and 3-methyl-1-butanol were detected (22).

The next step was to couple isobutanol and 3-methyl-1-butanol biosynthesis directly to formic acid generated in an electrochemical cell. The electrochemical cell used was devised and was powered by electricity, and it generated formic acid in the cathode, and oxygen in the anode (Figure 6a). The formic acid generated would then be transported into the *R. eutropha* cell, releasing CO<sub>2</sub>. The cells would then generate NADH that would be used to drive CO<sub>2</sub> fixation in the Calvin cycle and the end result would be the generation of biofuels (22). Although the overall integrated process was promising, as soon as the current was initiated the cells stopped growing (Figure 6b). This stunted growth was attributed to the production of toxic by-products that were generated from the overall light harvesting reaction (22).

To conclusively identify the toxic compounds three biologically derived sensors used to detect free radicals KadG, SodC, or NorA promoters were engineered into the *R. eutropha* cells. NorA is responsive to nitric oxide radical, SodC is responsive to superoxide, and KatG is responsive to hydrogen peroxide.<sup>30</sup> Coupling these promoters to the *lacZ* gene allowed for the detection of the presence of one of these radical byproducts (22). After turning on the current superoxide and nitric oxide radical were detected by up-regulation of SodC and NorA (Figure 6c), but not hydrogen peroxide. Since nitric oxide and superoxide are known to suspend cell growth, the reason the cells were unable to grow was attributed to the presence of the free radicals.

To circumvent this problem a new set-up of the device was devised, where the electrode was physically separated from the bulk of the solution where the cells reside, with a ceramic cup (Figure 7a) (22). The ceramic cup was permeable to most of the compounds, but not the cells. Its function was to separate the cells from the electrode and to effectively prevent the cells from coming into contact with the

free radicals, which were allowed to decay on their own. Gratifyingly, under these conditions the cells grew in the bioreactor and isobutanol and 3-methyl-1-butanol were generated from electricity and  $\text{CO}_2$  (Figure 7b) (22).

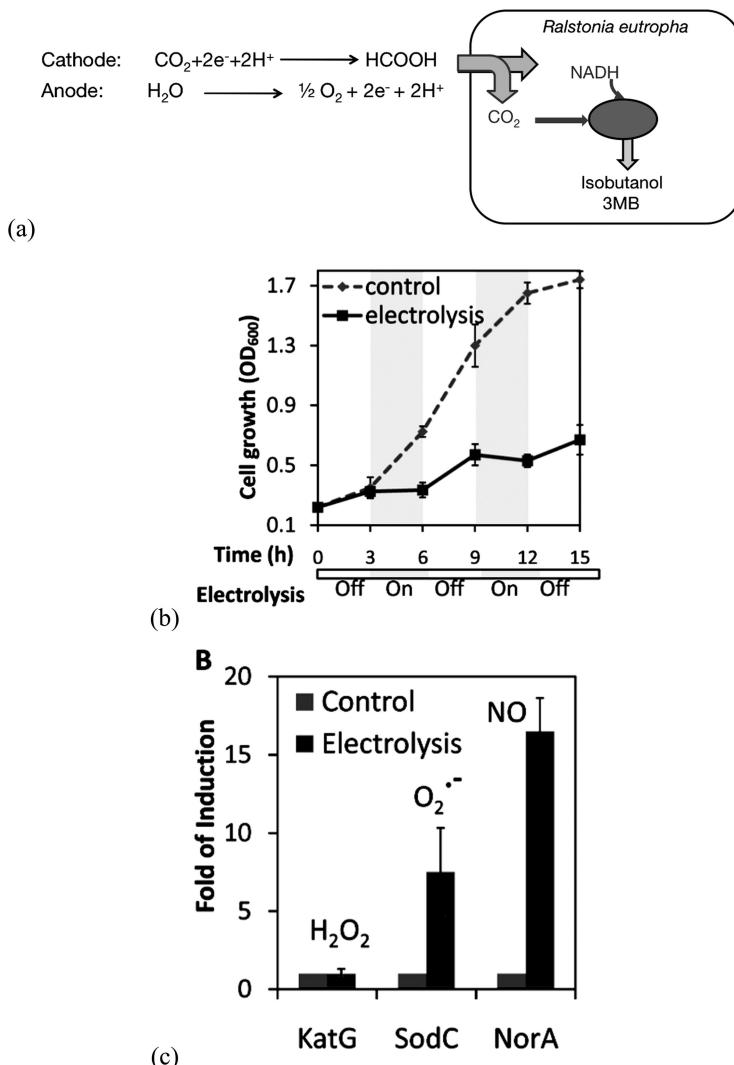
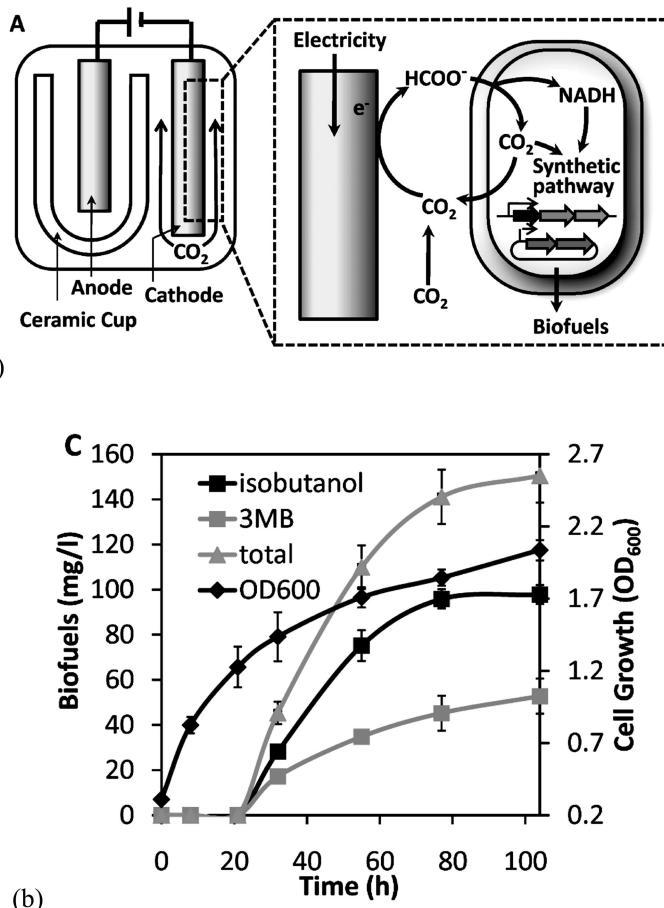


Figure 6. Electricity-powered  $\text{CO}_2$  fixation for isobutanol synthesis. (a) electrolysis reaction with formic acid generated at the cathode and  $\text{O}_2$  at the anode and coupling this to the production of isobutanol or 3-methyl-1-butanol. (b) Cell growth as a function of powering the electrolysis reaction. (Legend: ♦ control, no current; ▀, electrolyzed sample.) (c) Detection of *KatG*, *SodC* and *NorA* as a function of electrolysis. (Legend: grey bars, control; black bars, electrolyzed sample). Figures modified from reference (22).



*Figure 7. Electricity-powered  $CO_2$  fixation for isobutanol synthesis in a ceramic cup. (a) Overall reaction set-up separating the radicals formed in the light reaction from penetrating the solution that comes into contact with the cells. (b) Production of isobutanol (●), and 3-methyl-1-butanol (■) as a function of time after electrolytic formation of formic acid. Notice the cells also grew during electrolysis with increased  $OD_{600}$  (◆) over time (22).*

The process can be integrated to use energy from sunlight and fix  $CO_2$  from the atmosphere with a photovoltaic cell. The overall strategy to generate biofuels by combining electrochemical reactions with microbial reactions is promising (Figure 7). An estimation of the efficiency in each step for converting solar energy to isobutanol via electricity is outlined in Table 2. The overall efficiency is the product of the efficiency in each step (Table 2). The efficiencies values may appear low, however these values are 10 fold higher than the current schemes of biofuel production (22). It should be noted that these values only take into account a single criteria and do not account for cost. Other criteria must be considered for before scaling up.

**Table 2. Estimated energy efficiency from harvesting of solar energy to production of isobutanol (22)**

	<i>Solar to Electricity EFF. [Range=8-25%]</i>	<i>× Electrolysis EFF. [Range=56-80%]</i>	<i>Theoretical EFF.% Electricity to Isobutanol</i>	<i>× Theoretical yield [Range=30-90%]</i>	<i>Overall = % EFF.% Solar to Isobutanol</i>
Lower bound	8	56	50	40	0.9
<i>Likely</i>	15	62	50	60	2.79
Higher bound	25	80	50	90	9.0

EFF. = efficiency.

### 3. Conclusion

If we could look through a crystal ball our goal would be to develop a hybrid energy system that consists of manmade as well as biological systems that will be able to work together to harvest sunlight and CO<sub>2</sub> to generate next-generation biofuels. The remaining challenges seek to increase both the rate and efficiency of production, and need to include the costs and schemes associated with product isolation and scale-up. Finally, a way to harvest CO<sub>2</sub> from the atmosphere is still a challenging problem that will require a tremendous amount of further research. However, as demonstrated, progress has been made toward the identification of biological hosts that can be used to generate biofuels from glucose, light, hydrogen, and formic acid. Energy from electrochemical photovoltaic cells and the sun has been used to generate biofuels and research is ongoing as to integrate this with energy captured from the sun.

### Acknowledgments

This work is supported in part by ARPA-E Electrofuel program, NSF MCB-0903955 and MCB-1139318. The author thanks all the graduate students, postdoctoral researchers, and technician who participate in the work reported here.

### References

1. Department of Energy. *International Energy Outlook 2011*. U.S. Energy Information Administration Report DOE/EIA-0484. (2011). <http://www.eia.gov/forecasts/ieo/> (accessed on September 4, 2012).
2. Green, M. A.; Emery, K.; Hishikawa, Y.; Warta, W.; Dunlop, E. D. Solar cell efficiency tables (version 39). *Prog. Photovolt. Res.* **2012**, *20*, 12–20.

3. Zhu, X. G.; Long, S. P.; Ort, D. R. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr. Opin. Biotechnol.* **2008**, *19*, 153–159.
4. Advanced Research Projects Agency – Energy, U.S. Department of Energy. *Plants Engineered to Replace Oil (PETRO) Funding Opportunity Announcement*. Report: DE-FOA-0000470. (2011); <https://arpa-e-foa.energy.gov/> (accessed September 4, 2012).
5. Atsumi, S.; Cann, A. F.; Connor, M. R.; Shen, C. R.; Smith, K. M.; Brynildsen, M. P.; Chou, K. J.; Hanai, T.; Liao, J. C. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **2008**, *10*, 305–311.
6. Atsumi, S.; Hanai, T.; Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008**, *451*, 86–89.
7. Chen, C. K.; Blaschek, H. P. Acetate enhances solvent production and prevents degeneration in *Clostridium beijerinckii* BA101. *Appl. Microbiol. Biotechnol.* **1999**, *52*, 170–173.
8. Brynildsen, M. P.; Liao, J. C. An integrated network approach identifies the isobutanol response network of *Escherichia coli*. *Mol. Sys. Biol.* **2009**, *5*, 277.
9. Atsumi, S.; Wu, T. Y.; Eckl, E. M.; Hawkins, S. D.; Buelter, T.; Liao, J. C. Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Appl. Microbiol. Biotechnol.* **2010**, *85*, 651–657.
10. Smith, K. M.; Liao, J. C. An evolutionary strategy for isobutanol production strain development in *Escherichia coli*. *Metab Eng.* **2011**, *13*, 674–681.
11. Baez, A.; Cho, K. -M.; Liao, J. C. High-flux isobutanol production using engineered *Escherichia coli*: a bioreactor study with *in situ* product removal. *Appl. Microbiol. Biotechnol.* **2011**, *90*, 1681–1690.
12. Ingram, L. O.; Conway, T.; Clark, D. P.; Sewell, G. W.; Preston, J. F. Genetic engineering of ethanol production in *Escherichia coli*. *Appl. Environ. Microbiol.* **1987**, *53*, 2420–2425.
13. Hanai, T.; Atsumi, S.; Liao, J. C. Engineered synthetic pathway for isopropanol production in *Escherichia coli*. *Appl. Environ. Microbiol.* **2007**, *73*, 7814–7818.
14. Inokuma, K.; Liao, J. C.; Okamoto, M.; Hanai, T. Improvement of isopropanol production by metabolically engineered *Escherichia coli* using gas stripping. *J. Biosci. Bioeng.* **2010**, *110*, 696–701.
15. Shen, C. R.; Lan, E. I.; Dekishima, Y.; Baez, A.; Cho, K. M.; Liao, J. C. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* **2011**, *77*, 2905–2915.
16. Tucci, S.; Martin, W. A novel prokaryotic trans-2-enoyl-CoA reductase from the spirochete *Treponema denticola*. *FEBS Lett.* **2007**, *581*, 1561–1566.
17. Atsumi, S.; Higashide, W.; Liao, J. C. Direct recycling of carbon dioxide to isobutyraldehyde using photosynthesis. *Nat. Biotechnol.* **2009**, *27*, 1177–1180.
18. Sheehan, J. Engineering direct conversion of CO<sub>2</sub> to biofuel. *Nat. Biotechnol.* **2009**, *27*, 1128–1129.

19. Lan, E. I.; Liao, J. C. Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metab. Eng.* **2011**, *13*, 353–363.
20. Lan, E. I.; Liao, J. C. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl. Acad. Sci., U.S.A.* **2012**, *109*, 6018–6023.
21. Okamura, E.; Tomita, T.; Sawa, R.; Nishiyama, M.; Kuzuyama, T. Unprecedented acetoacetyl-coenzyme A synthesizing enzyme of the thiolase superfamily involved in the mevalonate pathway. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11265–11270.
22. Li, H.; Opgenorth, P. H.; Wernick, D. G.; Rogers, S.; Wu, T. Y.; Higashide, W.; Malati, P.; Huo, Y. X.; Cho, K. M.; Liao, J. C. Integrated electromicrobial conversion of CO<sub>2</sub> to higher alcohols. *Science* **2012**, *335*, 1596.

## Chapter 2

# Unfolded Protein Response (UPR) During CHO Cell Production Culture

**Zhimei Du,\* David Treiber, Rebecca E. McCoy, Feng He, Amanda K. Miller, Mei Han, Carole Heath, and Pranhitha Reddy**

**Amgen Inc., 1201 Amgen Court West, Seattle, Washington 98119**

**\*E-mail: zdu@amgen.com. Phone: +1 206-265-7367. Fax: +1 206-217-4692**

The unfolded protein response (UPR) is a complex and intricate network of signaling pathways in the endoplasmic reticulum (ER). The UPR is responsible for monitoring and controlling intracellular responses to the accumulation of unfolded or misfolded proteins under conditions of cellular stress. In mammalian cell cultures the expression of high levels of recombinant proteins, such as monoclonal antibodies, are associated with increased UPR levels. However, the dynamic changes of the UPR and its impact on cell performance as well as recombinant protein secretion are poorly understood. Herein a UPR-specific monitoring system was created that can be used to detect and quantify endogenous UPR activation levels in real-time during the production process. Using this UPR-specific monitoring system, it was found that recombinant Chinese hamster ovary (CHO) clones differed in their UPR induction patterns. Both the timing and the degree of UPR-induced transcriptional activation for the different clones were linked to the growth, viability, and productivity of the cells. In addition, the UPR-specific monitoring system revealed that cell culture conditions, such as the accumulation of metabolic end products and osmolarity, can also alter UPR levels independent of recombinant protein expression. Lastly, how a production process can be rescued from decreased cell viability and productivity by controlling the UPR in a live growing culture will be addressed. Some of findings below have been summarized in the manuscript submitted to a peer reviewed journal (revision under journal review).

# 1. Introduction

CHO cells are commonly used in the large-scale production of recombinant proteins, such as monoclonal antibodies (mAb). The unfolded protein response (UPR) pathway is a cell response pathway that is activated by an accumulation of unfolded or misfolded proteins in the endoplasmic reticulum (ER). Overall the UPR results in the activation of three different pathways: (1) the protein kinase RNA-like endoplasmic reticulum kinase pathway (PERK), (2) the inositol-requiring kinase 1 (Ire1) – X box binding protein 1 (Xbp1) pathway (Ire1/Xbp1), and (3) the activating transcription factor 6 pathway (ATF6) (1).

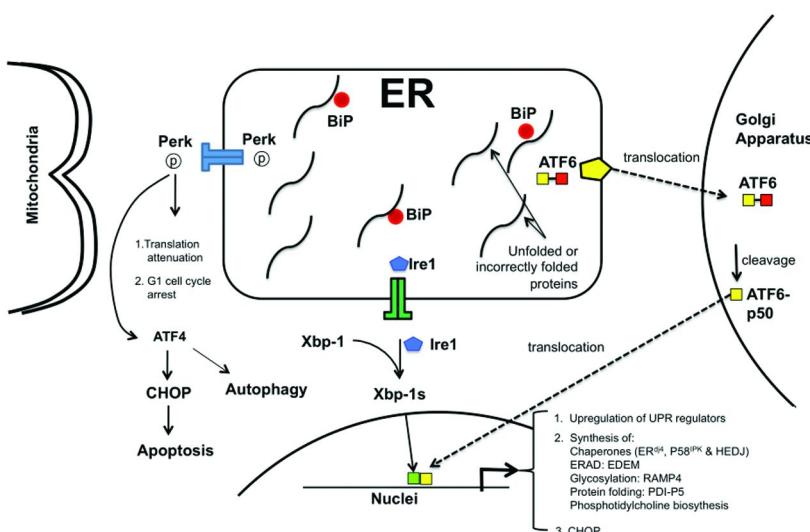


Figure 1. Schematic of the UPR pathway activated by the accumulation of unfolded or incorrectly folded proteins, by sequestering BiP. Legend: BiP, ; ATF6, ; Ire1, ; Perk, ; Xbp-1s, ; ATF6-p50, .

As previously stated the accumulation of unfolded or misfolded proteins in ER initiates the UPR pathway; this process occurs when the molecular chaperone binding immunoglobulin protein (BiP) (also known as glucose-regulated protein or heat shock 70 protein 5) binds to the hydrophobically exposed surface area of unfolded or misfolded proteins (Figure 1). Now otherwise engaged BiP can no longer bind to its ER membrane-binding partners Ire1, ATF6, and PERK triggering the activation of these proteins and their respective UPR related pathways (1). Activation of the PERK pathway mediates cell cycle arrest and protein translation attenuation (1). Activated Ire1 mediates mRNA alternative splicing of Xbp-1 (Figure 1). Once translated the spliced isoform Xbp-1s is translocated into the nucleus and activates a set of target genes that are involved in the UPR response (Figure 1) (2, 3). Activated ATF6 is translocated to the Golgi apparatus where it is subsequently cleaved by proteases into a smaller fragment,

called ATF6 p50 (50kDa, Figure 1) (4–6). ATF6 p50 is a potent transcription activator that is also translocated to the nucleus and activates another set of genes that may or may not overlap with Xbp-1 target genes (Figure 1) (4, 7). The UPR activated gene products include the UPR key regulators themselves (BiP, Ire1, ATF6, Xbp-1 and PERK) resulting in a positive feedback loop, as well as proteins that are involved in protein folding, glycosylation, degradation and lipid synthesis (1). Activated UPR-specific target genes have the ultimate goal to rapidly increase protein-folding capacity. If all these efforts still cannot rescue the cell from the protein folding mediated ER stress, the cell will undergo apoptosis (1, 8). Of these three pathways, only the Ire1/Xbp1 and ATF6 pathways are UPR-specific, whereas the PERK pathway is shared with other cellular stress pathways such as those involved in amino acid deprivation, infection with double-stranded RNA viruses, and mechanical stress. As the UPR pathway is centralized by the UPR-specific transcriptional events, detection of the overall UPR-specific transcriptional activation should provide a means to monitor or even quantify levels of UPR activation in cells under ER stress.

## 2. Materials and Methods

### 2.1. Cell Culture, Transfection, Cloning, and Stress Induction

For transient expression, HEK293 cells were transfected with UPR-inducible reporter plasmids using lipofectamine 2000 (Invitrogen), and cell lysates were harvested 24 hours later as described below. Stable CHO cell lines were obtained by transfecting a CHO host cell line or recombinant antibody expressing final clones with UPR-inducible reporter plasmids, followed by selection with G418.

### 2.2. DNA Constructs and Luciferase Assay

The UPR inducible promoter was synthesized and replaced Gal4 enhancers of the FIREfly luciferase reporter plasmid (Promega). For the luciferase assay cells were transfected and cell lysates were analyzed for luciferase activity according to the manufacturer's protocol (Promega).

### 2.3. Western Blot Analysis

Cells were lysed and subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis under reducing conditions, transferred to PVDF membrane, and probed with the indicated primary antibodies using standard procedures.

### 2.4. Antibody Production Process

The positive clones were analyzed in a 10-day fed-batch production assay. All of the production cultures were harvested on day 10. For all production assays, parameters such as cell growth, viability, antibody levels/expression titer, osmolarity, glucose levels, and lactate levels were collected on days 3, 6, 8 and 10.

## 2.5. Flow Cytometry Analysis and Fluorescence Microscopy

Cells from passaging and production cultures were collected, stained with PI, and analyzed by flow cytometry (Caliber, BD) for quantitative analysis of GFP fluorescence in the cell populations.

## 2.6. Cell Cycle Analysis

DNA content staining and FACS analysis were conducted using the Cell Cycle Analysis kit according to the instructions of the manufacturer (GenScript).

# 3. Results and Discussion

## 3.1. Monitoring of the Endogenous UPR Levels

In the present study, the levels of UPR induction during a CHO cell production process were investigated. Daily culture samples were taken from a CHO cell culture production batch and analyzed for the production and expression of UPR markers by Western Blot (Figure 2). The results clearly show that the levels of Chinese hamster ovary proteins (CHOP), proteins related to the overall CHO cell production, increased rapidly on day 1 and BiP was upregulated around day 6 (upper band Figure 2). As BiP plays a central role in the UPR pathway, these results suggest that UPR activation occurs as a result of the CHO cell culture production process. Although informative, quantification of overall endogenous levels of UPR activation by analyzing individual UPR markers was not possible. In order to monitor and quantify the comprehensive UPR activation in real time, a UPR-inducible reporter system was designed.

## 3.2. Quantification of UPR Induction Levels in Cells under ER Stress

To monitor and quantify the UPR levels in real time, a UPR-inducible reporter system was designed. The UPR-inducible reporter system contained an enhancer that responds to UPR-specific transcriptional activators, Xbp-1 and ATF6. The UPR-inducible reporter plasmid constructs also expressed a reporter gene, either luciferase or a degradable green fluorescent protein (dGFP, Figure 3). Cells were then transfected with the plasmids and monitored for the activity of the endogenous UPR-specific transcriptional activators during experimental conditions that simulate ER stress. As the plasmids contain a UPR responsive enhancer that binds to and is activated by UPR-specific transcriptional activators, its activation under ER stress will induce the downstream expression of the gene of interest. Thus UPR activation can be monitored using the reporter gene products, luciferase or dGFP (Figure 3). The half-life of dGFP is approximately 1 hour, a suitable timeframe for monitoring rapid changes in gene expression.

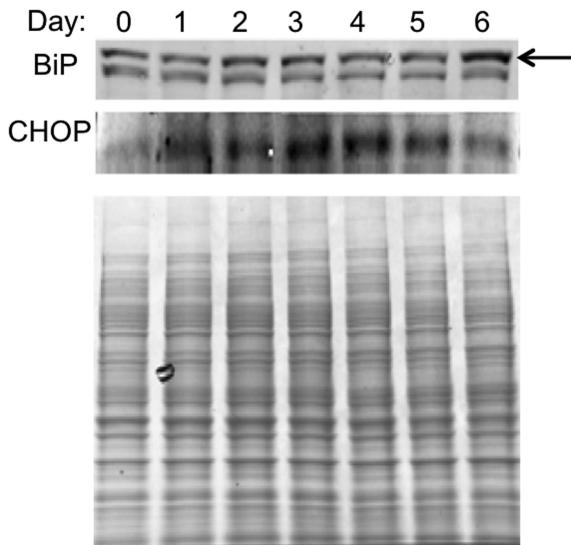


Figure 2. SDS-PAGE gel Western blot analysis of a CHO culture production batch days 0 through 7. Co-spotting against CHOP and BiP shown in the upper panels. Staining with Coomassie blue.

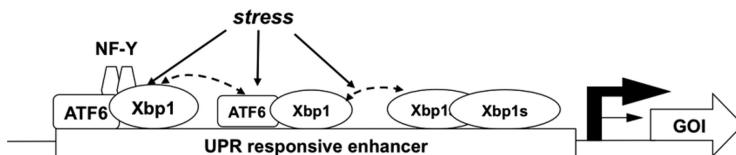
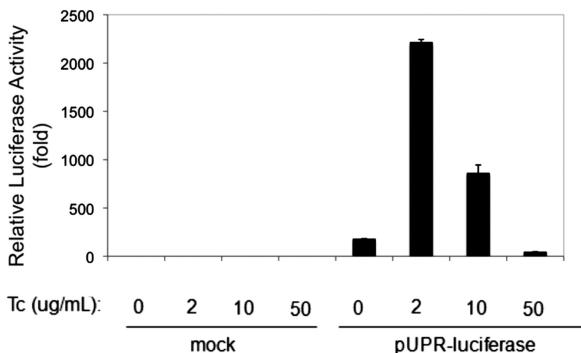


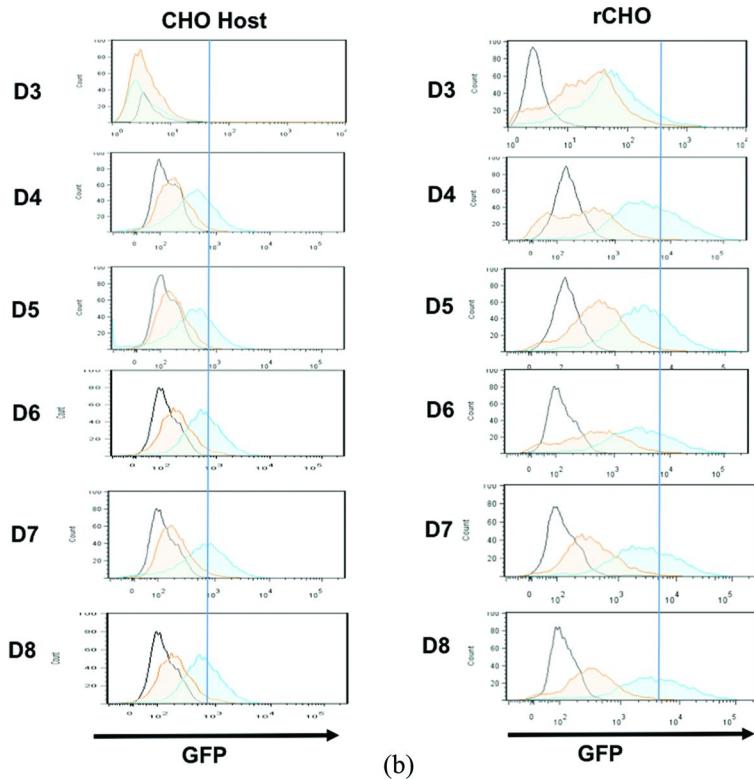
Figure 3. Schematic diagram of the UPR-dependent stress inducible reporter construct. Closer look at the UPR responsive enhancer and how ER stress causes its activation by ATF6 and Xbp-1s.

HEK293 cells were separately transfected with plasmids constructs containing or not containing a UPR-responsive enhancer and the cells were treated with increased doses of an ER stressor, tunicamycin (Tc). Tunicamycin induces ER stress by effectively blocking *N*-linked glycosylation of newly synthesized proteins, thereby disrupting protein folding in the ER (9). In response to 2  $\mu$ g/mL of tunicamycin the control construct displayed low levels of background luciferase activity, whereas the plasmid containing the UPR-responsive enhancer (pUPR-luciferase) achieved induction levels for luciferase of greater than 20 fold that observed in the control (Figure 4). Consistent with these results, *de novo* expression of ATF6 was also found to increase by Western blot (data not shown).



*Figure 4. UPR-responsive promoter is activated by increased concentrations (0-50  $\mu$ g/mL) of ER stressor tunicamycin in HEK293 cells. Relative luciferase activity (fold) for HEK293 cells transfected with a plasmid without the UPR responsive enhancer (control, mock), and HEK293 cells transfected with a plasmid with the UPR responsive enhancer (pUPR-luciferase) after exposure to increased concentrations (0-50  $\mu$ g/mL) of tunicamycin.*

Plasmid constructs containing the dGFP reporter gene under control of the UPR-inducible promoter were transfected into both antibody expressing (rCHO) and non-antibody expressing (CHO) CHO host cell lines to characterize the UPR induction levels in real time live CHO cell cultures. Stable cell lines were generated through drug selection. The cells were then subjected to a fed-batch (FB) production culture, a batch process that relies on feeding a cell culture a growth limiting nutrient substrate. UPR induced dGFP expression levels were measured by fluorescence-activated cell sorting (FACS). In CHO host cells a comparison of the induction of dGFP expression in controls without the UPR-inducible promoter (grey peak, Figure 5a) versus passaging cultures periodically diluted with fresh medium (orange peaks, Figure 5a) revealed UPR induction was minimal in the latter. On the other hand, UPR induction was higher in antibody (mAb) expressing rCHO cells in production culture versus the rCHO cells in passaging culture, which experiences little or no mAb expression (Figure 5b). These results suggested that recombinant antibody (mAb) expression activated the endogenous UPR pathway. In addition, the induction levels of dGFP were significantly higher in continued 10-day fed-batch production cultures (blue peak, Figure 5) in both CHO host cells and in rCHO cells (Figure 5). This increased UPR induction in fed-batch culture suggested that endogenous UPR is activated not only by recombinant protein expression, but also as a result of the conditions of the production process itself.

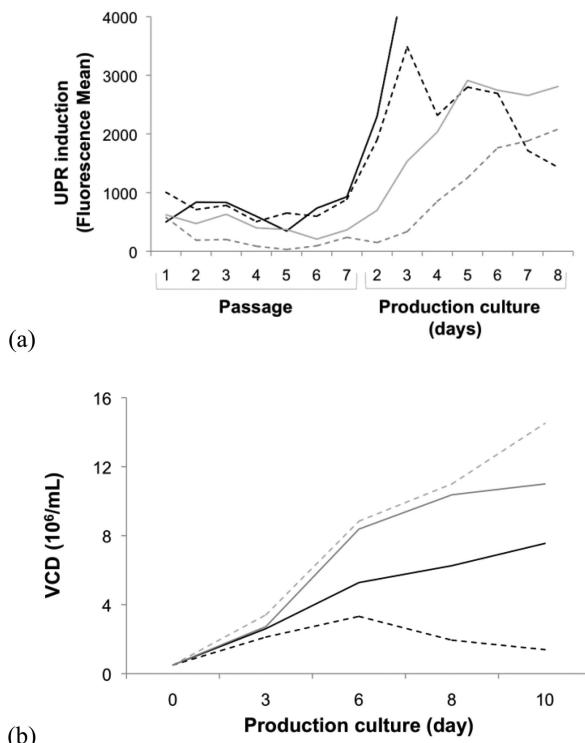


*Figure 5. FACS analysis of dGFP as a proxy for UPR upregulation in CHO production culture from Day 3 (D3) to Day 8 (D8). (a) FACS analysis of CHO culture without recombinant protein expression and (b) FACS analysis of CHO culture with recombinant protein expression (rCHO). Legend: control, grey; passaging culture, orange; 10-day fed-batch production cultures, blue.*

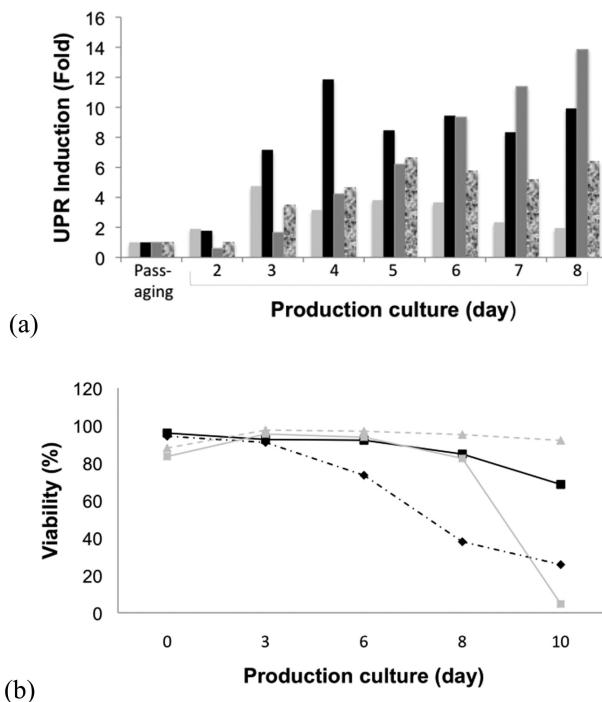
### 3.3. Using Fluorescence Reporters To Monitor Endogenous UPR Activation during Production Process

The output of dGFP fluorescence for four recombinant CHO cell lines, each encoding the UPR reporter system, over the entire production period was monitored to determine the dynamic changes in endogenous UPR activation levels in real time (Figure 6). The results confirm that UPR induction levels were relatively low during passage culture and also the early days of the production culture, but increased markedly on day 3 of production and reached a peak on day 6 to day 8 for the recombinant mAb producing rCHO clones (Figure 6a). Interestingly, different recombinant CHO cell clones displayed different patterns or levels of their respective UPR response, such as timing, duration, and relative

fold induction (Figure 6). As these parameters directly reflect the endogenous UPR and protein folding capacity, the differences in UPR induction levels for different CHO cell lines suggest that the UPR response is highly associated with cellular performance during the mAb production process. For example, the cell growth (viable cell density) for clones 1 and 2 were significantly impacted (Figure 6b) by their early and rapid UPR induction response (day 1, Figure 6a). Comparatively, the viable cell densities (VCD) for clones 3 and 4 (Figure 6b) are consistently higher throughout the production process, consistent with their late UPR induction (day 3, Figure 6a). These results suggested that UPR-induced transcriptional activation is either simultaneous with or directly involved in cell cycle arrest (recall that *perk* is a target gene of UPR activation).



*Figure 6. Kinetics of UPR activation is linked to production culture performance for four recombinant CHO cell clones. (a) Fluorescence mean as a proxy for UPR induction as a function of passage number and production culture day. (b) Viable cell densities for production culture from days 0-10. Legend: rCHO clone 1, dashed black line; rCHO clone 2, solid black line; rCHO clone 3, dashed grey line; rCHO clone 4, solid grey line.*

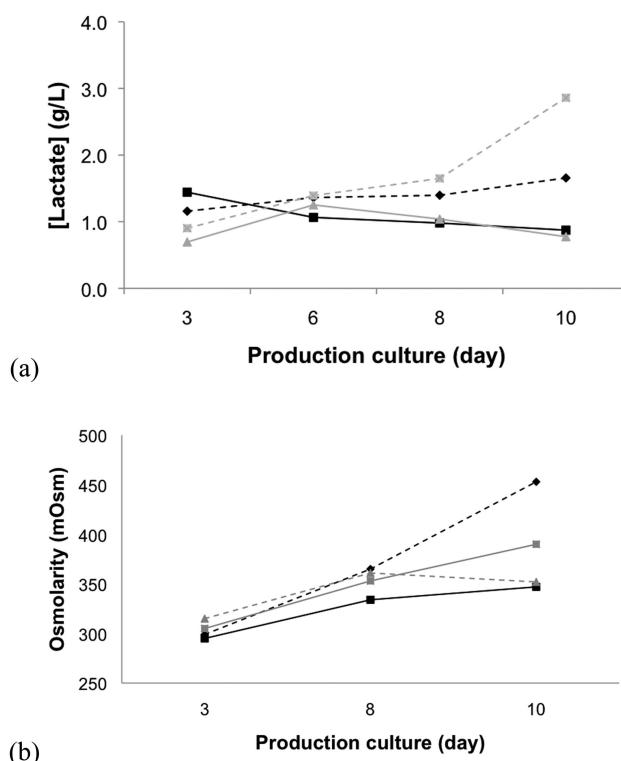


*Figure 7. Kinetics of UPR activation and production culture viability. (a) Real time production culture analysis for UPR induction levels in rCHOs from initial passaging to production days 2-8, rCHO clones 1- 4. Legend: rCHO clone 1, light grey bars; rCHO clone 2, black bars; rCHO clone 3, textured black and grey bars; rCHO clone 4, dark grey bars. (b) Real time rCHO clone cell viability in production culture from days 0-10 for rCHO clones 1-4. Legend: rCHO clone 1, dashed black line; rCHO clone 2, solid black line; rCHO clone 3, dashed grey line; rCHO clone 4, solid grey.*

The fold of UPR induction, through the expression of the UPR chaperons and regulators, represents the capability of the cell to rescue itself from ER stress and is thus likely to be directly related to cell viability (% viability) and cell productivity (titer, g/L). As shown in Figure 7, clones 2 and 4 display enhanced UPR induction, a result that is paralleled by improved productivities (as determined by mAb titer in g/L, Table 1) and viabilities (% viability, Figure 7b), suggesting that these clones contain a stronger UPR capacity. The rCHO cell clones with an enhanced UPR capacity (clones 2 and 4) also generated overall lower levels of lactate (Figure 8a) and lower osmolarities by the end of the production culture (Figure 8b). The results hinted at the possibility of crosstalk between the UPR and other metabolic pathways.

**Table 1. Corresponding mAb production titers for mAbs produced in rCHO clones 1-4 outlined in Figure 6-7**

<i>rCHO Clone</i>	<i>mAb Production Titer (g/L)</i>
1	1.2
2	2.1
3	1.7
4	2.4



*Figure 8. Recombinant CHO production cultures metabolite profiles for rCHO clones 1-4. (a) Lactate levels (g/L) for the production culture (day). (b) Cell culture osmolarity (mOsm) for the production culture (day). Legend: rCHO clone 1, dashed black line; rCHO clone 2, solid black line; rCHO clone 3, dashed grey line; rCHO clone 4, solid grey line.*

### 3.4. Defining the Association between UPR Activation and Processing Conditions

As demonstrated previously (Figure 5), CHO host cell lines not expressing recombinant proteins displayed elevated UPR activation during production, suggesting that the fed-batch production condition itself is able to cause a certain level of ER stress. It is possible that the accumulation of metabolic end-products that result from rich production base and feed media may negatively impact cell growth and productivity.

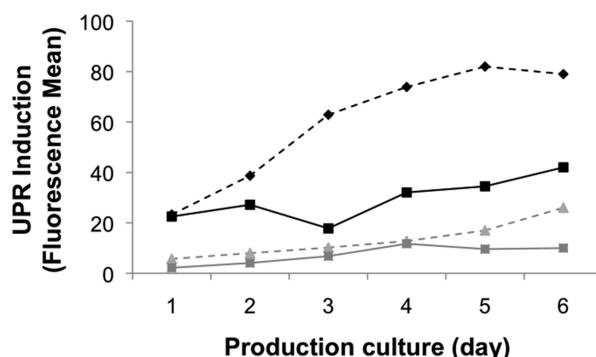


Figure 9. Real time fluorescence mean data (as a proxy for UPR induction levels) for rCHO cultures (5 and 6) allowed to grow under fed batch (FB) conditions (dashed lines) or daily media exchanged (solid lines) culture conditions for an entire production culture period (6 days). Legend: rCHO clone 5, black lines; rCHO clone 6, grey lines.

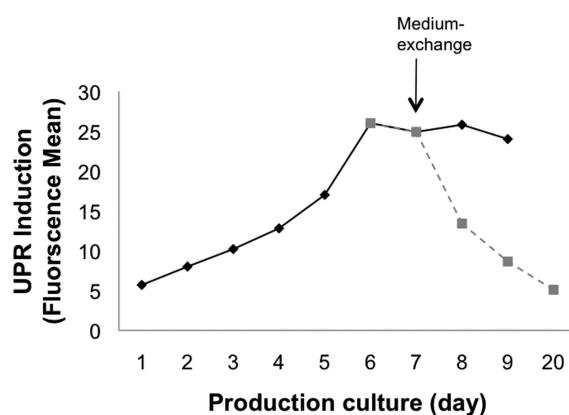


Figure 10. Impact of ME on UPR induction under FB media conditions. UPR induction levels in rCHO production culture from day 1-20. Legend: black line, control (no ME); grey line, ME batch.

To assess whether this fed-batch culture-mediated ER stress was the result of chemical stress from the accumulated metabolic end-products, the UPR induction levels in production cultures with or without daily media exchange were compared (Figure 9). The results reveal that cells cultured with daily media exchange (ME) displayed consistently lower UPR levels (Figure 9), suggesting that the accumulation of one or more metabolic end-products from fed-batch production culture caused ER stress.

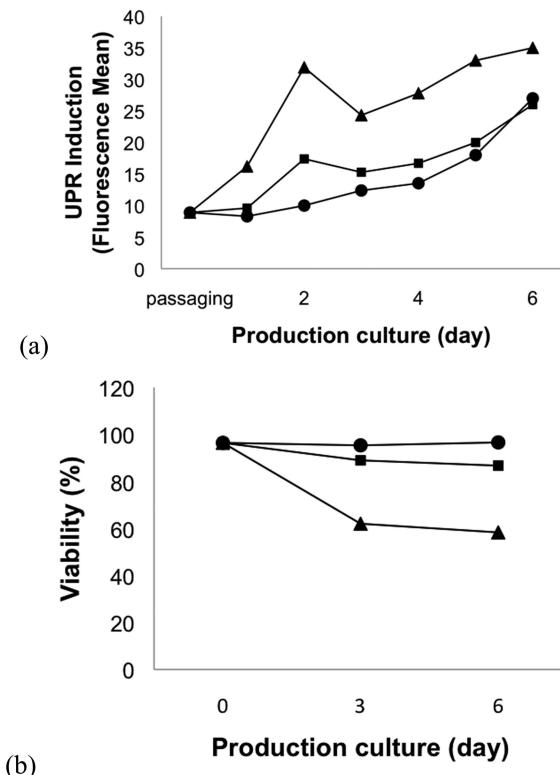


Figure 11. The impact of osmolarity on the UPR in rCHO cells. (a) UPR induction over the production culture period at 310 mOsm, 400 mOsm, and 500 mOsm media conditions. (b) Cell viability over the production culture period at 310 mOsm, 400 mOsm and 500 mOsm media conditions. Legend: 310 mOsm, ●; 400 mOsm, ■; 500 mOsm, ▲.

### 3.5. The Impact of Alleviating ER Stress Associated with Metabolite End Product Production

ER stress that results from the accumulation of metabolic end products can be alleviated by removal of the end products. To achieve this, a fraction of the fed-batch CHO production culture was exchanged for a fresh fed-batch culture media

(media exchange, ME) at day 7. The results revealed that cells treated with ME displayed significantly reduced UPR activation levels within 24 hours compared to controls (Figure 10a). The reduced UPR induction also correlated with an increase in cell viability and growth (data not shown).

Taken together the results indicated that the accumulated metabolic end products produced from cells during fed-batch culture caused ER stress. The results also suggest that to a certain point UPR activation could be dynamically regulated.

### 3.6. The Impact of Osmolarity on the UPR Pathway

Osmolarity increases during CHO cell production, and therefore is another form of chemical stress. Osmolarity increases either as a result of intervention (e.g., base addition), by the addition of compounds as feed, or because metabolites released from the cell accumulate in the culture supernatant. To determine whether increased osmolarity induces a UPR activation during CHO production, CHO cells were directly cultured in production media under hyperosmotic conditions (Figure 11).

It was found that hyperosmotic conditions by adding sorbitol into the production medium (400 and 500 mOsm) showed a more significant increase UPR induction than normal conditions (310 mOsm) (Figure 11a). Subsequently, UPR activity associated with hyperosmolarity was correlated with immediate cell cycle arrest (data not shown) with concomitant decrease in cell viability (Figure 11b).

## 4. Conclusion

In conclusion a UPR reporter system was constructed and used to detect and quantify the UPR response in real time live cells as a result of different kinds of stress signals, such as chemical stress, nutrient limitation (data not shown), and hyperosmolarity. Using this monitoring system, it was found that UPR activation is dynamic with the timing and fold of UPR induction directly correlating to the cell culture conditions. As the kinetics of UPR vary for different clones, monitoring levels of UPR induction can also provide information during cell line development, which might ultimately influence clone selection.

This UPR monitoring system has many potential applications. It can be used to monitor environmental changes in cell culture, such as raw materials or microbial/viral contaminations that result in changes in productivity and cell viability. This UPR monitoring system also allows for rapid detection of changes in ER stress that may be otherwise difficult to monitor by examining growth and viability alone. This UPR monitoring system also has the potential to be used to study signaling crosstalk between UPR and other cellular pathways, thereby identify potential routes to enhance the yield of therapeutic proteins generated in mammalian cell cultures. Lastly, this monitoring system could provide a means to screen potential therapeutic drugs inhibiting the UPR in the treatment of cancer (10).

## Acknowledgments

We thank Heather Nunn and Mirna Mujacic for technical help with clone imaging.

## References

1. Schroder, M.; Kaufman, R. J. The mammalian unfolded protein response. *Annu. Rev. Biochem.* **2005**, *74*, 739–789.
2. Tigges, M.; Fussenegger, M. Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. *Metab. Eng.* **2006**, *8*, 264–272.
3. Yoshida, H.; Matsui, T.; Yamamoto, A.; Okada, T.; Mori, K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* **2001**, *107*, 881–891.
4. Bommiasamy, H.; Back, S. H.; Fagone, P.; Lee, K.; Meshinchi, S.; Vink, E.; Sriburi, R.; Frank, M.; Jackowski, S.; Kaufman, R. J.; Brewer, J. W. ATF6alpha induces XBP1-independent expansion of the endoplasmic reticulum. *J. Cell Sci.* **2009**, *122*, 1626–1636.
5. Haze, K.; Yoshida, H.; Yanagi, H.; Yura, T.; Mori, K. Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. *Mol. Biol. Cell* **1999**, *10*, 3787–3799.
6. Ye, J.; Rawson, R. B.; Komuro, R.; Chen, X.; Dave, U. P.; Prywes, R.; Brown, M. S.; Goldstein, J. L. ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Mol. Cell* **2000**, *6*, 1355–1364.
7. Okada, T.; Yoshida, H.; Akazawa, R.; Negishi, M.; Mori, K. Distinct roles of activating transcription factor 6 (ATF6) and double-stranded RNA-activated protein kinase-like endoplasmic reticulum kinase (PERK) in transcription during the mammalian unfolded protein response. *Biochem. J.* **2002**, *366*, 585–594.
8. Schroder, M.; Kaufman, R. J. ER stress and the unfolded protein response. *Mutat. Res.* **2005**, *569*, 29–63.
9. Kuo, S. C.; Lampen, J. O. Tunicamycin--an inhibitor of yeast glycoprotein synthesis. *Biochem. Biophys. Res. Commun.* **1974**, *58*, 287–295.
10. Li, X.; Zhang, K.; Li, Z. Unfolded protein response in cancer: the physician's perspective. *J. Hematol. Oncol.* **2011**, *4*, 8.

## Chapter 3

# Case Studies in the Application of Aqueous Two-Phase Processes for the Recovery of High Value Biological Products

**Karla Mayolo-Deloisa, José González-Valdez,  
Celeste Ibarra-Herrera, Mirna González-González,  
Carolina García-Salinas, Oscar Aguilar, Jorge Benavides,  
and Marco Rito-Palomares\***

**Centro de Biotecnología-FEMSA, Departamento de Biotecnología e  
Ingeniería de Alimentos, Tecnológico de Monterrey,  
Ave Eugenio Garza Sada 2501-Sur, Monterrey, NL 64849, México  
\*E-mail: mrito@itesm.mx. Phone: +52 (81) 83284132.  
Fax: +52 (81) 83581400**

Mammalian, microbial (bacterial and yeast), and plant cells are effective expression systems used commercially to produce mass quantities of biological, pharmaceutical, or chemical products of interest. However, there is the need to establish selective and scalable methods of product recovery that integrate effectively with upstream cell cultures to rapidly yield products in a state suitable for validation operations. The current state of the art purification and recovery methods utilize well-established multi-step processes (e.g. product release, solid-liquid, concentration and chromatography steps) that usually result in low yield and high process cost. In this key note address a series of case studies were presented where an aqueous two-phase system (ATPS) extraction, an existing bioengineering strategy, was used to alleviate many of the previously mentioned existing process constraints. Specific case studies utilizing ATPS were presented for the purification of Rotavirus-like particles from insect cells, colorant proteins (i.e. C-Phycocyanin and B-Phycoerythrin) from microbial origin, human granulocyte-colony stimulating factor (hG-CSF)

from alfalfa, fractionation of PEGylated proteins, and stem cells. Conclusions will be drawn concerning the use of ATPS in downstream processes that can greatly simplify the current way in which bioproducts are recovered.

## 1. Introduction

The recovery and purification of bioproducts (antibody, protein, chemical compound, etc...) produced in a host system requires separation of the product of interest from components of the host system, cellular debris, and other contaminants after production. In general a bioprocess involves two major steps, production (fermentation) and bioseparation (primary recovery and purification). The process route in the bioseparation step depends upon the nature of the product of interest (extracellular or intracellular). Extracellular products are located in the fermentation media culture and processed in a multi-step purification process that includes a primary isolation procedure (usually involving solid-liquid separation, product concentration and major contaminant removal operations), a high-resolution purification procedure (usually involving chromatographic steps), and a final polishing step (usually a drying step). Intracellular components are liberated from the cell after cell lysis, and the debris is removed, the remaining intracellular components are concentrated and fed through the mentioned purification processes (Figure 1).

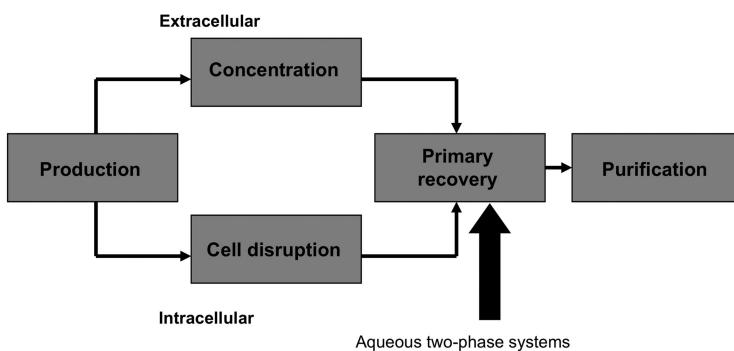
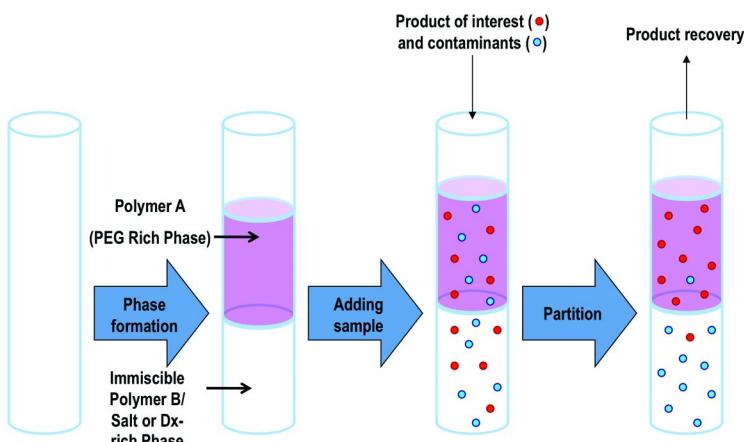


Figure 1. Schematic summary of the product production process with aqueous two-phase system (ATPS) added in primary recovery process.

In an effort to purify the target product more efficiently and effectively the primary recovery of the product was achieved using an aqueous two-phase system (ATPS) (Figure 1). ATPS can be exploited to separate the product of interest from major contaminants. ATPS is a well-known liquid-liquid fractionation technique that exploits the difference in solubility (or affinity) of the product of interest and the contaminant in two different immiscible aqueous phases (1). The two

immiscible aqueous phases are generated either by mixing two water-soluble immiscible polymers (e.g. polyethylene glycol (PEG) and dextran) or a single water-soluble polymer and a buffer (e.g. PEG and phosphate). As most biological products are water soluble, this allows a platform for extraction in an aqueous environment. Inserting ATPS as a unit operation integrates and intensifies a primary recovery process in an easily scaled-up, economical process that can be used in continuous or in batch mode. Integration refers to the combination of two or more unit operations (processing stage) into one that achieves the same goal. Intensification means more material can be loaded without altering already established process capability. ATPS can be used on pilot or industrial scale rather economically.



*Figure 2. Cartoon view of a simplified aqueous two-phase (ATP) extraction. Aqueous polymer phase (top phase) and aqueous salt/buffer phase (bottom phase) are immiscible. Exploiting the preferential solubility of the product of interest (red balls) versus contaminants (light blue balls) in opposite phases allows for an efficient extraction and isolation of each component. The affinity of both components in the particular aqueous polymer phase or the buffer phase is based on their partition coefficient,  $K_p$ .*

In ATPS two immiscible aqueous solutions are carefully selected based upon the thermodynamic partition coefficient ( $K_p$ ) of each component in each phase (Figure 2) (1, 2). In a general sense (in the absence of solute-solute or solute-solvent interactions) the  $K_p$  can be defined according to the concentration of a single molecular species in two phases if they are at equilibrium with one another, using equation (1) (1, 2).

$$K_p = C_T / C_B \quad (1)$$

Where  $C_T$  is the concentration of the un-ionized product of interest (or the un-ionized contaminant) in the top phase and  $C_B$  is the concentration of the un-ionized product of interest (or the un-ionized contaminant) in the bottom phase. To concentrate a product of interest (or contaminant) in the top phase its partition coefficient must be greater than 1 ( $K_p > 1$ ) (3). The partition behavior depends not only on the system parameters (polymer molecular weight, tie line length (TLL), volume ratio ( $V_R$ ), pH, sample loading, and temperature), but also the physicochemical properties of the product (or contaminant) such as its molecular weight, isoelectric point (pI) or pKa, and hydrophobicity (1). In an ATP extraction the  $K_p$  of each component can be optimized by manipulating parameters such as pH, concentration of a certain reagent, and ionic strength such that the product of interest can be forced into a phase (in this case the upper aqueous polymer phase) while the contaminants are simultaneously forced into the opposite phase (in this case the aqueous salt/buffer bottom phase, Figure 2). This type of partitioning can be seen graphically as mixture of two phases. The information in this binodal graph is critical when working with ATPS. In this phase diagram two major regions a monophasic region and a biphasic region are generated. The biphasic region is the region of interest for ATPS, with the area above the binodal curve providing conditions that generate two-phase system. Depending on the system parameters the composition of the coexisting phases and a tie line length (TLL) value can be determined. From the graph the tie line (TL) is the thermodynamic equilibrium between the two phases and is related to the concentration of polymer (or salt) forming phases (1). TLL is determined according equation (2) (1). In a particular TLL, different ATPS can be selected, these systems will have the same TLL but different volume ratio ( $V_R$ ). The  $V_R$  is defined according to equation (3):

$$TLL = \sqrt{[(\Delta C_1)^2 + (\Delta C_2)^2]} \quad (2)$$

$$V_R = V_T/V_B \quad (3)$$

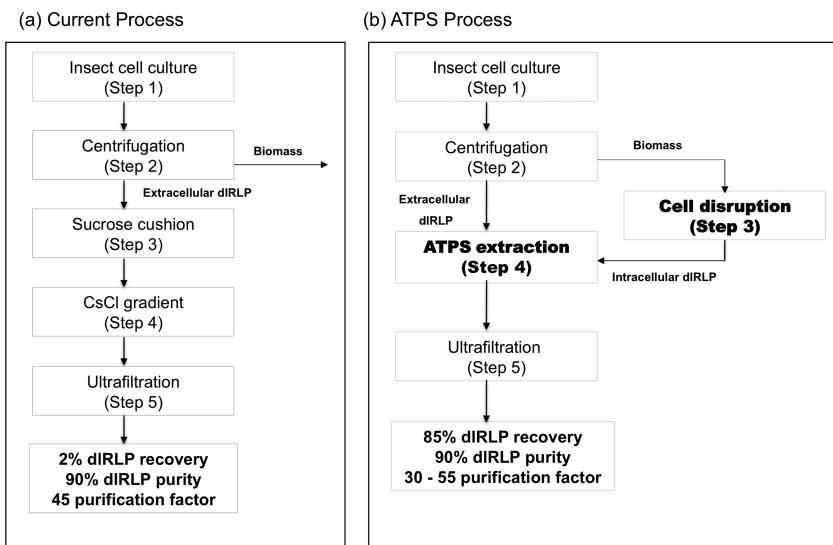
Where  $\Delta C_1^2$  and  $\Delta C_2^2$  are squares of the differences (absolute) in the phase forming constituents  $C_1$  and  $C_2$  in the top and bottom phases and  $V_T$  is the volume of the top phase and  $V_B$  is the volume of bottom phase.

Taken together when optimized ATPS offers a potentially efficient extraction protocol where one can simply select conditions where the target protein is, for example, more soluble in an upper aqueous polymer phase, while the contaminants are soluble in a lower salt/buffer phase. The current presentation presented a summary of case studies that utilize ATPS extraction as an *in process* step in the primary recovery stage to simplify purification of Rotavirus-like particles from insect cells (4), C-Phycocyanin (5) and B-Phycoerythrin (6) from microbes, human granulocyte-colony stimulating factor (hG-CSF) from alfalfa (7) (and additionally touched upon the potential use of combined 2D electrophoresis to identify process parameters for optimum ATP extraction - which will not be discussed here) (8), fractionation of PEGylated proteins (9), and stem cells (10).

## 2. Case Studies: Application of ATPS

### 2.1. ATPS in the Primary Recovery of Double-Layered Rotavirus-Like Particles (dIRLP)

Virus-like particles (VLP) are viral mimics that contain the main structural proteins of a virus of interest, but lack the genetic material to become infectious. Successful bioproduction of VLPs is achieved by expression of recombinant structural proteins without expression of non-structural proteins and genetic material in a host (such as in mammalian, yeast, plant, or insect cells) (11). Among other functions VLPs can be used in the development of vaccines (12) and nanomaterials (13).



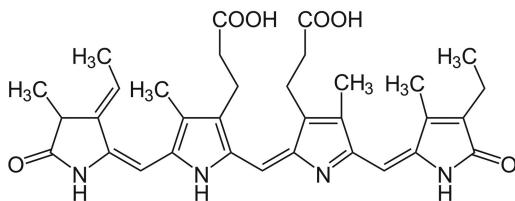
*Figure 3. Outline of the unit operations for the primary recovery of Rotavirus-like particles (dIRLP) from insect cells: (a) current process; (b) ATPS process. Modified from reference (4).*

In this particular project we studied the application of an ATP extraction system for the recovery of a double-layered Rotavirus like particle (dIRLP) that was produced on an 1000 ng/mL scale in the insect cell-baculovirus expression system. Like most VLPs, dIRLPs, are difficult to purify (14). DILRPs structurally contain two protein layers an inner layer composed of VP2 protein and an outer layer composed of VP6 protein. Purification of dIRLP with ATPS was thought to be an attractive alternative purification process to produce a vaccine

against Rotavirus infection, which kills 500,000 children a year as a result of acute gastroenteritis (15). Currently, the protocol in place for the isolation of Rotavirus-like particles is tedious and involves a five-step process (Figure 3, left). These steps include insect cell culture (Step 1), centrifugation (Step 2), a sucrose cushion (Step 3, where particle mixture is pelleted through a pad of sucrose leaving particles intact), a Cesium Chloride (CsCl) gradient (Step 4, a sedimentation method that separates particles by density), and ultrafiltration (Step 5, Figure 3) (4, 7). dIRLP recovered from this process has high product purity (90%), but suffers from a very poor yield (<2%) (4, 7). To improve the process ATPS was added as an *in process* step (Step 4, Figure 3), after centrifugation (Step 2) and cell disruption (Step 3) (4). In this modified process, unlike in previous protocols (4, 7), the intracellular and extracellular components were separated after centrifugation as previous studies revealed that approximately 60% of the total dIRLP from the insect cell culture are contained in the supernatant and the remaining 40% are intracellular, and both were carried through to the end (Step 3-5). The overall process provided dIRLP in a high purity (90%) with a dramatically improved recovery (85%) in comparison to previous non-ATPS methods. The process is not only efficient, but in contrast with the previous method, this prototype process is suitable to be scaled-up (unpublished results).

## 2.2. Recovery of Natural Colorants from Microbial Origin

Not only can ATPS be used to purify virus-like particles, but it can also be applied to the recovery of natural colorants or cyanobacterial bioproduced in microbes (5, 6, 16, 17). In fact C-phycocyanin, a naturally occurring protein-pigment complex was bioproduced in *Spirulina maxima* and isolated in a compact process that included only five non-chromatographic unit operations (5). The chromophore that associates with C-phycocyanin is a phycocyanobilin, this chromophore is present in the light-harvesting phycobiliprotein family that includes phycoerythrin (Figure 4) (18). The isolation of C-phycocyanin process (with its bound chromophore) included a step with ATPS, and provided C-phycocyanin in good yield (28%) and high purity (3.9, ratio of the absorbance intensities at 620nm to 280nm) (5). Commercially C-phycocyanin is valued at \$15,000 per gram of material. Similarly, B-phycoerythrin (BPE, Figure 4) a protein-chromophore covalent complex (also with bound phycocyanobilin, Figure 4) was also isolated using a modified ATPS procedure (6). Like C-phycocyanin, BPE also has various commercial applications as a colorant or marker in the food, cosmetic, pharmaceutical and chemical industries. BPE was recovered and purified from the red microalga *Porphyridium cruentum* after cell disruption using isoelectric precipitation (6). Subsequent polyethylene glycol (PEG) and phosphate buffer mediated ATPS extraction provided the pink colorant in high yield (72%) and purity (4.1) (6). The results were patented and have been scaled to a pilot plant process that is under evaluation and validation. ATPS extraction can be applied for the recovery and purification of many other bioproduced products such as *B*-carotene, lutein, or recombinant proteins (16, 17).



*Figure 4. Phycocyanobilin is the typical chromophore present in the light-harvesting phycobiliprotein family that includes phycoerythrin. The [Phycocyanobilin + C-phycocyanin] complex absorbs and emits light at  $\lambda_{abs}/\lambda_{em} = 620/650$  nm and the [Phycocyanobilin + Phycoerythrin] complex absorbs and emits light at  $\lambda_{abs}/\lambda_{em} = 550/575$  nm (18).*

### 2.3. Experience with ATPS Plant-Based Bioprocess Development

The ATP extraction system is a useful tool to isolate products from not only animal and microbial cells, but also from transgenic plants. Transgenic plants are a cheaper, and thus more attractive, hosts to scale-up the production of bioproducts (<\$100/gram) versus production of similar products in mammalian cell culture (\$10,000/gram) (19). As greater than 90% of the costs in a bioprocess are associated with the down-stream processes (costs associated initial capture and purification), to make plants an even more attractive bioproduction host the cost of downstream processes must also be economical (7, 20). The use of plants as a bioproduct host is not a novel idea; transgenic plants have been used as platform for the production of biopharmaceuticals from tobacco, maize, soy beans, and recently alfalfa (7, 20). Unfortunately, the bioproduction in plants does have its limitations. These limitations include a low yield of the target product, inconsistency in product quality, and a production system with a large amount of contaminant (7, 20). With these limitations in mind, and the added goal to limit costs, we sought to apply our ATPS in the purification and recovery of a bioproduct from a plant.

As a model system human granulocyte-colony stimulating factor (hG-CSF), a glycoprotein, was produced in alfalfa (7). Hg-CSF has been expressed in *Escherichia coli*, common yeasts, mammalian cells, and in tomato and tobacco (7). HG-CSF was chosen, as it is an important glycoprotein used in the treatment of neutropenia in cancer therapy, in bone marrow transplants, and in HIV-associated neutrophil defects (21). Unfortunately, hG-CSF treatment is expensive. The average cost per milligram is \$800 and a single dose costs \$250 (21). To lower these costs and to provide an attractive alternative production process, we investigated the production of hG-CSF in plants and utilized our ATPS technology during the recovery and purification process (7, 20).

Again the overall process strategy was simple, with three major steps in the isolation of hG-CSF from alfalfa. The first step was to understand and characterize the behavior (i.e. partitioning in a particular aqueous solution) of the major contaminants from the plants using ATPS. The second step was to identify

and characterize the partition behavior of the model protein using ATPS. The last step was to identify and characterize the aqueous partitioning behavior of the mixture using the optimize ATP phase conditions.

**Table 1. Summary of ATP extraction conditions screened for the recovery of host cell contaminants from the bioproduction of hG-CSF in alfalfa (7)**

Entry	PEG	%TLL (p/p)	$K_P$	% Recov.	% Recov.
	Size			Top Phase	Bottom Phase
1	600	32	ND	71	0
2	600	37	62.7	74	4
3	600	42	16.6	75	7
4	600	45	7.2	69	19
5	1450	27	5.0	54	14
6	1450	34	3.2	50	18
7	1450	42	3.0	44	17
8	1450	48	4.9	43	18
9	3350	42	1.7	19	17
10	3350	46	0.8	14	17
11	3350	51	1.1	21	18
12	3350	56	1.8	16	22
13	8000	21	ND	0	47
14	8000	36	0.1	7	65
15	8000	43	0.6	2	28
16	8000	48	0.6	18	23

Extraction of hG-CSF from its alfalfa host contaminants using an ATPS was achieved, but was highly dependent on the size of the PEG polymer used to generate the aqueous solution. For purposes of this discussion the term molecular weight (MW) refers to the size of the PEG polymer added to an aqueous solution to generate different immiscible aqueous polymer solutions. Four PEG sizes or MWs were investigated PEG600, PEG1450, PEG3350, and PEG8000. In addition to investigating different sizes of the polymer added into the aqueous phase, the concentration of the polymer was also increased, as indicated with the increasing %TLL (Table 1). Following the previous outlined steps we first concentrated the contaminant in the top phase using PEG600 and PEG1450, with a recovery ranging from 69% - 71% for the former and 43% - 55% for the latter. The contaminant could also be concentrated in bottom phase using PEG8000,

all be it with a significantly wider range in recovery 23% - 65%, and suboptimal  $K_p$  ( $K_p < 1$ ). The results give two options for how to force the contaminants in one phase (top or bottom), with Entries 1, 6 and 14 (Table 1) standing out as the highest recovery extraction conditions. Moving onto step 2 we were able to force the recombinant hG-CSF (rhG-CSF) into the top phase of PEG600, PEG1450, or PEG8000 ATPS-phosphate at pH of 7 and  $V_R = 1$  with a percent recovery of 72%, 98%, and 78%, respectively (Table 2). Quantification of protein was determined using a standard Bradford protein assay (22) and confirmed by SDS-PAGE electrophoresis (7). Based upon the combined results of step 1 and step 2, PEG8000 with a %TLL of 35 (Entry 14, Table 1 and Entry 3, Table 2) presented the optimum profile where the plant contaminants were forced to reside in the bottom phase (65% in the model) while a majority of the rhG-CSF was forced to reside in the top phase (78% in the model) (7). In the step 3, the behavior of the mixture (production contaminants + hg-CSF) was characterized using the optimized ATPS (PEG8000, %TLL 35,  $V_R = 1$ , pH 7, PEG 8000-Phosphate salts). The overall process was successful and rhG-CSF was extracted in the top phase from contaminants (7). SDS-PAGE of the two phases (layers) in comparison to contaminants confirmed this result (7).

**Table 2. Summary of ATP extraction conditions screened for the recovery of hG-CSF from host cell contaminants after its bioproduction in alfalfa (7).**  
**Not determined = ND,  $K_p$  not determined**

Entry	PEG	%TLL	%PEG	% $K_2PO_4$	% Recov.	% Recov.
	Size	(p/p)	(p/p)	(p/p)	Top Phase	Bottom Phase
1	600	32	15	18	72	ND
2	1450	34	16	14	99	ND
3	8000	36	16	11	78	ND

## 2.4. Separation of PEGylated Therapeutic Proteins Using ATPS

PEGylation is the process of covalently attaching a polyethylene glycol (PEG,  $H(OCH_2CH_2)_nOR$ ) group onto the surface of a drug molecule or protein (where R = the drug molecule or protein, Figure 5) (9). The covalent attachment is formed after the molecule of interest is treated with a methoxy-PEG (mPEG) that has only one reactive hydroxyl group (which avoids the formation of crossed linked products) and has been approved for use in pharmaceutical preparations (23). PEGylation of a drug molecule or protein can hide it from degradation enzymes, reduce immunogenicity, enhance physical and thermal stability, increase solubility or *in vivo* circulation time (24). PEGylated proteins have been approved by the US Food and Drug Administration (FDA) (24).

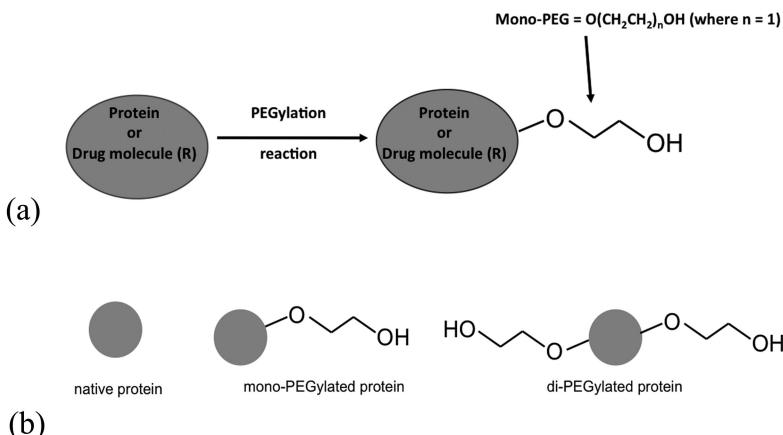


Figure 5. Cartoon view of PEGylation reaction (a) and its products (b).

In this case study ribonuclease A and  $\alpha$ -lactalbumin were PEGylated (9). Unfortunately, because a protein possess multiple reactive sites a population of over PEGylated conjugate species, such as mixture of native protein (non-PEGylated), mono- and di-PEGylated (Figure 5) that exhibit differences in biological activity were produced in the PEGylation reaction (24). For our purposes the mono-PEGylated proteins were of interest, thus we sought to apply our ATPS technology to isolate the desired mono-PEGylated proteins. The challenge was to recover and purify these products from other reaction products and then to sub-fraction the different PEGylated compounds. The separation of the mono- and di-PEGylated compounds was achieved chromatographically, in collaboration with researchers as Carnegie Mellon University (25).

The ATP partitioning behavior of the various PEGylated conjugates of ribonuclease A were investigated using four different aqueous PEG polymer sizes (PEG400, PEG1000, PEG3350 and PEG8000) at different TLLs (15, 25, 35, and 45 % w/w) and the results were plotted in a graph of  $\ln K_p$  as a function of %TLL (% w/w, Figure 6). The results clearly demonstrate that the various PEGylated ribonuclease A conjugates can be separated from the native protein using PEG8000, as demonstrated by their sub-fractionation in opposite phases at a %TLL of 35 (Figure 6). Similarly, sub-fractionation of native  $\alpha$ -lactalbumin from its PEGylated conjugates was also possible, although to a lesser degree than with ribonuclease A, with aqueous PEG8000 using an ATPS as demonstrated by their sub-fractionation in opposite phases at a %TLL of 35. An interesting observation in these experiments was that the native proteins tend to concentrate in bottom phase of the ATPS and the PEGylated compounds in the upper phase, under the conditions investigated. Unfortunately, separation of the mono- from the di-PEGylated ribonuclease and  $\alpha$ -lactalbumin has yet to be realized and is still ongoing. However, we present the isolated yields for the various mixtures

of the PEGylated conjugates of ribonuclease A and  $\alpha$ -lactalbumin versus their native proteins isolated from these ATP extraction experiments are shown in Figure 7. The results do demonstrate that ATP extractions also have the potential to separate native proteins from the PEGylated analogues (9).

## 2.5. Aqueous Two-Phase System Bioengineering Strategies for the Potential Recovery and Purification of Stem Cells

Recent efforts in our group have extended the application of ATPS technology into an entirely new area, the recovery and purification of live and viable cells, specifically stem cells (26). Currently, the isolation process is performed on bench scale, and suffers from many limitations (26, 27). Our goal in this area is to develop a scalable, rapid, economic and novel bioengineering strategy for the recovery and purification of stem cells exploiting non-conventional technologies, such as ATPS, that will allow for manipulation of high quantities of sample, reduce losses and processing times. Two strategies are currently being explored that include the use of a density gradient ATPS (Strategy 1) or the use of immunoaffinity ATPS, which includes free antibody or PEGylated antibody binding or immobilized antibody labeled microbeads (Strategy 2).

Efforts have focused on the isolation of CD133<sup>+</sup> stem cells from human umbilical chord blood using first a well known procedure, lymphoprep (28). CD133<sup>+</sup> are implicated in the treatment of many degenerative and chronic diseases, and thus a scalable isolation process is necessary. After isolated by lymphoprep the behavior of the stem cells was analyzed by flow cytometry and 7AAD. Flow cytometry and 7AAD steps are essential to ascertain cell viability, as all isolated stem cells must be alive and viable to be of any therapeutic use (29).

Efforts in the arena are ongoing and thus in progress. Briefly, novel ATPS exploiting the binding of antibodies and/or PEGylated antibodies to stem cells (Strategy 2) is the most promising way to direct (or force) stem cells into a particular phase (Figure 8). PEGylation of an antibody (that selectively binds to CD133<sup>+</sup> stem cells) forms a tightly associated complex with the stem cell and thus may influence the partition coefficient for complex in a particular phase of an ATPS (Figure 8). Recently, we disclosed that PEGylation of Biotin (H<sub>2</sub>NPEG-●-, where ● = Biotin) was achieved (30). Further, the PEGylate Biotin could bind to Streptavidin (◆) and to another Biotin-CD133<sup>+</sup> cell (● CD133) forming a complex of all three components (H<sub>2</sub>NPEG(-●-◆-CD133, Figure 8) (30). Whether or not the PEGylated complex to influences or enhances solubility is as yet unclear.

In last part of Strategy 2 combining this immunoaffinity guided ATP extraction process onto solid support, with immobilized affinity labeled microbeads may provide an alternate strategy for the selective isolation of stem cells from contaminants in two-phase system. Finally there is a recent review covers all uses of ATP technology in the purification of bioproduced products (26).

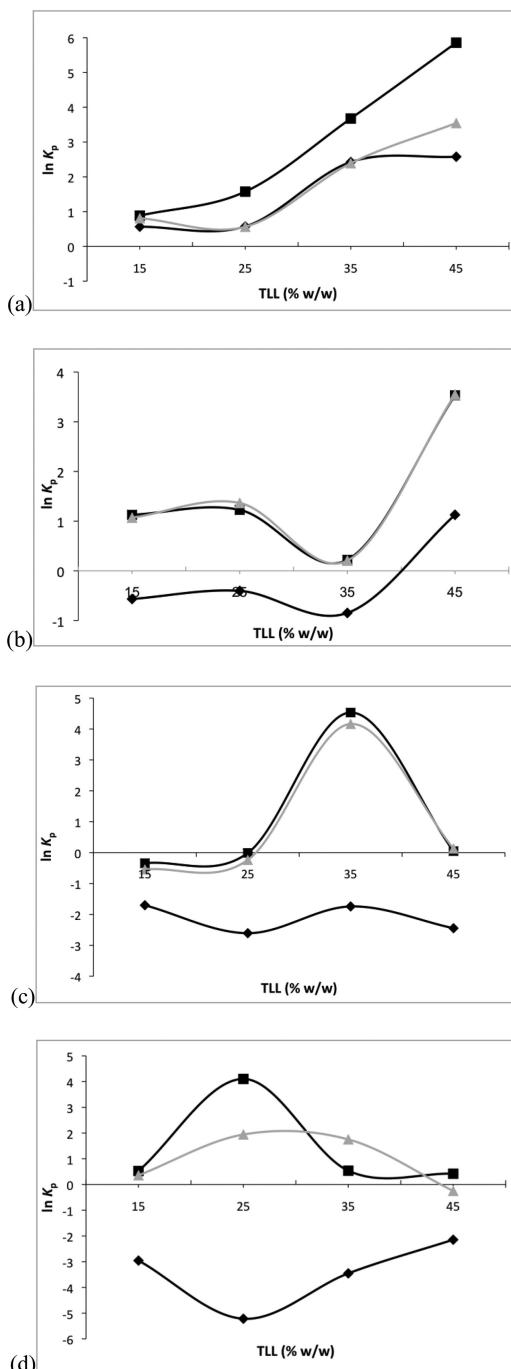


Figure 6. Plots of  $\ln K_p$  versus %TLL (p/p) for the ATP extraction of ribonuclease A. (a) PEG400 (top phase)-phosphate (bottom phase); (b) PEG1000 (top phase)-phosphate (bottom phase); (c) PEG3350 (top phase)-phosphate (bottom phase); (d) PEG8000 (top phase)-phosphate (bottom phase).  $K_p$  = partition coefficient and TLL = tie line length. Legend: ♦ = native ribonuclease A, ▀ = mono-PEGylated ribonuclease A, ▲ = di-PEGylated ribonuclease A (9).

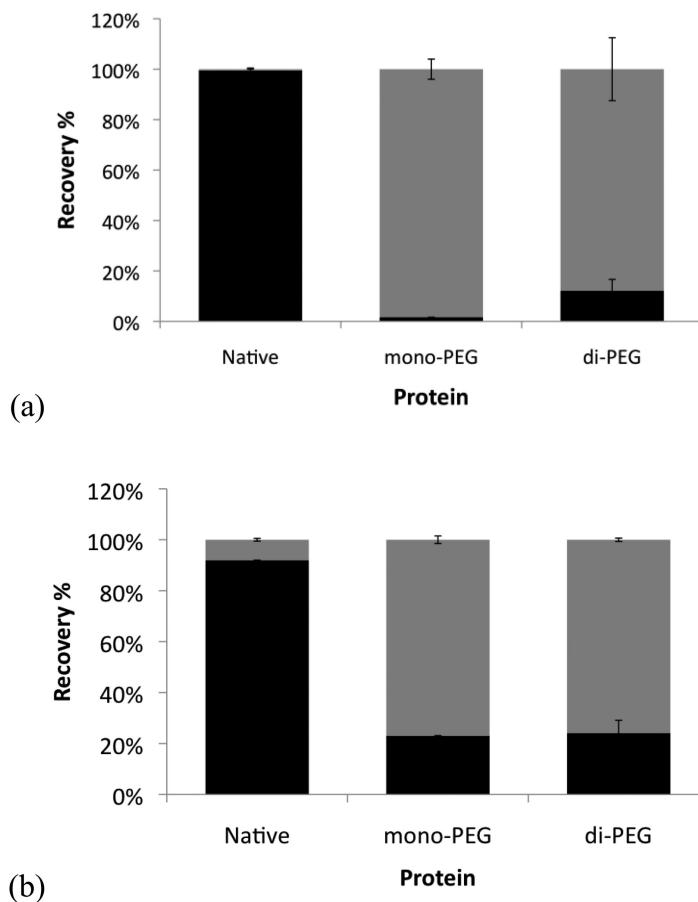
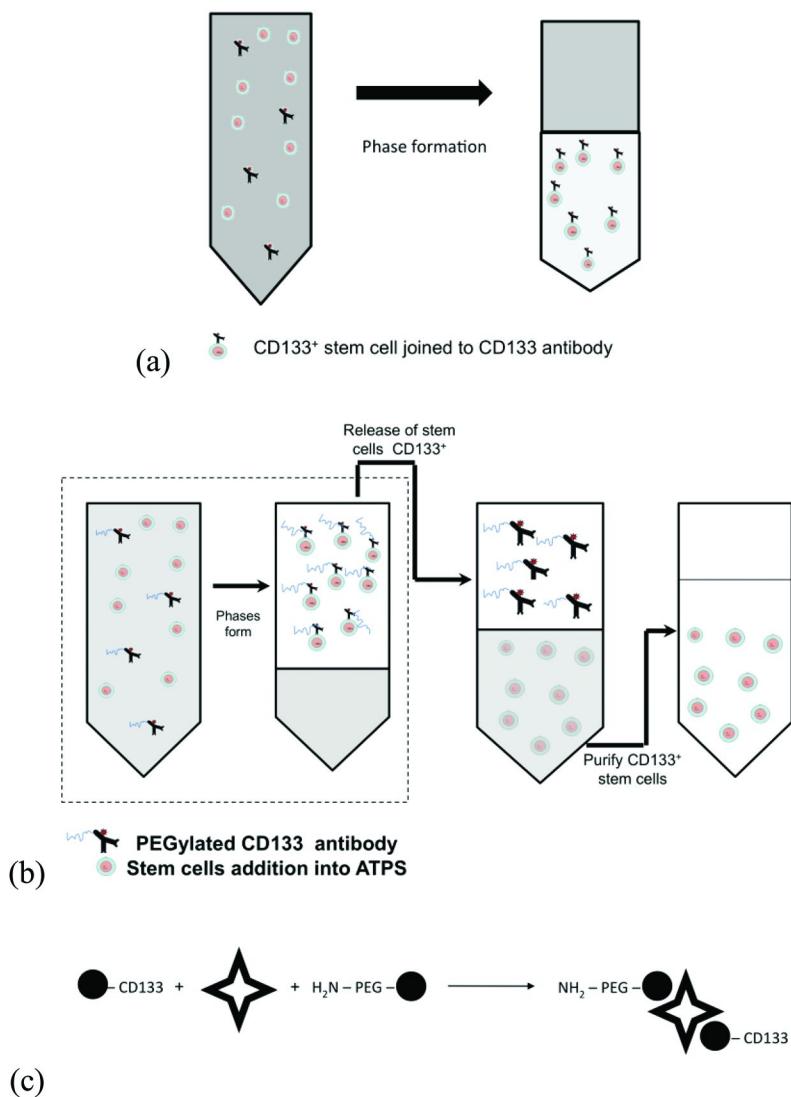


Figure 7. Recovered yields for the purification of ribonuclease A and  $\alpha$ -Lactalbumin using an ATPS of PEG8000-phosphate. (a) mono- and di-PEGylated ribonuclease A from native protein (b) and mono- and di-PEGylated  $\alpha$ -lactalbumin (9).



*Figure 8. Schematic look at the methods for isolating CD133<sup>+</sup> stem cells using ATPS according to Strategy 2. (a) Immunoaffinity based (CD133 antibody-stem cell) partitioning in ATPS; (b) PEGylated antibody ATPS based; and (c) Cartoon representation of the association of PEGylated Biotin ( $\text{H}_2\text{N-PEG-Biotin}$ ), Biotin ( $\bullet$ ) = Biotin) to Streptavidin ( $\blacklozenge$ ) and to another Biotin-CD133<sup>+</sup> cell ( $\bullet-CD133$ ) forming a complex of all three components (30).*

### 3. Conclusions

One of the major outcomes of all the strategies presented is that the product and contaminant partition characterization in an ATPS is key to proposing initial recovery conditions. An improved understanding of what drives ATP processes have resulted in the rapid establishment of strategies for the recovery of biological products. Further ATPS strategies are easy, scalable, and economical and as such are suitable vessels to help biotechnology engineers face the new challenges in downstream processing. ATPS can contribute with the existing techniques to address current and new problems of the biotechnology industry. However, challenges such as handling high numbers of cells or high products concentration still needs to be addressed (Product >10g/L, Cell concentration > 40% w/w).

### Acknowledgments

The authors wish to acknowledge the financial support of Tecnológico de Monterrey, Bioprocess research chair (Grant CAT161), the Mexican Research Council (CONACyT) and the Zambrano-Hellion Foundation.

### References

1. (a) Benavides, J.; Rito-Palomares, M.; Asenjo, J. A. Aqueous Two-Phase Systems. In *Comprehensive Biotechnology*; Moo-Young, M., Ed.; Academic Press: Burlington, MA, 2011; 2.49, pp 697–713. (b) Vázquez-Villegas, P.; Aguilar, O.; Rito-Palomares, M. Study of biomolecules partition coefficients on a novel continuous separator using polymer-salt aqueous two-phase systems. *Sep. Purif. Technol.* **2011**, *78*, 69–75.
2. (a) W. Nernst. Verteilung eines Stoffes zwischen zwei Lösungsmittel und Damfraum. *Z. Phys. Chem.* **1891**, *8*, 110–139. (b) Leo, A.; Hansch, H.; Elkins, D. Partition coefficients and their uses. *Chem. Rev.* **1971**, *71*, 525–616.
3. Rito-Palomares, M. Practical application of aqueous two-phase partition to process development for the recovery of biological products. *J. Chromatogr., B* **2004**, *807*, 3–11.
4. Benavides, J.; Mena, J. A.; Cisneros-Ruiz, M.; Ramírez, O. T.; Palomares, L. A.; Rito-Palomares, M. Rotavirus-like particles primary recovery from insect cells in aqueous two-phase systems. *J. Chromatogr., B* **2006**, *842*, 48–57.
5. Rito-Palomares, M.; Nuñez, L.; Amador, D. Practical application of aqueous two-phase systems for the development of a prototype process for C-phycocyanin recovery from *Spirulina maxima*. *J. Chem. Technol. Biotechnol.* **2001**, *76*, 1273–1280.
6. (a) Benavides, J.; Rito-Palomares, M. Bioprocess intensification: a potential aqueous two-phase process for the primary recovery of B-phycerythrin from *Porphyridium cruentum*. *J. Chromatogr., B* **2004**, *807*, 33–8. (b) Hernandez-Mireles, T.; Rito-Palomares, M. Improved recovery of B-phycerythrin produced by the red microalga *Porphyridium cruentum*. *J. Chem. Technol. Biotechnol.* **2006**, *81*, 989–996.

7. Ibarra-Herrera, C. C.; Aguilar, O.; Rito-Palomares, M. Application of an aqueous two-phase systems strategy for the potential recovery of a recombinant protein from alfalfa (*Medicago sativa*). *Sep. Purif. Technol.* **2011**, *77*, 94–98.
8. Aguilar, O.; Glatz, C. E.; Rito-Palomares, M. Characterization of green-tissue protein extract from alfalfa (*Medicago sativa*) exploiting a 3-D technique. *J. Sep. Sci.* **2009**, *32*, 3223–3231.
9. (a) González-Valdez, J.; Cueto, L. F.; Benavides, J.; Rito-Palomares, M. Potential application of aqueous two-phase systems for the fractionation of RNase A and  $\alpha$ -Lactalbumin from their PEGylated conjugates. *J. Chem. Technol. Biotechnol.* **2011**, *86*, 26–33. (b) González-Valdez, J.; Rito-Palomares, M.; Benavides, J. Advances and trends in the design, analysis, and characterization of polymer-protein conjugates for “PEGylated” bioprocesses. *Anal. Bioanal. Chem.* **2012**, *403*, 2225–35.
10. González-González, M.; Vázquez-Villegas, P.; García-Salinas, C.; Rito-Palomares, M. Current strategies and challenges for the purification of stem cells. *J. Chem. Technol. Biotechnol.* **2012**, *87*, 2–10.
11. (a) Crawford, S. E.; Labbé, M. J.; Cohen, M.; Burroughs, Y.; Zhou, M. K.; Estes, M. K. Characterization of virus-like particles produced by the expression of rotavirus capsid proteins in insect cells. *J. Virol.* **1994**, *68*, 5945–5952. (b) Mena, J. A.; Ramírez, O. T.; Palomares, L. A. Quantification of rotavirus-like particles by gel permeation chromatography, *J. Chromatog., B* **2005**, *824*, 267–276.
12. Fromantin, C.; Jamot, B.; Cohen, J.; Piroth, L.; Pothier, P.; Kohli, E. Rotavirus 2/6 Virus-Like Particles Administered Intranasally in Mice, with or without the Mucosal Adjuvants Cholera Toxin and *Escherichia coli* Heat-Labile Toxin, Induce a Th1/Th2-Like Immune Response. *J. Virol.* **2001**, *75*, 11010–11016.
13. Flynn, C. E.; Lee, S.-W.; Peelle, B. R.; Belcher, A. M. Viruses as vehicles for growth, organization and assembly of materials. *Acta Mater.* **2003**, *51*, 5867–5880.
14. (a) Vieira, H. L. A.; Estevao, C.; Roldao, A.; Peixoto, C. C.; Sousa, M. F. Q.; Cruz, P. E.; Carrondo, M. J. T.; Alves, P. M. Triple layered rotavirus VLP production: kinetics of vector replication, mRNA stability and recombinant protein production. *J. Biotechnol.* **2005**, *120*, 72–82. (b) Palomares, L. A.; Lopez, S.; Ramírez, O. T. Strategies for manipulating the relative concentration of recombinant rotavirus structural proteins during simultaneous production by insect cells. *Biotechnol. Bioeng.* **2002**, *78*, 635–644.
15. Kirkburk, C. D.; Buttery, J. *Expert Opin. Biol. Ther.* **2003**, *3*, 97–105.
16. Benavides, J.; Rito-Palomares, M. Practical experiences from the development of aqueous two-phase processes for the recovery of high value biological products. *J. Chem. Technol. Biotechnol.* **2008**, *83*, 133–142.
17. Chavez-Santoscoy, A.; Benavides, J.; Vermaas, W.; Rito-Palomares, M. Application of Aqueous Two-Phase Systems for the Potential Extractive Fermentation of Cyanobacterial Products. *Chem. Eng. Technol.* **2010**, *33*, 177–182.

18. Fisher, R. G.; Woods, N. E.; Fuchs, H. E.; Sweet, R. M. Three-dimensional structures of C-phycocyanin and B-phycoerythrin at 5-Å resolution. *J Biol. Chem.* **1980**, *255*, 5082–5089.
19. (a) Whitelam, G. C.; Cockburn, B.; Gandecha, A. R.; Owen, M. R. L. Heterologous protein production in transgenic plants. *Biotechnol. Gen. Eng. Rev.* **1993**, *11*, 1–29. (b) Krebbers, E.; Bosch, D.; Vandekerckhove, J. Prospects and progression the production of foreign proteins and peptides in transgenic plants. In *Plant Protein Engineering*; Shewry, P. R., Gutteridges, S., Eds; Cambridge University Press: London, England, 1992; pp 315–325 (c) Austin, S.; Bingham, E. T.; Koegel, R. G.; Matthews, D. E.; Shahan, M. N.; Strab, R. J. An overview of a feasibility study for the production of industrial enzymes in transgenic alfalfa. *Ann. N. Y. Acad. Sci.* **1994**, *721*, 235–244.
20. (a) Aguilar, O.; Rito-Palomares, M. Aqueous two-phase systems strategies for the recovery and characterization of biological products from plants. *J. Sci. Food Agric.* **2010**, *90*, 1385–1392. (b) Aguilar, O.; Rito-Palomares, M. Processing of soybean (*Glycine max*) extracts in aqueous two-phase systems as a first step for the potential recovery of recombinant proteins. *J. Chem. Technol. Biotechnol.* **2008**, *83*, 286–293.
21. Hoglund, M. Glycosylated and non-glycosylated recombinant human granulocyte colony-stimulating factor (rhG-CSF)-what is the difference? *Med. Oncol.* **1998**, *15*, 229–233.
22. Bradford, MM A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254.
23. (a) Harris, J. M.; Chess, R. B. Effect of PEGylation on pharmaceuticals. *Nat. Rev. Drug Discovery* **2003**, *2*, 214–221. (b) Roberts, M.; Bentley, M.; Harris J. Chemistry for peptide and protein PEGylation. *Adv. Drug Delivery Rev.* **2002**, *54*, 459–476. (c) Thanou, M.; Duncan, R. Polymer-protein and polymer-drug conjugates in cancer therapy. *Curr. Opin. Invest. Drugs* **2003**, *4*, 701–709.
24. (a) Lee, H.; Park, T. G. Preparation and characterization of mono-PEGylated epidermal growth factor: evaluation of *in vitro* biologic activity. *Pharm. Res.* **2002**, *19*, 845–851. (b) Morar, S.; Schrimsher, J. L.; Chávez, M. D. PEGylation of proteins: a structural approach. *BioPharm. Int.* **2006**, *19*, 34–49.
25. Cisneros-Ruiz, M.; Mayolo-Deloisa, K.; Przybycien, T. M.; Rito-Palomares, M. Separation of PEGylated from unmodified ribonuclease A using sepharose media. *Sep. Purif. Technol.* **2009**, *65*, 105–109.
26. González-González, M.; Vázquez-Villegas, P.; García-Salinas, C.; Rito-Palomares, M. Current strategies and challenges for the purification of stem cells. *J. Chem. Technol. Biotechnol.* **2012**, *87*, 2–10.
27. (a) Chalmers, J. J.; Zborowski, M.; Sun, L.; Moore L. Flow through, immunomagnetic cell separation. *Biotechnol. Progr.* **1998**, *14*, 141–148. (b) Namiri, M.; Baharvand, H.; Aghdami, N. Methods for isolation of bone marrow stem cells: comparative analysis. *Cell* **2011**, *12*, 439–446. (c) Bhagat A. A. S.; Bow, H.; Hou, H. W.; Tan, S. J.; Han, J.; Lim C.

T. Microfluidics for cell separation. *Med. Biol. Eng. Comp.* **2010**, *48*, 999–1014. (d) Thiel, A.; Scheffold, A.; Radbruch A. Immunomagnetic cell sorting- pushing the limits. *Immunotechnology* **1998**, *4*, 89–96. (e) Pappas, D. Wang K. Cellular separations: a review of new challenges in analytical chemistry. *Anal. Chim. Acta* **2007**, *601*, 26–35.

28. Cui, Y. X.; Kafienah, W.; Suleiman, M. S.; Ascione, R. A New Methodological Sequence to Expand and Transdifferentiate Human Umbilical Cord Blood Derived CD133(+) Cells into a Cardiomyocyte-like Phenotype. *Stem Cell Rev.* **2011**, [Epub ahead of print]. <http://www.springerlink.com/content/c0221013304t4012/>

29. (a) Melnik, K.; Nakamura, M.; Comella, K.; Lasky, L. C.; Zborowski, M.; Chalmers, J. J. Evaluation of eluents from separations of CD34+ cells from human cord blood using a commercial, immunomagnetic cell separation system. *Biotechnol. Progr.* **2001**, *17*, 907–916. (b) Liu, W.- H.; Li, R.; Dou, K.- F. Convenient and efficient enrichment of the CD133+ liver cells from rat fetal liver cells as a source of liver stem/progenitor cells. *Stem Cell Rev. Rep.* **2011**, *7*, 94–102.

30. González-González, M.; Mayolo-Deloisa, K.; Rito-Palomares, M. PEGylation, detection and chromatographic purification of site-specific PEGylated CD133-Biotin antibody in route to stem cell separation. *J. Chromatogr., B* **2012**, *893-894*, 182–186.

## Chapter 4

# Utilizing mAb Adsorption Isotherms To Direct Process Development of a Strong Cation-Exchange Chromatography Step

Christopher Gillespie\*

EMD Millipore, 80 Ashby Road, Bedford, Massachusetts 01464

\*E-mail: Christopher.Gillespie@merckgroup.com.

Phone: +1 (781) 533-5643. Fax: +1 (781) 533-5612

In spite of the development of alternative strategies to deal with downstream bottleneck concerns ion-exchange chromatography remains a useful tool that can be applied either in early or late-stage purification of monoclonal antibodies (mAb). Using two separate mAbs, adsorption isotherms and linear gradient elution studies were conducted. The goal was to understand the impact of loading, washing, and elution conditions on the dynamic binding capacity (DBC) and process yield obtained during preparative cation exchange (CEX) chromatography. Although adsorptive capacities were relatively salt insensitive at low pH values, optimal salt concentrations that maximized adsorption capacity were observed at higher pH values. Further, the yield of mAb from the preparative chromatographic runs was found to be dependent upon both its characteristic charge and the time the mAb was bound to the chromatographic resin. A theory is presented that will help explain the salt and pH dependent adsorption capacity trends, as well as the resulting mAb yield data. Finally, the use of adsorption isotherm information in the development of a strong CEX chromatography step will be discussed within the context of both direct capture and polishing of mAbs.

**Keywords:** Monoclonal antibodies; mAb purification; adsorption isotherms; dynamic binding capacity; static binding capacity; linear gradient elution; Langmuir isotherm

## 1. Introduction

Monoclonal antibodies (or mAbs) are identical colonies of monospecific antibodies that recognize only a single type of antigen. Mabs are useful tools in molecular biology and in medicine, where they can be used in diagnostics or as part of a monoclonal antibody therapy. In the latter, mAbs bind to a specific target or epitope, such as a tumor/cancer cell, and elicit an immune response. Some mAbs have been approved for clinical use by the FDA for the treatment of cancer (1), autoimmune diseases (2), and transplant rejection (3).

Industrially, mAbs are produced from mammalian cell cultures, e.g., hybridoma and Chinese Hamster Ovary (CHO) cells (4). *In vitro* mammalian cell cultures are grown in a nutrient rich growth media. Prior to clinical use of a therapeutic mAb, the product must be extracted from impurities such as host cell proteins (HCP), DNA, media components and other process impurities. As interest in the field has grown and methods to generate antibodies improved, antibody titers have steadily increased reaching 10 grams/liter with nearly 14 tons of monoclonal antibodies produced annually (5). This rapid increase in the output of mAbs has created a bottleneck within the downstream purification process.

Traditionally, mAb purification is performed following a platformed approach, where the first step is typically an affinity capture-step using Protein A chromatography followed by further polishing steps, commonly anion and/or cation exchange chromatography. Protein A affinity chromatography, although economically expensive, is extremely useful as it is very robust to processing conditions (e.g., salt insensitivity during loading and elution) and has been observed to produce mAbs to purity levels  $\geq 95\%$  in a single step. However, relatively harsh elution conditions can cause Protein A to leach/leak from the column into the mAb elution pool. As Protein A is a toxic compound (6), subsequent purifications are required to remove it from any biological sample.

To further reduce the impurity levels additional polishing steps, such as anion and/or cation exchange chromatography, have been employed. CEX chromatography is useful in mAb purification as it can be employed to remove specific impurities such as leached protein A, mAb aggregates (which are often removed concomitant with a yield compromise), and to a lesser extent HCP and DNA. CEX chromatography can be attractive as an alternative capture method for mAbs as it offers the potential of being a less expensive high capacity step. Unfortunately, the use of CEX chromatography for mAb purification is not without pitfalls, which include that it is not as robust as using Protein A affinity methods, and it requires more empirical development in comparison to the traditional Protein A affinity step.

Independent of the location within the downstream purification train the process development of a CEX step can be potentially time consuming. Here we describe how adsorption isotherms of mAbs can be employed as a predictive tool in the development of a purification step using a resin of choice. Adsorption isotherms are plots of the equilibrium adsorption capacity of a substance ( $q$ ) – in this case mAbs – as a function of its free concentration in solution ( $C^*$ ). Further, we report the development of a strong CEX polishing step to purify an in-house mAb. The adsorption isotherms were determined as a function of

different solution conditions - varying pH and salt – for a specific CEX resin, Fractogel® SO<sub>3</sub><sup>-</sup> (M). The results provided potential operating windows for the binding, washing, and elution conditions. Linear gradient elution studies were performed to assess mAb-resin interactions as a function of solution conditions while dynamic binding capacities (DBC) were measured and compared to the static binding capacities (SBC). The combined analysis provided insight into how to optimize the CEX chromatography step that was used as a post-Protein A polishing step, improving both HCP and DNA clearance.

## 2. Materials and Methods

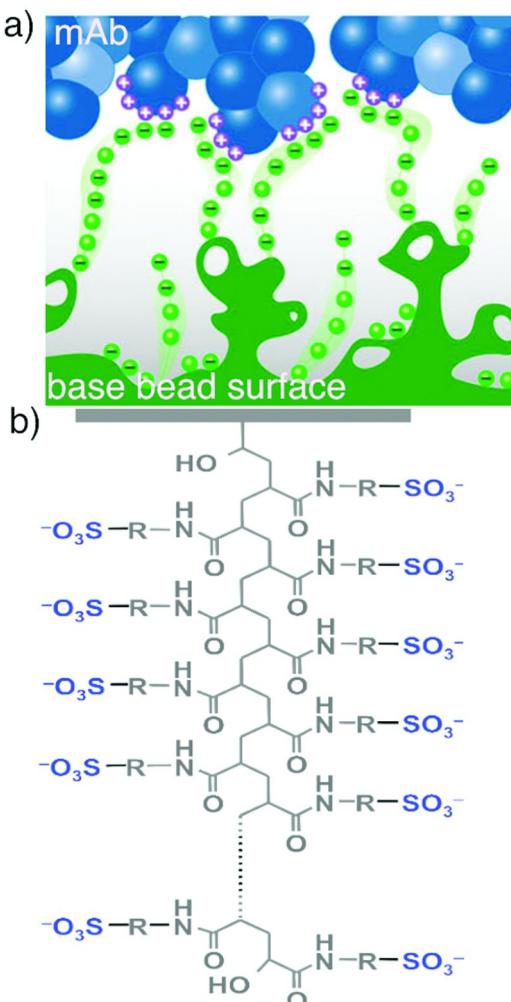
### 2.1. Materials

All solutions were prepared in ultrapure distilled water produced using a Milli-Q® system (EMD Millipore, Billerica, MA, U.S.A.) with a background buffer of 20 mM sodium acetate (Sigma Aldrich, Saint Louis, MO, U.S.A.) titrated to pH 4, 5 or 6 using 10 N NaOH and/or 6 M HCl. The ionic strength (Na<sup>+</sup> ion concentration) of the solutions were adjusted by adding NaCl (Fisher Scientific, Pittsburgh, PA, U.S.A.). Fractogel® SO<sub>3</sub><sup>-</sup> (M) (EMD Millipore), a strong cation exchange resin, was supplied in 20 % ethanol with 0.15 M NaCl. The resin was washed with water followed by equilibration into the buffer of interest. The mAb was produced in-house in CHO cells that were clarified by depth filtration followed by Protein A purification using ProSep®-vA High Capacity media (EMD Millipore). HCP concentrations were measured using a HCP ELISA (Cygnus, Southport, NC, U.S.A.) following the manufacturer's protocols (II), while DNA was quantified using a qPCR assay (Applied Biosystems, Foster City, CA, U.S.A.) according to the manufacturer's directions. MAb concentration was measured using either absorption at 280 nm with an extinction coefficient of 1.532 ml mg<sup>-1</sup> cm<sup>-1</sup> or Protein A HPLC using a POROS® A HPLC (Applied Biosystems) column on an Agilent 1260 HPLC (Agilent, Santa Clara, CA, U.S.A.).

#### 2.1.1. Solid Support and Chromatography

The CEX resin investigated here was Fractogel® SO<sub>3</sub><sup>-</sup> (M) (Figure 1) (7). The resin is a strong cation exchanger – meaning the negative charges it presents bind to cations in solution (Figure 1). Fractogel® SO<sub>3</sub><sup>-</sup> (M) was selected as it was experimentally determined to provide high capacity, high productivity, and a broad window of operation for our mAb, mAb1. A cartoon representation of the media is shown in Figure 1, where the flanking sulfonate (SO<sub>3</sub><sup>-</sup>) groups that represent the anionic binding locations are shown.

All chromatography steps were performed on an Äkta® Explorer 100 (GE Healthcare, Uppsala, Sweden) with UV-detection at 280 nm. DBC measurements were performed using a 6.6 mm ID Omnifit column (Omnifit, Danbury, CT, U.S.A.) packed to a bed height of 3 cm. Linear gradient elution experiments were performed using a 6.6 mm ID Omnifit column packed to a bed height of 15 cm.



*Figure 1. Schematic view of the Fractogel® solid support used for ion-exchange chromatography. a) Cartoon of the interaction of the positively charged surface of the mAb and the negatively charged bead surface. b) Cartoon representation of the sulfonated tentacle chemistry on the Fractogel® surface.*

## 2.2. Adsorption Isotherm and Static Binding Capacity Measurements

The adsorption isotherms for mAb1 with Fractogel® SO<sub>3</sub><sup>-</sup> (M) solid support were determined at pH 4, 5, and 6 under varying salt conditions ([NaCl] = 0 – 230 mM, Figures 2-4). The Langmuir isotherm model was fit to the measured static capacity data according to Equation (1)

$$q = \frac{q^*(KC^*)}{(1+KC^*)} \quad (1)$$

where  $q$  is amount of mAb adsorbed on to the resin in mg protein per mL of settled media,  $K$  is the strength of interaction in mL/mg,  $C^*$  is the equilibrium protein concentration in solution, and  $q^*$  is the maximum static binding capacity.

The adsorption isotherms were determined as follows. A known volume of resin was transferred to 1.5 mL eppendorf® tubes as a slurry and centrifuged. The buffer was decanted and a known volume of the mAb solution at known solution conditions (20 mM NaOAc with 0, 30, 80, 130, 180, 230 mM NaCl at pH 4, 5 or 6) was added. The tubes were left to rotate on a 360° rotisserie for approximately 24 hours. For each solution condition an eppendorf tube containing the same loading protein solution without resin was also rotated. Following incubation the tubes were centrifuged and the supernatant was decanted. The remaining soluble protein concentration was measured by absorbance at 280 nm as described in 2.1. The mass of protein bound to the resin was determined by mass balance as

$$q = \frac{V_l(C_o - C^*)}{V_r}, \quad (2)$$

where  $V_l$  and  $V_r$  is the volume of the protein solution and resin added, respectively.  $C_o$  is the initial protein concentration added.

### 2.3. Dynamic Binding Capacity Experiment

Dynamic binding capacities (DBC) were performed after column equilibration with 20 mM NaOAc at pH 4, 5 or 6, containing NaCl concentrations ranging from 0 - 230 mM. The protein in the equilibration buffer was loaded onto the column at 2 mg/mL using a 4 minute residence time with the DBC determined at 10 % of breakthrough. The bound protein was eluted with a step change to 1 M NaCl followed by column cleaning with 0.5 N NaOH.

### 2.4. Linear Gradient Elution Experiments

For all LGE experiments the column was equilibrated with 20 mM NaOAc at pH 4, 5 or 6. The column void volume ( $V_o$ ) was determined using blue dextran (GE Healthcare) pulse injections. The salt distribution coefficient ( $K'$ ) was measured using NaCl pulse injections. Protein in the equilibration buffer was loaded onto the column at a concentration of approximately 2 mg/mL to a loaded concentration of 5 g of mAb per L of packed bed. The column was then washed with five column volumes (CVs) of equilibrating buffer then eluted under a linear gradient of 10-40

CVs of 20 mM NaOAc, 1 M NaCl pH 4, 5 or 6. The superficial velocity was maintained at 100 cm/hr (0.57 mL/min) giving a 9.2 min residence time. After the elution was complete the column was washed with at least 3 CVs of 0.5 N NaOH prior to each subsequent run.

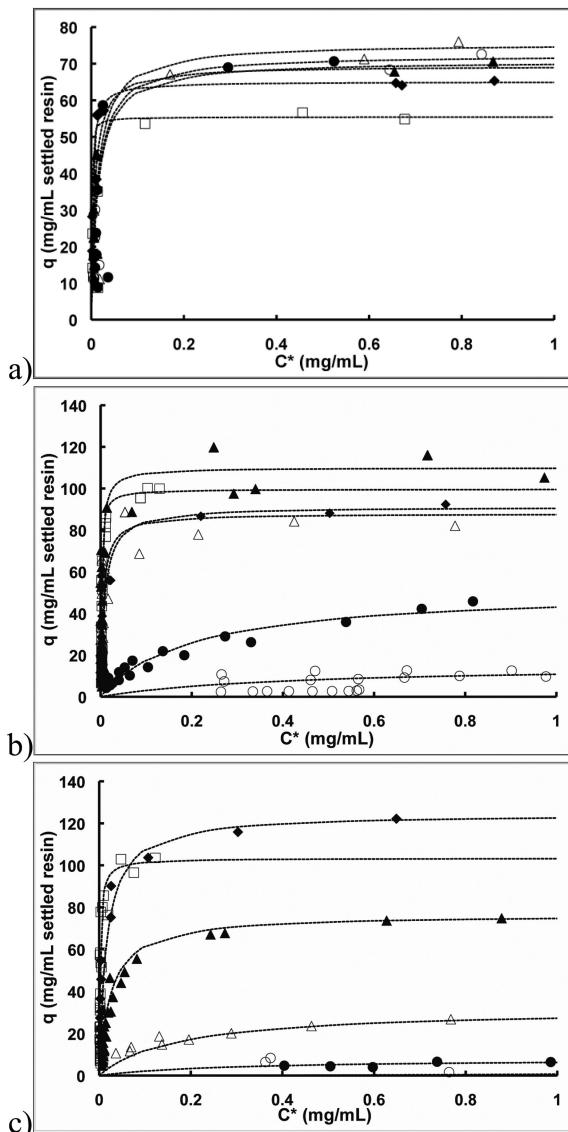


Figure 2. Adsorption isotherms for mAb1 using Fractogel®  $\text{SO}_3^-$  (M) with 20 (v), 50 (□), 100 (σ), 150 (Δ), 200 (●), and 250 (○) mM  $\text{Na}^+$  in 20 mM NaOAc at a) pH 4, b) pH 5, c) pH 6.

### 3. Results and Discussion

#### 3.1. Adsorption Isotherms and Static Binding Capacities

The adsorption isotherms for a particular monoclonal antibody (mAb1) of interest were studied at various pH values (pH 4, 5, and 6) as a function of increasing  $\text{Na}^+$  concentration (20-250 mM, Figures 2a-c). The resulting isotherms—saturation curves display the capacity or the maximum concentration of mAb that will bind to the resin under the particular conditions studied, visually this is the plateau in the graph,  $q^*$ .

At pH 4 it was observed that the maximum capacity (level of the plateau) decreased from approximately  $q = 60 \text{ mg/mL}$  to  $q = 50 \text{ mg/mL}$  ( $\Delta q = -10 \text{ mg/mL}$ ) upon increasing the concentration of  $\text{Na}^+$  ion from 20 to 50 mM (Figure 2a). This result is consistent with an ion exchange mechanism, where the increasing counterion concentration is expected to reduce the media's binding capacity. However, at higher concentrations of  $\text{Na}^+$  ( $> 50 \text{ mM}$ ) the capacity does not decrease further. Rather, it seems to stabilize at higher capacity than what was observed at 20 mM  $\text{Na}^+$  ( $q = > 60 \text{ mg/mL}$ , Figure 2a). Suggesting that at pH 4, the maximum binding capacities were relatively insensitive to salt. Further, the initial slopes of these curves indicate that the strength of binding is weakly dependent on ionic strength, which was not necessarily anticipated for a standard ion-exchange mechanism.

The opposite initial observation is seen at pH 5 (Figure 2b). Whereby increasing the concentration from 20, 50, 100 mM of  $\text{Na}^+$  caused a stepwise increase ( $\Delta q$  ca.  $+10 \text{ mg/mL}$ ) in the maximum capacity (Figure 2b). Again contrary to what would be expected for static binding capacities. However, at higher concentrations of  $\text{Na}^+$  from 150 mM up to 250 mM the traditional ion exchange pattern of decreasing capacity with increasing salt was observed. This can be seen in Figure 2b as the height of the plateau decreases significantly at 200 and 250 mM  $\text{NaCl}$ . Further the increased concentration (i.e. at 200 and 250 mM) caused significant changes in the initial slopes of these curves, which indicates that there is a corresponding decrease in the strength of interaction.

At pH 6 the entire salt range behaves more like a traditional ion exchanger in that at the lowest concentration of  $\text{Na}^+$  (20 mM) we see the highest capacity ( $q = 120 \text{ mg/mL}$ , Figure 2c). As the concentration of  $\text{Na}^+$  is increased there is a corresponding decrease in capacity and a decrease in affinity, as demonstrated by the change in initial slopes for these curves (Figure 2c).

These trends are summarized more clearly in the plot of the maximum capacity (from isotherms above),  $q^*$ , as a function of the sodium ion concentration ( $[\text{Na}^+]$ ) at the three different pH values evaluated 4, 5, and 6 (Figure 3a). At pH 4 the maximum capacity is relatively flat indicating that the capacity was independent of ionic strength/salt concentration (Figure 3a). At pH 5 there is an apparent maximum in static capacity at 100 mM  $\text{Na}^+$  ion concentration. Above 100 mM  $\text{Na}^+$  the traditional ion exchange mechanism of decreasing capacity with increasing salt is observed. At pH 6 there is a very clear trend of a continual decrease in the capacity with increasing salt concentrations. Because of the unexpected maximum

in static binding capacities with salt the adsorption isotherms for a second mAb, mAb2, were measured to determine if it behaved in a similar manner under the conditions investigated.

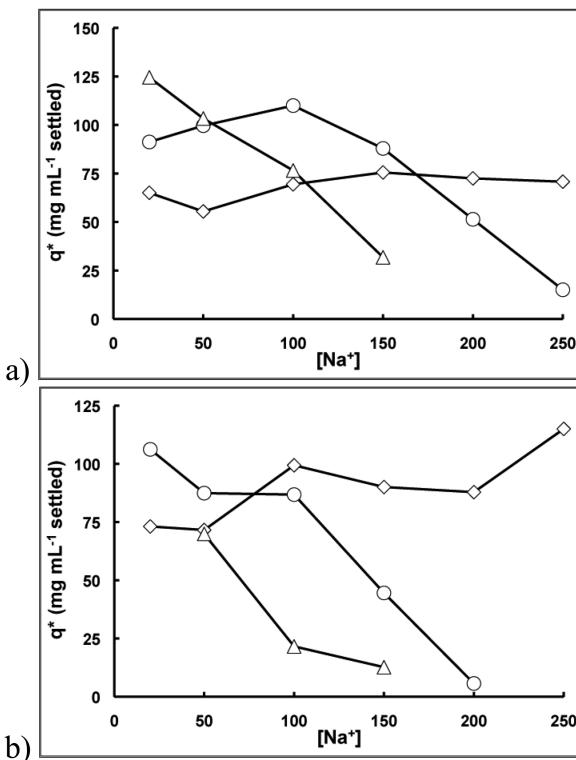


Figure 3. Static capacity trends ( $q^*$ ) as a function of ionic strength ( $\text{Na}^+$  concentration) maximum capacities at pH 4 ( $\diamond$ ), pH 5 ( $\circ$ ) and pH 6 ( $\triangle$ ) are shown for a) mAb1. b) mAb2.

The same adsorption isotherms were measured for a second monoclonal antibody, mAb2, under the same solution conditions at a pH of 4, 5, and 6 with varying the concentration of NaCl (Figure 3b). MAb2 was observed to behave more as a traditional ion exchanger, meaning there was no anomalous decrease in capacity at low salt, then increase as seen for mAb1 at pH 5 (Figure 3a). However, at pH 4 mAb2 appears to display somewhat anomalous behavior in that it seems more insensitive to salt, whereas at both pH 5 and 6 the normal trends of decreased capacity at increased salt concentrations was apparent. For mAb1 at pH 4, this insensitivity to salt at pH 4 could be exploited as it would enable a broad operating window for capacity that would not waver with varying salts. The results still beg the questions: why is this capacity independent of ionic strength at lower pH? What kind of concentrations of salt would be necessary to actually elute these mAbs binding under these conditions?

### 3.1.1. Linear Gradient Elutions

In an attempt to understand more clearly the cause of the anomalous static binding behavior of mAb1, linear gradient elution studies were performed following the method of Yamamoto, et al (8, 9). Pulse injections with loading at low conductivity (2-4 mS) and gradient elutions to 1.0 M NaCl were tested at 10, 20, 30, and 40 CVs. A plot of the normalized gradient slope ( $GH$ ) as a function of the ionic strength at the peak max ( $I_R$ ) of the elution is displayed in Figure 4, and was determined according to Equation (3). The results from these three pH values 4, 5, and 6 were fit to a model as:

$$GH = \frac{I_R^{(B+1)}}{A(B+1)} \quad (3)$$

where  $B$  is the characteristic charge and  $A$  is related to the actual strength of the interaction of the protein with the solid phase. At pH 6 mAb1 has a  $B$ -value of 11, while decreasing the pH to 5 increased the characteristic charge to 17.8 and at pH 4 the  $B$ -value increases to 28.7.

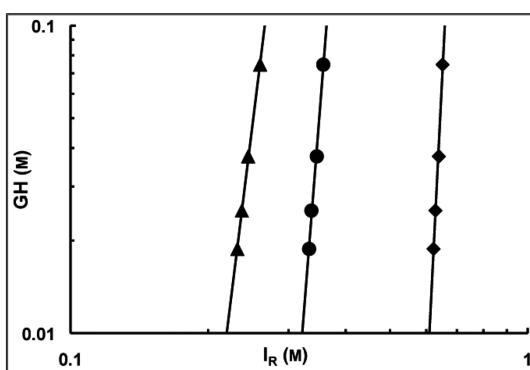
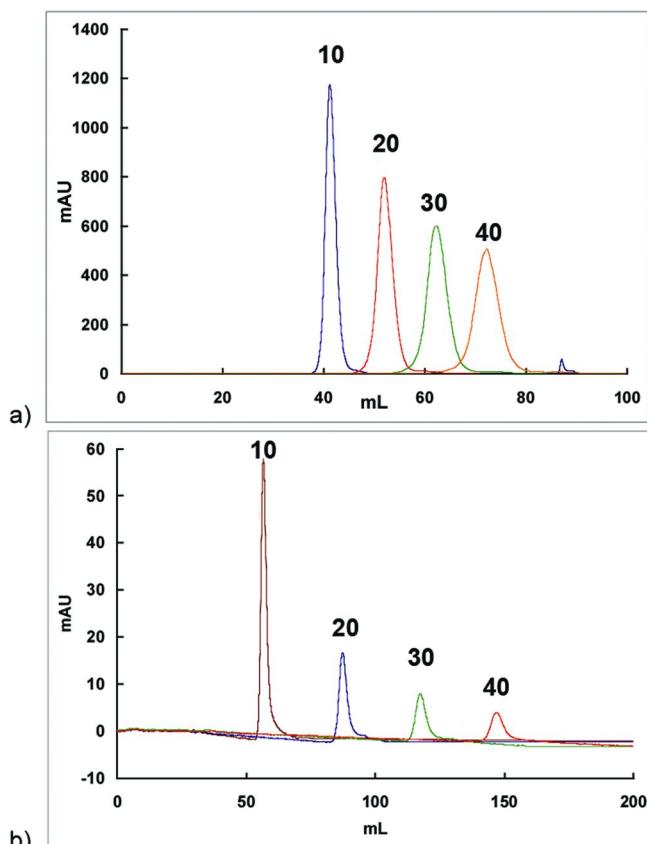


Figure 4. Log-log plot of linear gradient elution data as a function of ionic strength (salt concentration) for mAb1. The characteristic charge ( $B$ ) at pH 4 (◆)  $B = 28.7$ , pH 5 (●)  $B = 17.8$ , and pH 6 (▲)  $B = 11.0$  are determined from the slopes of the lines shown.

For comparison LGE studies of three other mAbs (mAb A, B and 01) were investigated under similar conditions at pH 5 with Fractogel®  $\text{SO}_3^-$  (M) media. The results are shown in Table 1 and it appears the strength of interactions, as given by the  $A$ -value are all very similar with the exception of mAb B, which is three orders of magnitude weaker than our current mAb ( $10^{-6}$  M), mAb1. According to the characteristic charge, our mAb1 seems to have the highest characteristic charge at pH 5, relative to the other mAbs evaluated, which may account for its anomalous behavior at lower pH levels.

**Table 1. Comparison of LGE parameters for different mAbs with Fractogel® SO<sub>3</sub><sup>-</sup> (M)**

<i>Monoclonal Antibody (mAb)</i>	<i>A-value (M)</i>	<i>B-value</i>	<i>Reference</i>
mAb1	12.72 x 10 <sup>-9</sup>	17.8	here in
mAb A	1.27 x 10 <sup>-8</sup>	13.5	(7b)
mAb B	3.21 x 10 <sup>-6</sup>	7.6	(7b)
mAb 01	6.77 x 10 <sup>-9</sup>	13.4	(10)



*Figure 5. Linear gradient elution peaks at pH 6 (a) and 4 (b) at 10, 20, 30, and 40 CVs.*

### 3.1.2. The Effect of pH from Linear Gradient Elutions on Recovery

When the plots used to generate the data shown in Figure 4 were revisited and the elution peaks more carefully examined a difference in percent (%) recovery between the pH levels became apparent. The two pH extremes, 4 and 6, were rerun and eluted from 0 to 1 M NaCl (Figure 5). Overlays of the elution peaks at 10, 20, 30, and 40 column volumes (CVs) at pH 4 and 6 are shown in Figure 5. The % recovery at pH 6 was good, at >90%. However that was not the case for pH 4, where not only is the intensity/absorbance level lower, despite the fact they were loaded at similar capacity, the peak structure is also comparatively different. In fact, for the 10 CV elution sample pH 4 (Figure 5b) the yield was only 2%, in comparison to >90% obtained for pH 6.

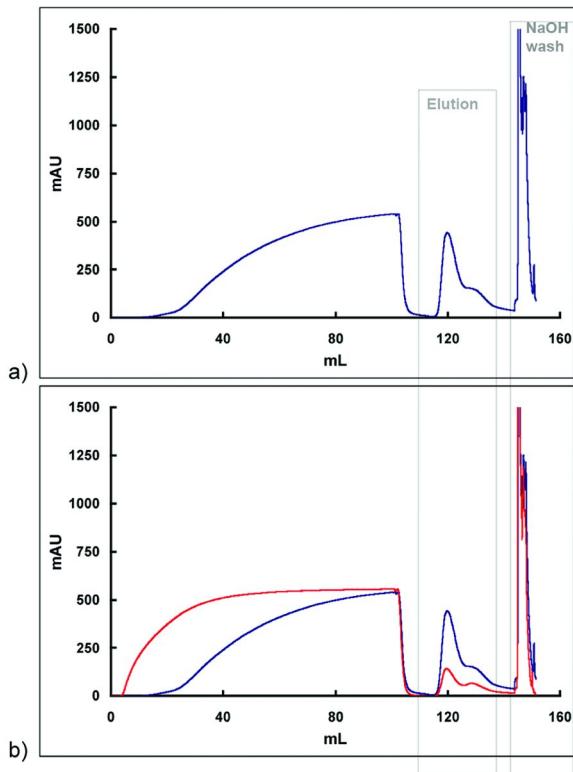


Figure 6. Dynamic binding experiments with mAb1, samples loaded at pH 4 20 mM NaOAc and conductivity of 2 mS at a 4 minute residence time. a) Elution 20 mM NaOAc, 0.25 M NaCl at pH 6, followed by 0.5 N NaOH wash. b) Overlay of repeat DBC experiment using same load, elution and cleaning conditions.

### 3.2. Dynamic Binding Capacities at Lower pH

Concerned about this yield discrepancy, DBC measurements were performed to ascertain if this effect was negligible at higher loadings, or if the lower pH identified a potential processing problem. The idea here was that higher loadings may cancel out the effect of the pulse injections, which are more on an analytical scale. The DBC experiments were performed by loading mAb1 at pH 4 (20 mM NaOAc) at a conductivity of  $\sim$  2 mS. The mAb was eluted using a step change to 20 mM NaOAc, 0.25 M NaCl pH 6 followed by 0.5 N NaOH cleaning step (Figure 6).

Several troubling observations arose from careful examination of the DBC curves (Figure 6a, blue line). First, the elution peak displayed a significant shoulder, indicative of an impurity or some other secondary effect, and the identity of this shoulder remains elusive. Second, there is a significant cleaning peak that eluted in the 0.5 N NaOH wash. Third, reloading that same column clearly demonstrates the loss of the majority of the binding capacity (Figure 6b, red line). The capacity at 10% break though (BT) was about 35 g/L for the first cycle while the DBC reduces to  $< 4$  g/L for the second. The elution conditions also provide two peaks, although as a result of the lower capacity there is a significant reduction in peak areas, while the cleaning peak is still very significant. Contrary to what would be expected for other solution conditions previously investigated the capacity could not be regained with NaOH wash. The results here clearly demonstrate that although the isotherms would predict a robust operating window the processing of this mAb at low pH is prohibitive.

### 3.3. Re-Evaluation of Static Binding Capacities To Prevent Loss of mAb1 in Wash Cycle

It was clear the results from the DBC experiments at pH 4 conditions caused significant loss in material. These results called for a re-evaluation of the operating window based on the measured static capacities. Careful examination of static capacities with a specific focus on the region of around 50 and 100 mM  $\text{Na}^+$  at pH 5 and 6 (Figure 3a) revealed something useful at higher salt concentrations. Specifically at pH 5, where there was a region of the graph with higher capacity at the higher salt concentration (ca.  $[\text{Na}^+] = 100$  mM,  $q^* = 100$  mg/mL). These higher salt and higher pH conditions could provide conditions where mAb would not be lost in a higher conductivity wash. Further, due to the fact that  $q^*$  is so high at the higher salt concentration we could also potentially use this as a direct capture method – loading the mAb directly from harvested cell culture fluid. This would reduce the amount of dilution necessary that would be required to lower the cell culture conductivity prior to loading.

### 3.4. Dynamic Binding Capacities with Higher pH and Higher Salt

Subsequent redetermination of the DBC at pH 5 and 6 was performed under the same conditions that were used to determine static capacities. A plot of dynamic and static capacities as a function of the concentration of  $\text{Na}^+$  at pH 5

and 6 show the trends are quite similar (Figure 7). At pH 5 the respective peak maximum were both observed to be around 100 mM, for dynamic and static capacities (Figure 7). At pH 6 the same decrease in capacity with increasing ionic strength was observed in both the dynamic and static capacities (Figure 7). As can be seen from the graph the dynamic capacities are significantly lower –almost 2 fold lower – than the static capacity. This result was not surprising as static capacities were performed with 24 h incubation, whereas the dynamic were performed under 4 min residence times.

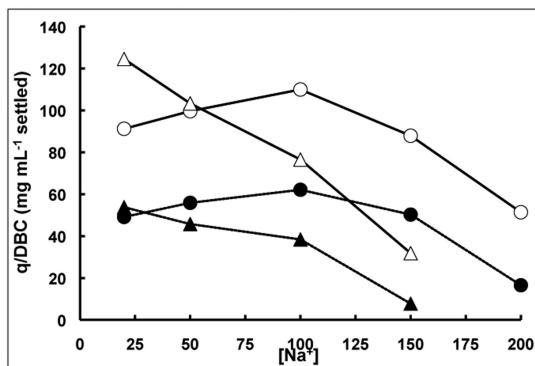


Figure 7. Comparison of dynamic (solid symbols) and static capacities (open symbols) for mAb1 at pH 5 (circle symbols) and pH 6 (triangle symbols). Lines are meant to guide the eye.

### 3.5. Cation-Exchange Chromatography as a Post-Protein A Polishing Step

Based on the combined data generated for mAb1 as a function of pH and Na<sup>+</sup> concentration an optimized post-protein A affinity polishing step was employed. MAb1 was loaded at pH 5 to > 83 g/L at a 10 minute residence time (to maximize binding capacity), and a low conductivity of 4-6 mS. The column was then washed with a solution of 20 mM NaOAc, pH 5, with 100 mM [Na<sup>+</sup>], which was used to remove any other impurities, i.e., a higher salt wash identified from isotherm data. These higher salt conditions ([Na<sup>+</sup>] = 100 mM) were chosen not only because of the higher capacity observed in the static capacities (Figure 3a); but also because of its potential to disrupt any intermolecular electrostatic interactions that may cause the formation of mAb1-impurity complexes. After the wash step mAb1 was recovered from the resin by a wash with 20 mM NaOAc, pH 6, with 250 mM [Na<sup>+</sup>], under these conditions q was essentially zero in the static capacities (Figure 3a).

When this intermediate wash is performed the improved HCP and DNA clearance is promising. It was found that the HCP log reduction value (LRV) increased from 0.5 to 0.8 and the DNA clearance improved from 0.3 to 0.9 LRV when compared to the same conditions in the absence of the intermediate wash. For mAb1 these LRV reductions correspond to HCP levels reduced from about 1500 ppm to about 840 ppm and DNA levels reduced from 0.2 ppm to 0.06 ppm.

This marked improvement in impurity clearance during the CEX step could allow for a more robust process with the potential for significant cost reductions in the later downstream steps. In the common mAb downstream paradigm the cost of flow through AEX steps may be enhanced by reductions in media, buffer and processing time.

## 4. Conclusion

In conclusion, adsorption isotherms can be useful tools during process development. They are easily leveraged and can aide in the identification of clear windows of operation that can direct further necessary experimentation. Here specific focus was placed on the optimization of the strong cation-exchange resin Fractogel® SO<sub>3</sub><sup>-</sup> to purify an in-house mAb. Interestingly, it was found that the maximum static binding capacity,  $q^*$ , was insensitive to ionic strength at pH 4, enabling a possible broad window of operation. Unfortunately, these conditions did not fare well when applied to linear gradient elutions, as the lower pH negatively impacted the yield. Dynamic binding capacity measurements confirmed that this yield compromise was significant, a result and could have potentially impacted mAb1 product qualities. Maximum static capacities were observed with moderate ionic strength (100 mM) at pH 5, which provided ideal sample loading and washing conditions. The mechanism of the maximum in the static capacity with increasing ionic strength remains elusive. However, the mechanism may be similar to that of the dynamic capacity: pore plugging (see ref. (7)b and references therein) and/or protein-protein surface repulsion (12). Lower capacities at pH 6 enabled the identification of sufficient elution conditions, as at pH 6 the common ion exchange behavior of decreased maximum capacity,  $q^*$ , with increased ionic strength was observed. Application of these conditions allowed for the development of a post-protein A polishing step that improved the impurity clearance of HCP (more than 1.5 fold) and DNA (3 fold). This improved impurity clearance may further improve the purification of mAb1 in later polishing steps, making the process more robust and potentially more economical.

## Acknowledgments

The author acknowledges the experimental contributions provided by Scott Bliss as well as the Analytical Technologies department. They also thank Michael Phillips, Shari Spector and Ajish Potty for useful discussions.

## References

1. (a) For a review see: Oldham, R. K. Monoclonal antibodies in cancer therapy: 25 years of progress. *J. Clinical Oncol.* **2008**, *26*, 1774–1777. (b) Textbook reference: Rang, H. P.; Dale, M. M.; Ritter, J. M. *Pharmacology*, 5th ed.; Churchill Livingstone: New York, 2003; p 241.

2. (a) Waldmann, T. A. Monoclonal antibodies in diagnosis and therapy. *Science* **1991**, *21*, 1657–1662. (b) Bruno, V.; Battaglia, G.; Nicoletti, F. The advent of monoclonal antibodies in the treatment of chronic autoimmune diseases. *Neurol. Sci.* **2011**, *31*, S283–S288. (c) Kieseir, B. C.; Lehmann, H. C.; zu Horste, G. M. Autoimmune diseases of the peripheral nervous system. *Autoimmun. Rev.* **2012**, *11*, 191–5.
3. (a) Pescovitz, M. D. Rituximab, an anti-CD20 monoclonal antibody: history and mechanism of action. *Am. J. Transplant.* **2006**, *6*, 859–866. (b) Breedveld, F. C. Therapeutic monoclonal antibodies. *Lancet*, **2001**, *355*, 735–740.
4. Farid, S. S. Process economics of industrial monoclonal antibody manufacture. *J. Chromatogr. B* **2007**, *848*, 8–18.
5. (a) Low, D.; O’Leary, R.; Pujar, N. S. Future of antibody purification. *J. Chromatogr. B* **2007**, *848*, 48–63. (b) Hacker, D. L.; Nallet, S.; Wurm, F. M. Recombinant protein production yields from mammalian cells: past, present, and future. *Biopharm. Int. Suppl.* June **2008**, *21*.
6. Terman, D. S.; Bertram, J. H. Antitumor effects of immobilized protein A and *staphylococcal* products: linkage between toxicity and efficacy, and identification of potential tumoricidal reagents. *Eur. J. Cancer Clin. Oncol.* **1985**, *21*, 1115–1122.
7. Specifications for resin: (a) Fractogel EMD SO<sub>3</sub><sup>-</sup> has methylacrylate with “tentacle” polymer functional groups. Particle size 40–90 μM, pore size 800 Å, lysozyme binding capacity 130 mg/mL. For additional information see: (b) Pabst, T. M.; Suda, E. J.; Thomas, K. E.; Mensah, P.; Ramasubramanyan, N.; Gustafson, M. E.; Hunter, A. K. Binding and elution behavior of proteins on strong cation exchangers. *J. Chromatogr. A* **2009**, *1216*, 7950–7956.
8. Ishihara, T.; Yamamoto, S. Optimization of monoclonal antibody purification by ion-exchange chromatography: Applications of simple methods with linear gradient elution experimental data. *J. Chromatogr. A* **2005**, *1069*, 99–106.
9. Ishihara, T.; Kadoya, T.; Endo, N.; Yamamoto, S. Optimization of elution salt concentration in stepwise elution of protein chromatography using linear gradient elution data: Reducing residual protein A by cation-exchange chromatography in monoclonal antibody purification. *J. Chromatogr. A* **2006**, *1114*, 97–101.
10. Urmann, M.; Graalfs, H.; Joehnck, M.; Jacob, L. R.; Frech, C. Cation-exchange chromatography of monoclonal antibodies: characterization of a novel stationary phase designed for production-scale purification. *MAbs* **2010**, *4*, 395–404.
11. Cygnus Technologies. *CHO Host Cell Proteins 3rd Generation*, Cat # : F550, [http://www.cygnustechnologies.com//images/file/Product\\_Inserts/f550\\_0\\_pi.pdf](http://www.cygnustechnologies.com//images/file/Product_Inserts/f550_0_pi.pdf).
12. Xu, X.; Lenhoff, A. M. A predictive approach to correlating protein adsorption isotherms on ion-exchange media. *J. Chem. Phys. B* **2008**, *112*, 1028–1040.

## Chapter 5

# Understanding the Role of Arginine and Citrate as Eluents in Affinity Chromatography

Diwakar Shukla and Bernhardt L. Trout\*

Department of Chemical Engineering,  
Massachusetts Institute of Technology, 77 Massachusetts Avenue,  
E19-502b, Cambridge, Massachusetts 02139, United States

\*E-mail: trout@mit.edu. Phone: (617) 258-5021. Fax: (617) 253-2272

Industrial scale purification of therapeutically useful antibodies (immunoglobulin, IgGs) typically rely on Protein A affinity chromatography, as a first step. Eluents i.e aqueous solution of small organic molecules, are typically used to elute the antibodies from the affinity column. The molecular interactions between the antibody and the Protein A affinity resin are exceedingly dependent on pH and eluent concentration. This complex mechanism is not well understood, beyond the application of general trends, for eluents such as arginine and citrate. We have performed molecular mechanics simulations of antibody and Protein A to rationalize the experimental observations of the effects of arginine and citrate on purification yield. It was found that the highly positively charged arginine reduced the binding free energy ( $\Delta\Delta G_B$ ) between the Protein A affinity resin and the antibody. While the negatively charged citrate increases the  $\Delta\Delta G_B$  for Protein A affinity resin and the antibody. The results provide a rationale for the observed fluctuations in the recovery of antibodies using citrate and arginine, and also provided insight into of the nature of the molecular interactions, which can be exploited for eluent design on Protein A affinity chromatography.

## 1. Introduction

Affinity chromatography is an important step in the purification of therapeutically useful proteins, especially immunoglobulins (IgGs, antibodies). As the production of therapeutically useful quantities of IgGs has increased over the last decade, and antibody titers have reached 10 g/L, efforts have focused on the optimization of downstream purification of the antibody to meet this rise (1). On an industrial scale antibodies are produced in mammalian cell cultures, such as Chinese Hamster Ovary (CHO) cells, where the cell has been genetically modified to produce large quantities of the desired antibody (2). Thus to obtain clinically useful quantities of pure antibodies, the desired antibody must be separated from the cellular components used for its preparation, such as host cell proteins (HCP), DNA, aggregates, growth media components, and other impurities.

The typical first step in the long purification process used for isolation of antibodies is affinity chromatography (3). Affinity chromatography is an economically expensive all be it robust purification process that can be employed to purify antibodies to levels  $\geq 95\%$  in a single step. As demonstrated in Figure 1 the first steps are the critical steps in the purification process, where the maximum loss of product occurs. By extension reduction in the total loss of protein in these steps will improve the total yield of the product.

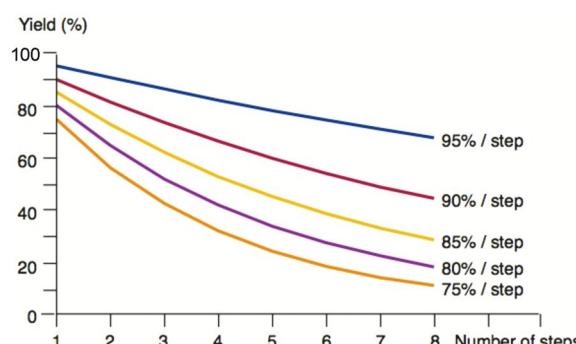
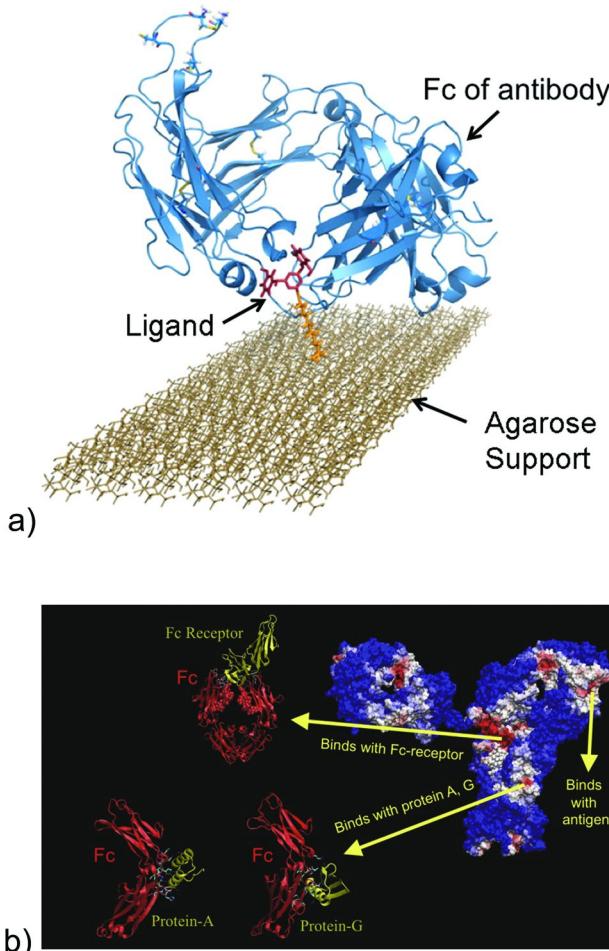


Figure 1. Graphical representation of the reduction in yield (recovery) of antibody/protein per step of a purification process.

Affinity chromatography columns contain a resin or a solid support with an attached ligand to which the antibody/proteins bind (Figure 2a). Typical ligands used include Protein A or Protein G, which have a high affinity (5) and selectivity for antibodies, specifically the Fc (fragment crystallizable) domain, and cause the antibody to be retained on the column (Figure 2b). The impurities are then eluted (removed or washed off) from the Protein A based on the lower affinity these components have for Protein A, while the antibody is left behind. However, the process is plagued with imperfections, such as damage to the protein that results of the harsh conditions used to elute tightly bound antibody, such as highly acidic solutions. At low pH, both arginine and sodium citrate are effective eluents, however highly acidic eluent solutions can denature and agglomerate antibodies,

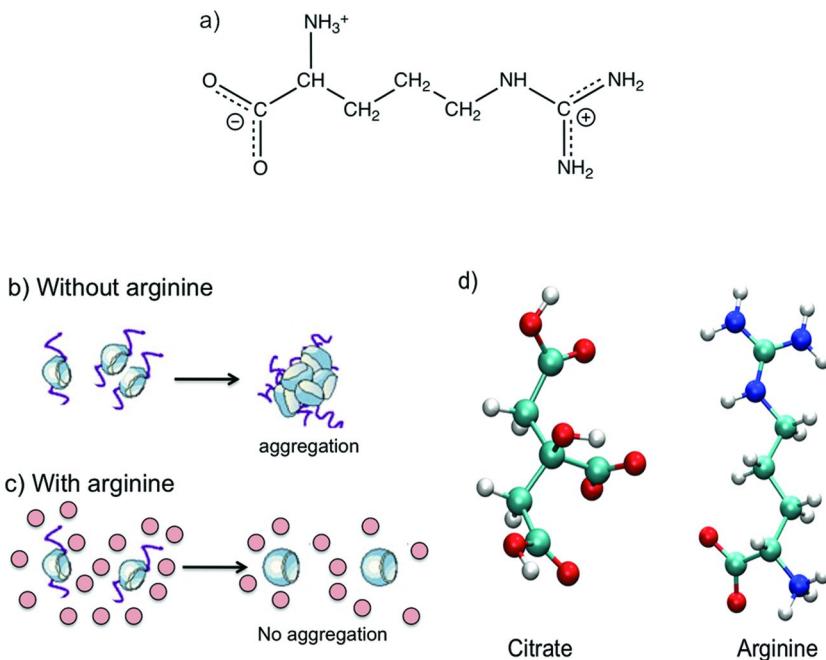
reducing the yield of monomeric forms. Further, these conditions can cause Protein A to leach from the column, which must be latter separated out of the sample as it is highly toxic (6). The result is the addition of further downstream purification steps to remove the leached Protein A.



*Figure 2. Schematic view of the interaction between Protein A and an antibody Fc domain. (a) Cartoon representation of Fc region of human IgG interacting with Protein A attached to a solid support; (b) Cartoon representation of the entire globular surface of an antibody (right), Fc domain (top left) and Protein A or G (bottom left). Figure prepared from X-ray crystal structure of Protein A and human IgG (PDB code: 1FC2) (4).*

The free energy of binding ( $\Delta\Delta G$ ), or the strength of the interaction between Protein A and antibodies is very high (5), and elution from an affinity column can be impeded the stronger the interaction. Typical elution conditions employed to

release the antibody from the Protein A include an acidic buffer, salt solutions, or the addition of an elution additive such as arginine or citrate. Fundamentally, all the elution conditions work by lowering the binding free energy of the antibody to the Protein A. All be it each eluent does this in a different way. However, the desired outcome is the same and that is to identify an eluent that can release the desired protein and not disturb or destabilize the proteins' integrity.



*Figure 3. (a) Structure of arginine in the pH range 2-9; (b) Cartoon representation of protein aggregation in the absence of arginine; (c) Cartoon representation of inhibition of protein aggregation in the presence of arginine in solution; (d) molecular model (ball-and-stick) representation of citrate and arginine (C-atoms, cyan; O-atoms, red; N-atoms, blue; H-atoms, white).*

As mentioned previously amino acid monomer arginine can be used as an additive in the elution solution. Arginine is a polar amino acid with three charged functional groups in solution. In solution arginine contains two positively charged amines (the *N*-terminal amine and the side chain guanidinium) as well as a negatively charged *C*-terminal carboxylic acid (Figure 3a). These charged groups play a key role in the utility of arginine as an eluent. In the 1980's arginine was serendipitously found to enhance protein refolding yields with 100 % refolding observed in a 1 M arginine solution (7). Later, it was found that arginine could prevent protein aggregation and reduce the rate of aggregation (8). Protein aggregation occurs when unfolded (denatured) proteins associate and form higher

oligomers, a process that commonly occurs as a result to exposure to acidic buffer or high ionic strength elution solutions (Figure 3b-c).

Tsumoto and coworkers, who compared another common eluent sodium citrate (Figure 3d), to arginine, documented the unique properties of arginine (9). Tsumoto reported poor recoveries for 0.1-0.5 M sodium citrate at pH 4.3. Product peak shape and recovery were markedly improved using 0.5 – 2.0 M arginine at pH 4.3 (9). In addition the authors proved that the properties of arginine do not extent to other amino acids such as lysine, proline, glycine and histidine (9). A result that suggests the properties of arginine are unique and worthy of further investigation. As effective eluents will not only reduce affinity between ligand and antibody, but also inhibit aggregation of eluted antibodies, or lab became interested in: Why does arginine suppress protein aggregation? Why arginine is an effective eluent?

The ability of arginine to suppress antibody aggregation has been established (Figure 3b-c) (8). The focus here became: How does arginine aide in the dissociation of an antibody from Protein A? To come up with an answer we used computational chemistry and molecular mechanics simulations to determine the free energy of the complex ( $\Delta G_b$ ) formed between Protein A and the antibody ([Protein A – Fc]) according to the standard  $\Delta G_b$  equation (1) that relates  $\Delta G_b$  to the Gibbs free energy for each individual reaction species. According to equation (1):

$$\Delta G_b = G_{\text{complex}} - G_{\text{IgG}} - G_{\text{protein-A}} \quad (1)$$

Further estimation of the effect of solute addition to the free energy provided the change in binding free energy ( $\Delta\Delta G_b$ ) between IgG and Protein A in the presence of cosolute as compared to water according to equation (2):

$$\Delta\Delta G_b = \Delta G_b^{\text{Cosolute}} - \Delta G_b^{\text{water}} \quad (2)$$

Analysis of the molecular dynamics simulations of the complex formed by Protein A, and the Fc domain of IgG, allowed for the identification of key residues involved in binding. Taken together the results provided a rationale for the role of arginine as an eluent in Protein A affinity chromatography and can be extended upon for the design of novel ligands or eluents.

## 2. Materials and Methods

### 2.1. Model Building: Reconstruction of the IgG-Protein A Complex

Molecular simulation on a reconstructed IgG – Protein A complex was performed as previously described (10, 11). Briefly, the X-ray structures were obtained from Protein Data Bank (PDB, [www.rcsb.org](http://www.rcsb.org)). The X-ray coordinates

for the complex of fragment B of a Protein A from *Staphylococcus aureus* to bind to the Fc fragment of a human IgG complex solved to 2.8 Å-resolution (PDB code: 1FC2) (4) and as well as the coordinates for a crystal structure of an entire human IgG (PDB code: 1HZH) (12) solved to 2.7 Å-resolution were used for the molecular mechanics and modeling studies (10). This complete IgG X-ray crystal structure was then compared to a X-ray crystal structure of the Fc domain bound to fragment B of Protein A. Superimposition and alignment of the two structures allowed us to build by reconstruction a model of the entire IgG – Protein A complex for our simulation (Figure 4).

## 2.2. Simulation Details

Molecular dynamics simulations were performed as previously described (10). Briefly, simulations were performed using the NAMD (13) package with CHARMM22 (14) force field and the TIP3P water model (15). NPT ensemble with periodic boundary conditions and full electrostatics computed using the particle-mesh Ewald (PME) (16) method, with the grid spacing on the order of 1 Å or less was used and pressure was maintained at 1 atm using the Langevin piston method (piston period 200 fs; damping time constant 100 fs; piston temperature 298 K). The protein residues and cosolute molecules were protonated to account for an experimental pH value of 4.3. Cosolute molecules and counter ions were added to obtain a net neutral charge. To set up the simulation systems for different cosolute concentrations, cosolute molecules were randomly placed within the simulation box, and the overlapping water molecules were subsequently removed. The system was then equilibrated for 1 ns at constant pressure and temperature, and snapshots of the simulation box were saved every picosecond.

## 2.3. Force Field

Force field parameters used were previously described (10). Briefly, the force field parameters for arginine with the N terminal and the side chain protonated and C terminal deprotonated were taken from the CHARMM force field (14). At a pH of 4.3, pKa for the carboxylic groups in citric acid at 25 °C are 3.128, 4.761, and 6.396. All citrate molecules were modeled as dihydrogen citrate ( $\text{H}_2\text{Cit}^-$ ) ions, as it is the dominant population (71%) (17). Structures of cosolutes considered in the present study, citrate and arginine, are shown in Figure 3d. Force field parameters for dihydrogen citrate were developed following a procedure defined by Kamath *et al.*, wherein the parameters are iteratively adjusted to reproduce both intramolecular and intermolecular target data (18).

### Intermolecular Data

Interaction energy and the minimum energy distance between dihydrogen citrate and water molecules – intermolecular data – are used in the first step of the optimization procedure to evaluate partial atomic charges of the molecule. CHARMM was used to calculate the hydrogen bonding of each dihydrogen citrate

polar group (both acceptors and donors) with a single water molecule and the target values for intermolecular data were obtained through *ab initio* simulations at the HF/6-31g\* level. Although H<sub>2</sub>Cit<sup>-</sup> is a symmetric molecule, the relatively high number of atoms in the molecule hampers the evaluation of its parameters. Consequently, to obtain a reliable first guess for partial atomic charges of H<sub>2</sub>Cit<sup>-</sup>, the initial analysis of intermolecular interactions and partial charge determination was limited to the central core of the molecule ((CH<sub>3</sub>)<sub>2</sub>-COHCOO<sup>-</sup>). Initial values of all other parameters were derived by analogy, using molecules with similar structures and groups.

### *Intramolecular Data*

Geometry, dipole moment, minimum energy structure vibrational frequencies, and potential energy surfaces (PES) of dihedral angular scans comprise the intramolecular data. The vibrational frequencies in CHARMM were computed with the MOLVIB program and PES fitting was carried out using the Guvench-MacKerell program (19). Target values for intramolecular data were obtained through QM computations at the MP2/6-31+g\* level of theory.

## **2.4. The Effect of Cosolutes on the Free Energy of Binding**

As the Fc region of the antibody is the only portion that binds to Protein A (Figure 2b), it was the only portion of the antibody used for the simulation, which simplified the calculations. The Fc region is the fragment crystallizable region of the antibody located on the tail of the antibody (Figure 2b) and it is known to interact with cell surface receptors. The effect of cosolutes on the  $\Delta\Delta G_B$  of Protein A and IgGs was previously described (10). Only a brief description of summary is presented here, for a more detail discussion see reference (10).

The addition of cosolute to the aqueous protein solution can affect the free energy of binding relative to binding free energy in water. This change in free energy is called ( $\Delta\Delta G_b$ ) between IgG and protein A in the presence of cosolute as compared to water according to equation (2) and  $\Delta\Delta G_B$  can be determined from the transfer free energy ( $\Delta\mu^{tr}$ ) for the complex, Protein A, and the Ig, according to equation (3) (10).

$$\Delta\Delta G_b = \Delta\mu_{complex}^{tr} - \Delta\mu_{IgG}^{tr} - \Delta\mu_{protein-A}^{tr} \quad (3)$$

The transfer free energy  $\Delta\mu^{tr}$  is the change in the interaction energy for a protein (Protein A and Fc) with a particular solvent when the solvent was changed from water to co-solute (10). It can be determined according to equation (4):

$$\Delta\mu_2^{tr} = \mu_2^{Cosolute} - \mu_2^{water} \quad (4)$$

The transfer free energy was then related to a preferential interaction coefficient ( $\Gamma_{23}$ ) to determine the total change in the chemical potential of the protein (antibody) that results from its transfer to a co-solute solution instead of water. The preferential interaction coefficient is the number of co-solute molecules per protein molecule added to the solution to keep the chemical potential of the co-solute constant. For a more detailed discussion of the mathematical derivation of  $\Gamma_{23}$  from  $\Delta\mu^t$  see reference (10). More simplistically  $\Gamma_{23}$  is also a measure of the excess local concentration around the protein as compared to the bulk cosolute solution. It can be measured experimentally using dialysis/ densimetry (20) and vapor-pressure osmometry (21). The preferential interaction coefficient was determined for every species involved in the reaction: the Fc region of the Ig ( $\Gamma_{23}^{IgG}$ ), for Protein A ( $\Gamma_{23}^{Protein\ A}$ ), and the [Fc-Protein A] complex ( $\Gamma_{23}^{complex}$ ). The preferential interaction coefficient of the complex [Fc – Protein A], defined as  $\Gamma_{23}^{complex}$ , was then used directly to determine the  $\Delta\Delta G_b$  for the complex according to equations (5) and (6) (10). Where  $m$  is molarity,  $T$  is temperature,  $P$  is pressure, and  $R$  is the gas constant.

$$\Delta\Delta G_b = RT \left( m_1 \frac{\partial \ln(a_1)}{\partial m_3} \right)_{T,P,m_2} \left( \Gamma_{23}^{complex} - \Gamma_{23}^{IgG} - \Gamma_{23}^{protein-A} \right) \quad (5)$$

$$\Delta\Delta G_b = RT \left( m_1 \frac{\partial \ln(a_1)}{\partial m_3} \right)_{T,P,m_2} \Delta\Gamma_{23} \quad (6)$$

Thus  $\Gamma_{23}^{complex}$  allows for the determination of the change the binding energy between the antibody and the Protein A. More detailed descriptions of the derivations see Ref. (10).

### 3. Results and Discussion

#### 3.1. Cosolute Effect on Preferential Interaction Coefficient

The preferential interaction coefficient,  $\Gamma_{23}$ , is a property of the individual proteins and their interaction with the cosolute, in this case arginine or citrate. By definition  $\Gamma_{23}$  is a measure of the local concentration of solute around the protein surface compared to the bulk concentration (the region far from the protein surface). In other words, the higher the affinity a molecule has for the protein the closer the molecule surface and greater the concentration of that molecule at the surface. Mathematically that means  $\Gamma_{23}$  will be greater than zero. In contrast, a molecule with a weaker affinity for the protein surface will be farther away from the surface and  $\Gamma_{23}$  will be less than zero. More detailed mathematical derivations

of  $\Gamma_{23}$  have been reported in several papers in literature (10, 22). Simulation of a protein in the presence of additives, such as arginine or sodium citrate in water, allowed for the determination of their respective concentrations around Protein A or IgG surface (10).

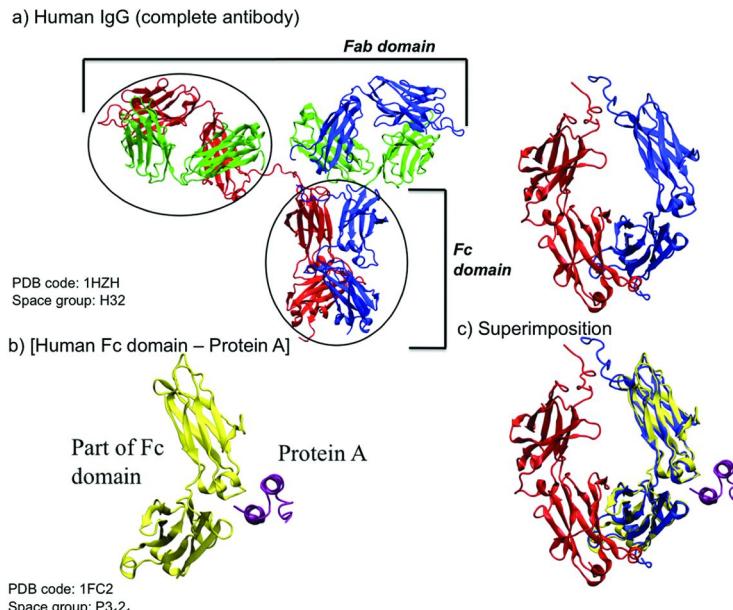


Figure 4. (a) Left: Cartoon representation of entire human IgG solved to 2.7 Å-resolution (PDB code: 1HZH) (4). Right: side view cartoon representation only the Fc domain IgG solved to 2.7 Å-resolution (PDB code: 1HZH) (4). Color code: heavy chains (red and blue); light chains (green); (b) Cartoon representation of the X-ray crystal structure of B fragment of Protein A from *S. aureus* bound to a part of the Fc domain of human IgG solved to 2.8 Å-resolution (PDB code: 1FC2) (12). Color code: Fc domain (yellow); Protein A (purple). (c) Superimposed structure of Fc domain from (a) and the co-crystal structure from (b) (PDB codes 1FC2 and 1HZH) (4, 10, 12). Figure modified from Ref. (10).

The approach used for the calculation of the preferential interaction coefficient has been used to estimate  $\Gamma_{23}$  values for a variety of proteins. The predicted values were in reasonable agreement with the experimental Dialysis/Densimetry and VPO data (1, 23). The results for the proteins RNase T1 in aqueous solution of additives arginine hydrochloride, glycerol, and urea are reported in a previous study from our group. A plot of the  $\Gamma_{23}$  for each additive and this unrelated small protein as a function of distance from protein surface is shown in the Appendix (Figure A1). The plot shows the variability in  $\Gamma_{23}$  for arginine hydrochloride,

glycerol and urea, with urea having the highest  $\Gamma_{23}$  and arginine a  $\Gamma_{23}$  close to zero. The results demonstrate urea has a higher affinity than arginine for the proteins' surface. More detailed mathematical derivations of  $\Gamma_{23}$  as a function of molarity ( $m$ ) or chemical potentials ( $n$ ) are presented in previous studies from our group (1, 24).

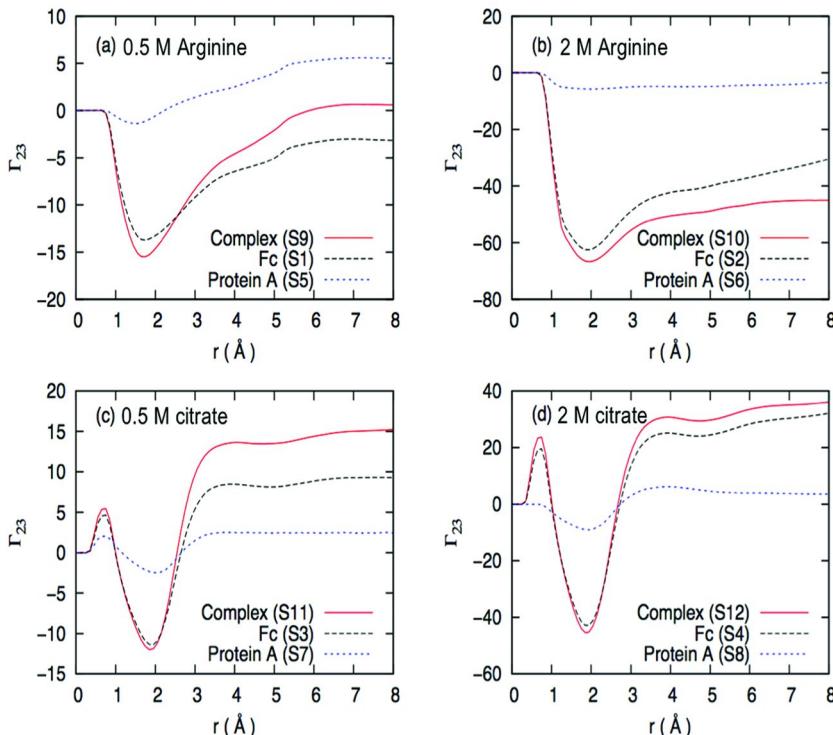


Figure 5. Plot of calculated preferential interaction coefficient ( $\Gamma_{23}$ ) as a function of distance. Higher affinity solutes have positive  $\Gamma_{23}$  and lower affinity solutes have negative  $\Gamma_{23}$ . a) Calculated  $\Gamma_{23}$  in 0.5 M arginine; b) Calculated  $\Gamma_{23}$  in 2 M arginine; c) Calculated  $\Gamma_{23}$  in 0.5 M citrate; d) Calculated  $\Gamma_{23}$  in 2 M citrate.

Figure modified from Ref. (10).

Once the simulation was validated on the model it was extended to the Protein A – IgG complex of interest. The results were previously described so will only be briefly summarized here (10). Again the calculation was simplified to include only the relevant interactions between the coordinates of an X-ray structure of Protein A (from *S. auerus*) and the Fc (from human IgG) domain of the antibody (Figure 4). The Fc domain of the complete IgG structure (PDB code 1HZH) solved to 2.7 Å-resolution in 2001 by Wilson and co-workers (12) was compared to a X-ray

crystal structure of a co-crystal structure fragment of human Fc domain bound to *S. aureus* Protein A solved by Deisenhofer in 1981 (4). Superposition and alignment of the two structures allowed for the reconstruction of a new model containing of the entire fragment B of Protein A and a complete Fc domain complex ([Protein A - Fc]) for molecular mechanics and modeling studies (Figure 4) (10).

The reconstructed [Protein A- Fc] complex was used in combination with the CHARMM (14) force field to determine  $\Gamma_{23}$  and thus the  $\Delta\Delta G_B$  for each eluent, citrate ion and arginine, in combination with the complex, Fc domain, and Protein A. The preferential interaction parameters are sensitive to the accurate force field parameters. Therefore, the force field parameters for citrate were developed for this study and optimized for the molecular dynamics simulations using the QM data. This allowed for accurate data that was comparable to experimentally determined data  $\Delta\Delta G_B$  (not shown). The outcome of the simulations were the focus of a recent disclosure and therefore only briefly summarized (10).

Briefly the results for arginine and citrate ion at 0.5 and 2 M for the [Protein A- Fc] complex, the Fc domain, and Protein A demonstrated the variability in  $\Gamma_{23}$  as a result of distance, but on average level off after an  $r > 4\text{-}6 \text{ \AA}$  (Figure 5). The results further demonstrated that  $\Gamma_{23}$  is highly dependent on the concentration of the additive (Figure 5). The preferential interaction coefficient of Protein A ( $\Gamma_{23}^{\text{Protein A}}$ ) for arginine and citrate is close to zero at concentrations 0.5 and 2 M at shorter distances ( $r$ ), however level out at  $r > 4 \text{ \AA}$ . The  $\Gamma_{23}^{\text{Protein A}}$  for arginine however is clearly slightly positive at lower concentration and slightly negative at the higher concentrations, and again leveling off after  $r > 4 \text{ \AA}$  (Figure 5a and b). The trend for arginine extends for the simulation with the Fc domain and the complex, with an increasingly negative  $\Gamma_{23}^{\text{Fc}}$  as concentration of the additive increases (Figure 5a and b). Conversely, according to the simulation citrate has a positive  $\Gamma_{23}$  with increasing concentration for both the Fc domain and the complex. Further,  $\Gamma_{23}$  increased with increasing concentration of citrate (Figure 5, Table 1).

This result seems counterintuitive as a positive  $\Gamma_{23}$  implies that more molecules reside on the protein surface, and a negative  $\Gamma_{23}$  the opposite. To rationalize these observations it was thought that if more molecules are at the surface then those eluent molecules effectively prevent protein aggregation. Citrate appears to be more effective at inhibition of aggregation than arginine, but additionally positive  $\Gamma_{23}$  also cause the destabilization the protein conformation. Typically, cosolutes such as urea and guanidinium chloride with high positive preferential interaction coefficient tend to reduce the native state conformational stability (22). Therefore, the citrate could also cause the protein to unfold – denature. Citrate is thus an effective eluent to remove antibodies from Protein A affinity resin, but as a side effect it can destabilize and unfold the antibody. Conversely, arginine imparts stability and does not cause denaturation, but inhibits aggregation.

### 3.2. Cosolute Effect on Free Energy of Binding

The free energy of binding,  $\Delta\Delta G_B$ , can also be expressed as the difference of the  $\Gamma_{23}^{\text{complex}}$  and  $\Gamma_{23}^{\text{Fc}}$  and  $\Gamma_{23}^{\text{Protein A}}$  domains. According to Figure 5, it would then appear that  $\Delta\Delta G_B$  is very small and almost approaches zero for each

concentration of cosolute. However, it is important to mention that the absolute numerical value is not relevant, rather it is the difference between absolute values of  $\Delta\Delta G_B$  for arginine and citrate for each protein or complex at the given concentration that is significant (Table 1).

**Table 1. Change in free energy of binding ( $\Delta\Delta G_B$ ) between IgG and Protein A, and the preferential interaction coefficients ( $\Gamma_{23}$ ) in the presence of arginine and citrate at pH 4.3. Table modified from Ref. (10)**

Additive	Conc. (M)	$\Gamma_{23}$ (Fc)	$\Gamma_{23}$ (Protein A)	$\Gamma_{23}$ (complex)	$\Delta\Gamma_{23}^a$	$\Delta\Delta G_B$ (kcal/mol)
arginine	0.5	-3	6	0	-3	1.5
arginine	2.0	-37	-3	-47	-7	3.6
citrate	0.5	9	3	15	3	-2.7
citrate	2.0	27	3	34	4	-3.3

<sup>a</sup>  $\Delta\Gamma_{23} = \Gamma_{23}^{\text{complex}} - \Gamma_{23}^{\text{Fc}} - \Gamma_{23}^{\text{Protein A}}$ .

From the changes in free energy of binding ( $\Delta\Delta G_B$ ) of Fc domain of IgG and Protein A in the presence of arginine and citrate solution (Table 1), it is clear that the cosolutes affect  $\Delta\Delta G_B$  in different ways. According to the results of the simulation (Table 1) at pH 4.3 arginine weakens the free energy of binding of Fc region of IgG for Protein A by 1.5 to 3.6 kcal/mol. This decrease in  $\Delta\Delta G_B$  would be expected to weaken the binding affinity of Protein A and Fc region of the IgG 10 to 100 fold (Table 1). This weakened affinity would also be expected to increase the rate of dissociation of the complex. In contrast at pH 4.3 citrate affect on  $\Delta\Delta G_B$  and thus antibody elution is the opposite, citrate enhances  $\Delta\Delta G_B$  from -2.7 to -3.3 kcal/mol. A more negative  $\Delta\Delta G_B$  suggests an enhanced the binding affinity of the Protein and IgG complex by 100 to 1000 fold (Table 1), which would be expected to hinder dissociation.

For the additive arginine the change in the preferential interaction coefficients at both concentrations at pH 4.3 for the complex ( $\Delta\Gamma_{23}^{\text{complex}}$ ) is less than zero (negative) and this implies the dissociated state is favored (Table 1). Thus arginine prefers to associate with the dissociated protein over the complex (Table 1). In contrast the  $\Delta\Gamma_{23}$  for citrate at both concentrations is greater than zero (positive) at pH 4.3 (Table 1), which implies citrate favors association to the complex. From a chromatography standpoint this means that in the presence of arginine antibody (IgG) elution (dissociation from Protein A) is facilitated and aggregation is disfavored. Conversely, in the presence of citrate elution (dissociation from Protein A) is not facilitated (as the complex association is enhanced) at pH 4.3 and induces aggregation.

### 3.3. Residues Involved in Complex Formation

As previously disclosed the residues involved in binding Protein A to the Fc region of the IgG have been identified (10). Briefly, the residues involved in binding the Fc region are concentrated in three distinct regions of the protein (see the Appendix Figure A2a). Whereas the residues from Protein A are glittered all over the surface of the protein (see the Appendix Figure A2b) (10). The key residues were defined as those that displayed a large change in solvent accessible surface area ( $\Delta$ SSA) upon binding, as the residue becomes more buried during complex formation (25).

The identification of residues involved in binding was determined using a calculation of the solvent accessible surface area (SAA) for the residues of the [Protein A – Fc] complex, were compared to the SAA residues for the Fc domain and Protein A. Providing residues at the interface of the Protein A – Fc complex that became buried upon association. Subtraction of the two areas (see the Appendix) provides the residues involved in binding (see the Appendix, Figure A2) (10).

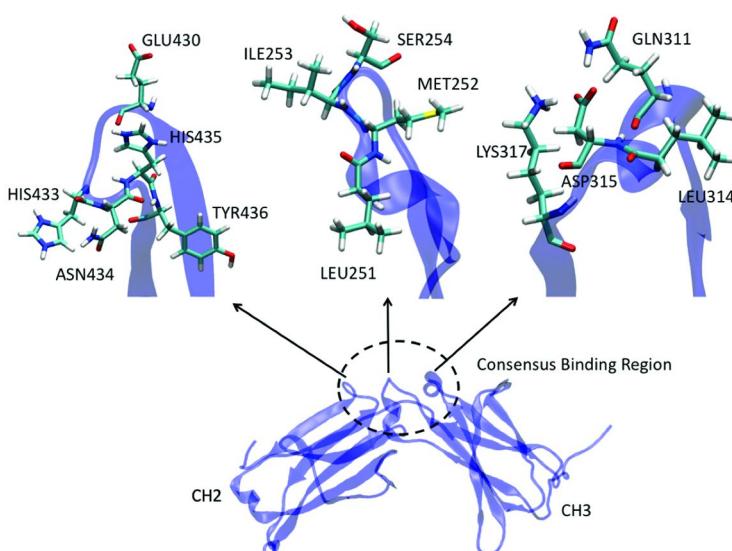
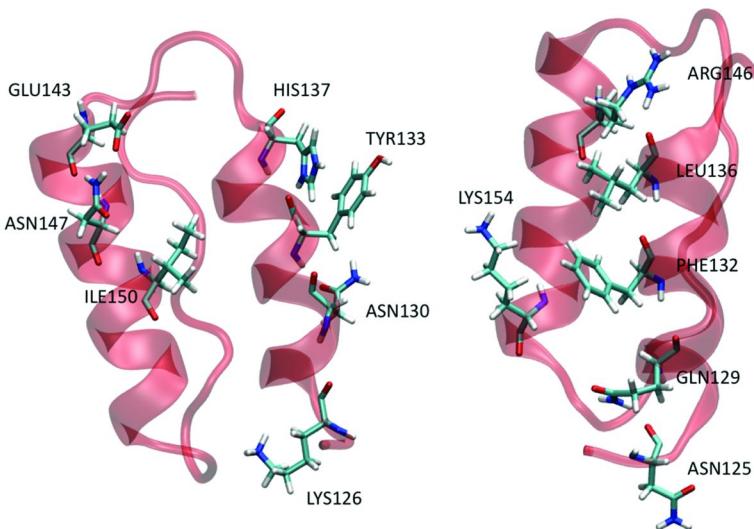


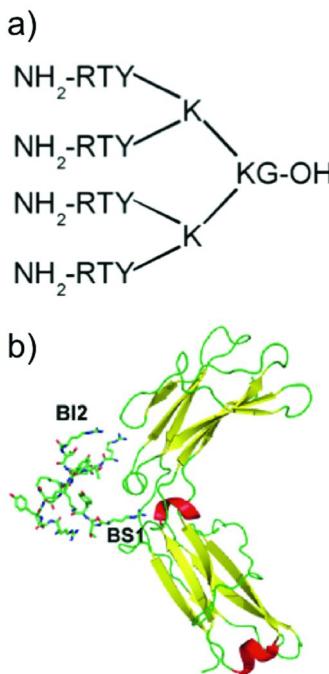
Figure 6. Closer look at the amino acid residues from the Fc domain of IgG that interact with Protein A. From reconstructed dimer (PDB code: 1FC2 and 1HZH) (4, 10, 12). Color code: ribbon-and-bow backbone, purple; C-atoms, cyan; N-atoms, blue; O-atoms, red; S-atoms, yellow; H-atoms, white. Figure modified from reference (10).

According to the X-ray crystal structure of the complex used for the molecular mechanics simulations, the residues in the three regions of the Fc domain (Figure 2a SI) that are involved in Protein A binding are all highly charged, polar, or aromatic (Figure 6). Charged residues in the binding site include Glu430, His433, His435, Lys317, and Asp315 (Figure 6). Polar residues include Asn434, Ser254, Met252, Gln311, and aromatic groups include Tyr436 and the fore mentioned histidines (Figure 6). Some residues are non-polar such as isoleucine and leucine.



*Figure 7. Closer look at the amino acid residues from fragment B of Protein A that interact with the Fc domain. From reconstructed dimer (PDB code: 1FC2 and 1HZH) (4, 10, 12). Color code: ribbon-and-bow backbone, pink; C-atoms, cyan; N-atoms, blue; O-atoms, red; S-atoms, yellow; H-atoms, white. Figure modified from reference (10).*

The key residues involved in Fc binding to the surface of Protein A have also been identified from experimental data to be polar, charged, and aromatic residues. These residues include Arg146, Lys154, Tyr133, Phe132, His137, and Glu143. Thus as would be anticipated the residues involved in binding from Protein A are also highly polar, charged or aromatic (Figure 7). In fact polar and aromatic groups were found to comprise 75% of the interfacial area. In total as these are the residues at the interface of the Protein A and Fc interaction, it is presumably these residues that are somehow involved in arginine and citrate binding during elution of the antibody. Analysis of the structure provides insights into which of the interactions are disrupted at the interface of Protein A and the antibody. As the residues were found to be mostly polar, charged or aromatic it would suggest, as would be anticipated, that the interactions to the charged cosolutes are most likely ionic in nature, however cation-pi interactions may also play a role in binding due to the presence of guanidinium group in the arginine.



*Figure 8. Protein A mimetics. (a) Schematic view of an arginine like dendrimer; (b) a known Protein A mimetic reported by Fassiana and co-workers binds with high affinity to human IgG (C $\gamma$ 2/C $\gamma$ 3 interface of the Fc domain). Figure modified from Ref. (23).*

### 3.4. Determination of the Contact Coefficient

To further identify a amino acid preference for each cosolute the contact coefficient (CC) was determined for each amino type exposed at the Protein A - Fc interface for both arginine and citrate (see the Appendix Figure A3) (10). The contact coefficient is the ratio of the local cosolute concentration to the bulk cosolute concentration around each amino acid (1, 26). A plot of contact coefficient as a function of amino acid demonstrate as expected that citrate prefers positively charge, and some aliphatic residues over negatively charged residues (see the Appendix, Figure A3) (10). As expected arginine prefers negatively charged groups as well as aromatics, presumably the latter form cation-pi interactions, and polar residues. The results further demonstrate that if the interface is comprised of polar and aromatic groups, arginine will associate at the interface and consequently lower the binding between Fc and Protein A.

### 3.5. Designing New Protein A Ligands and Eluents

The sum total of all the results presented suggest that ligands or eluents based on the arginine scaffold would be useful for the purification of proteins or antibodies using Protein A chromatography or new modified ligands. Specifically, using an arginine as a ligand instead of Protein A, one such example was reported by Fassiana and co-workers and is shown in Figure 8 (27). Where a dendrimer containing arginine arms is used as a modified Protein A mimetic dendrimer. Also the selectivity of Protein A could be further enhanced with the addition of an arginine moiety. To design an eluent with better elution properties than arginine would require that the eluent interact with polar, aromatic and charged amino acid residues and also reduce aggregation.

## Conclusions

In conclusion molecular dynamics simulations show that the cosolutes arginine and citrate work in opposite ways at pH 4.3 to elute antibodies from Protein A resin. Arginine lowers the free energy of binding between Protein A and antibody, thereby facilitating elution. Citrate strengthens the free energy of binding, thereby decreasing elution. On a molecular level, arginine interacts favorably with the dissociated state as compared to the associated state and citrate shows an opposite behavior. Arginine interacts favorably with residues at the interface between Fc and Protein A and citrate does not. Further, arginine interacts strongly with key residues involved in the binding between Protein A and antibody. The results shed light on optimization parameters that could be used for the rational design of eluents and new ligands.

## Appendix

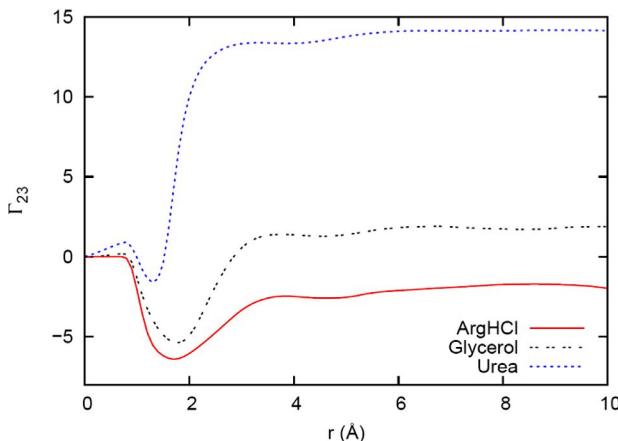


Figure A1. Preferential interaction coefficients for arginine hydrochloride (red line), glycerol (black dotted line), and urea (blue dotted line) as a function of distance from the protein surface as determined from molecular mechanics simulations1 with Protein RNase T1. For derivations see ref (10).

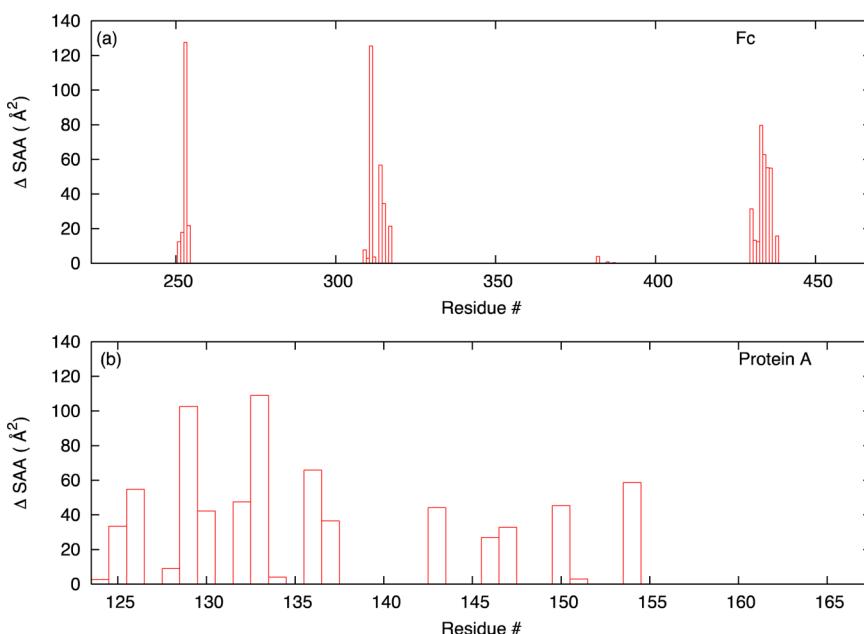


Figure A2. Identification of residues involved Protein A and Fc binding as determined from  $\Delta$ SSA simulations when they are in the dissociated state as compared to the associated state. (a) Fc and (b) Protein A.

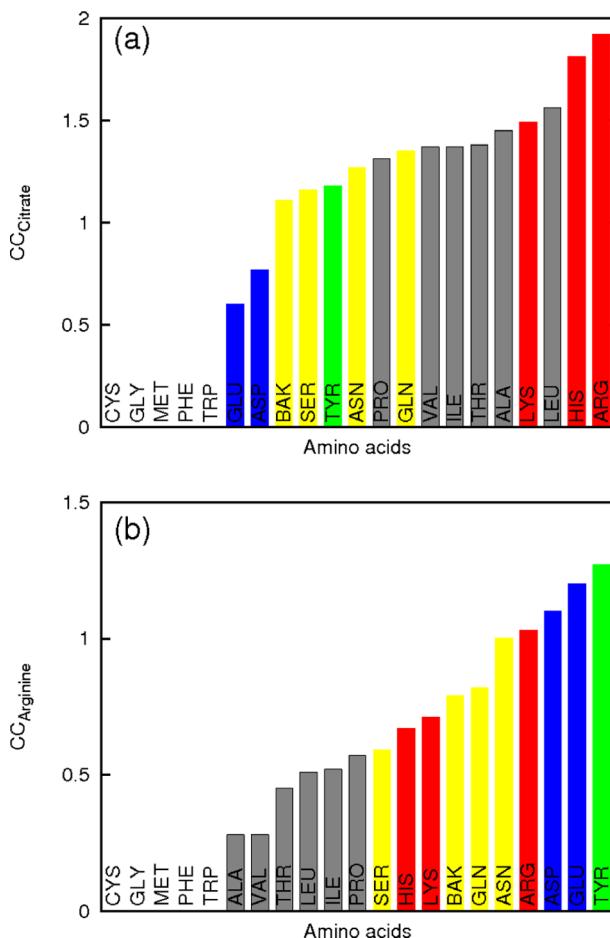


Figure A3. Contact coefficient for each amino acid at the Protein A – Fc interface. (a) 2.0 M citrate; (b) 2.0 M arginine solution. Color code: positively charged residues, red; negatively charged residues, blue; polar residues, yellow; aliphatic residues, grey; aromatic residues, green.

## References

1. (a) Low, D.; O’Leary, R.; Pujar, N. S. Future of antibody purification. *J. Chromatogr., B* **2007**, *848*, 48–63. (b) Hacker, D. L.; Nallet, S.; Wurm, F. M. Recombinant protein production yields from mammalian cells: past, present, and future. *Biopharm. Int. Suppl.* June 2008, *21*.
2. Farid, S. S. Process economics of industrial monoclonal antibody manufacture. *J. Chromatogr., B* **2007**, *848*, 8–18.
3. Clonis, Y. D. *J. Chromatogr., A* **2006**, *1101*, 1–24.
4. Deisenhofer, J. *Biochemistry* **1981**, *20*, 2361–2370.

5. Typical free energy of binding (DDG<sub>B</sub>) for the [Protein A – antibody] complex is -9.8 kcal/mol and that complicates subsequent antibody elution. For reference see: Arakawa, T. Philo; Tsumoto, J. S.; K. Yumioka, R.; Ejima, D. *Protein Expression. Purif.* **2004**, *36*, 244–248.
6. Terman, D. S.; Bertram, J. H. *Eur. J. Cancer Clin. Oncol.* **1985**, *21*, 1115–1122.
7. Rudolph, R.; Fishcer, S. U.S. Patent 4,933,434, 1994.
8. Lange, C.; Rudolph, R. *Curr. Pharm. Biotechnol.* **2009**, *10*, 408–414.
9. (a) Arakawa, T.; Philo, J. S.; Tsumoto, K.; Yumioka, R.; Ejima, D. *Protein Express. Purif.* **2004**, *36*, 244–248. (b) Ejima, D. Yumioka, R. Tsumoto, K.; Arakawa, T. *Anal. Biochem.* **2005**, *345*, 250–257. (c) Tsumoto, K. Umetsu, M. Kumagai, I. Ejima, D. Philo, J. S.; Arakawa, T. *Biotechnol. Prog.* **2004**, *20*, 1301–1308.
10. Shukla, D.; Zamolo, L.; Cavallotti, C.; Trout, B. L. *J. Phys. Chem. B* **2011**, *115*, 2645–2654.
11. (a) Shukla, D. Shinde, C.; Trout, B. L. *J. Phys. Chem. B* **2009**, *113*, 12546–12554. (b) Shukla, D.; Trout, B. L. *J. Phys. Chem. B* **2010**, *114*, 13426–13438.
12. Saphire, E. O.; Parren, P. W. H. I.; Pantophlet, R.; Zwick, M. B.; Morris, G. M.; Rudd, P. M.; Dwek, R. A.; Stanfield, R. L.; Burton, D. R.; Wilson, I. A. *Science* **2001**, *293*, 1155–1159.
13. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
14. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
15. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
16. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
17. Bates, R. G.; Pinching, G. D. *J. Am. Chem. Soc.* **1949**, *71*, 1274–1283.
18. Kamath, G.; Guvench, O.; MacKerell, A. D. *J. Chem. Theory Comput.* **2008**, *4*, 1990.
19. Guvench, O.; MacKerell, A. D. *J. Mol. Model.* **2008**, *14*, 667–679.
20. Inoue, H.; Timasheff, S. N. *Biopolymers* **1972**, *11*, 737–743.
21. Courtenay, E. S.; Capp, M. W.; Anderson, C. F.; Record, M. T. *Biochemistry* **2000**, *39*, 4455–4471.
22. (a) Baynes, B. M.; Trout, B. L. *J. Phys. Chem. B* **2003**, *107*, 14058–14067. (b) Timasheff, S. N. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7363–7367. (c) Timasheff, S. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9721–9726.
23. Vagenende, V.; Yap, M. G. S.; Trout, B. L. *J. Phys. Chem. B* **2009**, *113*, 11743–11753.
24. Vagenende, V.; Yap, M. G. S.; Trout, B. L. *J. Phys. Chem. B* **2009**, *113*, 16268–16275.
25. SAA is the surface area obtained from rolling a probe sphere on the van der Waals spheres of a CPK model of the molecule, and can be determined using the Lee and Richards surface area calculation method in the CHARMM package.<sup>14</sup> We used a probe sphere with radius equivalent to a water

molecule, 1.4 Å. Average surface areas from the 50 ns simulation of the complex, Fc, and Protein A in 0.5 M arginine solution are used to calculate ΔSAA values. In the Fc domain, the  $\beta$ -turns present at the junction of CH2 and CH3 chains (Figure 4), which form the consensus binding site, are involved in complex formation. In protein A, residues present at the interface are located at the two helices of the B-fragment of Protein A. Figure 2 SI shows Protein A residues with ΔSAA greater than 20 Å, the main contributions to the interfacial area. Although the exact surface areas are different, the key residues involved in complex formation identified using ΔSAA are the same as Li *et al.* characterized using X-ray crystallography. See: Li, R.; Dowd, V.; Stewart, D. J.; Burton, S. J.; Lowe, C. R. *Nat. Biotechnol.* **1998**, *16*, 190–195.

26. Stumpe, M. C.; Grubmüller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126–16131.
27. Moiani, D.; Salvalaglio, M.; Cavallotti, C.; Bujacz, A.; Redzynia, I.; Bujacz, G.; Dinon, F.; Pengo, P.; Fassina, G. *J. Phys. Chem. B* **2009**, *113*, 16268–16275.

## Chapter 6

# One Resin, Multiple Products: A Green Approach to Purification

**Ekta Mahajan,\* Jay Werber, Kapil Kothary, and Tina Larson**

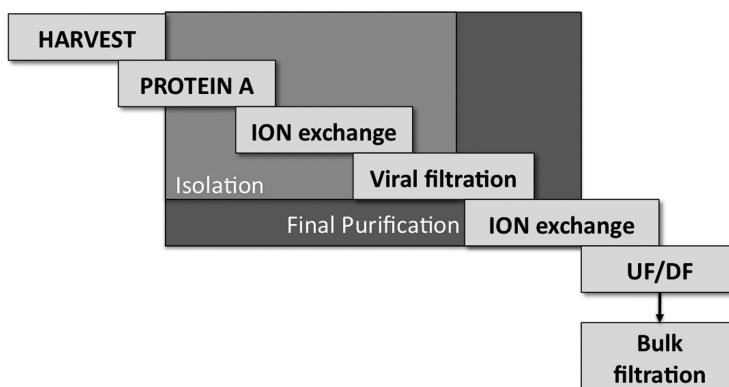
**Pharma Technical Development Engineering, Genentech, 1 DNA Way,  
South San Francisco, California 94080**

**\*E-mail: mahajan.ekta@gene.com. Phone: +1 (650) 467-5890.  
Fax: +1 (650) 225-4436**

Protein A affinity chromatography is a key purification step used during the purification of recombinant monoclonal antibodies (mAbs) harvested from cell culture fluid (HCCF). During this purification process typically a single Protein A resin is dedicated to purify a specific mAb of interest. For clinical manufacturing and pilot plant runs this can result in significant resin underuse, such that the Protein A resin is only used 10% of its potential lifetime. Herein we demonstrate that significant cost savings can be achieved (annually) if the Protein A resin is reused for multiple products. In this study, a cleaning procedure called the *MabSelect SuRe™ Campaign Changeover Procedure* (MSSCCP) was developed on lab-scale to reduce protein carryover during the reuse of the Protein A resin for purification of multiple products. Use of the MSSCCP cleaning procedure results in less than 1 ppm carryover of intact IgG into subsequent purification samples. This low protein carryover is  $10^3$  fold less protein carryover than that set in safety margins, and demonstrates that the same Protein A resins can be used to purify multiple products. The reuse procedure was successfully implemented on lab scale, and on pilot plant scale for the production of mAb drug substances.

## 1. Introduction

Recombinant monoclonal antibodies (mAbs) are used in medicine and diagnostics (1). Industrially, recombinant mAbs are bioproduced in living cells, such as Chinese Hamster Ovary (CHO) cells (2). Once produced, the mAb of interest must be isolated from the cellular and media components used for its production. This purification process has two main steps (a) the primary isolation process, which is followed by (b) the final purification process (Figure 1). The primary mAb isolation process begins after the cells are harvested for the mAb of interest. Once the mAb is harvested the product pool contains the mAb of interest as well as cellular components (media components, proteins, DNA) and viruses that may be present during the mAb production process. In the first chromatography step, the product pool is run through a Protein A affinity column (Step 1). The purpose of the Protein A affinity step is to remove media components, cellular debris, and putative viruses. Following Step 1, the product pool, containing the mAb, is further purified over an ion exchange column (IEX, Step 2). Step 2 serves to remove additional contaminants, such as aggregates and DNA. After IEX chromatography the last step of the primary isolation process includes a step to remove viruses using a virus reduction filter (Step 3). Typically after Step 3 a final purification of the mAb is performed with a second ion exchange step (Step 4), to remove any residual CHO proteins (CHOP). After final purification, Ultrafiltration/Diafiltration (UF/DF, Step 5) is performed to remove small molecules, concentrate the mAb, and exchange the buffer to formulate the purified mAb into its final formulation buffer. This is followed by a bulk filtration step, which ensures sterility of the mAb pool (Step 6).



*Figure 1. Outline of a Protein A-based mAb purification process. Protein A affinity chromatography (Step 1) is the first column purification step in the isolation process. Step 1 is followed by an ion exchange chromatography step (Step 2) then a viral filtration step (Step 3). In the final purification a last ion exchange column (Step 4) is run, which is followed by Ultrafiltration/Di-filtration (Step 5) and a final Bulk filtration step (Step 6).*

The use of Protein A affinity chromatography in industrial mAb purification is commonplace as it is efficient, scalable, and reproducible (3). However, Protein A resin costs are significant, comprising a substantial portion of the raw material costs in MAb manufacturing (4). This expense is further exacerbated by resin underuse, such that a single packed Protein A column is used only 10% of its potential lifetime (in the pilot plant and during clinical production) (2). In order to reduce these costs we sought to reuse our Protein A resin for multiple different mAb products. Protein A resin reuse for multiple products is not a common practice as reuse can result in protein carryover, not only from previous runs, but also from previously purified products. Thus we sought an efficient cleaning process that would enable reuse. The motivation for this work was to identify a process that would not only save money, space, and time, but also be environmentally friendly. The financial savings from reuse of Protein A resins was obvious, but time savings are obtained from the avoidance of repacking columns, for every new mAb synthesized. Reuse of Protein A resin is also cleaner for the environment, as there is less Protein A resin that is wasted, stored, or shipped. Note a typical MabSelect<sup>TM</sup> SuRe resin can be used up to 250 cycles (times) (4, 5). However, on pilot plant scale for a typical clinical or toxicology run, a Protein A column is only used at total 3-4 runs (18-30 cycles), wasting anywhere from 220-232 cycles (2). In the current project, the reuse of MabSelect<sup>TM</sup> SuRe resin for multiple CHO products on lab and pilot scale was optimized and enabled. To reduce levels of mAb carryover from previous purifications to acceptable levels an improved Protein A resin cleaning procedure to be used between mAb purification runs was identified and validated. This was achieved by addressing the following: (a) quantification of the amount of pre-cleaning protein carryover (if any) from previous purifications into subsequent purifications using the same Protein A resin, and (b) identification of a method to clean a Protein A affinity resin before or after use such that multiple products could be purified over the same Protein A resin with limited protein carryover and no safety concerns.

### 1.1. The Economic Impact of Reusing Protein A Resins

The economic impact of reusing our Protein A resins in a mAb purification process was determined to assess if an economic benefit existed. The assessment included cost savings for high demand, low demand, current use and reuse of in house Protein A resins for up to 250 cycles. Prior to implementation, it was estimated that two to five million US dollars would be saved over the next seven years by simply reusing the Protein A resins (Figure 1A). (In just two years since the implementation, actual savings have already reached approximately two million dollars, without accounting for the reduced labor costs. A cost saving that validates the use of a cleaning process to reuse Protein A resins). Since there was a significant economic benefit to Protein A resin reuse, the next step was to quantify the levels of pre-cleaning protein carryover.

## 1.2. Quantification of Protein Carryover and Risk

In an effort to quantify protein carryover from sample to sample a pre-cleaning test run was performed on lab scale using a standard Protein A affinity column ( $0.66 \times 20$  cm). This run was called a ‘mock run’ as the process was run according to a *Standard Purification Procedure* (Appendix) except the load cycle was simulated with the equilibration buffer. The elution pool was collected, as per a typical Protein A process, and analyzed to determine the presence of protein. The analysis revealed that 20-30 ppm of protein carryover is present in the ‘mock elution’ (Figure 2) in the absence of any additional column cleaning. The result was confirmed with a second run (Figure 2).

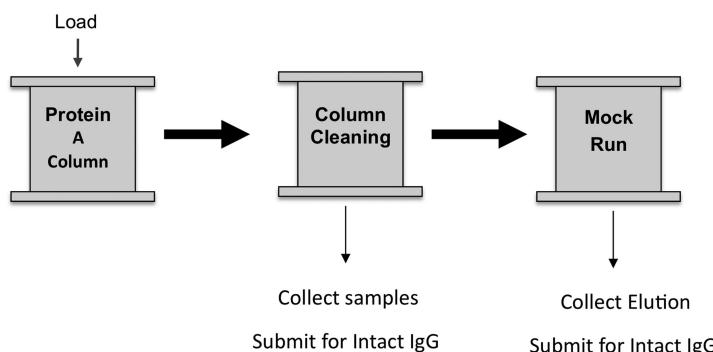


Figure 2. Schematic view of the experimental protocol for the determination of protein carryover after sample loading and purification on Protein A column followed by column cleaning.

In order to determine safe carryover levels, a risk assessment was conducted to determine acceptable immunoglobulins (IgG) and protein carryover levels in Genentech mAbs, and a substance-specific Acceptable Daily Exposure (ADE) for IgG was calculated. An ADE is defined as a substance-specific dose that is unlikely to cause an adverse health event or undesirable physiological effect if an individual is exposed to this dose or to a lower dose over a lifetime (6). In addition to an ADE, an Estimated Daily Intake (EDI) for IgG was also determined based on the amount of IgG administered per dose. A comparison of the ADE to EDI, resulted in a “worst case” x-fold safety margin (7). The “worst-case” safety margin is the highest value of IgG carryover allowed from a previous sample, and this value is set at  $10 \mu\text{g mAb A/ ml mAb B}$  or 1000 ppm. Where mAb A is the carried over mAb and mAb B is the desired mAb of interest (6, 7).

The pre-cleaning carryover results (20-30 ppm) obtained in the ‘mock run’ without additional cleaning was well below the established safety limit of 1000 ppm, set by the risk assessment. However, it was decided to err on the side of

caution, as the limit for clinical manufacturing would be lower than 1000ppm. The goal of the project was to develop cleaning procedure that can be transferred to clinical manufacturing. Hence, a cleaning procedure was identified to minimize carryover to less than 1 ppm. After careful optimization the optimum cleaning strategy was based on (a) a static hold and (b) pH cycling. The addition of a static hold procedure in the cleaning process allowed for extra residence time without using extra buffer. Increased residence time likely aides with mass transfer, and effectively serves to extract any remaining protein on the column into the buffer. Alternation between an acidic and basic buffer called pH cycling, enhances protein extraction and thus effectively washes the column. The optimal cleaning conditions included those buffers that were already used as elution and regeneration buffers. The 'Elution buffer' was 0.15 M AcOH (pH 2.9) and 'Regeneration buffer' for cleaning was 0.1 N NaOH (pH 12). The choice of buffers was not serendipitous, rather based on their respective properties (8). For example, the 'Elution buffer' (0.15 M AcOH, pH 2.9) was used to wash bound IgG from the Protein A resin. Sodium hydroxide solubilizes proteins and nucleic acids (all components of the production process) by denaturation and cleavage of the proteins into small fragments. In addition, sodium hydroxide destroys endotoxins and regenerates the resin. As neither of these conditions are incompatible with the resin, and were already used for purification of in house mAbs, their use was also economical.

## 2. Materials and Methods

### 2.1. Resin Selection

MabSelect<sup>TM</sup> SuRe Protein A affinity resin was chosen for the optimization as it has a large working pH range (pH 3-12), and is stable under basic conditions, without loss of binding capacity. Thus it was compatible with current Elution (0.15 M AcOH) and Regeneration (0.1 N NaOH) buffers for cleaning.

Other resins, such as ProSep<sup>®</sup> vA, were previously investigated. Briefly, several different cleaning agents were investigated to clean the ProSep<sup>®</sup> vA columns. However the majority of the conditions showed similar performance (Figure 2A). Screening of a variety of buffers followed by consecutive 'mock runs,' as outlined in the *Standard Purification Procedure* (Appendix), resulted in decreased protein carryover from sample to sample (Figure 2A). Increased elution flow rate also effectively cleaned the column (data not shown). The major findings of this study were that static holds and pH cycling contributed more significantly to the reduction of protein carryover compared to other variables tested. Although some of the cleaning procedures did reduce protein carryover on ProSep<sup>®</sup> vA, the reduction was not sufficient to warrant its usage on pilot, or larger scale. Nevertheless, the results from the pH cycling and static hold experiments proved useful in the optimization of the cleaning procedure on MabSelect<sup>TM</sup> SuRe resin.

**Table 1. Cleaning procedures investigated to reuse a MabSelect<sup>TM</sup> SuRe column<sup>a</sup>**

Entry	Condition	<i>Intact IgG carryover (ng carryover/mg product)<sup>f</sup></i>
1	Regeneration buffer <sup>b</sup> (6 CV) Equilibration buffer <sup>c</sup> (5 CV)	0.98 ppm
2	Regeneration buffer <sup>b</sup> (5 CV) Elution buffer <sup>d</sup> (3 CV) Regeneration buffer <sup>b</sup> (5 CV) Equilibration buffer <sup>c</sup> (5 CV)	1.80 ppm
3	Elution buffer <sup>d</sup> (3 CV) Regeneration buffer <sup>b</sup> (5 CV)	<1 ppm
4	Elution buffer <sup>d</sup> (2 CV) 30-minute static hold <sup>e</sup> Elution buffer <sup>d</sup> (2 CV) Regeneration buffer <sup>b</sup> (4 CV)	< 10 ppm
5	Elution buffer <sup>d</sup> (2 CV) 30-minute static hold <sup>e</sup> Elution buffer <sup>d</sup> (2 CV) Regeneration buffer <sup>b</sup> (2.5 CV) 30-minute static hold <sup>e</sup> Regeneration buffer <sup>b</sup> (2.5 CV)	<1 ppm
6	Equilibration buffer <sup>c</sup> (2 CV) 30-minute static hold Equilibration buffer <sup>c</sup> (2 CV) Elution buffer <sup>d</sup> (2 CV, pH 2.8) 30-minute static hold Elution buffer <sup>d</sup> (2 CV) Regeneration buffer <sup>b</sup> (2 CV) 30-minute static hold Regeneration buffer <sup>b</sup> (2 CV)	<3 ppm
7	Equilibration buffer <sup>c</sup> (4 CV) 6 cycles of the following: Elution buffer <sup>d</sup> (3 CV) 10-minute static hold Elution buffer <sup>d</sup> (1 CV) Regeneration buffer <sup>b</sup> (3 CV) 10-minute static hold Regeneration buffer <sup>b</sup> (1 CV)	<0.3 ppm
8	6 cycles: Elution buffer <sup>d</sup> (3CV) 15-minute static hold Elution buffer <sup>d</sup> (1 CV) Regeneration buffer <sup>b</sup> (3 CV) 15-minute static hold Regeneration buffer <sup>b</sup> (1 CV) Storage buffer <sup>g</sup> (3 CV)	<0.5 ppm

*Continued on next page.*

**Table 1. (Continued). Cleaning procedures investigated to reuse a MabSelect<sup>TM</sup> SuRe column<sup>a</sup>**

Entry	Condition	Intact IgG carryover (ng carryover/mg product) <sup>f</sup>
	15-minute static hold Storage buffer <sup>b</sup> (1 CV)	

CV = column volume. <sup>a</sup> 0.66 × 20 cm. <sup>b</sup> Regeneration Buffer = 0.1 N NaOH (pH 12). <sup>c</sup> Equilibration buffer = 25 mM Tris, 25 mM NaCl (pH 7.1). <sup>d</sup> Elution buffer = 0.15 M Acetic acid (pH 2.9). <sup>e</sup> static hold = holding the buffer in the column at 0 mL/min flow rate. <sup>f</sup> Carryover determined in a ‘mock run.’ <sup>g</sup> Storage buffer = 100 mM sodium acetate and 2% Benzyl alcohol (pH 5).

## 2.2. Analytical Methods

The antibody concentration of HCCF was determined using a 2.1 × 30 cm POROS column (Applied Biosystems, Foster City, CA) on an Agilent 1100 HPLC (Agilent Technologies, Santa Clara, CA). Buffer A (100 mM sodium phosphate, 250 mM sodium chloride, pH 6.3), Buffer B (2% acetic acid, 100 mM glycine), and Buffer C (0.1 M phosphoric acid, 205 CAN) were used, and the total run time was 4.5 min. The protein concentration in the purified pool was measured using the Agilent 8453 (Agilent Technologies, Santa Clara, CA) spectrophotometer at 280 nm. Multi-product enzyme-linked immunosorbent assay methods were used for CHOP, and leached ProA analysis. TaqMan polymerase chain reaction was used for CHO DNA analysis. Total protein was measured using a Capillary Zone Electrophoresis/Laser-Induced Fluorescence Detection (CZE-LIF) assay. Intact antibody and fragmented antibody parts were measured using a generic ELISA assay. SDS/PAGE was performed on 18% Tris-HCl gel.

## 2.3. Quantification of Protein Carryover without Cleaning

The experimental protocol for the determination of mAb carryover is outlined in Figure 2. First, 18 load cycles of a mAb were loaded onto a Protein A affinity column (0.66 × 20 cm, Volume = 6.8 mL) at 30 g/L, and the sample was eluted. The Protein A affinity column was subsequently cleaned, following one of the column cleaning procedures outlined in Table 1. After cleaning, a ‘mock run’ was performed (see Appendix for ‘mock run’). To determine levels of protein or impurity carryover, analytical samples were taken either during the ‘mock run’ at specific time points or during the column wash procedure. Analytical samples collected were adjusted to pH 5 - 5.5 (1.5 M Tris base buffer) and then treated with a detergent (0.1% polysorbate, 0.05% sodium azide) to prevent protein surface adhesion (as this would give a false negative result).

In a first experiment the carryover in the elution pool was first determined for three different mAbs (mAbA, mAbB, and mAbC) purified sequentially on a MabSelect<sup>TM</sup> SuRe Protein A column without intermittent cleaning. The three

purification cycles were loaded at 30 g/L on a Protein A affinity column (0.66 × 20 cm, Volume = 6.8 mL), and the results are shown in Figure 3A. The data are graphed as the amount of intact IgG protein (ng carryover/mg product) carried over from the previous run as a function of elution. According to the graph without intermittent cleaning the highest carryover of the three load cycles was 30–40 ppm (Figure 3A). The results clearly showed that in order to stay below 1 ppm of protein carryover, additional cleaning cycles are needed to recycle the column.

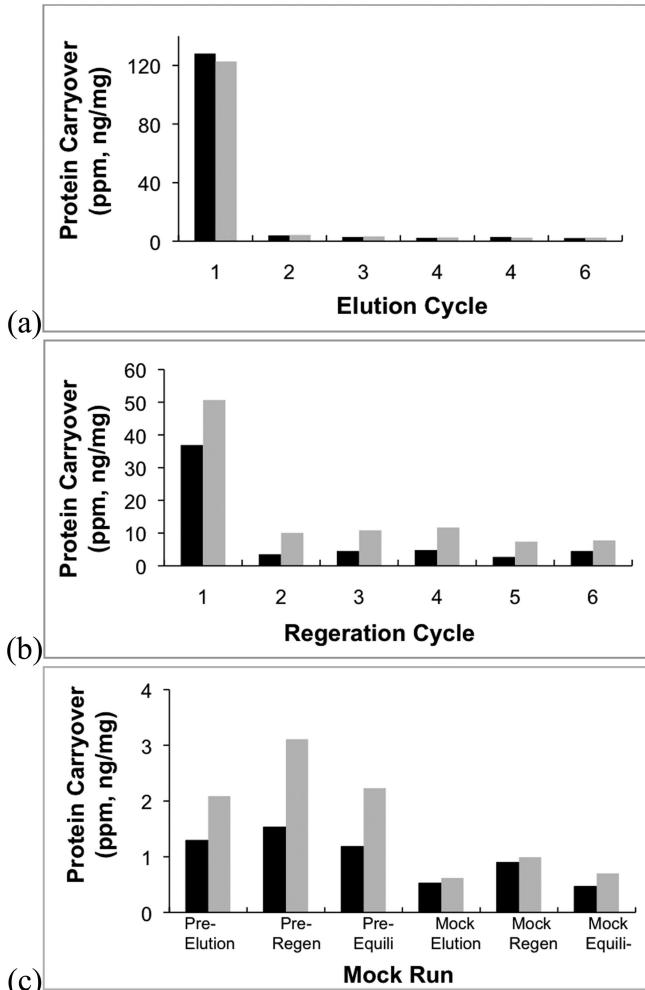
#### 2.4. Optimization of the MabSelect<sup>TM</sup> SuRe Cleaning Procedure (CP)

In an attempt to simplify cleaning procedures by reducing buffer consumption and cleaning times, different combinations of buffers and run times were investigated (Table 1, Entries 1–3). As the levels of carryover never fell below the limit of 1 ppm, it was clear that more rigorous cleaning procedures needed to be identified for lab-scale recycling on the MabSelect<sup>TM</sup> SuRe columns. More rigorous cleaning conditions included the addition of static holds, where the column was held for a defined period of time in a buffer and run at zero flow (Table 1, Entry 4–5). It was found that static holds effectively washed more protein off the resin than flushing with a buffer. Static holds effectively increased the amount of intact IgG washed off the column 5-fold after an elution buffer static hold, and intact IgG was not detected after a Regeneration buffer static hold (Figure 4A). Further, the amount of carryover was significantly reduced in a ‘mock elution’ to less than 10 ppm of intact IgG for Entry 4 (Table 1) and less than 1 ppm of intact IgG is carried over for Entry 5 (Table 1).

Increasing the number of cleanings with static hold time were then investigated with Entries 6–7 (Table 1). Clearly, static hold times with additional cleaning cycles more effectively cleaned the resin than all other previously investigated conditions. Such that after ‘mock runs’ were carried out, less than 3 ppm of carryover for Entry 6 (Table 1) and less than 0.3 ppm of carryover for Entry 7 (Table 1) were detected (Table 1, Figure 5A). The increase in pH cycles eluted more protein during the sharp pH transitions. However, aggregate time with 0.1N NaOH is increased by increasing number of cycles with long static hold which could be detrimental to the resin binding capacity. Since, majority of protein is eluted after first cycle (Figure 3, 5A) for both 30 minute and 10 minute static hold time, the extra static hold time had limited additional benefit. Hence, shorter static hold times with increased cycles was preferred over longer hold times. In addition to our ELISA assays, capillary electrophoresis-Sodium Dodecyl Sulfate analysis (CE-SDS) was performed to ensure fragment clearance through the cleaning cycles. CE-SDS analysis of a ‘mock elution’ after cleaning of a MabSelect<sup>TM</sup> SuRe column using the procedure outlined in Entry 7 (Table 1) for a 94 ng/mL ‘mock elution’ sample revealed that the isolated mAb was over 90% fully intact (Figure 6A).

As a ‘proof of concept’ using the cleaning procedure in Entry 7 (Table 1) an Akta chromatogram (generated during purification run) of a ‘mock elution’ suggested efficient cleaning of the column was achieved when a shift from Elution buffer (0.15 M acetic acid) to Regeneration buffer (0.1 N NaOH) was made. This

was evidenced by successive spikes in the UV-intensity during this pH cycling for each of the 6 cycles (Figure 7A). Taken together, the pH cycling and static holds provide the most ideal cleaning procedures.



*Figure 3. Results for a lab scale purification of mAbA using the optimized cleaning procedure (Entry 7, Table 1) before 'mock runs' to determine protein carryover. (a) Protein carryover (ng/mg protein) as a function of Elution buffer cycle wash; (b) protein carryover (ng/mg protein) as a function of Regeneration buffer cycle wash; (c) protein carryover (ng/mg protein) as a function of the stage in the 'mock run.' Legend: intact IgG (black), Fc fragments (grey).*

Scaling of the optimized cleaning procedure (Entry 7, Table 1) on lab scale for the purification of mAbA on a MabSelect<sup>TM</sup> SuRe Omni fit column (0.66 × 20 cm, Volume = 6.7 mL, 18 cycles of HCCF at 30 g/L) did in fact minimize protein carryover (Figure 3). The results clearly demonstrate that after cleaning the column using the six cycles of Elution buffer (0.15 M Acetic acid), significantly less intact IgG or Fc fragment was detected after each cycle, such that by cycle 6 less than 5 ppm was detected (Figure 3). Similarly, even less intact IgG and Fc fragments (<10 ppm) were detected after each of the six cycle washes with Regeneration buffer (0.1 M NaOH, Figure 3). Further, by the time the ‘mock elution’ was carried out (after pre-elution, pre-regeneration, and pre-equilibration) less than 1 ppm of protein carryover was detected in the ‘mock elution’ sample (Figure 3).

During the course of the cleaning procedure optimization, small amounts of protein came off the column after periods of storage in storage buffer (100 mM Sodium Acetate, 2% benzyl alcohol at pH 5.0). This observation suggested that perhaps the storage buffer could also serve as an efficient cleaning buffer for the MabSelect<sup>TM</sup> SuRe resin. However, subsequent ‘mock runs’ after intermittent column cleaning with the storage buffer did not result in more efficient column cleaning than previously optimized conditions (Entry 8, Table 1, Figure 8A). Further, the addition of this cleaning buffer in the process would make the overall process longer, without additional benefit. Hence, it was decided to proceed with the existing optimized cleaning procedure.

The optimized cleaning procedure (Entry 7, Table 1) was then implemented on a pilot scale for the purification of mAbZ (14 × 20, Volume = 3.23L) as a final test prior to extending the procedure onto mAbs of interest. The results were promising as expected with decreasing intact IgG and Fc detected after each cleaning such that less than 1 ppm protein carryover was detected in the ‘mock elution’ of a ‘mock run.’ This particular pilot run was performed with mAbZ on a MabSelect<sup>TM</sup> SuRe column that had previously been used in nine purification cycles (Figure 4).

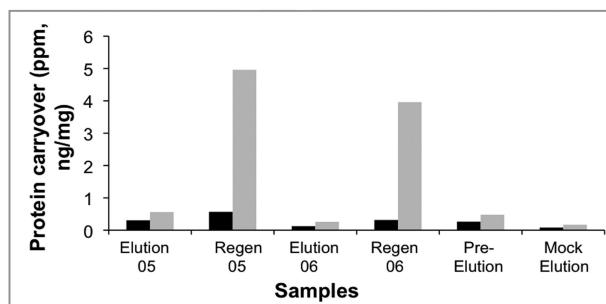


Figure 4. Plot of protein carryover as a function of sample from a 3.23 L pilot scale purification of mAbZ on MabSelect<sup>TM</sup> SuRe column followed by column cleaning with optimized 6 cycle cleaning procedure (Entry 7, Table 1). Legend: intact IgG (black), Fc fragments (grey).

As this cleaning procedure was so effective a total of 5 pilot scale columns were cleaned after implementation. Further analysis of all these pilot scale samples to determine the amount of other impurities such as: leached Protein A (Leached Protein A assay) (9), other proteins (CZE LIF- Total protein, assay) (10), Chinese Hamster Ovary Proteins and DNA (CHOP assay (11) and CHO DNA assay (12)), Fc fragments of an antibody (Human Fc ELISA) (13), and total antibody (Intact Human IgG ELISA) (13) were also performed to verify that the process performed similarly at pilot scale (Table 2). As expected, all detected impurities were well below acceptable limits, and the cleaning procedure can be used for purification of mAbs (Entry 6, Table 2).

**Table 2. Analysis of all samples from pilot scale purification of mAbZ using optimized cleaning**

Entry	Sample	Leached Protein A ppm	CZE-LIF ( $\mu$ g/mL)	CHOP (ppm)	CHO DNA (pg/mL)
1	Elution 05	2.817	<0.25	0.51	<1.00
2	Elution 06	3.06	<0.25	<0.5	<1.00
3	Regeneration 05	17.7	>2.5	<0.5	<1.00
4	Regeneration 06	17.6	>2.5	<0.5	67.93
5	Pre-Elution	<1	>2.5	<0.5	<1.0
6	Mock Elution	<1	<0.25	0.74	<1.0

### 3. Results and Discussion

The previously optimized extended cleaning condition (Entry 7, Table 1) was modified slightly for future studies and incorporated a 15 minute static hold time, instead of a 10 minute one (Figure 4). This new protocol is called *MabSelect SuRe™ Campaign Changeover Procedure* (MSSCCP). The overall process of cleaning the resin took 4.5 hours at 20 column volume (CV)/hour flow rate, and it was run for 6 cycles (6 times). The MSSCCP conditions included pH cycling, between the Elution and Regeneration buffer, and static holds to effectively wash the column. Briefly, the procedure is detailed in Figure 5. The entire process is run a total of 6 cycles in order to thoroughly clean the resin. Finally, the resin was washed with Equilibration buffer (3CV) before storage in storage buffer (5CV, Storage buffer). In order to effectively monitor resin-cleaning, samples were collected after the 15-minute hold times to analyze carryover at each cycle and determine how much protein was removed from the resin at each step and each cycle (Figure 5).

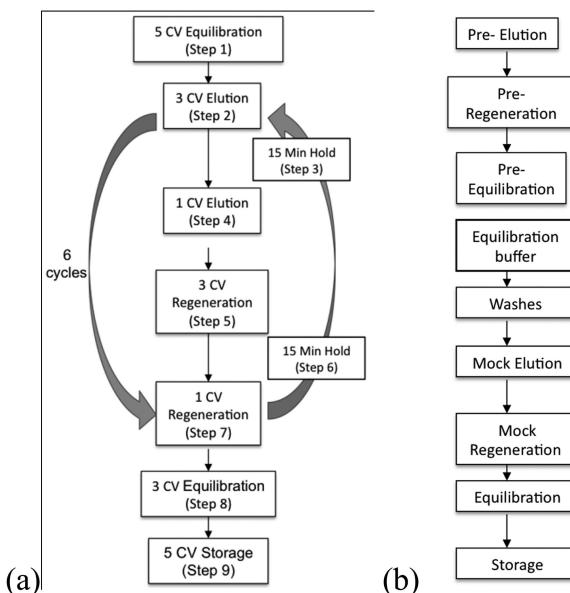


Figure 5. Schematic outline of MSSCCP method, 'optimized cleaning protocol' (a) and a 'mock run' (b). Equilibration buffer = 25 mM Tris, 25 mM NaCl (pH 7.1). Elution buffer = 0.15 M sodium acetate, Regeneration buffer = 0.1 N NaOH.

After the resin was cleaned, a 'mock run' was performed to verify the protein carryover (Figure 5). The 'mock elution' was collected and assayed to determine the amount of carryover and the presence of other impurities (9–13). These other impurities include host cell components, proteins, viruses, or DNA. These assays include a test for intact human immunoglobulin (IgG) using an ELISA (13), human Fc fragment in another ELISA (13), any other protein using a Capillary Zone Electrophoresis/Laser-Induced Fluorescence Detection assay (CZE/LIF) (10), Chinese Hamster Ovary Proteins in a CHOP assay (11), and leached Protein A assay (9). In the 'Intact Human IgG ELISA' and the 'human Fc ELISA', the amount of entire antibody or antibody fragment on the column is quantified; where the former binds to both the fragment antigen-binding region (Fab) and the Fragment crystallizable (Fc) regions, and the latter binds only to human Fc region (13). The CZE-LIF assay can confirm those results by quantifying the total amount of protein in a sample (10). Finally, it is known that Protein A can leach off the resin during runs or during harsh cleaning, negatively impacting binding capacity, thus it is important to determine the amount of leached Protein A (4, 10, 14).

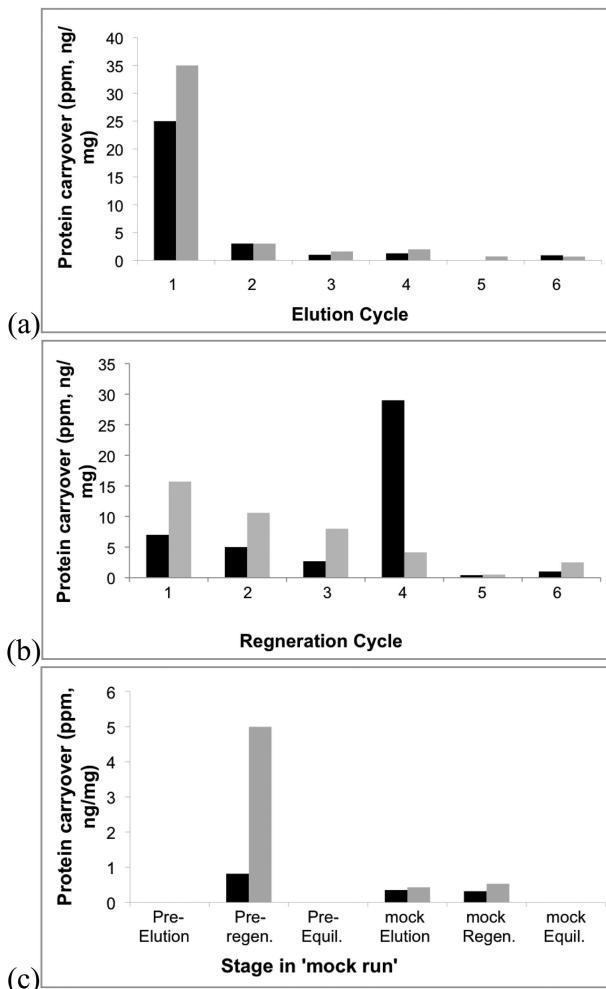


Figure 6. Results for a lab scale purification of mAbC using MSSCCP cleaning procedure (Figure 5) before 'mock run' to determine protein carryover. (a) Protein carryover (ng/mg protein) as a function of Elution buffer cycle wash; (b) protein carryover (ng/mg protein) as a function of Regeneration buffer cycle wash; (c) protein carryover (ng/mg protein) as a function of the stage in the 'mock run'.

To test the efficiency of the cleaning procedure a mAb of interest, mAbC was purified on lab scale on the MabSelect<sup>TM</sup> SuRe column (0.66 × 20 cm, Volume = 6.8 mL) with an AKTA Explorer 100 (as described in 2.2). The protein carryover during and after several cleaning cycles was measured and the results

demonstrated that protein carryover decreased after each cleaning cycle (Figure 6). The intact IgG protein and Fc fragment carryover of mAbC in a subsequent ‘mock run’ decreased from 25 ng/mg intact IgG and 35 ng/mg Fc fragments in the first cycle with Elution buffer to less than 5 ng/mg for both after the sixth cycle with Elution buffer (Figure 6). During and after Regeneration (which followed Elution), significantly more intact IgG and Fc fragments washed off the column until the sixth cycle was reached where levels were reduced cumulatively to less than 5 ng/mg carryover. To test carryover a ‘mock run’ (run after the entire cleaning cycle) was performed and additional IgG and Fc fragments washed off the column in the pre-regeneration, but by the time the ‘mock elution’ process began (where a second mAb would be expected to come off in the reuse process) less than 1 ppm of IgG and Fc fragments were detected (Figure 6). Taken together these results confirm that the MSSCCP conditions are effective cleaning conditions and the total amount of protein carried over from previous runs will be less than 1 ppm when reusing the resin, after using the MSSCCP cleaning procedure.

In order to get a better idea of what type of protein fragments are present per cycle wash, a sample from each cycle was run on a 10% Tris-HCl gel (mAbC) (Figure 7) (15). From cycle 1 to cycle 6 less protein was observed in each successive cycle (decreased band intensity in each lane, Figure 7). Further, lanes with samples from the static hold cycles were more concentrated than during cycle 1-6, with more protein removed after each static hold cycle. These results demonstrate that extended residence times help remove residual protein off the column.

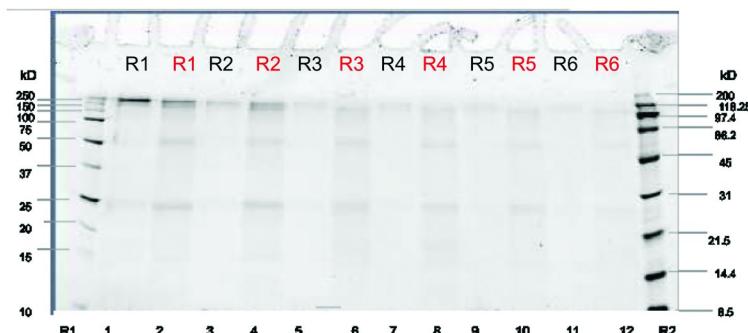


Figure 7. 10% Tris-HCl gel taken at different stages of an optimized cleaning protocol of a MabSelect™ SuRe column (Figure 5). Samples were taken after the purification of mAbC. Note: regeneration samples are concentrated (25 fold) and lanes marked in red contain samples after the 15 min static holds.

This optimized cleaning process was extended to the purification of mAbX at pilot scale on a MabSelect™ SuRe column (13.8 × 20, Volume = 3 L). The results were similar to the results previously seen at lab scale (Figure 3). As seen

on the lab scale runs, protein impurities are removed from the resin in the initial cleaning stages, and their overall concentration decreases after each elution and regeneration cycle, until the sixth cycle is reached (Figure 8). During the resin regeneration the amount of protein that is initially washed off the column is much greater than after the sixth cycle, such that by the time the sixth regeneration cycle is reached, less than 1 ppm of protein impurities are detected (Figure 8).

Purification of mAbY on a MabSelect<sup>TM</sup> SuRe column (20 × 20, Volume = 6.28 L) and subsequent column cleaning with the MSSCCP cleaning protocol (Figure 5), followed by a ‘mock run’ resulted in less than 1 ppm of leached Protein A, less than 0.25 mg/mL (limit of quantification) CZE-LIF, less than 0.5 ppm CHOP, and less than 1.0 pg/mL CHO DNA after the sixth regeneration cycle in the mock elution (Table 3). All impurities were comparable to historical data, and within acceptable limits.

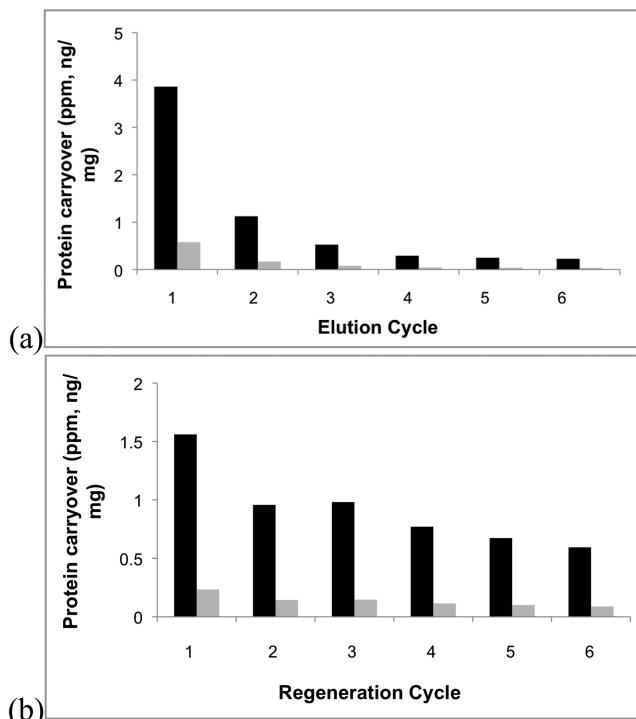


Figure 8. Results for a pilot scale column (3 L) purification of mAbC using the optimized cleaning procedure (Entry 7, Table 1) before ‘mock run’ to determine protein carryover. (a) Protein carryover (ng/mg protein) as a function of Elution buffer cycle wash; (b) protein carryover (ng/mg protein) as a function of Regeneration buffer cycle wash. Legend: intact IgG (black), F<sub>c</sub> fragments (grey).

**Table 3. Analysis of all samples from large scale column (6.28 L) purification of mAbY on a MabSelect™ SuRe column using the optimized cleaning procedure MSSCCP to verify minimal carryover<sup>a</sup>**

Entry	Sample	Leached Protein A (ppm)	CZE-LIF (µg/mL)	CHOP ppm	CHO DNA (pg/mL)
1	Elution 02	6.24	0.35	0.74	<1.00
2	Elution 03	2.72	0.28	0.63	<1.00
3	Elution 04	3.08	0.35	<0.5	<1.00
4	Elution 05	3.01	0.34	<0.5	<1.00
5	Elution 06	3.03	<0.25	<0.5	<1.00
6	Regeneration 06	23.72	>2.5	1.1	27.05
7	Pre-Elution	1.4	<0.25	<0.5	1.12
8	Mock Elution	<1	<0.25	<0.5	<1.0

<sup>a</sup> For information on a particular assay see references (10–13).

In a final test of the robustness of the optimized cleaning protocol MSSCCP (Figure 5), a 6.28 L MabSelect™ SuRe resin that had previously undergone 153 multiproduct load cycles was used to purify a mAb of interest. The carryover from the previously used resin into a ‘mock run’ (Entry 1, Table 4) was compared to the carryover observed from three other different MabSelect™ SuRe resins. The results are outlined in Table 4. Briefly, the old MabSelect™ SuRe large scale multiproduct resin (Entry 1, and 2, Table 4), behaved comparably to a new mAb specific (not a multiproduct) MabSelect™ SuRe resin (Entry 3, Table 4) and a new mAb specific lab scale MabSelect™ SuRe resin (Entry 4, Table 4). According to the results all the Protein A resins provided mAbs in greater than 90% yield, with comparable CHOP, percentage aggregate and leached Protein A (ng/mg). Taken together a multiproduct resin has no negative impact on product impurity profile or the step yield. Pilot scale and lab scale results are also comparable (Entry 4, Table 4).

In addition to the work presented, several more in house mAbs have been purified on multiproduct resins that were cleaned using the MSSCCP procedure. The results have all been very similar and reproducible with total protein levels below 0.25 ppm (assay detection limit) (Table 1A). The results suggest the optimized cleaning procedure MSSCCP is an efficient, reproducible and robust way to clean, regenerate, reuse and recycle multiproduct MabSelect SuRe™ Protein A resins. Use of the MSSCCP cleaning procedure for intermittent Protein A resin cleaning reduces protein carryover to well below established safety margins.

**Table 4. Comparison of mAb product yields and impurity profiles of different MabSelect™ SuRe columns after using the optimized cleaning procedure<sup>a</sup>**

Entry	Column /Scale	Step Op.	Yield (%)	Leached Protein A (ppm)	% ag. <sup>f</sup>	CHOP ppm	CHO DNA (pg/mL)
1	A/400 L <sup>b</sup>	HCCF	-	-	-	1,161,000	59,700
		Protein A	>90	8	3.17	<12000	62
2	B/100 L <sup>c</sup>	HCCF	-	-	-	2,591,520	-
		Protein A	>90	12	2.6	<12000	80
3	C /400 L <sup>d</sup>	HCCF	-	-	2.2	1,115,320	-
		Protein A	>90	10	1.6	<12000	-
4	D <sup>g</sup>	Protein A	>95%	15	N/A	<12000	960

<sup>a</sup> For information on a particular assay see the Appendix. <sup>b</sup> Column A was an old MabSelect™ SuRe multiproduct column used previously in 153 load cycles. <sup>c</sup> Column B was a lab scale run on a newer MabSelect™ SuRe column. <sup>d</sup> Column C was a pilot scale purification on a MabSelect™ SuRe mAb-specific (not multiproduct) column. <sup>e</sup> Exchg. = Exchange. <sup>f</sup> ag. = aggregate. <sup>g</sup> Column D was a newer lab scale mAb specific MabSelect™ SuRe column.

#### 4. Conclusions

We have developed a highly effective *MabSelect SuRe™ Campaign Changeover Procedure* (MSSCCP) cleaning method that allows MabSelect SuRe™ Protein A resin to be used for multi-product purification with no impact on product purity and or loss of resin binding capacity. Data from lab as well as pilot scale experiments suggest that a MSSCCP cleaning protocol that includes 6 cycles of 0.15 M Acetic Acid (Elution buffer) and 0.1 N Sodium Hydroxide (Regeneration buffer) washes and 15 min hold times clean the MabSelect SuRe™ resin to less than 5 ppm of protein carryover in this first mAb cleaning step. The MSSCCP process was successfully implemented on a multi-product Protein A resin (MabSelectSuReTM) on pilot scale, giving further credence of the usefulness of the strategy.

#### Acknowledgments

The authors would like to thank resin reuse team Sam Dalsin, Bijee George, Jimmy Sugahara, Ron Massicotte, Nuno Fontes, Aurelia Safta, Deborah O'Connor, Michelle Butler, Amy Lim, Marc Wong, Ailen Sanchez. We would also like to thank Kevin Ford and Stephen Gomez for providing toxicology assessment, the analytical group for supporting the assays, Terry Hudson, Benedicte Lebreton, Adeyma Arroyo, Philip Lester and Paul Bezy for supporting the project.

## Appendix

### Materials and Methods.

**Equipment and Materials.** **AKTA explorer 100 system** A standard ÄKTA explorer 100 chromatography system from GE Healthcare (Uppsala, Sweden) was used for experimentation (Fig. 1). 0.66cm diameter x 20cm bed height columns (Omnifit) packed with MabSelect SuRe™ (GE Healthcare) Protein A media were used for system evaluation. The system was controlled using UNICORN software (v 5.11). **Affinity resins** MabSelect SuRe resin (GE Healthcare, Uppsala, Sweden) was used in this project as resin of choice because it is composed of a rigid, high-flow agarose matrix and alkali-stabilized protein A-derived ligand. This ligand provides greater stability than conventional protein A-based media under the alkaline conditions used in cleaning-in-place (CIP) protocols. Cleaning can be performed with cost-effective reagents such as sodium hydroxide that may help improve process economy.

**Standard Purification Procedure ('mock elution').** Protein A cycles were run using the following parameters (a) MabSelect SuRe resin with a load capacity of 30g/L resin. (b) The HCCF was loaded at 15 °C (12–18 °C) (all other phases at room temp.) on a 0.66 × 20 cm column. (c) The pool pH was adjusted to pH 5.0 by the addition of 1.5 M Tris base. The buffers used were similar to those used for the batch process. The columns were equilibrated with 25 mM Tris and 25 mM sodium chloride, washed 0.4 M potassium phosphate, eluted with 0.1 N acetic acid (pH 2.9), and regenerated with 0.1 N sodium hydroxide for MabSelect SuRe (2, 4, 15).

### List of buffers

**Buffers.** (a) 0.15 M Acetic Acid (also the elution buffer); (b) 0.1 M Sodium Hydroxide (also the regeneration buffer).

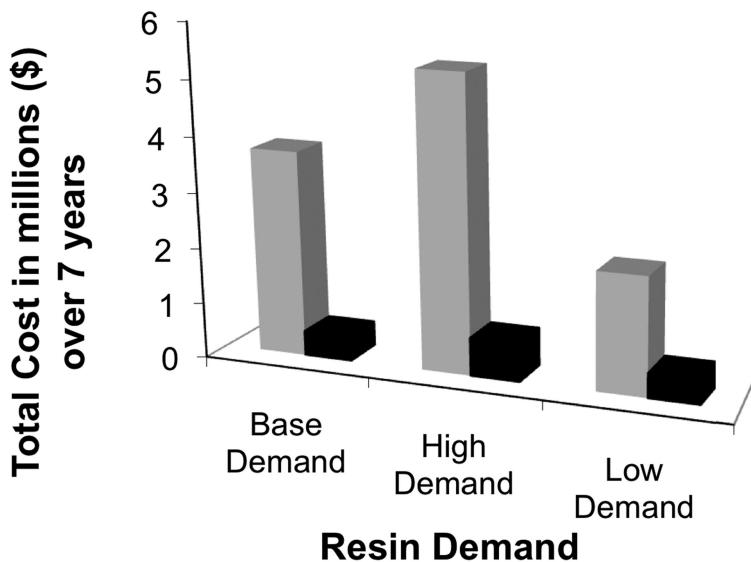
**Cleaning Strategy.** The cleaning procedure is performed at 20CV/hr flow rate. The cleaning procedure was developed based on two factors (a) pH cycling and (b) Static hold times.

**Cleaning Procedure (MSSCCP).** The MSSCCP is performed in between any two runs. Typically for Protein A resin each run consists of several cycles. It should be noted that the MSSCCP procedure is not performed in between cycles (only between runs). The following steps are involved in the resin cleaning procedure: First the column is conditioned with 5CV equilibration buffer. Then the following sequence of steps (1-6) is repeated 6 ×:

- (1) 3CV with 0.15M Acetic Acid
- (2) 15 minutes of static hold time,
- (3) 1CV with 0.15M Acetic Acid,
- (4) 3CV of 1M Sodium Hydroxide
- (5) 15 minutes of static hold time,
- (6) 1CV of 1M Sodium Hydroxide

Once the 6x cleaning is done the column is again equilibrated with 3CV of equilibration buffer and finally stored by running 5CV of storage buffer.

**Analytical Methods.** The antibody concentration of HCCF was measured using a  $2.1 \times 30$  cm POROS column (Applied Biosystems, Foster City, CA) on an Agilent 1100 HPLC (Agilent Technologies, Santa Clara, CA). Buffer A (100 mM sodium phosphate, 250 mM sodium chloride, pH 6.3), Buffer B (2% acetic acid, 100 mM glycine), and Buffer C (0.1 M phosphoric acid, 205 CAN) were used, and the total run time was 4.5 min. The protein concentration in the purified pool was measured using the Agilent 8453 (Agilent Technologies, Santa Clara, CA) spectrophotometer at 280 nm. Multi-product enzyme-linked immunosorbent assay methods were used for HCP, and leached ProA analysis. TaqMan polymerase chain reaction was used for CHO DNA analysis. Total protein was measured using a CZE-LIF assay. Intact antibody and fragmented antibody parts were measured using a generic ELISA assay.



*Figure 1A. Cost analysis for reusing resin over the next seven years. Legend: current use (grey), cost if resin is reused (black). For MabSelect™ Sure resin reuse limit = 250 protein load cycles.*

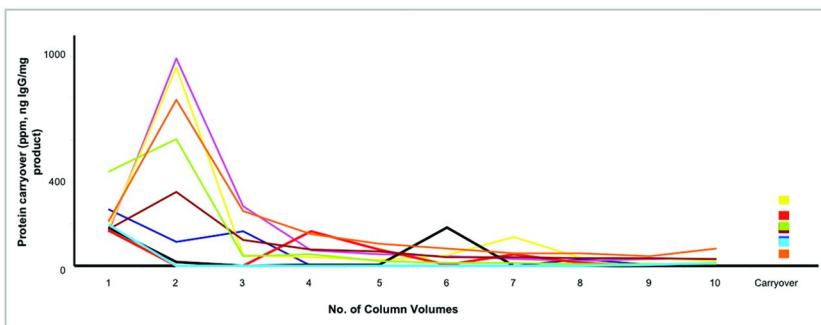


Figure 2A. Plot of protein carryover as a function of CV washes (resin cleaning efforts) with different buffer solutions on a ProSep®vA column. Legend: 6 M guanidine HCl: magenta; 19% ethanol: red; 2 M arginine HCl: brown; 20% hexene glycol: grey; 8 M urea/1 M NaCl: orange; equilibration buffer: blue; 1% v/v phosphoric acid: yellow; 0.1 M Imidazole/19% Ethanol: black; 0.1 M Acetic acid: green; 2 M potassium phosphate: turquoise.

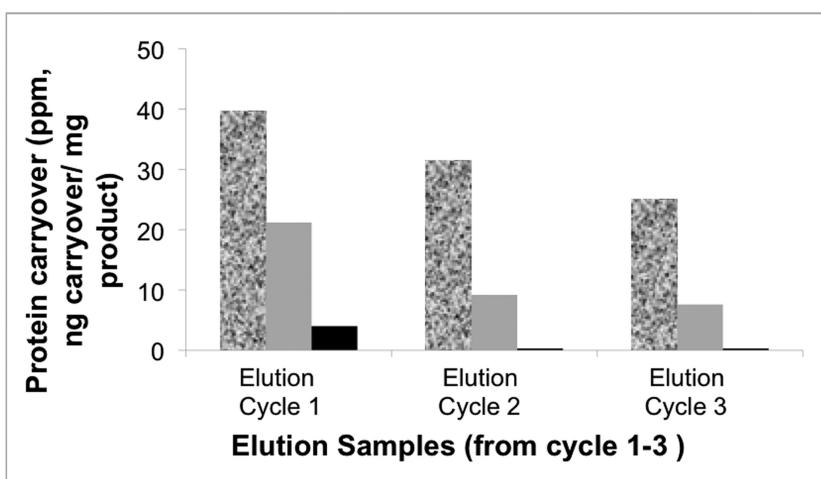


Figure 3A. Plot of total protein carryover (intact IgG + Fc fragment) as a function of Elution sample cycle from a sequential lab scale purification of mAbA, mAbB, and mAbC on a MabSelect™ SuRe column without additional resin cleaning. Legend: mAbA carryover in mAbB elution (black & grey), mAbB carryover in mAbC elution (grey), and mAbA carryover in mAbC elution (black).

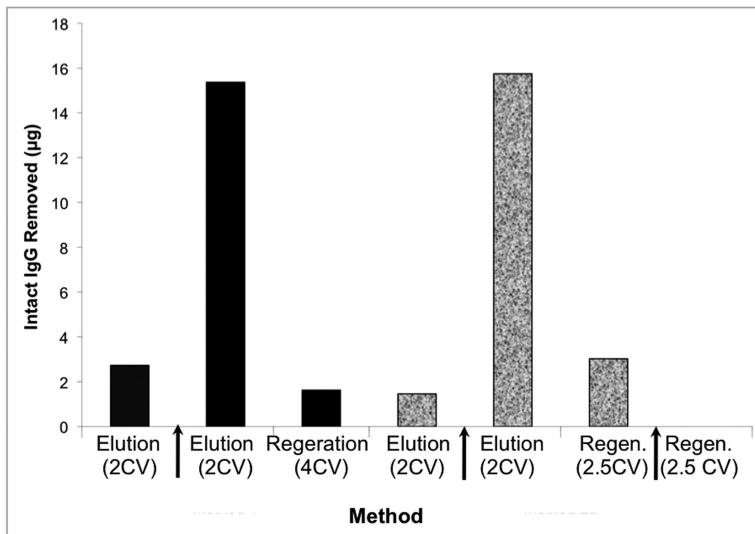


Figure 4A. Protein removal seen in initial cleaning cycles as a function of elution buffer (0.15 M Acetic acid) or Regeneration buffer (0.1 N NaOH) CV washing. Arrows point to 30 min static holds. Note there is a 5 fold increase in the amount of protein that is washed off the column after a 30 min static hold and no protein was detected after static hold with Regeneration buffer. Legend: Method 4 (Table 1): black; method 5 (Table 1): black and grey.

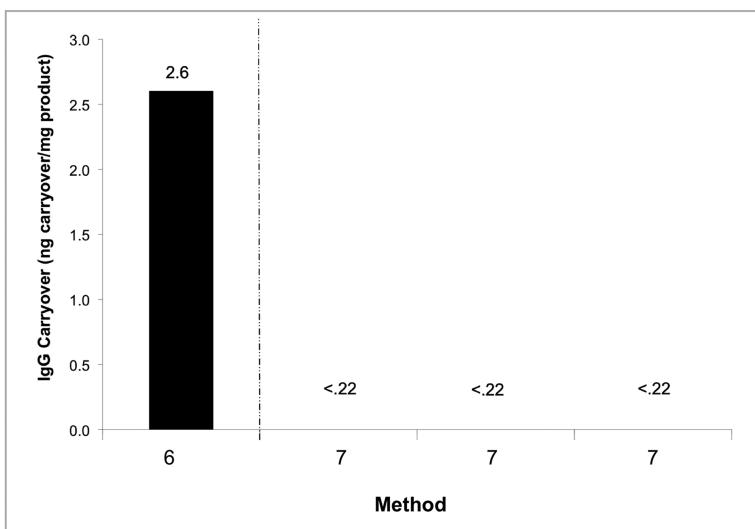
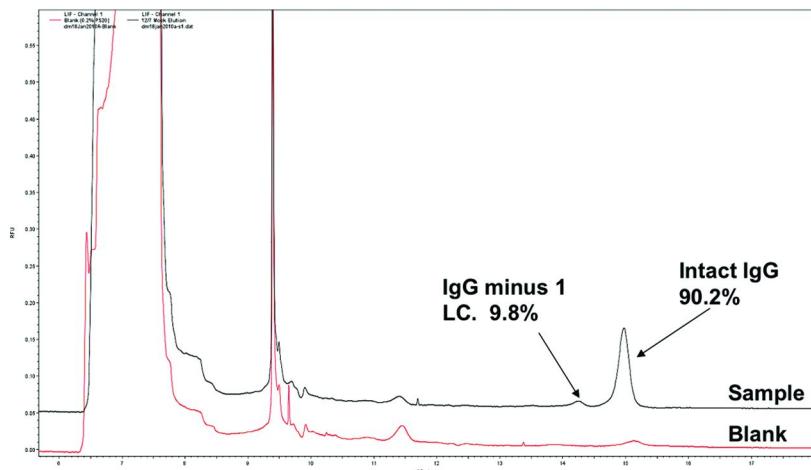
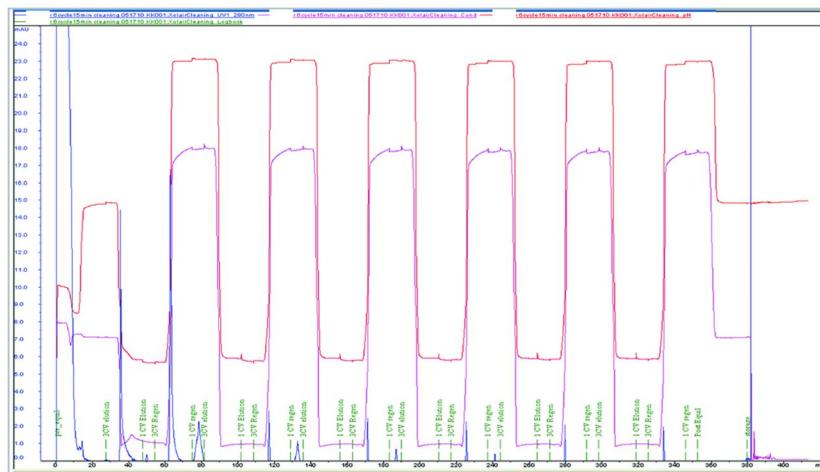


Figure 5A. Intact IgG protein detected after 'mock runs' were performed on a MabSelect<sup>TM</sup> SuRe column after cleaning using Method 6 (black bar; Table 1) and Method 7 (none detected, Table 1). Three 'mock runs' were performed using Method 7 (Table 1) and results were reproducible.



**Figure 6A.** Capillary Electrophoresis-Sodium Dodecyl Sulfate (CE-SDS) analysis of a 'mock elution' after cleaning of a MabSelect™ SuRe column following the Method 7 cleaning procedure (Table 1) for a 94 ng/mL mock elution sample revealed that the mAb was over 90% fully intact.



**Figure 7A.** Akta chromatograms using the Method 7 cleaning procedure (Table 1) a chromatogram of a 'mock elution' suggests efficient cleaning of the column, as demonstrated with spikes in the UV-intensity when a shift from Elution buffer (0.15 M acetic acid) to Regeneration buffer (0.1 N NaOH) is made for each of the 6 cycles. Suggesting that alternating strongly acidic and basic buffers with 15 min buffer hold times is an ideal cleaning protocol. Blue line = UV 280 nm, Red line = pH, Magenta line = conductivity.

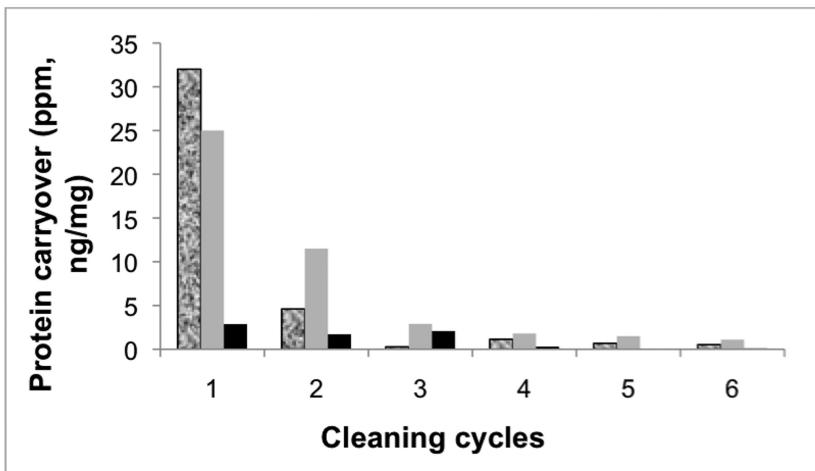


Figure 8A. Plot of total protein carryover as a function of cleaning cycle following conditions outlined in Method 8, Table 1 with Storage buffer (100 mM sodium acetate and 2% Benzyl alcohol (pH 5)). Legend: Elution cycles (black & grey), Regeneration cycles (grey), Storage buffer cycles (black).

**Table 1A. Analysis of in house mAbs that have been purified on Protein A resin that was cleaned using the optimized cleaning procedure MSSCCP on a multiproduct column**

<i>mAb</i>	<i>scale</i>	<i>CV (mL)</i>	<i>Total protein (mg/mL)</i>
1	Lab	6.8	<0.25
2	Lab	6.8	0.46
3	Pilot	3000	<0.25
4	Pilot	6280	<0.25
5	Pilot	6280	<0.25
6	Pilot	3230	<0.25
7	Pilot	6280	<0.25
8	Pilot	1730	<0.25

<sup>a</sup> For information on a particular assay see the Supporting Information. <sup>b</sup> CV = column volume.

## References

1. (a) Albrecht, H.; Radosevich, J. A.; Babich, M. Fundamentals of antibody-related therapy and diagnostics. *Drugs Today* **2009**, *45*, 199–211.  
(b) Takimoto, C. H. Principles of Oncologic Pharmacotherapy Calvo, E. In *Cancer Management: A Multidisciplinary Approach Medical, Surgical & Radiation Oncolog*, 11th ed.; Pazdur, R., Wagman, L. D.; Camphausen, K. A., Hoskins, W. J., Eds.; CMP Healthcare Media LLC: Lawrence, KS, 2008
2. Fahrner, R. L.; Knudsen, H. L.; Basey, C. D.; Galan, W.; Feuerhelm, D.; Vanderlaan, M.; Blank, G. S. Industrial purification of pharmaceutical antibodies: development, operation, and validation of chromatography processes. *Biotechnol. Genet. Eng. Rev.* **2001**, *18*, 301–327.
3. (a) *Affinity Chromatography Principles and Methods*; Amersheim Biosciences: Uppsala, Sweden, 2002; <http://www.gelifesciences.com/webapp/wcs/stores/servlet/productBy/1/en/GELifeSciences-us/18102229> (accessed on July 13, 2012) (b) Fahrner, R. L.; Iyer, H. V.; Blank, G. S. The optimal flow rate and column length for maximum production rate of protein A affinity chromatography. *Bioprocess Eng.* **1999**, *22*, 287.
4. (a) Fahrner, R. L.; Whitney, D. H.; Vanderlaan, M.; Blank, G. S. Performance comparison of protein A affinity-chromatography sorbents for purifying recombinant monoclonal antibodies. *Biotechnol. Appl. Biochem.* **1999**, *30*, 121–128. (b) Kelley, B. Very Large Scale Monoclonal Antibody Purification: The Case for Conventional Unit Operations. *Biotechnol. Prog.* **2007**, *23*, 995–1008.
5. (a) *MabSelect™ Sure resin*; Application note 28-9872096 AA; Lifetime performance study of MabSelect Sure™ LX during repeated cleaning-in-place; GE Healthcare: Piscataway, NJ, Feb. 2011. [https://www.gelifesciences.com/gehcls\\_images/GELS/Related%20Content/Files/1314807262343/litdoc28987296AA\\_20110831222625.pdf](https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1314807262343/litdoc28987296AA_20110831222625.pdf) (accessed on June 27, 2012).
6. (a) Teschner, W.; Butterweck, H. A.; Auer, W.; Muchitsch, E. M.; Weber, A.; Liu, S. L.; Wah, P. S.; Schwarz, H. P. A new liquid, intravenous immunoglobulin product (IGIV 10%) highly purified by a state-of-the-art process. *Vox Sang.* **2007**, *92*, 42–55. (b) Food and Drug Administration, HHS. *Guidance for Industry Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers*; Rockville, MD, July 2005; <http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CE8QFjAA&url=http%3A%2F%2Fwww.fda.gov%2Fdownloads%2FDrugs%2F...%2FGuidances%2FUCM078932.pdf&ei=f4QhUJv4K9Ov6gGQ-4DgAg&usg=AFQjCNFbTE75U0nDbFpfduxK85uWXT8frg> (accessed on August 7, 2012) (c) European Medicines Agency. *Impurities: Residual Solvents, Note for Guidance on Impurities: Residual Solvents* (CPMP/ICH/283/95); London, U.K., Sept. 1997; [http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general\\_content\\_000431.jsp&mid=WC0b01ac0580029593](http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000431.jsp&mid=WC0b01ac0580029593) (accessed on August 7, 2012).

7. (a) *OCTAGAM®*; Product Approval Information Summary Basis of Approval OCTAGAM® 5%; OCTAPHARMA Pharmazeutika: Vienna, Austria, August, 2002; <http://www.fda.gov/downloads/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/LicensedProductsBLAs/FractionatedPlasmaProducts/ucm064955.pdf> (accessed on August 7, 2012).
8. Other cleaning buffers such as chaotropes 6M Guanidine HCl, were not investigated. Although guanidine increases solubility of proteins the solvation reaction is endothermic. Further guanidine is corrosive and it can reduce the binding capacity of the resin.
9. Zhu-Shimon, J.; Gunawan, F.; Thomas, A.; Vanderlaan, M.; Stults, J. Trace level analysis of leached Protein A in bioprocess samples without interference from the large excess of rhMAb IgG. *J. Immunol. Methods* **2009**, *341*, 59–67.
10. Michels, D. et al. Total protein was measured using an in house Capillary Zone Electrophoresis/Laser-Induced Fluorescence Detection (CZE-LIF) assay; in preparation.
11. Fahrner, R. L.; Whitney, D. H.; Vanderlaan, M.; Blank, G. S. Performance comparison of protein A affinity-chromatography sorbents for purifying recombinant monoclonal antibodies. *Biotechnol. Appl. Biochem.* **1999**, *30*, 121–128.
12. An in house TaqMan polymerase chain reaction was used for CHO DNA analysis.
13. Intact antibody and fragmented antibody parts were measured using a generic in house developed sandwich ELISAs.
14. Fahrner, R. L.; Knudsen, H. L.; Basey, C. D.; Galan, W.; Feuerhelm, D.; Vanderlaan, M.; Blank, G. S. Industrial purification of pharmaceutical antibodies: development, operation, and validation of chromatography processes. *Biotechnol. Genet. Eng. Rev.* **2001**, *18*, 301–327.
15. Trexlar-Schmidt, M.; Sze-Khoo, S.; Kothran, A. R.; Thai, B. Q.; Sargis, S.; Lebreton, B.; Kelley, B.; Blank, G. Purification Strategies to Process 5 g/L Titers of Monoclonal Antibodies. Altering the order of operations, using new resins, and increasing dynamic binding capacity can obviate the need for major facility changes. *Biopharm. Intl.* **2009**, March2.

## Chapter 7

# How To Tie a Peptide Knot

**Si Jia Pan, Wai Ling Cheung, Siyan Zhang, Mikhail Maksimov, and  
A. James Link\***

**Departments of Chemical Engineering and Molecular Biology,  
Princeton University, A207 Engineering Quadrangle,  
Princeton, New Jersey, 08544**

**\*E-mail: ajlink@princeton.edu. Phone: +1(609) 258-7191.  
Fax: +1(609) 258-0211**

Lasso peptides are unique ribosomally-synthesized natural products with extraordinary stability. The focus of the present studies was on the role of the leader peptide in the biosynthesis of lasso peptides in bacteria as well as efforts to engineer the lasso peptide scaffold for new purposes.

**Keywords:** Lasso peptide; peptide knot; Microcin J25; lasso peptide libraries; antimicrobial activity

### 1. Introduction

The lasso peptide Microcin J25 (MccJ25) is a member of a class of low molecular weight antibacterial agents called microcins. Microcins are secreted by Enterobacteriaceae to kill other competing bacteria; and are biosynthesized ribosomally, exhibiting many post-translational modifications (1). MccJ25 is a 21 amino acid peptide (2) (Figure 1) that displays potent antibiotic activity against Gram-negative (Gram-) bacteria, and is especially active against enteric (intestinal) bacteria, *Escherichia coli* (*E. coli*), strains of *Salmonella*, and strains of *Shigella* (1). MccJ25 garners its antimicrobial activity via the inhibition of RNA polymerase (3). Other than MccJ25 only one other clinically used small molecule antibiotic, rifampicin, is known to target and inhibit RNA polymerase (4). The biosynthesis of most of the lasso peptide class remains elusive, with the exception of MccJ25 and capistruin (5, 6) and most recently lariatin (7) and astexin-1 (8). Precise molecular details about the biosynthesis of lasso peptides still need to be elucidated.

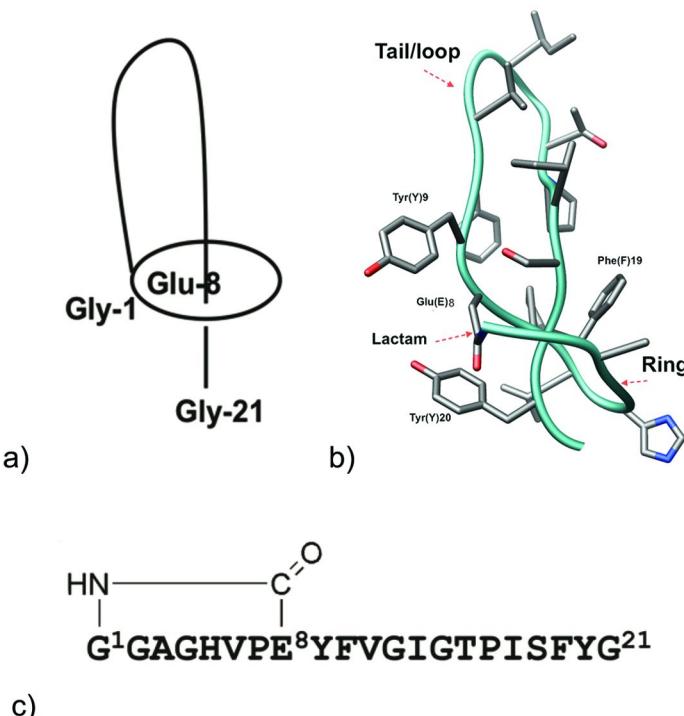


Figure 1. Different representations of the lasso peptide structure of MccJ25. (a) Schematic view of the lasso. (b) Ribbon and stick representation of MccJ25 (PDB code 1Q71) (9). (c) Amino acid sequence of MccJ25, with the isopeptide bond that forms the 8-membered lactam or ring framework. Figure 1b prepared using Chimera (10). Legend: C-atoms: gray; N-atoms: blue, O-atoms: red; ribbon: cyan.

To date, nine different lasso peptides have been discovered from bacterial producers and each peptide represents an attractive drug target (11–18). This attractiveness is not only because of the diverse biological functions and activities, but also because of three additional properties. First, the larger surface area of a lasso peptide allows the peptide to engage in a greater number of intermolecular interactions, such as van der Waals contacts and hydrogen bonds, to a given drug target relative to its small molecule counterparts. Second, lasso peptides (like all peptides composed of natural amino acids) are intrinsically evolvable; since the peptide framework is encoded with a gene, a large series of analogues can be generated using recombinant DNA techniques and directed evolution. The ability to rapidly evolve a peptide therapeutic is attractive considering a major problem with drugs is the inherent ability of microbes and viruses to evolve themselves. Third, the lasso framework holds the peptide in a rigid and well-defined confirmation imparting incredible heat and protease resistance

(Figure 1) (2b). In contrast short linear peptides lack a well-defined secondary structure and this lack of pre-organization can hinder a peptide from readily recognizing its target. Further, oral delivery of linear peptides is complicated by the fact that peptides are substrates for digestive proteases and thus are degraded before they can reach their targets. One way to overcome both these limitations is to rigidify the peptide framework by introducing conformational constraints. Such constraints can be added into peptides in a number of ways, but the focus here will be on the generation of the lasso peptide framework.

The backbone of all lasso peptides, including MccJ25 resembles a lasso or a slipknot, the namesake of the class (Figure 1a). There are two regions of lasso peptide natural product: the ring and the loop/tail. The ring portion of the lasso results from the formation of an isopeptide bond between the *N*-terminal amine and a carboxylic acid side chain (Figure 1c). The ring size varies amongst the class from 8-9 amino acids (11). In the case of MccJ25 this carboxylic acid is of a glutamic acid side chain (Glu8, E8, Figure 1), making its ring size 8 amino acids. The remaining *C*-terminal amino acids (13 in the case of MccJ25) are then threaded through the central 8-membered ring, resulting in a completed lasso structure (Figure 1b). The amino acid sequence of the lasso peptide MccJ25 is shown in Figure 1c. All lasso peptides identified overall have diverse structures despite resulting from the same type of chemical connectivity. This can be seen in a comparison of NMR structures of MccJ25 and capistruin (Figure 2). Capistruin's tail is pulled through the ring substantially less than MccJ25. Thus although lasso peptides occupy the same chemical space, there are numerous possible topologies. This diversity in chemical space may be indicative of the many different functions this class of peptides have in nature (11).

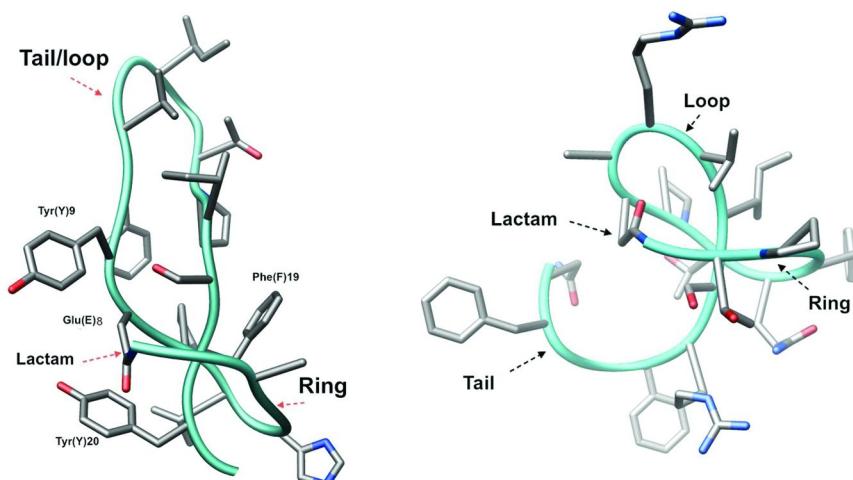


Figure 2. Comparison of the tertiary structure in ribbon and bow representation of microcin J25 (PDB code 1Q71) (20) and capistruin (BMRB entry 20014) (18).

## 1.1. Lasso Peptide Stability

MccJ25 is a remarkable example of the inherent stability of lasso peptides (2b). The protein is stable to thermal and chemical denaturation and exhibits tremendous protease stability and remains undigested when treated with the typical digestive proteases (19). The stability of the lasso structure is in part the result of the presence of a phenylalanine 19 (F19) and tyrosine (Y20) on either side of the 8-membered lactam ring (20). These bulky residues prevent the tail from slipping back through the ring providing a “lock” mechanism. Thus the MccJ25 peptide is trapped in this entropically disfavored state affording it tremendous thermal and chemical stability.

## 1.2. Biosynthesis

The biosynthesis of MccJ25, like capistruin, is encoded in 4 genes (Figure 3). The genes are organized on a single cluster (*mcjABCD*) that is found on a plasmid pTUC100 in certain strains of *E. coli* (Figure 3) (2, 5b). A previously constructed a plasmid, pJP3, has been generated that contains the MccJ25 biosynthetic machinery and allows for inducible control over the production of MccJ25 (21). The *mcjA* gene encodes for the 58 amino acid peptide precursor to MccJ25, a protein called McjA. The *mcjB* and *mcjC* genes encode two enzymes (McjB and McjC) responsible for the conversion of McjA into the mature 21 amino acid lasso peptide MccJ25 (Figure 4) (22). Bioinformatic studies suggest that the McjB protein is a protease, responsible for cleavage of the 37 amino acid leader sequence, and McjC catalyzes the formation of the isopeptide bond (Figure 4) (22a). The latter was hypothesized due to its sequence homology to other amide-bond forming enzymes, including asparagine synthetases. The *mcjD* gene encodes a membrane protein, McjD, which serves as pump (22a). McjD is responsible for pumping the mature lasso peptide MccJ25 from the cytoplasm of the cell, where it is generated, into the extracellular space. McjD is also an immunity factor that ensures the host is not poisoned by the antibiotic it generates by efficiently pumping it out the host cell.

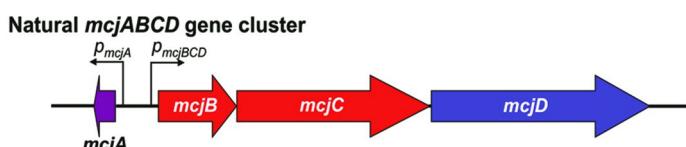


Figure 3. Schematic view of the natural gene cluster *mcjABCD* on plasmid pTUC100 in *E. coli*. Figure adapted from reference (24).

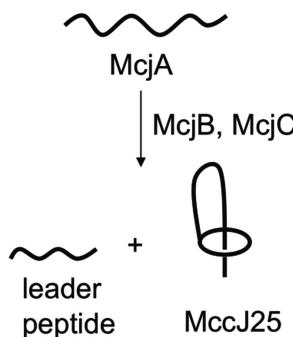


Figure 4. Schematic representation of the function of *McjB* and *McjC* proteins in the biosynthesis of the mature 21 amino acid lasso peptide *MccJ25* from *McjA*.

Figure adapted from reference (23).

With *MccJ25* as the model peptide, the present work uses the principles of protein engineering to answer these questions about lasso peptides: How much of the leader peptide on *McjA* is needed to generate an active *MccJ25*? What residues within the truncated leader sequence are essential to generate an active *MccJ25*? What is the effect of engineering *MccJ25* via amino acid substitutions?

## 2. Material and Methods

### 2.1. Spot-on-Lawn Assay

To assess the level of *MccJ25* produced by each of the truncation variants, alanine mutant or library mutant, a spot-on-lawn assay was performed essentially as described elsewhere (23). Briefly, the culture supernatant was obtained via centrifugation and boiled to ensure that *MccJ25* was the only bactericidal component in the supernatant. These supernatants were spotted on an agar plate overlaid with soft agar impregnated with exponentially growing *Salmonella newport* (*Salmonella enterica* subsp. *enterica* serovar Newport), a strain of bacteria hypersensitive to *MccJ25* (2a). Finally, the plates were incubated overnight at 37 °C. In this assay, the size of the zone of *S. newport* growth inhibition semi-quantitatively reflects the level of *MccJ25* production.

### 2.2. Truncation

The truncation variants were prepared as previously described (24). Removal of 5 amino acid segments was from the *N*-terminus of the leader peptide. These variants were cloned into pJP3, and the resulting plasmids were transformed into the *E. coli* expression host DH5α. Transformants were grown to mid-exponential phase in LB medium, induced with 1 mM IPTG, and allowed to produce *MccJ25* overnight (20 h) at 37 °C.

### 2.3. Alanine Scan

Single alanine variants of McjA(26-57) were expressed and culture supernatants were spotted on *S. Newport* spot-on-lawn assay (23), as described above (24).

### 2.4. Lasso Peptide Library

To construct the NNT saturation mutagenesis library of mcjA mutants, the “mcjA ring F” and “mcjA ring R” oligonucleotides were used to synthesize the mature peptide coding region of mcjA ring mutants by primer overlap PCR, and likewise the “mcjA tail F” and “mcjA tail R” oligonucleotides were used to synthesize the mature peptide region of the mcjA tail mutants. The fragment coding for the signal peptide of mcjA and its upstream region extending to the XhoI site from pJP3 was amplified using upstream primer “mcjA signal F” and downstream primer “mcjA signal R”. Next, the fragment containing the mcjA signal and the mature peptide coding fragment of mcjA mutants were joined using assembly PCR to make full-length mutant mcjA. Both the ring and tail mutant mcjA fragments were digested with XhoI and HindIII, and pWC8 was digested with XhoI and HindIII to remove wild type mcjA, then the mutant mcjA fragments were ligated into the place of wild type mcjA in the pWC8 plasmid, creating the mcjA ring and tail NNT libraries (25).

For the ring and tail libraries, plasmids carrying the libraries were transformed into XL-1 Blue cells and the cells were spread onto LB plates. About 12,000 colonies were obtained for each library, and the colonies were scraped from the plates and grown in 100 mL of fresh LB broth for an hour. Plasmids carrying the libraries were extracted from the culture by midiprep (Qiagen Plasmid Midi Kit), and subsequently used to transform cells containing pJP31 for library screening (25).

For the tail arabinose (tail ara+) library, plasmids carrying the tail library were transformed into *E. coli* cells containing pJP31. The cells were spread onto LB plates containing 0.2% arabinose. About 20,000 colonies were obtained and harvested by scraping and grown in 100 mL of LB broth supplemented with arabinose for an hour. Aliquots of the culture containing the library were mixed with equal parts of a 65% glycerol solution and frozen at -80 °C until needed for screening (25).

### 2.5. Replica Plate Assay

XL-1 Blue cells harboring the mcjA ring and tail libraries prepared as described above were grown on LB plates (25 g/L agar) overnight. The colonies on each plate were transferred onto two fresh plates (25 g/L agar) by replica plating. During replica plating, the Petri dish carrying the original colonies was inverted and lowered onto a block (Scienceware Replica-Plating Tool) covered with velveteen square. After carefully lifting the original plate, an LB plate and subsequently an LB plate with IPTG/glucose were inverted and lowered onto the velveteen to create two plates carrying identical colony replicas. Following

overnight incubation of the two replica plates, clones that were growth-inhibited on the IPTG/glucose plate were identified as producers of functional MccJ25 variants (25).

### 3. Results and Discussion

#### 3.1. Truncation Studies

McjA is a 58 amino acid protein precursor to MccJ25, of those only 21 are present in the mature lasso peptide. Thus the first 37 amino acids at the *N*-terminus of McjA are excised during MccJ25 biosynthesis. Thus the following questions were posed: Are these extra amino acids needed to correctly process MccJ25? Is this extra rope needed at the beginning to tie the peptide knot?

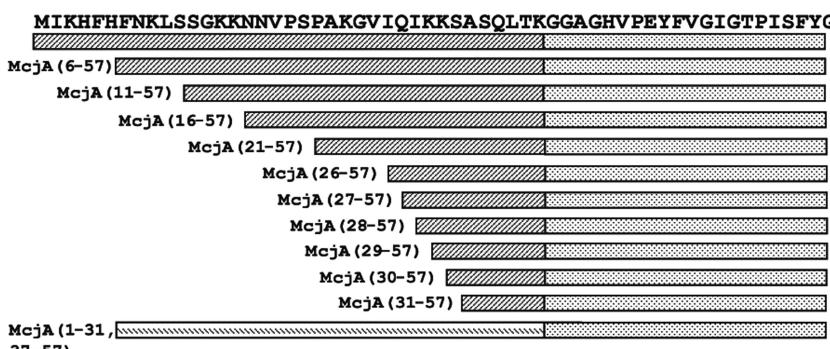
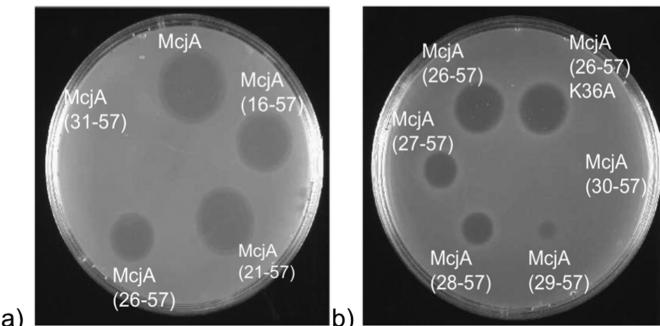


Figure 5. Amino acid sequence of McjA and schematic of truncated variants prepared biosynthetically. Figure modified from reference (23).

To answer this question a series of truncated versions of McjA were generated by sequential removal of five amino acid blocks from the *N*-terminus of the leader peptide of McjA (Figure 5) (15b). An additional control was prepared by removal of the final five amino acids of the leader peptide (Figure 5).

To access the impact of truncation on the formation of active MccJ25 each variant was evaluated using a previously well-established spot-on-lawn assay (23). The results of this study have been reported elsewhere (24), so they will be briefly summarized here. The results of the spot-on-lawn assay are shown in Figure 6, reading the plate counterclockwise starting from the top where the native McjA protein was tested. Not surprisingly there is a large zone of inhibition for wild type precursor sequence McjA (Figure 6a). Sequential removal of the first 5 (6-57 construct), 10 (11-57 construct) amino acid blocks off the *N*-terminus are not shown for simplicity. However, for each successive construct where 15 (16-57 construct), 20 (21-57 construct), 25 (26-57 construct), amino acids were removed from the *N*-terminus, a zone of inhibition, which is indicative of an antibiotic being produced,

was observed - albeit to a lesser extent for each successive construct (Figure 6a). However, if 30 amino acids are removed, 31-57 construct, no antibiotic activity was detected (Figure 5a). The result suggests a minimal amount of leader peptide is required to tie the knot.



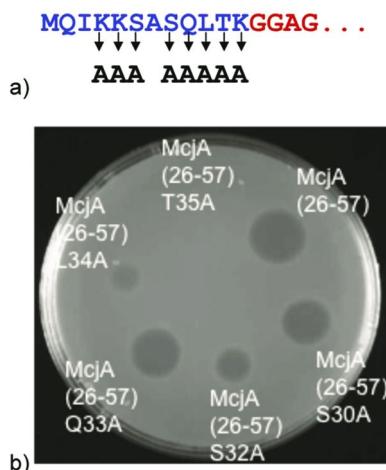
*Figure 6. S. Newport spot-on-lawn assay (23), or zone of inhibition assay, for the determination of antimicrobial activity for the N-terminal McjA truncated variants. (a) Zones of clearance for the native 15, 20, 25, and 30 amino acid McjA truncations. (b) Zones of clearance for narrow range of variants 25, 26, 27, 28, and 29 amino acid truncated from the N-terminus of McjA. Results demonstrate that the 29-57 construct is the minimum tail length. Figure modified from reference (24).*

To narrow in on the exact leader peptide length needed to tie the knot, a second series of sequence variants were prepared (Figure 6b). For these 25 (26-57 construct), 26 (27-57 construct), 27 (28-57 construct), 28 (29-57 construct), and 29 (30-57 construct) amino acids were removed from the *N*-terminus of McjA (Figure 6b). The results clearly demonstrate a decrease in the spot size for each single and sequential amino acid removed, through to the 29-57 construct. This construct represents the minimal leader peptide required for MccJ25 biosynthesis. Thus 28 out of 37, or more than 75%, of the leader peptide is dispensable with regards to MccJ25 biosynthesis. By analogy, only a little extra rope is needed to tie the MccJ25 knot.

### 3.2. Sequence Variation in the Leader Peptide

Only nine amino acids were needed in leader peptide to tie the MccJ25 knot. Next a probe of the function of the different amino acids within that minimal leader peptide was undertaken employing an alanine scan (24) on the 26-57 construct. For this study each amino acid that was not already an alanine was individually mutated to an alanine (Figure 7a) (24). Again the mutated 26-57 constructs were screened in the spot-on-lawn (23) assay to ascertain the effect of mutation on antimicrobial activity (Figure 7). The results have been published elsewhere so they will only be summarized here (24).

Reading the plate clockwise starting from the McjA T35A mutant (26-57 construct), it is clear that this residue, T35, is essential for conversion of McjA into the mature MccJ25 (Figure 7b), as it was the only mutant to lose antimicrobial activity completely. Two other mutants, S32A and L34A (Figure 7b) did show reduced antimicrobial activity – smaller zones of inhibition in comparison to wild type McjA. Interestingly, a threonine residue is also present in the second to last position of the capistruin leader peptide (24), suggesting it may play a conserved role in lasso peptide biosynthesis. We have examined the role of this conserved threonine residue in our subsequent work (8, 11).



*Figure 7. Alanine scan for the determination of essential residues in the leader peptide on the 26-57 constructs. (a) Outline of selected mutations along the leader peptide McjA truncated variants starting from K28A to K36A. (b) *S. newport* spot-on-lawn assay (23) for the determination of antimicrobial activity for the alanine scan of the 26-57 constructs. Figure modified from reference (24).*

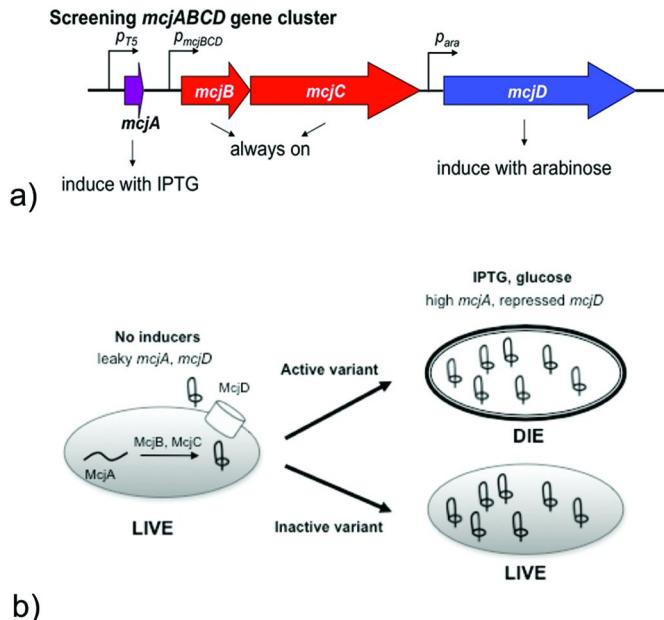
### 3.3. Molecular Engineering of Lasso Peptides

Molecular engineering of MccJ25 was performed to answer the following questions: Can the lasso peptide scaffold be reprogrammed with any amino acid sequence? What is the effect of altering the sequence of MccJ25 to engineer a chimeric molecule with dual properties? For instance, one could imagine designing a chimera in which a peptide is grafted onto a lasso peptide scaffold. The resulting chimera could have the properties of the parent peptide, but with the enhanced stability afforded by the lasso framework.

To answer these questions the ability of MccJ25 to withstand multiple amino acid substitutions needed to be addressed. To do this in a systematic fashion libraries of MccJ25 variants were prepared in a high-throughput (HT) fashion,

and these analogues were screened for antimicrobial activity (25). The assumption made was that if the variants displayed antimicrobial activity then a lasso peptide is still formed.

As a screen to rapidly test for the lasso structure was not in place and a screen for antimicrobial function was, the latter was used as proxy for structure. As the work was published in detail elsewhere (24) it will only be summarized here.

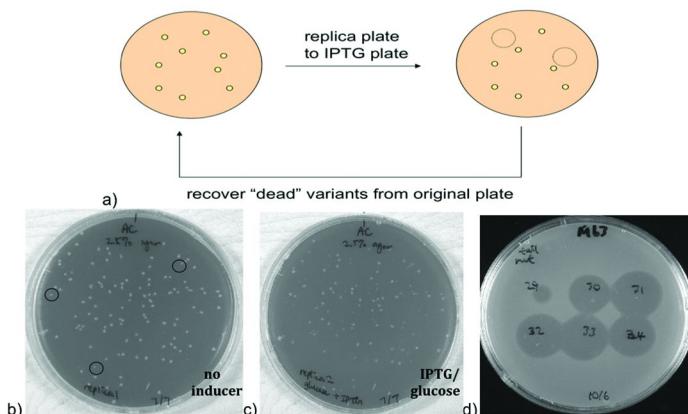


*Figure 8. An orthogonally regulated system for MccJ25 production and export. a) Schematic illustration of the plasmid used to generated the MccJ25 library that was under dual control with an inducible isopropyl- $\beta$ -D-thio-galactoside (IPTG) and arabinose promoter. b) Schematic cartoon representation of the orthogonally inducible system used for the identification of active MccJ25 mutants. Figures modified from reference (25).*

Two different MccJ25 libraries were constructed. The libraries were generated through saturation mutagenesis that targeted either the  $\beta$ -hairpin loop/tail or the ring (24). Three amino acid residues were modified concurrently in each library and the mutants were prepared using standard recombinant DNA techniques (see Materials and Methods). To control the phenotypic outcome and identify active mutants the *mcjABCD* gene cluster was manipulated to elicit separate control over the production and export of MccJ25 (Figure 8). As the *mcjA* gene encodes the production of the lasso peptide framework, it was put under the control of and isopropyl- $\beta$ -D-thio-galactoside (IPTG) inducible T5 promoter (Figure 8a). Thus in the presence of IPTG in the growth media the bacteria would produce McjA. A second and orthogonal promoter, which was

under the control of arabinose, was used to control the *mcjD* gene (Figure 8a). In the absence of arabinose, McjD is not produced and the cells die (Figure 8b) (26).

Typically, genetic selection assays are used to find the mutants that keep cells alive, not dead as in this assay set-up. To overcome this limitation a replica plating protocol was employed (Figure 9a) (27). The replica plating protocol was used as a Stage I assay to identify colonies from the MccJ25 tail and ring mutant library that do not survive under restrictive conditions. Those colonies that disappeared in the replica restrictive plate were isolated in the parent plate. In Figure 9b three examples of colonies that disappeared are circled in the parent plate.



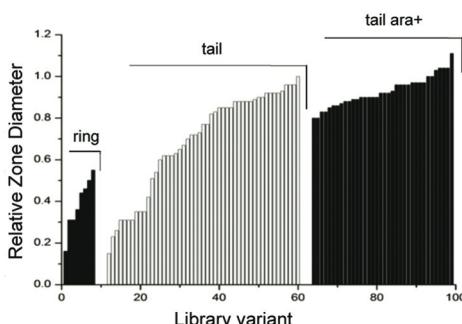
*Figure 9. Characterization of cells expressing orthogonally inducible MccJ25 production. (a) Schematic cartoon representation of the process behind formation of a permissive plate (without inducer) and the generation of an identical daughter plate (with inducer). Stage I: identification of functional mutants: (b) Example of a library screen of *E. coli* XL-1 Blue harboring the screening gene cluster and producing variants of MccJ25 (no inducer). Colonies that disappeared in plate (c) are circled in (b). (c) library screen of *E. coli* XL-1 Blue harboring pWC8 and pJP31 incubated with inducer (IPTG and glucose). Stage II Secondary screen: (d) zone of inhibition assay against *S. newport* to evaluate library. Modified from reference (25).*

The hits generated from Stage I were then verified in Stage II assay where the culture supernatants were screened for their ability to inhibit *Salmonella* growth. Stage II is a validation assay was performed to reexamine the hits generated in the primary screen. A *Salmonella* spot-on-lawn assay (23) was performed to examine the size of the zones of inhibition that are formed by the different mutants.

In the library screen 13,000 different clones were studied. Of the 13,000 clones screen 400 clones made it through to Stage II. Of those 400 clones, nearly 100 distinct mutants were identified to maintain antimicrobial activity. Recall that prior to this work only 8 lasso peptides were known, and in the course of a few months employing laboratory evolution techniques 100 new lasso peptides were identified.

A bar graph representation of the relative zone of inhibition diameters (normalized to control) as a function of the library that clone came from clearly demonstrated a trend (Figure 10). The trend observed is that mutants from the ring library behave poorly (smaller relative diameters). Further there are not only significantly more functional mutants from the tail library they are also significantly more potent (larger relative diameter). Collectively this means the tail portion is much more tolerant to amino acid modification than the ring (Figure 10).

A careful look at the distribution of the particular amino acid substitutions at each position along the tail reveals there is no discernable pattern (25). In other words, there is no particular amino acid substitution, or residue, that is favored, suggesting this region of the antibiotic is highly mutable. Further, since the screening assay for antimicrobial activity is more stringent than screening only for the formation of a lasso peptide, it is likely this series represents only a small fraction of the sequences that can form lasso peptides (25).



*Figure 10. Relative zone diameters measured in the *Salmonella* spot-on-lawn zone of inhibition assay as a function of library variant for all ring and tail clones. The zone diameters are normalized to wild type MccJ25 zone diameter. Zones of inhibition for the majority of variants are much smaller in comparison to wild type and the tail mutants outperformed the ring mutants. Modified from reference (25).*

## Conclusions

This manuscript seeks to answer the question: How to tie a peptide knot? It was demonstrated through sequential removal of the amino acids along the *N*-terminus of the leader peptide that only 9 amino acids of the leader peptide are required to form a functional MccJ25 lasso peptide. By analogy only a little bit of extra rope on the precursor protein is needed to tie the peptide knot. Site-directed mutagenesis studies using the well-known alanine scanning technique revealed that only a single residue inside the 9 amino acid leader sequence is deemed essential, it is T35. Two other residues, S32 and S34, are important, but not crucial. Further, directed evolution was employed to generate a diverse

library of  $\beta$ -hairpin tail/loop and ring variants to ascertain the effect of amino acid substitution in these regions. The results revealed numerous different peptide sequences can be tied into lasso peptides in the tail region using the two enzymes that are required for the biosynthesis of functional MccJ25, McjB and McjC. In other words, many different ropes can be tied into lasso peptide knots.

## References

1. For references on Microcins see: (a) Severinov, K.; Semenova, E.; Kazakov, A.; Kazakov, T.; Gelfand, M. S. *Mol. Microbiol.* **2007**, *65*, 1380. (b) Duquesne, S.; Petit, V.; Peduzzi, J.; Rebuffat, S. *J. Mol. Microbiol. Biotechnol.* **2007**, *13*, 200. (c) Baquero, F.; Moreno, F. *FEMS Microbiol. Lett.* **1984**, *23*, 117. (d) Duquesne, S.; Destoumieux-Garzon, D.; Peduzzi, J.; Rebuffat, S. *Nat. Prod. Rep.* **2007**, *24*, 708. (e) Pavlova, O. A.; Severinov, K. V. *Russ. J. Genet.* **2006**, *42*, 1380.
2. Microcin J25 peptide see: (a) Salomon, R. A.; Farias, R. N. *J. Bacteriol.* **1992**, *174*, 7428. (b) Rebuffat, S.; Blond, A.; Destoumieux-Garzon, D.; Goulard, C.; Peduzzi, J. *Curr. Protein Pept. Sci.* **2004**, *5*, 383.
3. Delgado, M. A.; Rintoul, M. R.; Farias, R. N.; Salomon, R. A. *J. Bacteriol.* **2001**, *183*, 4543.
4. McClure, W. R.; Cech, C. L. *J. Biol. Chem.* **1978**, *253*, 8949.
5. For the biosynthesis of MccJ25 see: (a) Salomon, R. A.; Farias, R. N. *J. Bacteriol.* **1992**, *174*, 7428. (b) Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Gonzalez-Pastor, J. E.; Moreno, F.; Salomon, R. A. *J. Bacteriol.* **1999**, *181*, 2659.
6. Knappe, T. A.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 11446.
7. Inokoshi, J.; Matsuhamma, M.; Miyake, M.; Ikeda, H.; Tomoda, H. *Appl. Microbiol. Biotechnol.* **2012**, *95*, 451.
8. Maksimov, M. O.; Pelczer, I.; Link, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 15223.
9. Rosengren, K. J.; Clarck, R. J.; Daly, N. L.; Goransson, U.; Jones, A.; Cralk, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 12464.
10. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by Grant NIH P41 RR-01081). For details, see the following: (a) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF ChimeraA Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. (b) Chimera home page can be found at <http://www.cgl.ucsf.edu/chimera>.
11. Maksimov, M. O.; Pan, S. J.; Link, A. J. *Nat. Prod. Rep.* **2012**, *29*, 996.
12. For leading references for Anantin see: (a) Iwatsuki, M.; Tomoda, H.; Uchida, R.; Gouda, H.; Hirono, S.; Ōmura, S. *J. Am. Chem. Soc.* **2006**, *128*, 7486-7491. (b) Wyss, D. F.; Lahm, H. W.; Manneberg, M.; Labhardt, A. M. *J. Antibiot.* **1991**, *44*, 172–180.

13. For leading references for RES-701-1 see: (a) Chokekijchai, S.; Kojima, E.; Anderson, S.; Nomizu, M.; Tanaka, M.; Machida, M.; Date, T.; Toyota, K.; Ishida, S.; Watanabe, K. *Antimicrob. Agents Chemother.* **1995**, *39*, 2345–2347. (b) Kimura, K.; Yamazaki, M.; Sasaki, N.; Yamashita, T.; Negishi, S.; Nakamura, T.; and Koshino, H. *J. Antimicrob.* **2007**, *60*, 519–523.
14. For leading references for Siamycin see: (a) Detlefsen, D. J.; Hill, S. E.; Volk, K. J.; Klohr, S. E.; Tsunakawa, M.; Furumai, T.; Lin P. F. Nishio M, Kawano K, Oki T. *J. Antibiot.* **1995**, *48*, 1515–1517. (b) Tsunakawa, M.; Hu, S. L.; Hoshino, Y.; Detlefsen, D. J.; Hill, S. E.; Furumai, T.; White, R. J.; Nishio, M.; Kawano, K.; Yamamoto, S. *J. Antibiot.* **1995**, *48*, 433–434.
15. For leading references for Propeptin see: (a) Ezumi, Y.; Suzuki, Y.; Itoh, Y.; Uramoto, M.; Kimura, K.; Goto, M.; Yoshihama, M.; Ichikawa, T. *J. Antibiot.* **2002**, *55*, 296–300. (b) Kimura, K.; Kanou, F.; Takahashi, H.; Ezumi, Y.; Uramoto, M.; Yoshihama, M. *J. Antibiot.* **1997**, *50*, 373–378. (c) Kimura, K.; Yamazaki, M.; Sasaki, N.; Yamashita, T.; Negishi, S.; Nakamura, T.; Koshino H. *J. Antibiot.* **2007**, *60*, 519–523.
16. For leading references for BI-32169 see: Knappe, T. A.; Linne, U.; Xie, X.; Marahiel, M. A. *FEBS Lett.* **2010**, *584*, 785–789. (b) Potterat, O.; Wagner, K.; Gemmecker, G.; Mack, J.; Puder, C.; Vettermann, R.; Streicher R. *J. Nat. Prod.* **2004**, *67*, 1528–1531.
17. For leading references for Lariatin see: Iwatsuki, M.; Koizumi, Y.; Gouda, H.; Hirono, S.; Tomoda, H.; Omura, S. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2888–2890.
18. For leading references for Capistruin see: (a) Knappe, T. A.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 11446–11454. (b) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. *Chem. Biol.* **2009**, *16*, 1290–1298.
19. Rosengren, K. J.; Blond, A.; Afonso, C.; Tabet, J. –C.; Rebuffat, S.; Craik, D. J. *Biochem.* **2004**, *43*, 4696–4702.
20. (a) Bayro, M. J.; Mukhopadhyay, J.; Swapna, G. V. T.; Huang, J. Y.; Ma, L. C.; Sineva, E.; Dawson, P. E.; Montelione, G. T.; Ebright, R. H. *J. Am. Chem. Soc.* **2003**, *125*, 12382. (b) Rosengren, K. J.; Clark, R. J.; Daly, N. L.; Goransson, U.; Jones, A.; Craik, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 12464. (c) Wilson, K. A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A. *J. Am. Chem. Soc.* **2003**, *125*, 12475.
21. Pan, S. J.; Cheung, W. L.; Link, A. J. *Protein Expression Purif.* **2010**, *71*, 200–206.
22. (a) Duquesne, S.; Destoumieux-Garzón, D.; Zirah, S.; Goulard, C.; Peduzzi, J.; Rebuffat, S. *Chem. Biol.* **2007**, *14*, 793. (b) Clarke, D. J.; Campopiano, D. J. *Org. Biomol. Chem.* **2007**, *5*, 2564.
23. The spot-on-lawn is an antimicrobial assay performed on agar plates impregnated with exponentially growing *Salmonella* *Newport*, a serotype of *Salmonella enterica* that is hypersensitive to microcin J25. The zones of clearing, or zones of inhibition, are areas of inhibited bacterial growth that appear where the culture supernatants from each of the different constructs were spotted

24. Cheung, W. L.; Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2010**, *132*, 2514.
25. Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 5016.
26. This assay allowed for independent control the production and the export of MccJ25 (Figure 7b). The reason why this is useful is that in the different states there are conditions where, in the absence of inducer (IPTG or arabinose), the bacteria cells should live; as “leaky” production of small amounts of the mature peptide and transporter are possible. However, under restrictive conditions where IPTG and glucose are added, cells will produce lasso peptide but be unable to export it. Cells that are producing an active MccJ25 variant will die under these conditions while cells producing inactive variants will survive (Figure 8b).
27. The general procedure for the generation of the a replica plate. The first step is to grow the cells on a permissive plate – where all mutants will survive. Then an exact replica of that plate is prepared, except IPTG and glucose are added to the growth media (Figure 9b and 9c). As observed on the first permissive parent plate, Figure 9b, all the colonies have a similar size and opacity. Under restrictive conditions, there is a marked variation in the colony sizes and opacities (Figure 9c). In fact, some colonies completely disappeared (Figure 9c).

## Chapter 8

# Freeze-Dry Mediated Formation of 5-(Hydroxymethyl)furfural

**Kelly Fitzpatrick,\* Brent Kendrick, Camilla Santos, Pat Green,  
Bing Zhang, Deanna Hunt, Michael Ronk, and Ying Luo**

**Amgen, 4000 Nelson Road, Longmont, Colorado 80503**

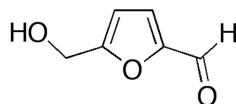
**\*E-mail: fitzpatk@amgen.com. Phone: 303-401-1636. Fax: 303-401-4403**

During the development of a formulation for a protein therapeutic, 5-(hydroxymethyl)furfural (HMF) was discovered as a by-product of the freeze-dehydration (lyophilization) process. It was discovered that HMF formed as a result of decomposition of carbohydrates in the buffered formulation solution ( $\text{pH} < 7$ ) in both placebo and drug product solutions. HMF is a known by-product of heat-induced carbohydrate breakdown *via* the *Maillard* reaction. The potential toxicological impact of this impurity was assessed, and it was found that there was no impact for the levels present in therapeutic formulation. Further, the impact of the impurity on product characterization and process testing strategy were assessed. While no modifications of the protein therapeutic were detected the process testing strategy was impacted by the presence of HMF. Due to interference by the strong UV response factor of HMF with the standard UV based concentration determination, in combination with the low product concentration, there was a need to develop a new product concentration determination method for drug product in-process and final drug product testing. This new method utilized reverse phase-high performance liquid chromatography (RP-HPLC) to separate the product from the HMF compound. The implications of this finding suggest the impurity, HMF, may be present in other marketed freeze-dried drug product formulations prepared in a similar manner.

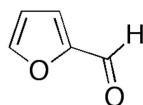
**Keywords:** 5-(Hydroxymethyl)furfural; lyophilization-induced reactions; formulation by-products; carbohydrate decomposition; sugar breakdown; protein therapeutic

## 1. Introduction

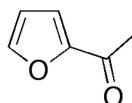
The decomposition of mono-and disaccharides such as sucrose, glucose, and fructose is a natural result of processing, such as cooking or heating. Some of the products of this decomposition are a concern in the food industry, as well as the pharmaceutical industry, as many of these ingredients are in food stocks such as milk (1), fruit juices (2), caramels (3), tomato paste (4), infant formulas (5), and infusion fluids (IVs) (6), just to name a few. The stability of the carbohydrates within these consumables is important, as degradation affects not only the quality and taste of the food or pharmaceutical drug, but also its shelf-life. Furthermore, within the pharmaceutical industry the presence of such impurities can affect the quality of the marketed drug from a toxicological standpoint.



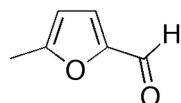
5-(hydroxymethyl)-2-furfural (**1**)



2-furfuraldehyde (**2**)

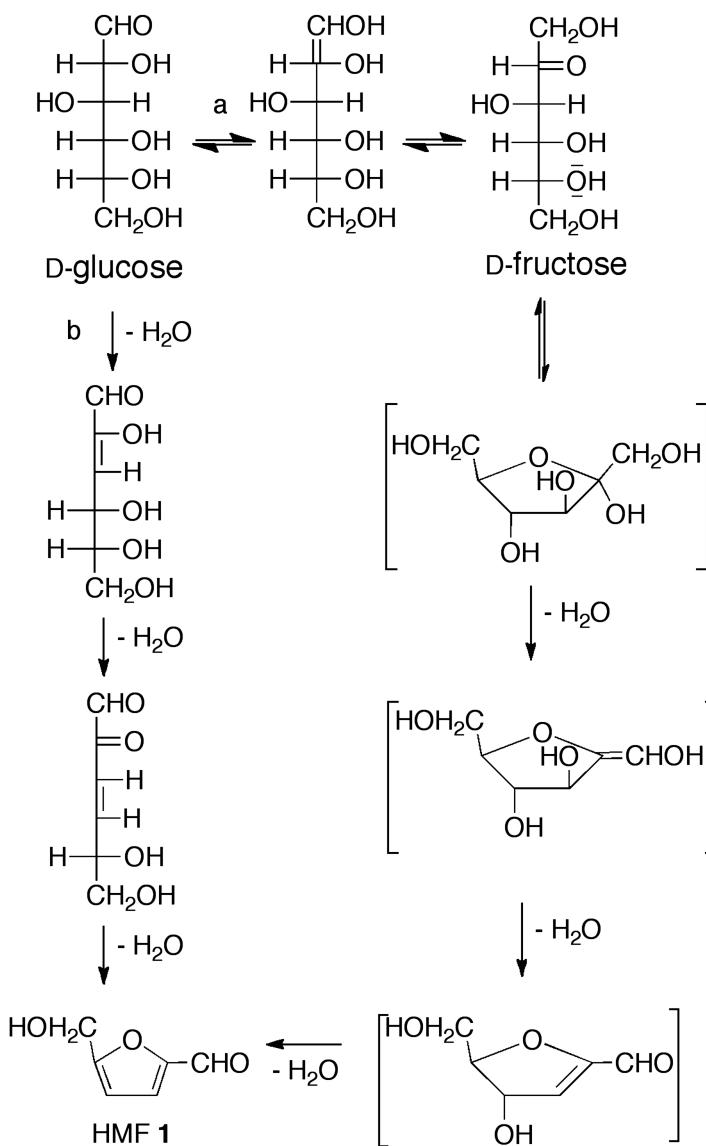


2-furfyl methyl ketone (**3**)



5-methyl-2-furfural (**4**)

Figure 1. By-products of non-enzymatic breakdown of carbohydrates (**1-4**).



*Scheme 1. Representative schematic of the formation of furfural from fructose. According to the: a) Loby De Bruyn-Alberta van Ekenstein transformation and b) the Maillard reaction (11–13)*

Some common well known by products of non-enzymatic breakdown of carbohydrates (Figure 1) in processed foods and pharmaceuticals are 5-(hydroxymethyl)-2-furfural (**1**), 2-fufuraldehyde (**2**), 2-furyl methyl ketone (**3**), and 5-methyl-2-furfural (**4**). Compound **1** (or HMF) and

its related impurities are formed in the presence of amino acids (e.g. lysine) and reducing sugars such as sucrose, fructose, dextrose (a.k.a glucose) (7), or by heat-induced degradation from long-term storage (at ambient temperature, 25°C) (1, 8), or treatment with acid (9), and in some cases refluxing water alone from glucose (10). The reaction mechanism for the degradation of saccharides to HMF and F is not well understood. However, several mechanisms have been put forth, and generally each mechanism consists of a condensation or rearrangement (or isomerization) followed by loss (or dehydration) of three water molecules to provide the conjugated furan ring. In the *Maillard* reaction pathway, HMF is believed to be formed upon reaction of amino acids, like lysine, with reducing sugars (Scheme 1b) (11). The acid catalyzed *Loby De Bruyn-Alberta van Ekenstein* transformation (Scheme 1a) also proceeds via rearrangement of glucose to fructose, followed by dehydration and cyclization to the furan (12, 13).

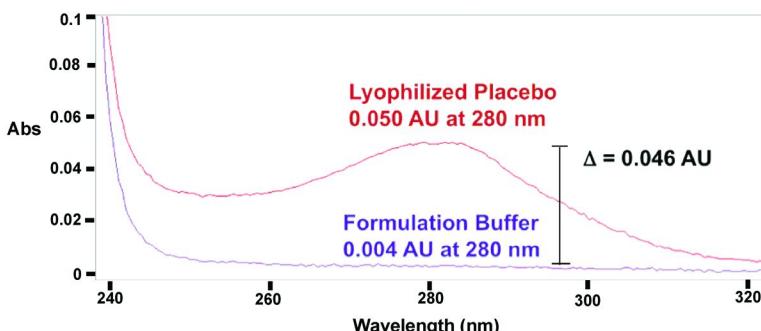


Figure 2. UV-absorbance spectra of placebo. The placebo contains formulation buffer, sugars, and an amino acid and is at a pH < 7. Purple spectra: formulation buffer prior to lyophilization; red spectra: reconstituted placebo after lyophilization.

During the formulation process for a clinical lot of a placebo (ie. in the absence of active drug) an impurity was found during a routine quality control (QC) UV-analysis of a reconstituted sample of the lyophilized placebo. The lyophilized placebo contained the same concentration of formulation components, which included formulation buffer (pH < 7), sugars and amino acid, as the sample containing the active drug. Prior to lyophilization, the formulation solution itself did not contain any impurity, as judged by UV-analysis (Figure 2, purple trace). However, upon lyophilization, the UV-spectrum displayed a pronounced peak at 280 nm (Figure 2, red trace), which indicated that the impurity was formed during the lyophilization process. The probable cause was identified as the decomposition of the sugars in the formulation solution. As outlined in Figure 3, we sought to identify the impurity and the lyophilization cycle parameter that caused the impurity, accurately quantify the amount of the impurity in drug

product lots, perform a toxicological assessment to determine if the impurity was of toxicological concern, ensure accurate protein concentration in drug product and, finally, use this method during any site or processes changes to ensure comparability.

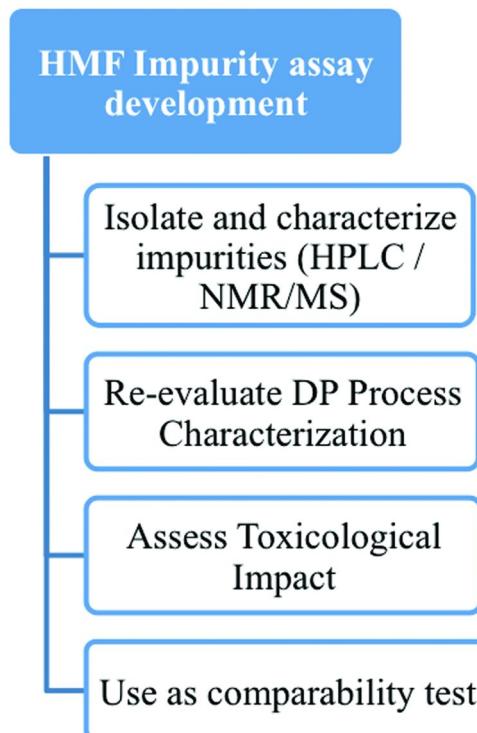


Figure 3. Analytical solution diagram to address drug product impurity issue.

## 2. Results and Discussion

### 2.1. Identification of Impurity

The impurity was analyzed in the reconstituted placebo and drug product solutions by RP-HPLC. Subsequently, the identity of the impurity was confirmed as HMF by HPLC, UV, <sup>1</sup>H-NMR, <sup>13</sup>C-NMR and mass spectral analysis (see the Supporting Information) (17). The HPLC retention time ( $R_t$ ), UV absorbance pattern, <sup>1</sup>H-NMR, and <sup>13</sup>C-NMR of the isolated sample matched that of an HMF standard (see the Supporting Information). The 1D proton (<sup>1</sup>H) NMR spectrum in D<sub>2</sub>O of the isolated impurity revealed an aldehyde proton, two protons in the aromatic region, and two methylene protons (Table 1A, SI) (17). The coupling

constants ( $J$  values) between the two aromatic protons was 3.6 Hz, which is in agreement with a furan ring structure. In addition, the observed  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts ( $\delta$ ) and coupling constants ( $J$ ) for the impurity match those of authentic HMF. The  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectrum was also obtained in  $\text{DMSO-d}_6$  to ensure the exchangeable proton signals were detected.

The definitive assignment of all protons and carbons was achieved by 1D and 2D NMR experiments (Figure 1A and 2 SI). Nuclear overhauser effect (NOE) analysis demonstrated that the following proton pairs are in close proximity: H-C(8)-H-C(4), H-C(4)-H-C(3) and H-C(3)-H-C(6). The 2D HSQC spectrum revealed four carbons directly attached to protons (data not shown). The 2D HMBC analysis detected numerous long-rang  $^1\text{H}$ - $^{13}\text{C}$  correlations for the compound (data not shown). Due to sample matrix effects observed in the isolated compound, spiking experiments with commercially available HMF were performed in order to eliminate any ambiguity resulting from the sample matrix (Table 1A and 2A SI). The spiking spectra in  $\text{D}_2\text{O}$  and  $\text{DMSO-d}_6$  displayed the coalescence of the two sets of  $^1\text{H}$  signals into one. It was demonstrated that the different peak shape of the hydroxyl proton was a result of the difference in sample matrix. In summary, chemical shifts, multiplicities,  $^1\text{H}$ - $^1\text{H}$  through-space interactions and  $^1\text{H}$ - $^{13}\text{C}$  through-bond correlations are consistent with the proposed HMF structure (Tables 1A and 2A SI) (17).

Analysis using mass spectrometry revealed the  $m/z$  value of the main peak in the RP-HPLC chromatogram was 127, corresponding to the elemental composition of HMF ( $\text{MH}^+$ ,  $\text{C}_6\text{H}_7\text{O}_3$ ; Exact Mass: 127.0395) and provided definitive confirmation that the impurity in the drug product formulation was HMF (Figure 4). Interestingly, there is no literature precedence for the formation of HMF, or decomposition of sugars upon freeze-drying/lyophilization.

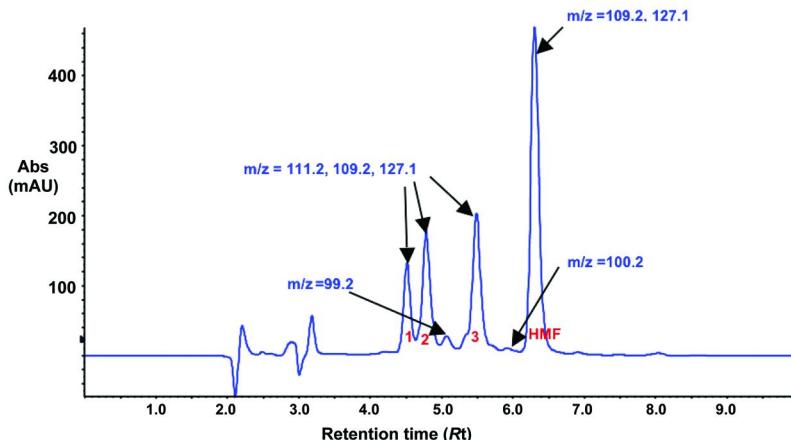


Figure 4. Mass spectral information for impurities separated by RP-HPLC (MS data obtained by APCI-Q-Trap). XWC of DAD spectral data from  $\lambda = 280-282 \text{ } \nu\mu$ .

## 2.2. Lyophilization Process Analysis

Due to the fact that HMF formation occurs during lyophilization, an attempt was made to isolate the portion of the lyophilization cycle during which HMF forms by simulating the lyophilization cycle using small scale models. From these studies, the HMF formation was isolated to the primary and secondary drying portion of the lyophilization cycle and was dependent on the length of total drying time. The first lyophilization cycle had a combined primary (22 H) and secondary (16.5 H) drying time of about 39 hours (H) and the second lyophilization cycle had a combined primary (77 H) and secondary (25.5 H) drying time of about 103 H. When the drug product was lyophilized with the extended drying time higher amounts of HMF were produced. Taking one step further, the aqueous portion of the lyophilization buffer was eliminated by SpeedVac™, and the product was stored in a dessicator at 30°C for 6 days, which resulted in an impurity profile by RP-HPLC nearly matching that of the large scale drug product. The RP-HPLC chromatograms of the products of the drying experiments are captured in Figure 5, where the black chromatogram is the product resulting from the first lyophilization cycle (about 39 H total drying time), the blue chromatogram is the product resulting from additional drying in the chamber (about 103 H total drying time), and the top pink trace is the product resulting from the lyophilization buffer after SpeedVac™ for 15 hrs, then storage in a dessicator at 30°C for 6 days. Using excessive drying conditions, the impurity profile of the drug product was reproduced. In short, the drying time influences the formation of HMF and its related impurities in the placebo and drug product.

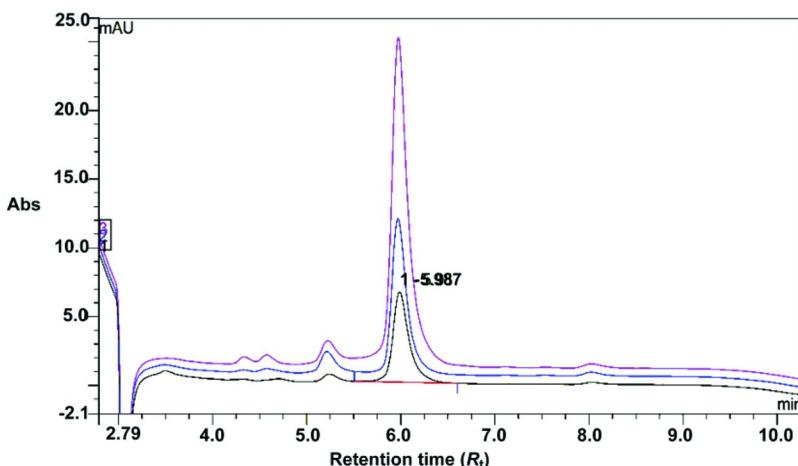


Figure 5. Effect of drying conditions on the formation of HMF and its related substances.

## 2.3. Toxicology

### 2.3.1. HMF Quantitation Method

As this impurity was identified prior to commercialization, a rigorous toxicology assessment was performed to determine not only the levels of the HMF impurity in the drug product samples, but also to ensure that the level of HMF was at levels that would not cause a safety concern. Further, assessment of the impurity levels in all DP lots was performed, including lots used in clinical studies and validation, as well as stability lots stored at different temperatures. To do this, a quantitative RP-HPLC method was developed using a modified C18 column and an acetonitrile/trifluoroacetic acid gradient monitored by UV at 280 nm. The method was then qualified and used to identify and quantitate the presence of HMF and related impurities in the drug product samples and in samples subjected to various storage conditions.

### 2.3.2. Toxicology Assessment

The toxicology of HMF has been well documented in the literature, as it is a well-known breakdown product of sugars. HMF is present in many food stocks (honeys, fruit juices, and caramels) (2, 3, 18) and pharmaceutical compounds (heat sterilized invertose, glucose and fructose) (15). For example, it has been identified in samples of morphine (16), and in infusion solutions (6). An in-house toxicology assessment was performed to determine the acceptable human exposure limits for HMF. The USP-NF daily dosage limit for HMF is 1 - 2.5 mg per kg body weight (15), which was estimated from 100 mL of hexose parenteral preparation (19). The lyophilization process robustness design of experiments (within acceptable in-process control limits) was evaluated to determine the effect of allowable changes of the process parameters on the formation of HMF (data not shown). All drug product lots had comparable levels of HMF, indicating that the lyophilization process was robust with respect to forming consistent levels of HMF. The levels of HMF analyzed in clinical lots and validation lots, as well as lots that were stored at recommended storage conditions, were found to be approximately 3 orders of magnitude ( $10^3 = 0.001$ ) below the recommended safety limits established by the toxicology assessment which was 1 to 2.5 mg/day as defined by United States Pharmacopeia–National Formulary (USP-NF). Stability samples at recommended storage temperature and accelerated temperatures were also assessed to determine the impact of temperature on the lyophilized drug product and further formation of HMF. Under the recommended storage temperature, the change in HMF was minimal over time. Under accelerated stability conditions, there was a non-linear increase in HMF levels that plateaus over time, and at the highest level the concentration of HMF found was 500-fold below the recommended safety limit. The quantitative RP-HPLC method for HMF that was developed was also used for the comparability testing required for the clinical to production site change, and will be used for any future drug product manufacturing or site changes.

### 3. Analytical Impact of HMF to Quantitation of Drug Product

#### 3.1. Quantification of DP Protein Concentration

As stated in section 2.3.1, the formulation buffer after lyophilization contained an impurity which had absorbance at 280 nm. This resulted in an overestimate of about 5% in calculated protein concentration for those impacted drug product lots. In order to remove HMF, and thus to obtain accurate values of protein concentration, a quantitative RP-HPLC method was developed. This proposed analytical solution led to activities which are outlined in Figure 6. First, a RP-HPLC method developed for protein concentration must pass required qualification and validation parameters, including demonstration of specificity to ensure that the impurities are separated from the protein. Second, all in house lots had to be re-analyzed using the new RP-HPLC method and the protein concentrations had to be recalculated. This led to changes for the “Certificate of Analysis” for these lots. Finally, the drug product specifications needed to be updated to reflect the new RP-HPLC for analysis of protein concentration for the release of the product (Figure 6).

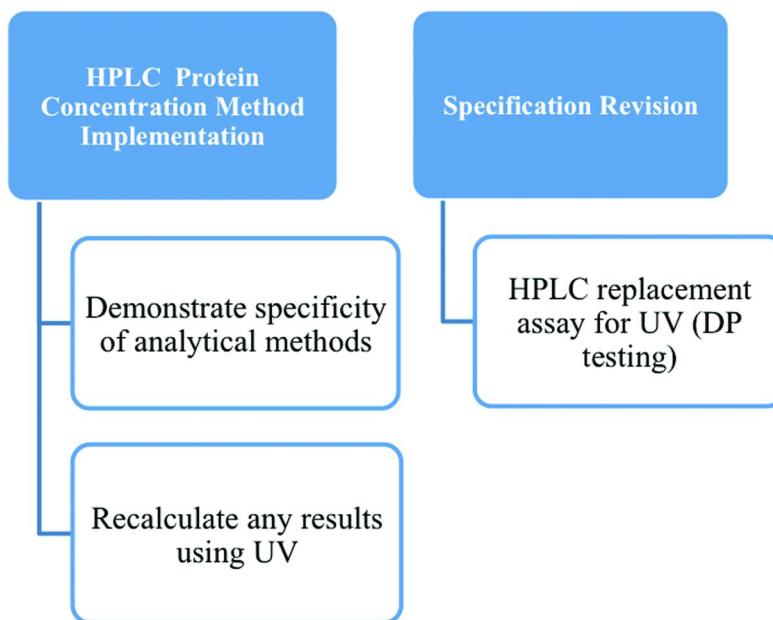


Figure 6. Analytical solution diagram to address drug product protein concentration issue.

### 3.2. Improving Analytical Methods To Improve Production Efficiency

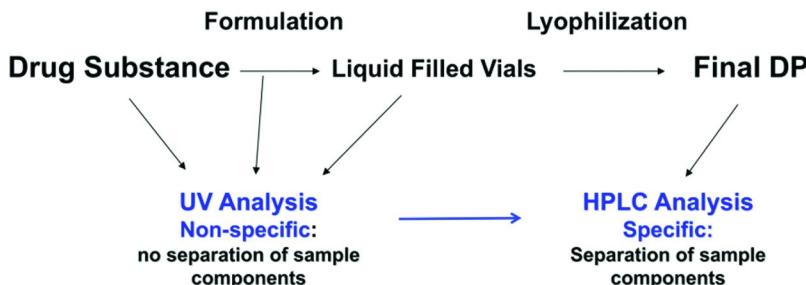


Figure 7. Schematic representation of the formulation and lyophilization process for the drug product and protein concentration analysis.

Transitioning to the use of the new RP-HPLC method for determining protein concentration in the process was another challenge that needed to be addressed. An outline of the formulation and lyophilization process for the drug product, as well as the analytical methodologies employed for concentration determination, are shown in Figure 7. Briefly, drug product is formulated to the desired protein concentration with the formulation buffer. An in-process direct UV-analysis (sample tested directly with no chromatographic separation) was used on samples upstream from the lyophilization step, since no HMF is present in the sample at this point (see Figure 2). For samples post-lyophilization, the RP-HPLC method was employed for protein concentration determination.

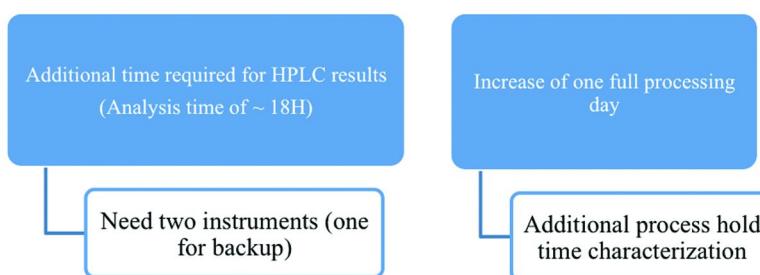


Figure 8. Schematic view of additional challenges in product analysis resulted in longer process times.

The RP-HPLC method used for determining the protein concentration for release was a long method (75 minute run per sample). The time to analyze one sample was approximately 18 hours in order to complete the standard curve, system suitability and assay acceptance injections. In order to implement this RP-HPLC into the “in-process phase” of formulation for drug product production, two instruments needed to be up and running - in case one went down - to ensure

that sample analysis would complete with enough time to avoid impacting process hold time excursions (Figure 8). Additional hold times were characterized, and the process was validated to incorporate a “worst-case processing” to ensure that this time would not impact product quality. To reduce possible process excursion risks, as well as to meet method efficiency design goals, a faster RP-HPLC method was developed and put in-line for testing these “in-process samples”. Implementation of the faster method allowed adequate operational timing to provide for adherence to the original process hold times.

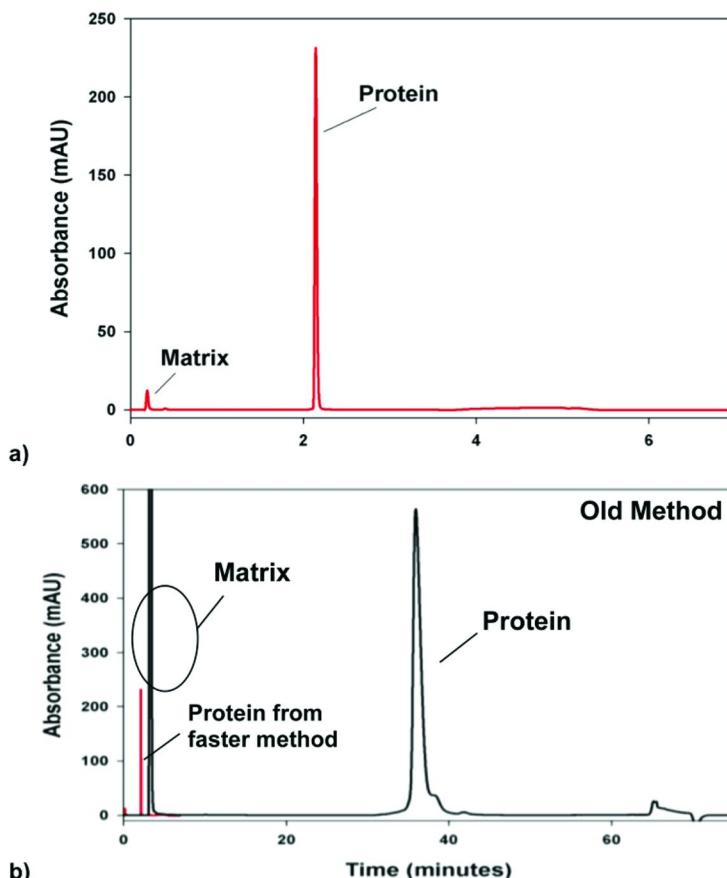


Figure 9. RP-HPLC trace of protein lyophilized drug product. Using a) new faster RP-HPLC method; b) overlay of new faster and original longer method, notice in the new method the protein elutes before the matrix in the old method.

The new faster RP-HPLC method reduced sample run time from 75 minutes to 7 minutes. Using the faster RP-HPLC method, the matrix (or impurities such as HMF) could still be separated from active protein (Figure 9). In both the new and old RP-HPLC methods, the matrix compounds elutes at the beginning of the run,

and the protein elutes as a single peak at a retention time ( $R_t$ ) of about 2 minutes under the new conditions (Figure 9a, red line), and as a single peak at just over 35 minutes in the old method (Figure 9b, black line). Statistically the two methods were found to be equivalent, yielding the equivalent protein concentration results; therefore, there was no impact to product quality, and there is a positive impact to the process by adding the faster RP-HPLC method to perform the in-process testing. The faster RP-HPLC method proved useful, as it was less expensive due the reduced analysis time, the columns were cheaper, and less mobile phase was needed. A comparison of the original RP-HPLC method (black trace) and the faster RP-HPLC method (red trace) shows the gain in efficiency (Figure 9). Using this method, the processing time circumvented the addition of 24 H, required for analysis with the old method, thus, there was a net production efficiency gain.

There was also consistent evaluation of the protein concentration throughout the process, and at drug product release, thereby “centering” the process around the target protein concentration.

## Conclusion

In conclusion, we have the first reported example of the formation of 5-(hydroxymethyl)-2-furfural (**1**) from sugars in our formulation buffer upon lyophilization (freeze-drying). We have shown that the presence of this impurity does not pose a health concern, as its overall content in the drug product units is 3-5 orders of magnitude below than the USP-daily dosage limits. Furthermore, we have also demonstrated how one small molecule can influence multiple facets of a therapeutic protein drug product development process - from the discovery of the impurity to ultimately better control of protein concentration in the process.

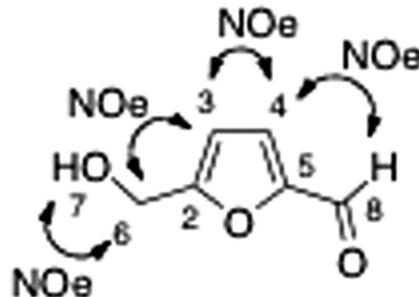
## Acknowledgments

We would like to acknowledge Julie Lippincott and Ryan Petitt for general guidance and analytical testing support.

## Appendix

### Materials and Methods.

**Determination of  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR data for HMF impurity and authentic HMF.** High resolution 1D and 2D NMR spectroscopy was employed to elucidate the structure of the compound isolated from placebo on a Bruker Avance 600 MHz spectrometer equipped with a 2.5 mm inverse broadband microprobe. NMR samples were prepared using two different solvents, deuterium oxide ( $\text{D}_2\text{O}$ ) and deuterated dimethylsulfoxide ( $\text{DMSO}-d_6$ ). Trimethylsilane (TMS) was used as internal standard. The two solvents were used in order to differentiate exchangeable and non-exchangable protons. 1D  $^1\text{H}$  NMR, 1D nuclear overhauser effect (NOE), 2D  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear single quantum correlation (HSQC) and heteronuclear multiple bond correlation (HMBC) were recorded. All NMR experiments were run at 22°C.

**Table 1A. NMR spectral data of the compound isolated from placebo in comparison to commercially available HMF in D<sub>2</sub>O<sup>a</sup>**

ID	HMF (isolated)		HMF (authentic)		Spike <sup>1</sup> H ( $\delta$ , ppm)
	<sup>1</sup> H ( $\delta$ , ppm)	<sup>13</sup> C ( $\delta$ , ppm)	<sup>1</sup> H ( $\delta$ , ppm)	<sup>13</sup> C ( $\delta$ , ppm)	
2	-	161.2	-	161.4	
3	6.55 (d, $J$ = 3.6 Hz, 1H)	110.9	6.44 (d, $J$ = 3.6 Hz, 1H)	110.8	6.55 (d, $J$ = 3.6 Hz, 1H)
4	7.40 (d, $J$ = 3.6 Hz, 1H)	126.7	7.28 (d, $J$ = 3.6 Hz, 1H)	126.8	7.41 (d, $J$ = 3.6 Hz, 1H)
5	-	151.8	-	151.6	-
6	4.57 (s, 2H)	56.0	4.47 (s, 2H)	56.0	4.57 (s, 2H)

*Continued on next page.*

**Table 1A. (Continued). NMR spectral data of the compound isolated from placebo in comparison to commercially available HMF in D<sub>2</sub>O<sup>a</sup>**

ID	HMF (isolated)		HMF (authentic)		Spike
	<sup>1</sup> H ( $\delta$ , ppm)	<sup>13</sup> C ( $\delta$ , ppm)	<sup>1</sup> H ( $\delta$ , ppm)	<sup>13</sup> C ( $\delta$ , ppm)	<sup>1</sup> H ( $\delta$ , ppm)
7	-	-	-	-	-
8	9.33 (s, 1H)	180.4	9.20 (s, 1H)	180.1	9.34 (s, 1H)

<sup>a</sup> <sup>1</sup>H chemical shifts are referenced to HOD peak at 4.70 ppm. Integration, multiplicity and coupling constants are shown in parenthesis.

**Table 2A. NMR spectral data of the compound isolated from placebo in comparison to commercially available HMF in DMSO-d<sub>6</sub><sup>a</sup>**

ID	HMF ( <i>isolated</i> )		HMF ( <i>authentic</i> )		<i>Spike</i> <sup>1</sup> H ( $\delta$ , ppm)
	<sup>1</sup> H ( $\delta$ , ppm)	<sup>13</sup> C ( $\delta$ , ppm)	<sup>1</sup> H ( $\delta$ , ppm)	<sup>13</sup> C ( $\delta$ , ppm)	
2	-	161.2	-	160.0	
3	6.61 ( <i>d</i> , <i>J</i> = 3.6 Hz, 1H)	109.6	6.61 ( <i>d</i> , <i>J</i> = 3.9 Hz, 1H)	109.9	6.61 ( <i>d</i> , <i>J</i> = 3.6 Hz, 1H)
4	7.51 ( <i>d</i> , <i>J</i> = 3.6 Hz, 1H)	124.4	7.51 ( <i>d</i> , <i>J</i> = 3.9 Hz, 1H)	124.9	7.51 ( <i>d</i> , <i>J</i> = 3.6 Hz, 1H)
5	-	151.8	-	151.6	-
6	4.51 (s, 2H)	55.8	4.51 (s, 2H)	56.6	4.51 (s, 2H)
7	5.61 (br s, 1H)	-	5.61 (t, 1H)	-	5.61 (br s, 1H)
8	9.56 (s, 1H)	180.4	9.56 (s, 1H)	178.6	9.56 (s, 1H)

<sup>a</sup> Chemical shifts are reported in ppm and referenced to internal standard (TMS) at 0.00 ppm. Integration, multiplicity and coupling constants are shown in parenthesis.

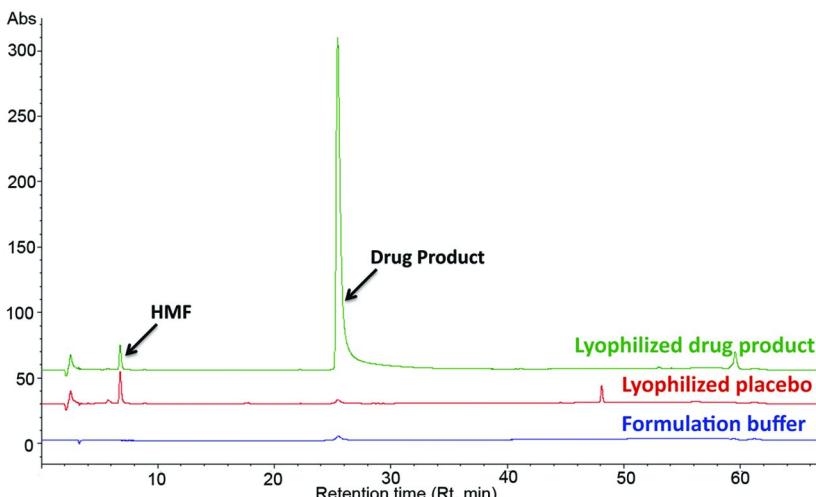


Figure 1A. Superposition of HPLC trace of a) non-lyophilized formulation buffer; b) lyophilized formulation buffer; c) lyophilized drug product. The appearance of HMF impurity ( $R_t = 7$  min) appears after lyophilization.

## References

1. Morales, F. J.; Romero, C.; Jimenez-Pérez, S. An enhanced liquid chromatography method for 5-hydroxymethylfurfural determination in UHT milk. *Chromatographia* **1992**, *33*, 45–48.
2. (a) Mijares, R. M.; Park, G. L.; Nelson, D. B.; McIver, R. C. HPLC analysis of HMF in orange juice. *J. Food Sci.* **1986**, *51*, 843–844. (b) Fuleki, T.; Pelayo, E. Sugar, alcohols, and hydroxymethylfurfural in authentic varietal and commercial grape juices. *J. – Assoc. Off. Anal. Chem.* **1993**, *76*, 59–66.
3. Alfonso, F. C.; Martin, G. E.; Dyer, R. H. High pressure liquid chromatographic determination of 5-(hydroxymethyl)-2-furfuraldehyde in carmel solution. *J. – Assoc. Off. Anal. Chem.* **1980**, *63*, 1310–1313.
4. Allen, B. H.; Chin, H. B. Rapid HPLC determination of hydroxymethylfurfural in tomato paste. *J. – Assoc. Off. Anal. Chem.* **1980**, *63*, 1074–1076.
5. Ferrer, E.; Alegría, A.; Farré, R.; Abellán, P.; Romero, F. High-performance liquid chromatographic determination of furfural compounds in infant formulas: changes during heat treatment and storage. *J. Chromatogr. A* **2002**, *947*, 85–95.
6. Cook, A. P.; Macleod, T. M.; Appleton, J. D.; Fell, A. F. Reversed-phase high-performance liquid chromatographic method for quantification of 5-hydroxymethylfurfural as the major degradation product of glucose in infusion liquids. *J. Chromatogr.* **1989**, *467*, 395–401.

7. a) BeMiller, J. N; Whistler, R. L. Carbohydrates. In *Food Chemistry*, 3rd ed.; Fennema, O. R., Eds.; Marcel Dekker, Inc.: New York, 1996; pp 157–224.  
b) A lead reference for reaction of amino acids with reducing sugars: Ashoor, S. H.; Zent, J. B. Maillard browning of common amino acids and sugars. *J. Food. Sci.* **1984**, *49*, 1206–1207.
8. Örsi, F. On the kinetics of the formation of 5-hydroxymethylfurfural. *Nahrung* **1971**, *34*, 43–55.
9. Shallenberger, R. S.; Mattick, L. R. Relative stability of glucose and fructose at different acid pH. *Food Chem.* **1983**, *12*, 159–165.
10. Scallet, B. L.; Gardner, J. H. Formation of 5-hydroxymethylfurfural from D-glucose in aqueous solution. *J. Am. Chem. Soc* **1945**, *67*, 1934–1935.
11. Martins, S. I. F. S.; Van Boekel, M. A. J. S. A kinetic model for the glucose/glycine Maillard reaction pathways. *Food Chem.* **2005**, *90*, 257–269.
12. (a) van Boekel, M. A. J. S. Effect of heating on Maillard reactions in milk. *Food Chem.* **1998**, *62*, 403–414. (b) Morales, F. J.; Jimenez,-Perez, S. Study of hydroxymethylfurfural formation from acid degradation of the Amadori product in milk-resembling systems. *J. Agric. Food. Chem.* **1998**, *46*, 3885–3890.
13. Chávez-Servín, J. L.; Castellote, A. I.; López-Sabater, M. C. Analysis of potential and free furfural compounds in milk-based formulae by high-performance liquid chromatography evolution during storage. *J. Chromatogr. A* **2005**, *1076*, 133–140.
14. Mrochek, J. E; Rainey, W. T. Identification and biochemical significance of substitute furans in human urine. *Clin. Chem.* **1972**, *18*, 821–828.
15. Ulbricht, R. J.; Northup, S. J.; Thomas, J. A. A review of 5-hydroxymethylfurfural (HMF) in parental solutions. *Fundam. Appl. Toxicol.* **1984**, *4*, 843–853.
16. Kelly, S. S.; Glynn, P. M.; Madden, S. J.; Grayson, D. H. Impurities in a morphine sulfate drug product identified as 5-(hydroxymethyl)-2-furfural, 10-hydroxymorphine and 10-oxomorphine. *J. Pharm. Sci.* **2003**, *92*, 485–493.
17. Characterization for HMF isolated from placebo.  $^1\text{H}$  NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz)  $\delta$  9.56 (s, 1H), 7.51 (d,  $J$  = 3.6 Hz, 1H), 6.61 (d,  $J$  = 3.6 Hz, 1H), 5.61 (br s, 1H), 4.51 (s, 2H);  $^{13}\text{C}$  NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz)  $\delta$  180.4, 161.2, 151.8, 124.4, 109.6, 55.8; APCI-MS  $m/z$  127.1 ( $M\text{H}^+$ , C<sub>6</sub>H<sub>7</sub>O<sub>3</sub>; Exact Mass: 127.0395).
18. Salinas, F.; Mansilla, A. E.; Nevado, J. J. B. Flow-injection determination of HMF in honey by the Winkler method. *Fresenius' J. Anal. Chem.* **1991**, *340*, 250–252.
19. USP-NF Online: Available at <http://www.uspnf.com/uspnf>, cited 7/2006.

## Chapter 9

# Reprogramming the Biosynthesis of Natural Products by Directed Evolution

**Gavin Williams,\* Irina Koryakina, John McArthur,  
Matthew Draelos, Shan Randal, and David Muddiman**

**Department of Chemistry, Campus Box 8204,  
North Carolina State University, Raleigh, North Carolina 27695-8204**

**\* E-mail: gavin\_williams@ncsu.edu. Phone: + 1 (919) 515-3562.  
Fax: + 1 (919) 515-5079**

Directed evolution is a powerful biochemical tool used to alter the function and properties of enzymes. Modification of the activity of key enzymes responsible for the biosynthesis of complex, biologically active natural products may provide routes to increase the diversity of natural products or generate novel and more potent analogues. Here we describe our combined efforts aimed to utilize and improve upon combinatorial biosynthesis techniques to reprogram the biosynthesis of polyketide natural products. We describe a synthetic biology approach that uses directed enzyme evolution to broaden the substrate specificities of enzymes used in natural product biosynthesis. These engineered mutant enzymes will be used to construct artificial biosynthetic pathways for the synthesis and installation of a broad array of natural and non-natural building blocks into polyketides.

### 1. Introduction

Natural products are secondary metabolites produced predominantly in plants and microorganisms, and often display potent biological activities (1), including anticancer, antimicrobial, and/or antiviral activity. Natural products are an important source of drugs and drug scaffolds as a result of their broad and potent activity. In fact 50% of all drugs are ultimately derived or inspired by natural products (1).

The potent and broad biological activity of natural products is an evolutionary adaptation brought about by their ability to recognize various biological targets, and as such natural products are so called ‘privileged scaffolds’ (2). Natural product scaffolds occupy a very unique chemical space, when compared to purely synthetic molecules such as those produced through high-throughput synthetic methods (3). Thus a key challenge becomes how to expand upon this impressive chemical and biological diversity to make ‘non-natural’ analogues of natural products, on a large enough scale, to identify new and potentially more active analogues. Research in our lab seeks to expand the diversity of natural products by harnessing the natural product’s own biosynthetic machinery.

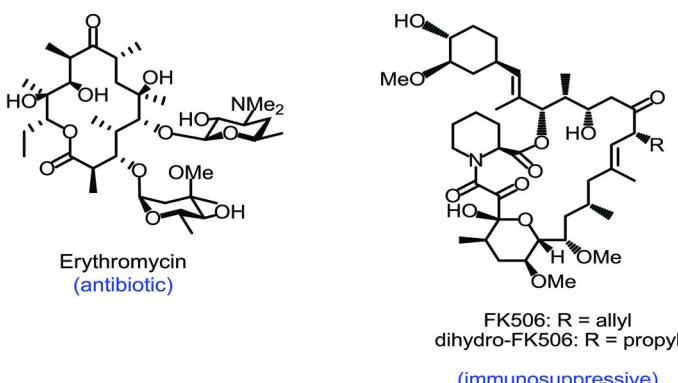
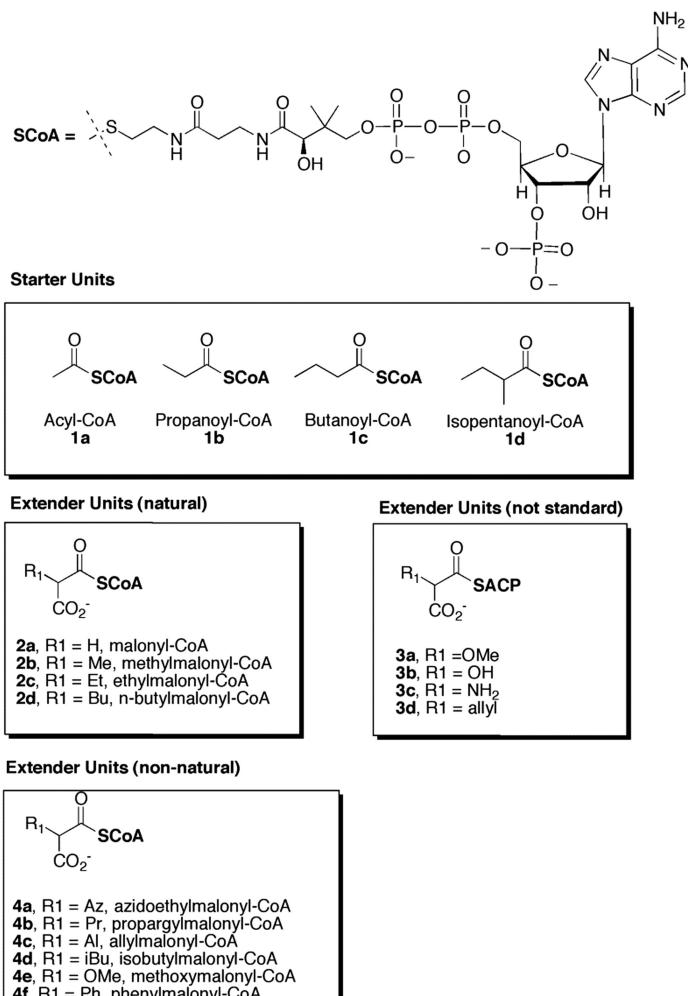


Figure 1. Examples two of structurally diverse polyketide natural products with broad and potent biological activities (1).

Polyketides are a large and structurally diverse class of natural products with extensive and powerful biological activities (Figure 1). Surprisingly, in spite of this tremendous structurally diversity, many polyketides are biosynthesized *via* a common, elegant and efficient strategy by means of multi-domain enzymes called polyketide synthases (PKSs) (4). PKSs catalyze the condensation of small molecule building blocks referred to as starter units and extender units (Figure 2) (5). The starter units (e.g. **1a-1d**, Figure 2), are used to initiate the biosynthesis of a polyketide scaffold, and the extender units (Figure 2) are used to extend a growing polyketide chain. The polyketide building blocks are then added one-by-one to furnish the complete scaffold. These scaffolds can then be further modified by glycosylation, hydroxylation, methylation, etc.

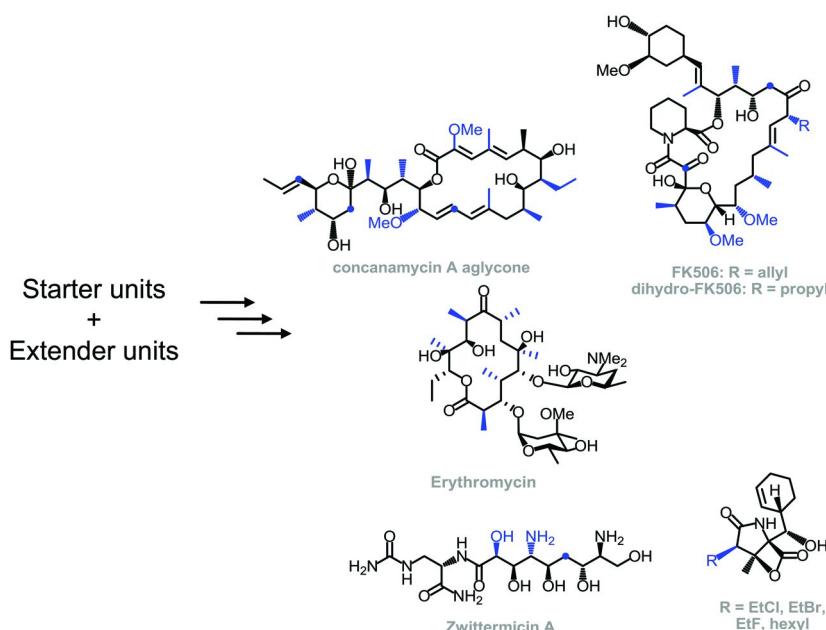
Many PKSs are multi-modular (and multi-domain) enzyme complexes (4). Each module selects and incorporates a designated small molecule building block into its polyketide structure. Each module has several enzymatic domains and each domain has a specific function, such as acyl transferase (AT), acyl carrier protein (ACP), keto synthase (KS), keto reductase (KR), or thioesterase (TE), amongst others (4).



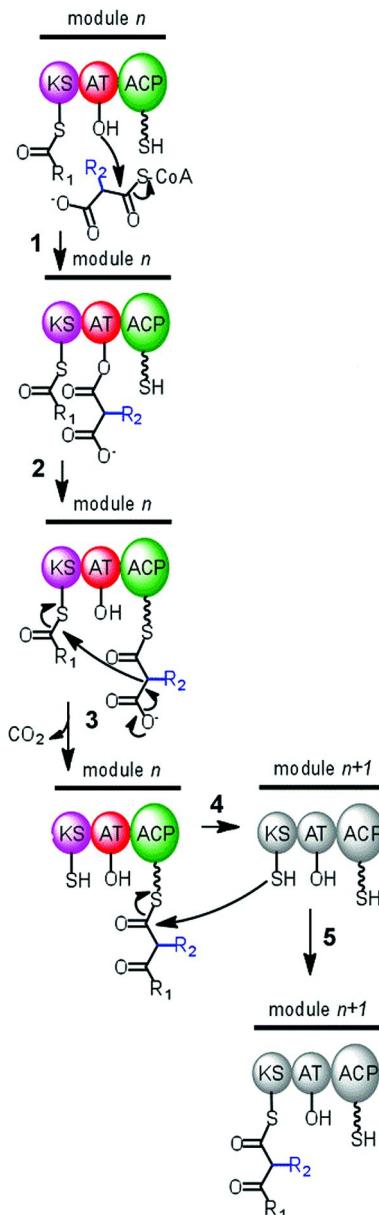
*Figure 2. Summary of natural starter units (**1a-1d**) and natural (**2a-2d**) and non-natural (**4a-4f**) extender units activated as the CoA thioesters and non-standard extender units activated with an acyl carrier protein thioester (ACP, **3a-3d**) as substrate for PKSs.*

The contribution that extender unit building blocks make to polyketide structure are highlighted in Figure 3, which shows the contribution from a side chain's carbon 2 (C<sub>2</sub>) of a derivatized malonyl-Coenzyme A (malonyl-CoA, **2a**) extender unit (Figure 2). In general, there is a limited selection of extender units that are available to PKSs in the natural extender unit pool (Figure 2) (4, 5). The most common include malonyl-CoA (**2a**), methylmalonyl-CoA (**2b**), and

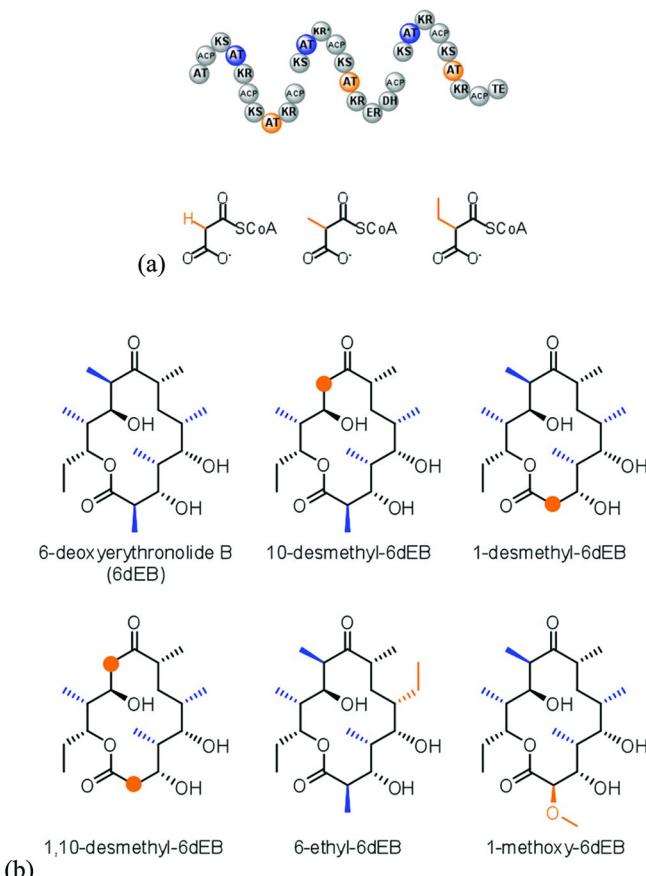
ethylmalonyl-CoA (**2c**) (Figure 2). Non-standard extender units are also available to PKSs in the natural pool, but they are attached to an acyl carrier protein (ACP), instead of CoA, as their respective thioesters (**6**). These non-standard ACP thioesters include methoxymalonyl-ACP (**3a**), hydroxymalonyl-ACP (**3b**), allylmalonyl-ACP (**3c**), and aminomalonyl-ACP (**3d**) (**6**). As mentioned previously, the carrier protein is part of the polyketide biosynthetic apparatus. Additional extender units are available such as choroethylmalonyl-CoA and hexylmalonyl-CoA, however, these originate from more specialized pathways (**7**). For the generation of non-standard extender units, polyketide producing organisms typically provide only two or three of these specialized pathways. Consequently, PKSs have available only a modest choice of extender units, limiting the side chain diversity seen in many polyketide natural products. Thus we aimed to (a) introduce a larger non-natural pool of extender units not previously available in natural biosynthetic systems, (b) introduce this non-natural chemical functionality into polyketides using their own biosynthetic machinery, and (c) engineer modified PKSs with broad extender unit specificity for the regioselective installation of these non-natural functionalities into a growing polyketide.



*Figure 3. Extender unit contribution (blue) to the overall structure of some selected polyketide natural products. The aim is to modify polyketide biosynthesis to produce analogues regioselectively modified with non-natural functionality at these highlighted positions.*



*Figure 4. Elongation process for the biosynthesis of a polyketide (PK). (1-2) AT domain recognizes and attaches an acyl-CoA substrate onto the AT domain (red) of PKS module  $n$ , then transfers the acyl group onto the ACP domain. (3) The newly added malonyl derivative is decarboxylated and undergoes Claisen condensation with the ketide unit(s) in the KS domain, elongating the PK chain and releasing carbon dioxide (CO<sub>2</sub>). (4) The KS domain of the neighboring module  $n+1$  picks up the growing PK chain from the ACP. (5) Entire process begins again on for a new building block (grey module).*



*Figure 5. Erythromycin and its PKS. (a) Erythromycin PKS and limited natural pool of extender units. Erythromycin PKS is made up of three separate polypeptide chains, comprising six 'extender' modules in sequence. (b) Erythronolide B aglycon precursor of erythromycin (Figure 1). Analogues made by genetic modification to the erythromycin PKS AT-domain have only conservative substitutions. Color code: AT domains colored (blue or orange, a) match corresponding conservative modification in the natural product by color (blue or orange, b) (9–11).*

Acyltransferase (AT) domains are the 'gatekeeper' domains within the PKS module, and they are responsible for recognizing and selecting the extender unit in the producing microorganism (4). The AT domain selects and then attaches the extender unit to an acyl carrier protein (ACP) of the same module (Figure 4). In a PKS, KS domains are responsible for catalyzing carbon-carbon formation within growing polyketides. Emerging research suggests that unlike AT domains, KS domains are promiscuous with respect to processing extender units that are

not naturally provided by the producing organism (8). Although the variety of extender units explored had only conservative changes, compared to those naturally accepted. With respect to AT domains, genetic substitution with AT domains of different specificity has been achieved, albeit with reduced yields (9–11). As each AT domain is responsible for the introduction of a single ketide unit, further extending the AT domain substrate scope by altering AT specificity was thought to provide a regioselective route to alter side chains of a polyketide. The general idea is referred to as combinatorial biosynthesis, which in some respects is still in its infancy.

Consider the PKS responsible for the biosynthesis of 6-deoxyerythronolide B (the precursor to erythromycin, Figure 5). Erythromycin PKS contains six modules that are distributed across three separate polypeptides with a total of seven AT domains. In this particular case the PKS produces the erythromycin scaffold, 6-deoxy-erythronolide B (Figure 5). Genetic substitution of the AT domains with AT domains from other polyketide pathways took advantage of the co-linearity in organization of the erythromycin PKS domain and the final erythronolide structure (Figure 5) (9–11). This genetic modification of the erythromycin PKS enzyme provided modest promiscuity, such that the PKS could recognize **2a–2c** (Figure 2 and Figure 5) (9–11). Although successful, such genetic experiments are not efficient, not only because genetic substitution of AT domains is inherently difficult (due to imprecise AT boundary selection and/or improper protein:protein interactions), but also because such studies rely on *in vivo* extender units (which are inherently limited).

In an effort to make more significant changes in resulting polyketide structure we sought to expand and extend upon the extender unit building block pool to provide a large array of non-natural extender units for use by PKSs. For example, extender units with appended azides ( $\text{N}_3$ ) or alkynes (**4a–4f**, Figure 2). In addition to providing diversity, these particular functional groups provide the potential for subsequent downstream chemical ligation chemistry. Thus the structural changes afforded by expanding the extender unit pool are potentially much greater than those provided by genetic modification of the AT domain alone.

## 2. Results and Discussion

### 2.1. Overview of Synthetic Biology Philosophy for Polyketide Modification

An overview of the synthetic biology behind this research is shown in Figure 6. Not only do non-natural building blocks need to be supplied to PKSs, but also a method must be generated to incorporate these extender units into growing polyketides using highly permissive PKSs or accessory enzymes. To generate the non-natural acyl-CoAs *in vitro*, a series of non-natural malonate diacids building blocks will be fed to growing bacterial cultures. These non-natural diacids freely diffuse into the microbial cell, and are activated as their acyl-CoA thioesters via a suitable promiscuous synthetase (Figure 6) (12). We recently completed a program of directed enzyme evolution to create a highly promiscuous acyl-CoA synthetase for the synthesis of a broad array of tailor-made extender units (acyl-CoAs) to be later installed into growing polyketides (Figure 6). With this broad panel of

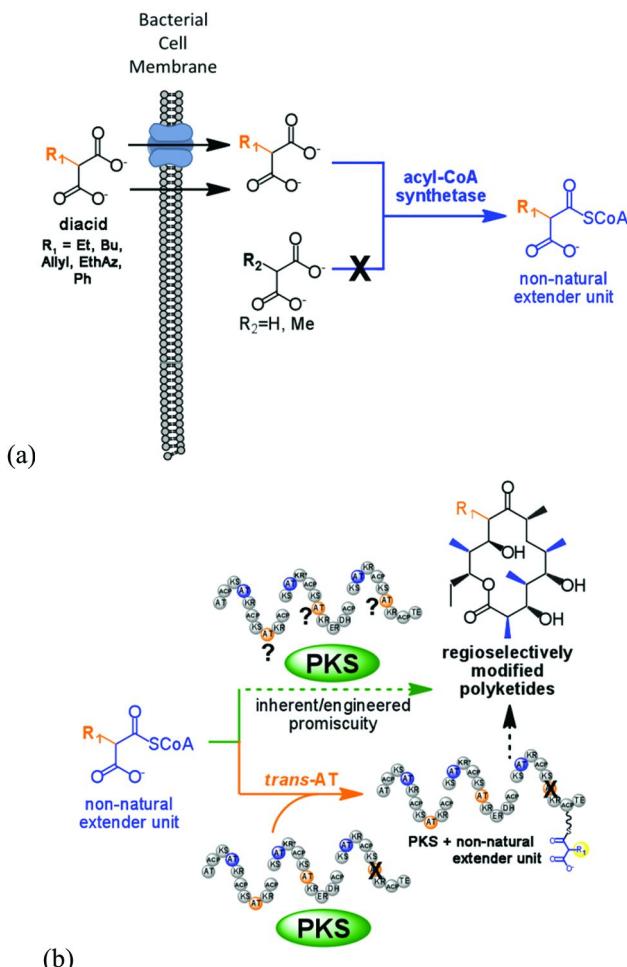
non-natural extender units in hand we moved onto the next challenge to probe the promiscuity of several PKS systems, to identify naturally occurring PKSs that would process these non-natural building blocks.

It is possible that naturally occurring AT domains of PKSs cannot process our library of non-natural building blocks, thus a novel approach is proposed employing PKS machinery called *trans*-acyl transferases (*trans*-ATs) that are found in so-called “AT-less PKSs” (PKSs that lack AT domains). Such *trans*-ATs are not embedded within a PKS module as with canonical PKS machinery, rather they are freestanding enzymes that act in *trans* with a cognate ACP and replace the *cis*-AT usually found within a PKS module (13, 14). As *trans*-ATs are exogenous these enzymes could more easily traverse distinctive biosynthetic systems, as compared to their *cis* counterparts, and thus are interesting from a combinatorial biosynthesis perspective. Directed enzyme evolution can then be employed to shift the specificity of a PKS AT-domain, *cis* or *trans*, toward non-natural extender units (Figure 6).

## 2.2. Engineering a Promiscuous Acyl-CoA Synthetase To Process Non-Natural Extender Unit Diacids

The acyl-CoA synthetase MatB from *Rhizobium trifolii* was used as a template for engineering an enzyme for the generation of a library of non-natural acyl-CoAs (12). Wild-type *R. trifolii* MatB displays moderate promiscuity to non-natural extender units (12), thus it is an ideal starting point for engineering. In addition, an X-ray crystal structure of a MatB homologue from *Streptomyces coelicolor* had been solved (15). We used the structure to guide a semi-rationale random mutagenesis strategy to provide libraries of enzyme variants. This strategy required the development of a high-throughput screen (HTS) to identify, from large libraries of mutants, which mutants shifted MatB extender unit specificity or activity. The logic behind this HTS is shown in Figure 7. In the enzymatic reaction under study MatB catalyses the reaction of the CoA thiol with the corresponding diacid to generate a thioester. The residual concentration of unreacted CoA thiol could then be determined using Ellman’s reagent – a well-known thiol sensor – that reacts with thiols forming a UV active species that is monitored at 420 nm (12, 16, 17). The amount of unreacted CoA thiol that subsequently reacts with Ellman’s reagent is determined by UV/Vis spectroscopy. A decrease in the absorbance at 420 nm is expected as CoA is consumed in the CoA ligase reaction.

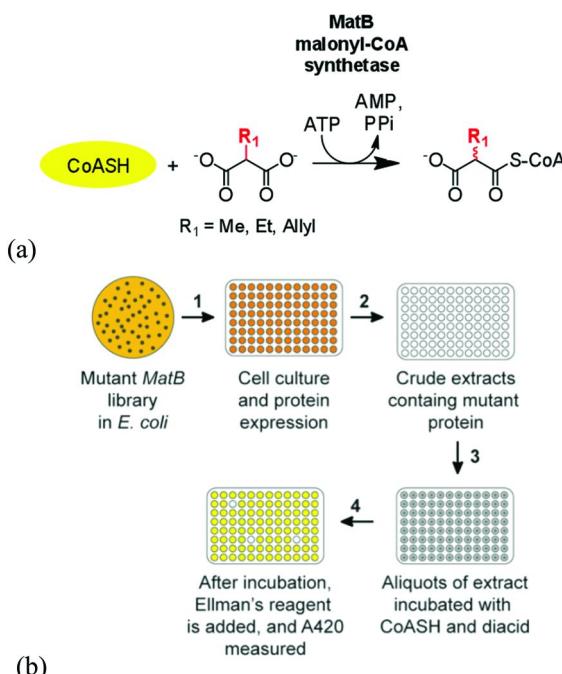
Saturation libraries of several hundred MatB random mutants were generated in *Escherichia coli* (*E. coli*) colonies over-expressing the MatB mutants (12). After production in microtiter plates, the cells were lysed, and thiol CoA consumption was monitored with Ellman’s reagent. Two residues, identified from the crystal structure of the MatB homologous enzyme (15), were ultimately confirmed to be intimately involved in **2b** specificity. These residues were threonine 207 (Thr207) and methionine 306 (Met306) (12). In this initial screen only **2b** and **4c** building blocks were used, as the former is the natural substrate of wild type (*wt*) MatB, and **4c** represents a very poor substrate for *wt*MatB.



*Figure 6. Overview of the synthetic biology approach proposed for polyketide diversification. (a) Diffusion or protein-assisted transport of select diacids across the bacterial cell wall and CoA-activation catalyzed by promiscuous acyl-CoA synthetases provided a library of non-natural extender units. (b) Feeding the non-natural extender units to a PKS in a bacterial cell over-expressing exogenous trans-ATs provides the acylated PKS. PKSs carrying non-natural extender unit could then be used to generate regioselectively modified polyketides.*

A larger panel of ‘non-natural’ diacids were then screened against a selection of active mutants from the T207X and M306X library, as well as a select set of double mutants (T207X/M306X) that were purified to homogeneity and characterized using a robust HPLC-based assay (12). The results are shown in

the 3D plot shown in Figure 8, with specificity constant ( $k_{cat}/K_m$ ) as a function of mutant and building block diacid used. On the *x*-axis are the ten diacid building blocks screened against the mutant enzyme screened on the *y*-axis. The activities of all the mutants were compared to the *wt*MatB enzyme. As demonstrated in the table, the double mutants were more promiscuous, in terms of enzyme activity, in comparison to the single mutants. As mentioned previously, *wt*MatB does display weak promiscuity toward several building blocks, but it does not utilize five of the non-natural extender units (Figure 8). Further, some of the single mutants display significant changes in specificity such that they turnover one or two more non-natural extender units with several hundred-fold shifts in specificity away from the natural **2b** analogue. The double mutant T207G/M306I was the most promiscuous mutant and was able to utilize every member of the non-natural extender unit diacid panel (12).



*Figure 7. Assay used to screen acyl-CoA synthetase activity. (a) Reaction catalyzed by malonyl-CoA synthetase. (b) Ellman's reagent based microtiter plate assay to screen random mutant MatB activity (12).*

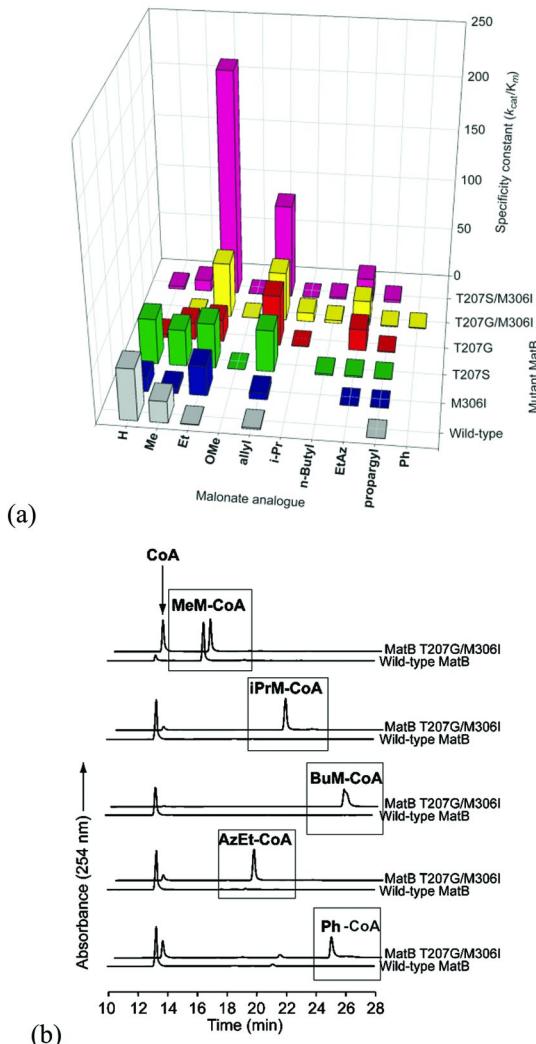


Figure 8. Assay of 'non-natural' diacids against a selection of active MatB mutants from the T207X and M306X library and T207X/M306X double mutant purified to homogeneity and characterized. (a) Activity-based assay to determine specificity ( $k_{cat}/K_m$ ). (b) HPLC-based assay to verify synthetic conversion (12).

To illustrate the synthetic utility of the T207G/M306I double mutant MatB enzyme (Figure 8a, yellow bars) *in vitro* we incubated it (along with the *wt*MatB enzyme as a control) with the entire panel of non-natural extender units for 3 h and analyzed the product mixture by high-performance liquid chromatography (HPLC, Figure 8b). The results show that *wt*MatB does not produce the corresponding acyl-CoAs with isopropylmalonate, ethylmalonate,

and phenylmalonate non-natural extender building blocks (Figure 8b). In contrast, the T207G/M306I double mutant MatB produces product with all the non-natural malonate diacids studied, as revealed by the corresponding peaks in the HPLC traces (Figure 8b). This data shows that the MatB mutant T207G/M306I can be used to access non-natural polyketide synthetase building blocks. The observation that these double mutant MatB enzymes are significantly less active toward **2a** and **2b** suggest the mutant enzymes are orthogonally selective toward the non-natural building blocks versus the *wt*MatB substrates. The results hint to the possibility that a non-natural polyketide can be built up *in vivo* in a microbial host over-expressing the mutant MatB enzyme by feeding it these non-natural building blocks. Finally, the result demonstrates the synthetic potential as part of our overall synthetic biology platform (Figure 6) (12).

### 2.3. A *trans*-AT Can Incorporate Non-Natural Extender Units onto a PKS

The chemo-enzymatic possibilities provided by the mutant MatB enzyme and novel building blocks led us to investigate a way to install and/or process these new building blocks into PKS systems. To do this, *trans*-AT systems were initially investigated. Recall *trans*-AT enzymes take the acyl group of an acyl-CoA and transfer it to a cognate ACP domain within a PKS module that lacks its own embedded (*cis*)-AT (13, 14). Specifically, we focused on three well-known model *trans*-ATs. Prior to this work, *trans*-ATs have only been probed with a small set of extender units that include **2a**-**2c**, and the previous studies revealed that each *trans*-AT utilizes only a single extender unit. Stringent *trans*-ATs include *Sorangium cellulosum* disorazole synthase (DSZS) and *S. coelicolor* fatty acid/polyketide biosynthesis enzyme MCAT; as their natural function is just to transfer **2a** onto a PKS (18, 19). The DSZS and MCAT producers have extender units with larger side chains available to them within the cell, but don't use them due to this presumed high selectivity. Another *trans*-AT, KirCII, involved in kirromycin biosynthesis is known to transfer an extender unit with a relatively large side chain, **2c** (20), and we hypothesized it would be promiscuous to other malonyl side chains for several reasons. First, it is known that KirCII discriminates against smaller side chains than **2c** (e.g. Me and H), otherwise kirromycin analogues with these modifications would be known. Second, the kirromycin producer does not have available to it extender units with side-chains larger than ethyl. Consequently, KirCII was likely never under evolutionary pressure to discriminate against other larger acyl-CoAs.

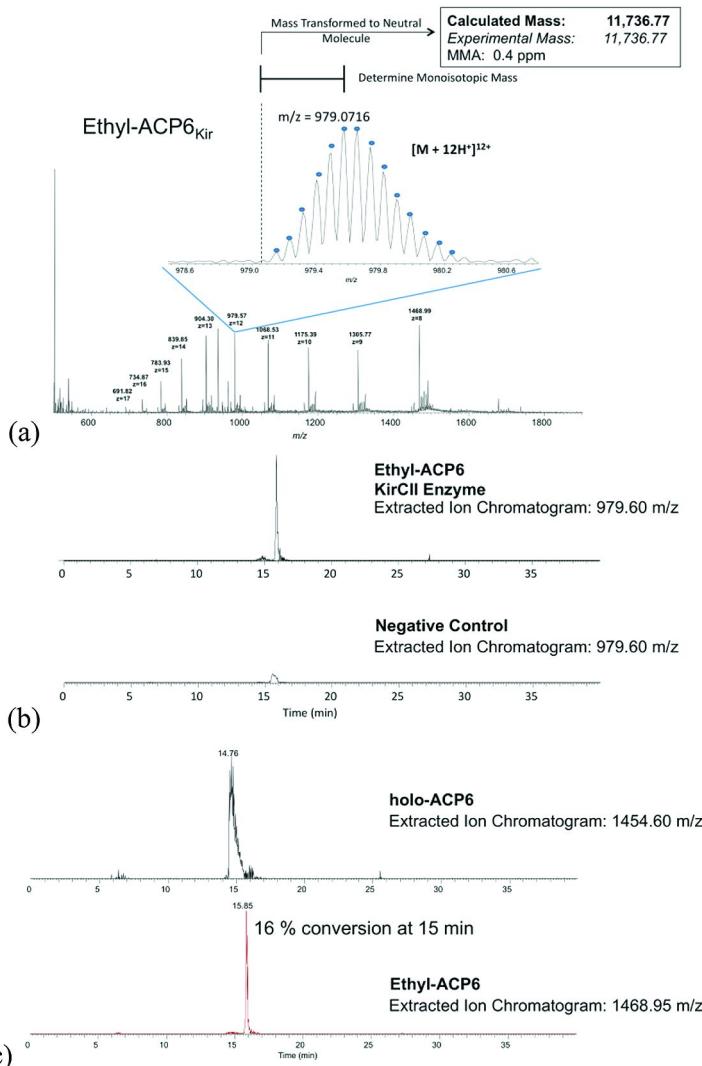
The acyl-CoA promiscuity of KirCII, MCAT and DSZS with the non-natural extender units was probed. To do this the excised acyl carrier proteins (ACPs) were provided and the corresponding acylated ACP was detected by mass spectrometry (MS). A Thermo Fisher™ Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer equipped with a 7 Tesla magnet was used to acquire spectra of intact proteins and reacted proteins using liquid chromatography-MS (LC-MS) at 50,000 resolving power. FT-ICR and LC-MS were used to detect ACP, using DSZS, MCAT or KirCII as the *trans*-AT. In the example shown (Figure 9) the charge state distribution for the ethylmalonyl-ACP

(Ethyl-ACP6) was detected. Employing a higher resolving power the 12+ charge state was resolved showing the expected  $^{13}\text{C}$ -isotopic distribution. As the molecular formula of the product can be calculated experimentally (Ethyl-ACP6 MW<sub>calc</sub> = 11736.77 amu), as well as the natural abundances of different isotopes (Figure 9), the experimental  $^{13}\text{C}$  isotopic distribution of Ethyl-ACP6 for the 12+ charge state can be fit the theoretical distribution (Ethyl-ACP6 MW<sub>expt</sub> = 11736.77 amu). Thus the monoisotopic mass can be determined. The monoisotopic mass is calculated using the exact mass of the most abundant isotope of each element in the molecule and is well-defined allowing greater confidence in protein assignments. As shown in Figure 10 the experimental mass for the 12+ charge state ( $m/z$ ) matches the calculated mass, confirming that ethyl-ACP6 was formed. Negative control reactions were carried out in parallel that lack the *trans*-AT to establish that ACP acylation is completely dependent on the presence of the *trans*-AT enzyme. The areas under the MS curve were used to calculate the percent acylation or conversion (Figure 9) (21).

Not surprisingly, the highly stringent *trans*-ATs MCAT and DSZS were very active toward **2b** (100% conversion) and displayed only trace activity towards **2c** and **4f** (1% and 2% conversion, respectively). In contrast, acylation of the ACP in the presence of KirCII was very active toward **2c** (natural substrate) as well as **4a**, **4b**, **4c**, and **4f** extender units (16%, 52%, 100%, 10% conversion, respectively). This is the first example of a poly-specific *trans*-AT that displays activity toward non-natural extender units. Taken together the results offer a novel potential strategy to incorporate diversity into polyketides biosynthetically, without chemical synthesis, by harnessing *trans*-AT promiscuity efficiently (21).

#### 2.4. Can PKSs Process Non-Natural Acyl-CoAs?

In a final step, the specificity of a complete PKS module was studied employing a well-known assay based on the terminal module and thioesterase domain (TE) of the erythronolide B synthase (DEBS) called Mod6TE (15). The S-*N*-acetylcysteamine ester (**5**), a diketide mimic of the natural module 6 substrate, was provided to Mod6TE. S-*N*-acetylcysteamine is an electrophile that mimics what would normally be provided as an ACP-linked thioester by the preceding DEBS module (Figure 10). With this substrate mimic in hand the promiscuity of the complete module was screened with the entire panel of non-natural acyl-CoA extender units generated using our optimized mutant MatB (Figure 2, Figure 10). The formation of the expected triketide lactone was then monitored by high-performance liquid chromatography (HPLC) and LC-MS. The results (not shown) confirmed that this Mod6TE (as part of an entire erythromycin PKS) was able to utilize, as expected, **2b** and **2c** and it also turned over **4a**, **4b**, **4c**, and **4e** (data not shown) (21). Again the azido and propargyl functional groups can be further utilized for chemoselective ligation. The results hint that the AT domains - responsible for extender unit selection - and more importantly the KS domains that actually catalyze carbon-carbon bond formation can utilize a diverse array of non-natural extender units (21).



*Figure 9. FT-ICR MS spectra used to detect acylated CP using DSZS, MCAT or KirCII (15). The ACPs require post-translational modification with phosphopantetheine; Apo = non-modified, holo = modified. Note calculated mass Ethyl-ACP6 = 11,736.77, mass/charge ratio (m/z) = 11,736.77/(12+ charge state) = 979.716. (a) FT-ICR MS spectra of Ethyl-ACP6, comparison of calculated versus experimentally determined MS. (b) Comparison of MS spectra obtained for the acylation of a CP in the presence of 2c and KirCII and its negative control without KirCII. (c) Spectral comparison used to determine % conversion (21).*

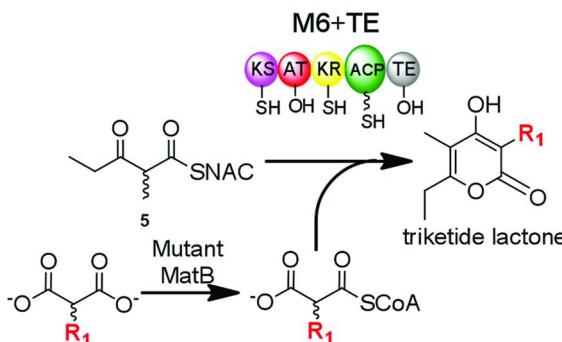


Figure 10. Scheme for the *S*-*N*-acetyl cysteamine ester (**5**) and Mod6TE-based assay to screen for non-natural extender unit PKS incorporation. Formation of the corresponding triketide lactone with the incorporated non-natural extender unit can be monitored by HPLC, LC-MS or MS for quantification (21).

### 3. Conclusion

A robust mutant MatB synthetase was developed and utilized to provide a library of ten natural and non-natural extender unit acyl-CoAs. This library was in turn used to probe the promiscuity of several PKS systems, including the *trans*-ATs DSZS, MCAT and KirCII, and an intact PKS module. The results revealed significant promiscuity in the erythromycin PKS system toward extender units that are not found in natural systems and led to the discovery of the first polyspecific *trans*-AT, KirCII. These two discoveries imply this novel promiscuity could potentially be harnessed for the regioselective modification of polyketides. Thus directed evolution aimed to shift the specificity of acyl transferases (ATs) and *trans*-ATs towards certain sets of extender units may provide non-natural polyketide analogues.

### Acknowledgments

The authors would like to thank the Mass Spectrometry Facility at NC State University. This study was supported by an NSF CAREER Award to G.J.W (CHE-1151299), and the W. M. Keck Foundation.

### References

1. Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75*, 311–335.
2. Newman, D. J.; Cragg, G. M. Natural product scaffolds as leads to drugs. *Future Med. Chem.* **2009**, *1*, 1415–1427.
3. Dobson, C. M. Chemical space and biology. *Nature* **2004**, *432*, 824–828.

4. Ames, B. D.; Tsai, S. C. Structural enzymology of polyketide synthases. *Methods Enzymol.* **2009**, *459*, 17–47.
5. (a) Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.* **2009**, *26*, 90–114. (b) Wilson, M. C.; Moore, B. S. Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/ reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* **2012**, *29*, 72–86.
6. Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.* **2009**, *26*, 90–114.
7. Wilson, M. C.; Moore, B. S. Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* **2012**, *29*, 72–86.
8. Bonnett, S. A.; Rath, C. M.; Shareef, A. R.; Joels, J. R.; Chemler, J. A.; Håkansson, K.; Reynolds, K.; Sherman, D. H. Acyl-CoA subunit selectivity in the pikromycin polyketide synthase PikaIV: steady-state kinetics and active-site occupancy analysis by FTICR-MS. *Chem. Biol.* **2011**, *18*, 1075–1081.
9. McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Ashley, G. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1846–1851.
10. Stassi, D. L.; Kakavas, S. J.; Reynolds, K. A.; Gunawardana, G.; Swanson, S.; Zeidner, D.; Jackson, M.; Liu, H.; Buko, A.; Katz, L. Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7305–7309.
11. Kato, Y.; Bai, L.; Xue, Q.; Revill, W. P.; Yu, T. W.; Floss, H. G. Functional expression of genes involved in the biosynthesis of the novel polyketide chain extension unit, methoxymalonyl-acyl carrier protein, and engineered biosynthesis of 2-desmethyl-2-methoxy-6-deoxyerythronolide B. *J. Am. Chem. Soc.* **2002**, *124*, 5268–5269.
12. Koryakina, I.; Williams, G. J. Mutant Malonyl-CoA synthetases with altered specificity for polyketide synthase extender unit generation. *ChemBioChem* **2011**, *12*, 2289–2293.
13. Cheng, Y. Q.; Tang, G. L.; Shen, B. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3149–3154.
14. Musiol, E. M.; Weber, T. Discrete acyltransferases involved in polyketide biosynthesis. *Med. Chem. Commun.* [Online early access]. DOI: 10.1039/C2MD20048A. Published Online: Mar 6, 2012. <http://pubs.rsc.org/en/Content/ArticleLanding/2012/MD/c2md20048a> (accessed Jul 20, 2012).
15. Hughes, A. J.; Keatinge-Clay, A. Enzymatic extender unit generation for *in vitro* polyketide synthase reactions: structural and functional showcasing of *Streptomyces coelicolor* MatB. *Chem. Biol.* **2011**, *18*, 165–176.
16. Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol.* **1999**, *6*, 117–125.

17. Sharma, K. K.; Boddy, C. N. The thioesterase domain from the pimaricin and erythromycin biosynthetic pathways can catalyze hydrolysis of simple thioester substrates. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3034–3037.
18. Wong, F. T.; Jin, X.; Mathews, II; Cane, D. E.; Khosla, C. Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase. *Biochemistry* **2011**, *50*, 6539–48.
19. Revill, W. P.; Bibb, M. J.; Hopwood, D. A. Purification of a malonyltransferase from *Streptomyces coelicolor* A3 (2) and analysis of its genetic determinant. *J. Bacteriol.* **1995**, *177*, 3946–52.
20. Musiol, E. M.; Härtner, T.; Kulik, A.; Moldenhauer, J.; Piel, J.; Wohlleben, W.; Weber, T. Supramolecular templating in kirromycin biosynthesis: the acyltransferase KirCII loads ethylmalonyl-CoA extender onto a specific ACP of the trans-AT PKS. *Chem. Biol.* **2011**, *18*, 438–44.
21. Koryakina, I.; McArthur, J.; Williams, G. J. Submitted for publication.

## Chapter 10

# Image Analysis Method for Evaluating Heterogeneous Growth and Differentiation of Embryonic Stem Cell Cultures

Megan M. Hunt,<sup>1,2</sup> Guoliang Meng,<sup>3</sup> Derrick E. Rancourt,<sup>3</sup>  
Ian D. Gates,<sup>2</sup> and Michael S. Kallos\*,<sup>1,2</sup>

<sup>1</sup>Pharmaceutical Production Research Facility (PPRF), Schulich School of Engineering, University of Calgary, Alberta, Canada

<sup>2</sup>Department of Chemical and Petroleum Engineering, Schulich School of Engineering, University of Calgary, Alberta, Canada

<sup>3</sup>Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Calgary, Alberta, Canada

\* E-mail: [mskallos@ucalgary.ca](mailto:mskallos@ucalgary.ca). Phone: 403-220-7447. Fax: 403-284-4852

Embryonic stem cells (ESCs) not only have the ability to readily proliferate in culture, but also the potential to differentiate into all adult cell types and as such they are emerging as a promising cell source for tissue engineering applications. The spatial arrangement and temporal evolution of ESCs in laboratory cultures can be complex and often leads to heterogeneous cell densities and viabilities. This spatial heterogeneity influences factors such as mass transfer, creating nutrient gradients within the culture medium, which may in turn lead to variations in cell growth, viability, and differentiation. Methods which are available to accurately assess cell pluripotency (or lack thereof) include fixing and staining cells with antibodies for flow cytometry or immunocytochemistry, but these methods render the cells useless for further applications. The present research focuses on optimization of a novel non-invasive alternative tool to assess ESC colony pluripotency using fractal analysis. Here, we report preliminary results on the use of this image analysis technique to index cell growth and differentiation in human and mouse ESCs in static dishes to determine changes in the heterogeneity and complexity of a proliferating cell culture.

## 1. Introduction

Embryonic stem cells (ESCs) are derived from the inner cell mass of a fertilized embryo at the blastocyst stage (1, 2). In contrast to somatic or specialized adult cells that suffer from limited proliferation capabilities, ESCs divide almost indefinitely and retain the ability to differentiate into all adult cell types - including those cells that do not naturally regenerate. These pluripotent and self-regenerative behaviors have made ESCs an attractive cell source for potential use in regenerative medicine (3) to treat degenerative diseases and in tissue engineering (4).

Tissue engineering, or tissue replacement therapy, is a process whereby tissue constructs, in this case generated from ESCs, could be used to regenerate and thus restore function to dead or damaged tissues. Tissue engineering or replacement has been implicated as a the treatment for diseases such as osteoarthritis (5), osteoporosis (6), Parkinson's disease (by transplantation with regenerated dopaminergic neurons) (3), or to regenerate injured or damaged tissues such as those sustained in spinal cord damage (7). However, to study potential therapeutic uses, clinically relevant numbers of ESCs must be generated, and this is estimated to be in the of billions to tens of billions of cells (8). A general procedure is used for the expansion of different ESC lines on a clinically relevant scale, and starts with static tissue cultures (5 mL) of ESCs that are aliquoted (passaged) into smaller scale suspension bioreactors (100 mL), allowed to proliferate as aggregates of cells, and subsequently passaged into another larger computer controlled bioreactor ( $> 1\text{L}$ ) (3, 9–12). In comparison to traditional static culture flasks, this ESC expansion procedure in large bioreactors results in improved cell proliferation. This is because large bioreactors eliminate many problems associated with limited growth surface area seen in tissue culture flasks as well as aiding in efficient nutrient transfer to the cells through continuous mixing (11). Further, large bioreactors eliminate the need to manage hundreds of tissue culture flasks at each passage. With continued research and improvement of suspension-based cell expansion technology, new culture protocols are developed and at each step and proliferated cell batches must be compared to the parent static control culture; always under the assumption that static cultures are standardized, which is not necessarily the case (3, 9–12).

Quality control (QC) data to determine the health and viability of a growing cell culture typically includes live (viable) and dead cell counts and analysis of specific proteins (referred to as markers) linked to the pluripotent nature of the cells (3, 10, 11). An example of a plot of the viable cell density and viability as a function of time for a static culture of mouse ESC (mESC) line D3 is shown in Figure 1. In these particular assays viable cell number and viability were determined at specific time intervals over one passage. In addition to cell viability assays, ESCs from the culture can also be harvested and fixed and stained for immunocytochemistry or flow cytometry to show expression of specific ESC markers, such as the transcription factor Oct4, or early differentiation markers, such as  $\alpha$ -fetoprotein (Table 1) (10, 13–22). These analytical tools although useful, require significant processing, large quantities of cells, take long time

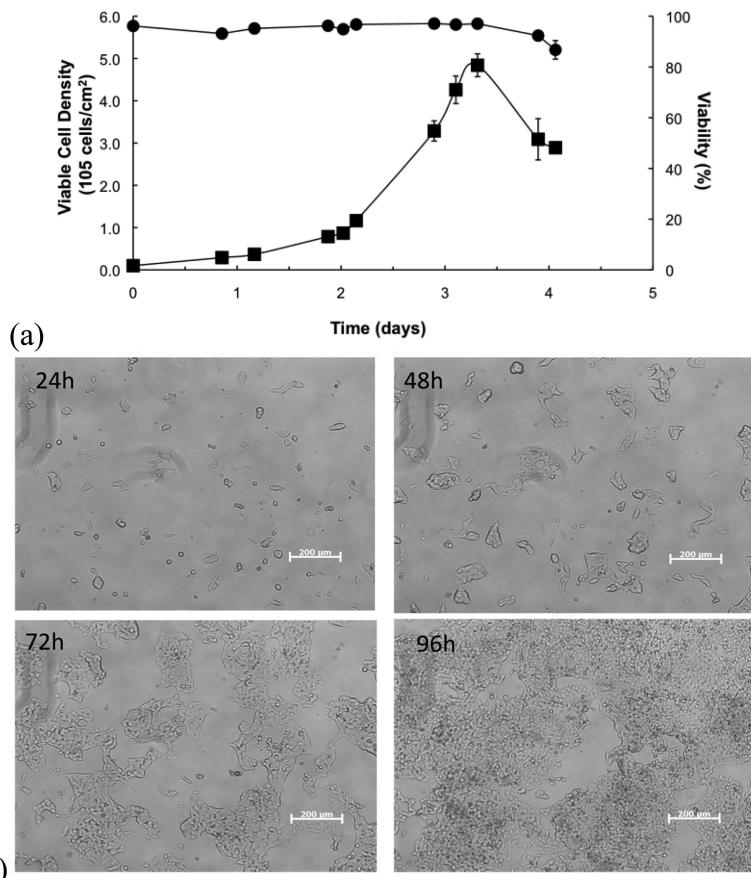
frames to run (hindering real time monitoring), and ultimately result in cell destruction (13). As such, the majority of day-to-day static culture is carried out and monitored via morphological analysis.

**Table 1. Standard markers of pluripotency and early differentiation for mouse and human ESCs (13–22)**

		<i>Pluripotency marker</i>	<i>Early differentiation markers</i>			
		<i>Intracellular</i>	<i>Surface</i>	<i>Endoderm</i>	<i>Ectoderm</i>	<i>Mesoderm</i>
Human	Oct4		SSEA3	$\alpha$ -fetoprotein	$\beta$ -III-tubulin	Actin
		Nanog	SSEA4			
		Sox2	TRA-1-60			
			TRA-1-81			
Mouse	Oct4		SSEA1	Sox17	Sox1	Brachyury
		Nanog		Foxa2		
		Sox2				

Analysis of culture morphology is a way to track a growing culture of ESCs without destruction of the cells (13). Qualitative analysis of pictures or time lapsed videos of growing colonies under a microscope reveals a great deal of information about the health of a culture as it is based on shape of the colonies themselves (Figure 1). Unfortunately, morphological evaluation is subjective, as it is based on an observation of the heterogeneity within cell cultures. As a subjective process properly trained individuals are able to assess culture health, and whether stem cells have started to undergo differentiation based on cell shape and orientation within the cultures. However, training and experience are variables that result in widely differing opinions on what a “healthy” (i.e. viable, undifferentiated) culture should in fact look like. Thus the question becomes: how to take the subjectivity and guesswork out this qualitative morphological evaluation?

The goal of the present research was to develop a robust, repeatable, and reliable method to assign numerical values to cell culture morphology analysis. If a numerical value could be assigned it would allow for quantitative assessment, rather than a qualitative assessment, of cell cultures. To do this we sought to apply the principle of fractal geometry and fractal shapes to growing cell cultures.



*Figure 1. Standard culture data for mESC line D3 cells in static batch 2D cultures. (a) Plot of viable cell density (■,  $10^5$  cells/cm $^2$ ) and viability (●, %) as a function of time (days). (b) Time-lapsed images of a growing culture showing random heterogeneous growth of cell colonies.*

## 2. Material and Methods

Unless otherwise stated, all cell handling procedures were conducted in a sterile laminar flow hood using aseptic techniques. All cell cultures were incubated in standard conditions (100% humidity, 37°C and 5% CO<sub>2</sub>).

### 2.1. Mouse ESC Culture

For all mouse ESC cultures, the D3 line was utilized (ATCC, Rockville, MD). Standard mouse ESC maintenance medium consisting of Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) supplemented with 15% fetal bovine serum

(FBS, Invitrogen), 1% non-essential amino acids (NEAA, Invitrogen), 0.2%  $\beta$ -mercaptoethanol (Invitrogen) and 1% penicillin/streptomycin (Pen/Strep, Invitrogen) was used. Mouse ESC cultures were grown in gelatinized 6-well plates (BD Falcon) containing maintenance medium supplemented with 1000 U/mL Leukemia inhibitory factor (10<sup>6</sup>U LIF, ESGRO®, EMD Millipore) and passaged with 0.25% trypsin (Invitrogen) every 4 days or when they were judged to be approximately 80% confluent.

## 2.2. Human ESC Culture

The human ESC line H1 (WiCell Research Institute) was maintained on Matrigel (BD Biosciences) coated 35 mm or 60 mm dishes (Nunc) in complete mTeSR1 medium (StemCell Technologies) following manufacturers protocols. Cultures were passaged as small clumps via Dispase (1 mg/mL) (StemCell Technologies) every 5 days or when they were judged to be approximately 70% confluent.

## 2.3. Human ESC Differentiation via Embryoid Body Formation

Human ESCs were aggregated in agar coated 60 mm dishes (Nunc) in differentiation medium consisting of DMEM (Invitrogen), 10% FBS (Invitrogen), 1% non-essential amino acids (NEAA, Invitrogen), 0.1mM  $\beta$ -mercaptoethanol (Invitrogen) and 1% penicillin/streptomycin (Pen/Strep, Invitrogen). The embryoid bodies were transferred on Day 4 onto gelatin coated 4-well plates (Nunc) to allow attachment and further spontaneous differentiation over an additional 10 days. Medium was replaced daily.

## 2.4. Cell Counts

To monitor growth of mouse ESCs, two wells of a 6 well plate were harvested by dissociating the cell colonies with 1 mL/well of 0.25% trypsin for 10 minutes. Medium containing FBS was added at 1 mL/well to inactivate the trypsin after 10 minutes and two 100  $\mu$ L samples were taken from each well to determine cell numbers and viability using a hemocytometer and the trypan blue exclusion method. Human ESC cultures were dissociated from 35 mm plates using 0.25% trypsin in a similar manner, however only one 35 mm dish was harvested and two samples of 100  $\mu$ L counted.

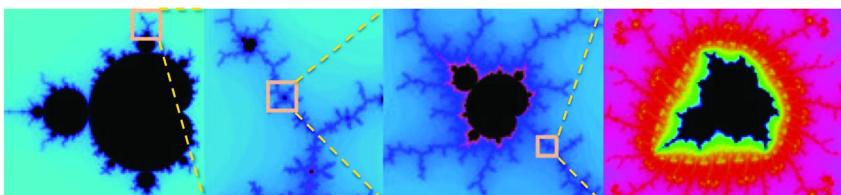
## 2.5. Immunocytochemistry of Human ESCs

For assessment of specific marker expression for both pluripotency and early differentiation the following protocol was used. Cell cultures were fixed with 4% paraformaldehyde then permeated with 0.1% Triton-X 100 (in PBS) and

incubated with PBS containing 10% FBS. For pluripotent cultures, the primary antibody used was anti-Oct4 (Santa Cruz) at the recommended dilution of 1:100. For differentiating cultures, markers against  $\beta$ -III-tubulin (ectoderm) and  $\alpha$ -fetoprotein (endoderm) (1:400 dilution, Sigma Aldrich) were used. Primary antibodies were added to cultures and stored at 4°C overnight. After washing, an appropriate secondary antibody (in this case, Alexa Fluor 546 goat anti-mouse, 1:200 dilution, Invitrogen) was added to each culture for 2 hours at room temperature.

## 2.6. Image Generation and Fractal Dimensions

Greyscale images of mouse ESC cultures were captured on a Zeiss Axiovert 200M microscope. Greyscale and fluorescent human ESC culture images were captured using a Zeiss LSM 510 confocal microscope. Conversion of all images to binary format was carried out via the “make binary” command within the free ImageJ (version 1.46) software from NIH (23). The open source plug in FracLac (version 2.5) for ImageJ was then used to determine the fractal dimension of the binary images using the box counting method command (24).



*Figure 2. Benoit fractal set showing self-similarity at increasing magnifications (35). Notice the repeating self-similar simple shape that is revealed at increased magnification.*

## 3. Results and Discussion

### 3.1. Fractal Geometry

Fractal geometry has been used to describe patterns in natural occurring objects from fern leaves to collagen fibers (25–28). As a tool to analyze shapes, fractal geometry more accurately describes naturally occurring objects than traditional Euclidian geometry (27). Fractal shapes are patterns that repeat at increasingly finer or coarser resolution, thus patterns are self-similar, repetitive, and reveal greater detail as they are enlarged (Figure 2) (27). Application of

the principles of fractal geometry to growing cell cultures may assign order to seemingly random structures (29–34). Objects that exhibit fractal geometry are referred to as fractals with an associated fractal dimension (27). A fractal dimension is a non-integer or an integer ratio that indicates the statistical degree of complexity of a shape or how a shape fills the space it occupies; and how the detail is impacted by scale. Briefly, for a two-dimensional object, if Euclidean, the area of the object scales with the length scale of the object to the power 2, for example, for a square, its area  $A = l^2$  where  $l$  is the length of the square edge. For a fractal object, the area may scale with a length scale to a fractal dimension between the limits 1 and 2, where 1 can be thought of as a straight line and 2 indicates an object is filling the area. The higher the fractal dimension, the more space filling is the object. Thus by assignment of a fractal dimension value to an object, the result is a systematic method to describe, predict, and simulate the object's heterogeneity and complexity.

### 3.2. Determining a Fractal Dimension

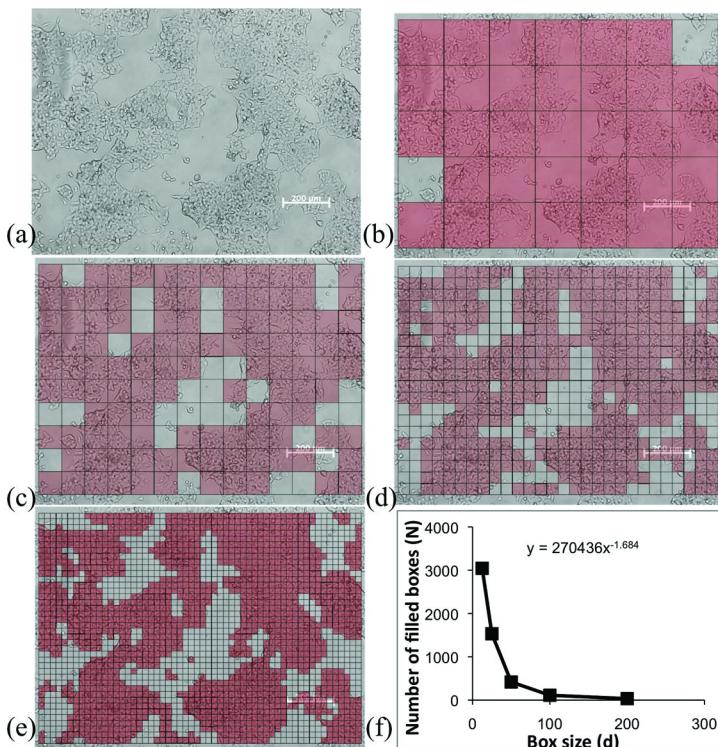
The most straightforward method to determine the fractal dimension of a two-dimensional (2D) image is using the box counting method (27). In this technique, a 2D image of an object of interest (in this case a cell culture) is gridded at length scale  $d$  (27). A count of the number of boxes ( $N$ ) necessary to cover the objects in the image or those that filled by the image of interest (in this case cell colonies) in the entire grid is performed (Figure 3). This process is repeated for several iterations over a range of grid sizes (smaller or larger,  $d$ ) that are changed by a constant factor (in this case a factor of 2, Figure 3). A plot of the number of filled boxes versus the size of the boxes is plotted (Figure 3). A plot of the logarithm of  $N$  versus the logarithm of  $d$  provides a relationship between the size of the grid and the number of cell colonies in the grid (Figure 4). The fractal dimension is derived from the slope of that line.

Application of the box counting method to estimate the fractal dimension of D3 mouse ESCs at 74.5 hours is shown in Figure 3, a grid scale from 200  $\mu\text{m}$  to 25  $\mu\text{m}$  was used. The results from box counting of this one cell photograph plotted on a logarithm scale gave a straight line, suggesting a fractal relationship (Figure 3). In the case shown the fractal dimension was calculated to be 1.68. As this is a 2D image, the fractal dimension is expected to be between 1 and 2, because the cells do not fill the area completely (there are barren regions).

### 3.3. Automation of the Box Counting Method

Fractal analysis and box-counting can be automated with computer software programs. Automation reduces error associated with manual box counting, increases output, and allows for faster analysis of images over one entire growth curve. The free National Institute of Health (NIH) software FracLac (24) from ImageJ (23) can be used to determine fractal dimensions of binary images. Binary images, digital images with only two possible numbers associated with a pixel

one number for background and another number for foreground (for example white and black), are needed to use the FracLac software. For example, cell colonies can be assigned black and background can be assigned white, in a black and white image.



*Figure 3. Illustration of the box counting method for the determination of a fractal dimension for mouse ESC line D3 at 74.5 hours. (a) The original grey scale photo was (b) gridded to  $N = 33$ ,  $d = 200 \mu\text{m}$  then gridded further (c) to  $N = 111$ ,  $d = 100 \mu\text{m}$  then gridded further (d)  $N = 418$ ,  $d = 50 \mu\text{m}$  and then further (e)  $N = 1532$ ,  $d = 25 \mu\text{m}$ . (not shown,  $N=3044$ ,  $d=12.5 \mu\text{m}$ ) (f) A plot of the number of filled boxes (N) for each box size (d) was plotted to provide the relationship between cells and size of boxes.*

Prior to running additional experiments, we confirmed that ESC colonies of pluripotent cells grow in a fractal manner. To do this images of human H1 ESCs in static 2D cultures were taken at specific time points (24, 48, 72, 96, 120 and 144 h), the images were investigated at increasing magnifications (10x, 20x and 40x), and the fractal analysis was carried out to calculate a fractal dimension at each time point. Recall that fractals are repeating patterns, and a principle of the

fractal geometry is that at higher magnification more self-similar fractal patterns appear (27). To prove a shape is a true fractal shape mathematically, and thus a fractal dimension, it should remain constant at all magnification levels. As the fractal dimension calculated at each magnification was essentially the same, the results suggested that human ESCs for the most part grew in a self-similar way (Figure 5). The fractal dimension was calculated to be 1.85 after 48 hours (Figure 5). The result confirms that the cells appear to be growing according to a fractal power law (27).

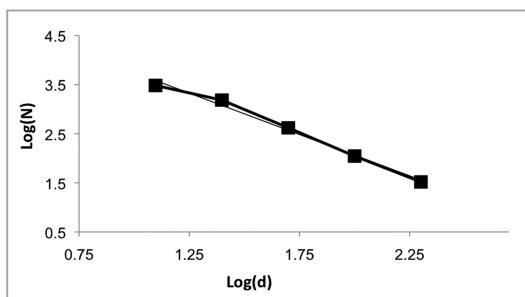


Figure 4. Plot of the logarithm of  $N$  as a function of the logarithm of  $d$ . Log  $N$ -log  $d$  plot gives a straight line that has an equation of  $y = -1.68x + 5.43$ ,  $R^2 = 0.98992$ . The slope of the line gives the fractal dimension, which in this case is 1.68 for mouse ESCs line D3 at 74.5 hours.

Once a fractal relationship was confirmed the next step was to investigate if there was a correlation between cell number and fractal dimension. To do this, cell counts and images of ESCs were taken during growth at specific time points (24, 48, 72, and 96 h) and the number of cell colonies were counted and analyzed by FracLac (24). The results are shown in Figure 5 for human ESCs and Figure 6 for mouse ESCs. Surprisingly, the relationship was linear rather than exponential, and leveled off at a fractal dimension of 1.85 for human and 1.8 for mouse. Thus a similar trend of leveling off the fractal dimension was observed in both human (Figure 5) and mouse (Figure 6) ESC growth, despite the significant differences in their respective culture morphologies.

The plateau in the growth of pluripotent ESCs (non-differentiated) may hint to a possible relationship between a change in fractal dimension and a change in the ESC's pluripotent state. Recalling Figure 5 and 6, we see that the plateau in fractal dimension is approached as the cells enter into the exponential growth phase, at 48 hours. At this point there is still space for cells to grow on the culture plate, which indicates that the plateau of fractal dimension is not related to cell number or culture saturation (confluence). Rather, we suggest that as the cells enter into their exponential growth phase they tend to grow in a more uniform manner, as evidenced by a consistent fractal dimension over time.

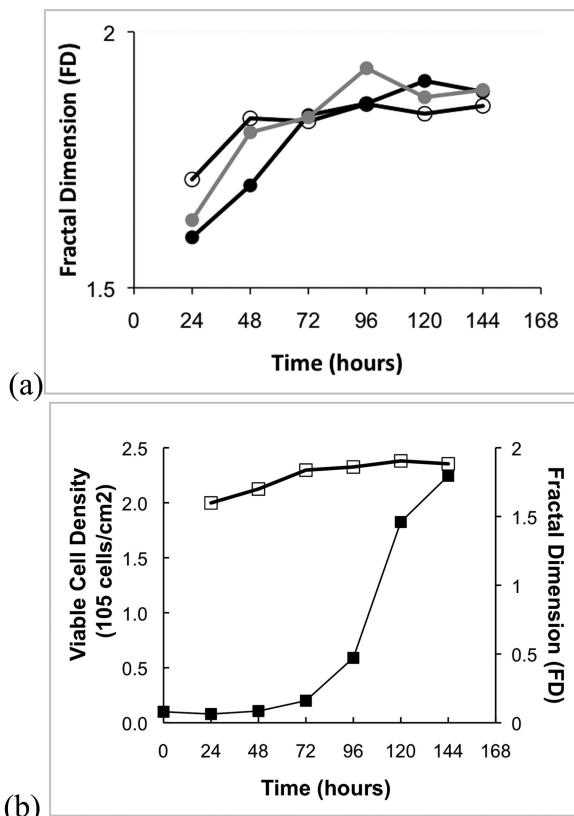
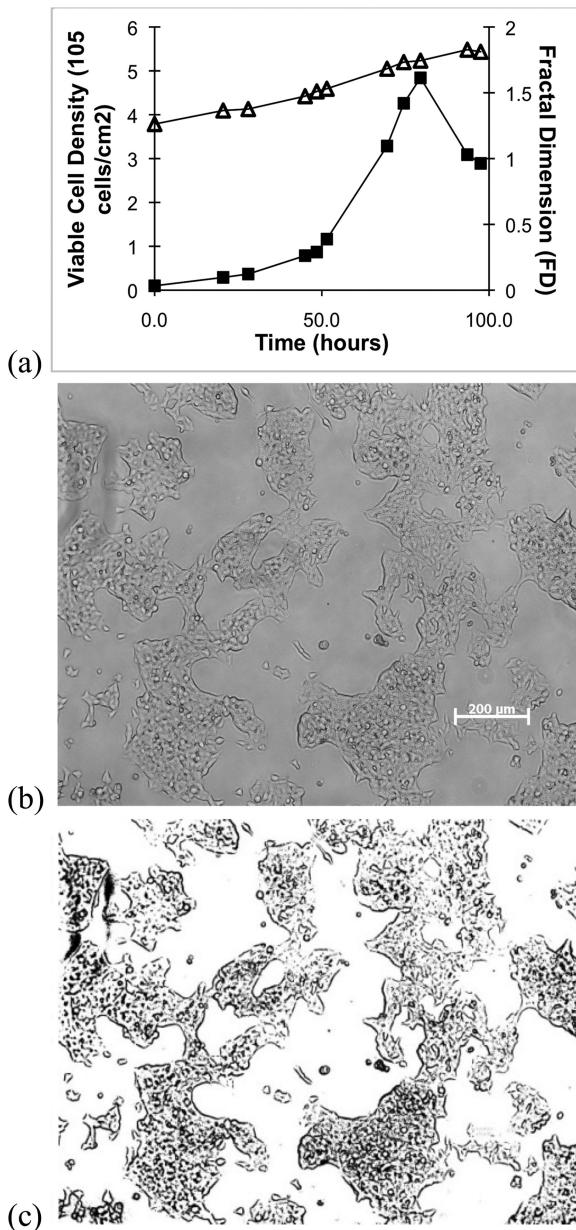
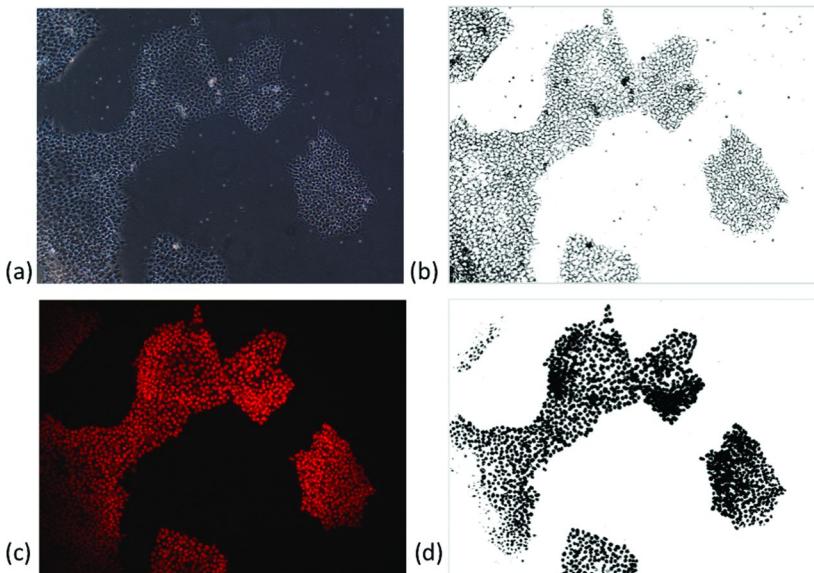


Figure 5. Fractal dimensions and viable cell density as a function of time or different image magnifications for human H1 ESCs. (a) Fractal dimensions for H1 in static culture at increasing magnification. (b) Viable cell density and fractal dimension ( $10^5$  cells/cm $^2$ ), fractal dimension (FD). Legend: 10x magnification (● (black)), 20x magnification (● (gray)), 40x magnification (○), viable cell density (■,  $10^5$  cells/cm $^2$ ), fractal dimension (□). Fractal dimension at increasing magnification tend towards a value of 1.85 illustrating self-similarity of the cultures on different magnifications.

The next step was to investigate whether fractal dimensions were also related to cell marker expression, specifically whether cell differentiation would result in a significant change in fractal dimension. Pluripotent H1 human ESCs stained for Oct4 at 48 hours showed comparable fractal dimensions in the grey scale image (fractal dimension equal to 1.63) and fluorescent Oct4 image (fractal dimension equal to 1.61) (both images were converted to binary images prior to analysis, Figure 7). The identical fractal dimensions for the overall cell culture versus the Oct4 expression sample, suggests a relationship at this stage in ESC growth.



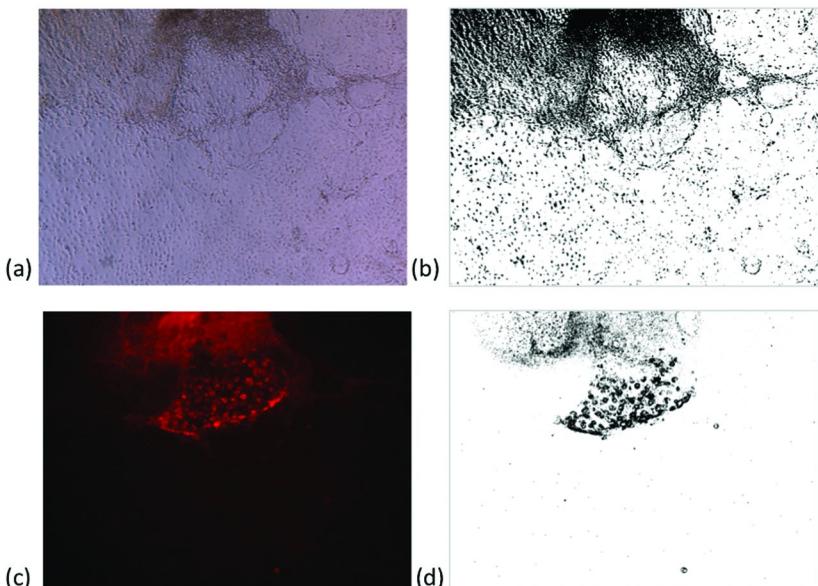
*Figure 6. Automated results using FracLac (24) of cell counts (viable cell density and fractal dimension) and image processing for murine D3 ESCs. (a) Average viable cell density and fractal dimension versus time (hours) in static culture of D3 mouse ESCs. (b) Grey scale image of murine ESCs at time  $t = 74.5$  hours (c) converted to binary black and white image for analysis. Values of fractal dimension increase in a linear manner, while cell numbers increase exponentially – cell culture is filling area more efficiently.*



*Figure 7. Human H1 ESCs at 48 hours (growth curve fractal dimension at 48 hours was 1.65) stained for Oct4. (a) grey scale image and (b) binary black and white image of entire colony under normal microscope fractal dimension 1.63. (c) Fluorescent image of Oct4 positive cells and (d) black and white binary image of fluorescent Oct4 positive cells fractal dimension 1.61.*

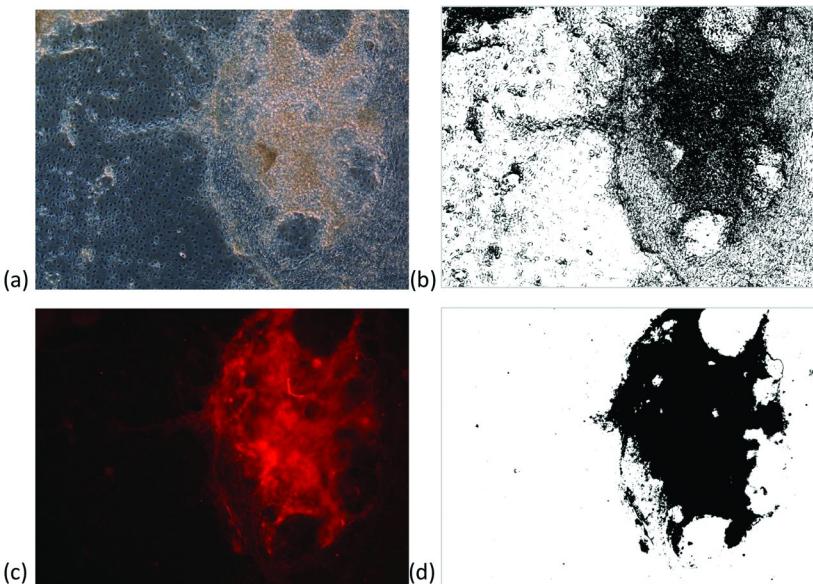
Staining the same H1 ESC line on Day 8 of differentiation (in vitro via embryoid body (EB) formation followed by spontaneous differentiation) for  $\alpha$ -fetoprotein (endodermal marker, Table 1) and comparison of the grey scale versus the fluorescent images revealed a markedly different fractal dimension. Fractal analysis revealed cells that were positive for  $\alpha$ -fetoprotein had a fractal dimension of 1.41, in comparison to the overall culture that had a fractal dimension of 1.78 (Figure 8).

Similarly, staining the H1 ESC line at Day 8 for  $\beta$ -III-tubulin (ectodermal marker, Table 1) and comparing the grey and the fluorescent image reveal a markedly different fractal dimension for the cells that were positive for  $\beta$ -III-tubulin (fractal dimension equal to 1.59) versus the overall culture (fractal dimension equal to 1.85, Figure 9).



*Figure 8. Human H1 ESCs on Day 8 of differentiation (via EB formation) cells stained for  $\alpha$ -fetoprotein to determine if cell morphology may indicate expression levels. (a) grey scale image and (b) binary black and white image of entire colony under normal microscope fractal dimension equal to 1.78 (c) Comparison to brightfield image, fluorescent  $\alpha$ -fetoprotein positive cells and (d) binary black and white image fluorescent  $\alpha$ -fetoprotein positive cells fractal dimension 1.41.*

The results of fluorescent staining reveal that cells expressing specific differentiation markers have a very different fractal dimension value than what was determined for the pluripotent cultures. However, the unstained grey scale images of these same differentiating cells do not reflect the same fractal dimension. It is hypothesized that this discrepancy is the result of only partial differentiation of the culture towards the specific germ lineage that was analyzed. Further analysis of the greyscale images is required, as the intention of the method is to be able to pick out areas of differentiation without having to fix and stain the cells, as the latter overall process is cell destructive. Thus, rather than analyze the entire culture morphology in greyscale images, we are now looking at analyzing images at higher magnifications to evaluate specific cells or small groups of cells within the culture.



*Figure 9. Human H1 cell line on day 8 of differentiation (via EB formation)-stained for B-III-tubulin. (a) Grey scale image and (b) binary black and white image of entire colony under normal microscope fractal dimension 1.85. (c) Comparison to brightfield image, fluorescent beta-III-tubulin positive cells and (d) binary black and white image fluorescent beta-III-tubulin positive cells fractal dimension 1.59.*

#### 4. Conclusions

Non-invasive images taken of growing cultures of mouse and human ESCs reveal there is a significant difference in the morphology of differentiating versus pluripotent ESCs. We have demonstrated that pluripotent human ESCs do grow in a fractal manner, with self-similar images showing similar fractal dimensions at various magnification levels. Improving the gridding and re-running the analysis for ESCs stained for more cell-surface markers of differentiation may provide a better understanding of the relationship between differentiation and fractal dimension. Initial analysis does show a difference in fractal dimension between cells with high Oct4 expression and cells undergoing differentiation, however brightfield photographs do not follow the same trend, thus further investigation by analyzing the multifractal spectrum is underway. Preliminary multi-fractal analysis by using the multi-fractal analysis (27) tool in FracLac (24) reveals more pronounced differences between growth curve time points. Multi-fractal analysis may also provide a better understanding of the relationship between differentiation and fractal dimension, as this process can give far more detail with

regard to factors such as cell clustering. The goal is to advance this tool to a stage where scientists can quickly analyze growing static 2D cultures of ESCs in a high-throughput, nondestructive and quantitative manner.

## Acknowledgments

The authors acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Alberta Innovates Technology Futures (AITF) (graduate scholarship to MMH).

## References

1. Reubinoff, B. E.; Pera, M. F.; Fong, C. Y.; Trounson, A.; Bongso, A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat. Biotechnol.* **2000**, *18*, 399–404.
2. Thomson, J. A.; Itskovitz-Eldor, J.; Shapiro, S. S.; Waknitz, M. A.; Swiergiel, J. J.; Marshall, V. S.; Jones, J. M. Embryonic stem cell lines derived from human blastocysts. *Science* **1998**, *282*, 1145–1147.
3. Behie, L. A.; Kallos, M. S.; Sen, A. Bioprocessing Aspects of Neural Stem Cell Production in Bioreactors. *BioProcess J.* **2004**, *3*, 27–42.
4. (a) Li, Z.; Leung, M.; Hopper, R.; Ellenbogen, R.; Zhang, M. Feeder-free self-renewal of human embryonic stem cells in 3D porous natural polymer scaffolds. *Biomaterials* **2010**, *31*, 404–412. (b) Toh, W. S.; Lee, E. H.; Cao, T. Potential of human embryonic stem cells in cartilage tissue engineering and regenerative medicine. *Stem Cell Rev.* **2011**, *7*, 544–559.
5. Gong, G.; Ferrari, D.; Dealy, C. N.; Kosher, R. A. Direct and progressive differentiation of human embryonic stem cells into the chondrogenic lineage. *J. Cell Physiol.* **2010**, *244*, 664–671.
6. (a) zur Nieden, N. I. Embryonic stem cells for osteo-degenerative diseases. *Methods Mol. Biol.* **2011**, *690*, 1–30. (b) Jethva, R.; Otsuru, S.; Dominici, M.; Horwitz E. M. Cell therapy for disorders of bone. *Cyotherapy* **2009**, *11*, 3–17.
7. Kumar, A. A.; Kumar, S. R.; Narayanan, R.; Arul, K.; Baskaran, M. Autologous bone marrow derived mononuclear cell therapy for spinal cord injury: A phase I/II clinical safety and primary efficacy data. *Exp. Clin. Transplant* **2009**, *7*, 241–248.
8. Ouyang, A.; Yang, S. T. A two-stage fibrous bed bioreactor system for mass production of embryonic stem cells. *Expert Opin. Biol. Ther.* **2008**, *8*, 895–909.
9. Cormier, J. T.; zur Nieden, N. I.; Rancourt, D. E.; Kallos, M. S. Expansion of undifferentiated murine embryonic stem cells as aggregates in suspension culture bioreactors. *Tissue Eng.* **2006**, *12*, 3233–3245.
10. Oh, S. K.; Choo, A. B. Human embryonic stem cell technology: large scale cell amplification and differentiation. *Cytotechnology* **2006**, *50*, 181–90.

11. zur Nieden, N. I.; Cormier, J. T.; Rancourt, D. E.; Kallos, M. S. Embryonic stem cells remain highly pluripotent following long term expansion as aggregates in suspension bioreactors. *J. Biotechnol.* **2007**, *129*, 421–432.
12. Sen, A.; Kallos, M. S.; Behie, L. A. Neural Stem Cells: Bioprocess Engineering. In *Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology*; Wiley-Blackwell: New York, 2009; pp 1–16.
13. Kirouac, D. C.; Zandstra, P. W. The systematic production of cells for cell therapies. *Cell Stem Cell* **2008**, *3*, 369–381.
14. Yu, J.; Thomson, J. A. Pluripotent stem cell lines. *Genes Dev.* **2008**, *22*, 1987–1997.
15. Wilkinson, D. S.; Ogden, S. K.; Stratton, S. A.; Piechan, J. L.; Nguyen, T. T.; Smulian, G. A.; Barton, M. C. A direct intersection between p53 and transforming growth factor beta pathways targets chromatin modification and transcription repression of the alpha-fetoprotein gene. *Mol. Cell. Biol.* **2005**, *25*, 1200–1212.
16. Osafune, K.; Caron, L.; Borowiak, M.; Martinez, R. J.; Fitz-Gerald, C. S.; Sato, Y.; Cowan, C. A.; Chien, K. R.; Melton, D. A. Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat. Biotechnol.* **2008**, *26*, 313–315.
17. Murry, C. E.; Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **2008**, *132*, 661–80.
18. Pan, G.; Thomson, J. A. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res.* **2007**, *17*, 42–9.
19. Draper, J. S.; Pigott, C.; Thomson, J. A.; Andrews, P. W. Surface antigens of human embryonic stem cells: changes upon differentiation in culture. *J. Anat.* **2002**, *200*, 249–58.
20. Trounson, A. The production and directed differentiation of human embryonic stem cells. *Endocr. Rev.* **2006**, *27*, 208–19.
21. Sperger, J. M.; Chen, X.; Draper, J. S.; Antosiewicz, J. E.; Chon, C. H.; Jones, S. B.; Brooks, J. D.; Andrews, P. W.; Brown, P. O.; Thomson, J. A. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13350–13355.
22. Roskams, A. J. I.; Cai, X.; Ronnett, G. V. Expression of neuron-specific beta-III tubulin during olfactory neurogenesis in the embryonic and adult rat. *Neuroscience* **1998**, *83*, 191–200.
23. References for ImageJ: (a) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. (b) Rasband, W. S. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, U.S.A., <http://imagej.nih.gov/ij/>, 1997–2012. (c) Abramoff, M. D.; Magalhaes, P. J.; Ram, S. J. Image Processing with ImageJ. *Biophotonics Int.* **2004**, *11*, 36–42.
24. FracLac can be found at Karperien, A. FracLac for ImageJ, version 2.5. <http://rsb.info.nih.gov/ij/plugins/fraclac/FLHelp/Introduction.htm>. 1999–2012.

25. Campbell, R. D. Describing the shapes of fern leaves: A fractal geometrical approach. *Acta Biotheor.* **1996**, *44*, 119–142.
26. Frisch, K. E.; Duenwald-Kuehl, S. E.; Kobayashi, H.; Chamberlain, C. S.; Lakes, R. S.; Vanderby, R. Quantification of collagen organization using fractal dimensions and Fourier transforms. *Acta Histochem.* **2012**, *114*, 140–4.
27. Lopes, R.; Betrouni, N. Fractal and multifractal analysis: a review. *Med. Image Anal.* **2009**, *13*, 634–649.
28. Smith, T. G.; Lange, G. D.; Marks, W. B. Fractal methods and results in cellular morphology-dimensions, lacunarity and multifractals. *J. Neurosci. Methods* **1996**, *69*, 123–136.
29. Losa, G. A.; Di Ieva, A.; Grizzi, F.; De Vico, G. On the fractal nature of nervous cell system. *Front. Neuroanat.* **2011**, *5*, 45.
30. Losa, G. A.; Castelli, C. Nuclear patterns of human breast cancer cells during apoptosis: characterisation by fractal dimension and co-occurrence matrix statistics. *Cell Tissue Res* **2005**, *322*, 257–267.
31. Waliszewski, P.; Konarski, J. Neuronal differentiation and synapse formation occur in space and time with fractal dimension. *Synapse* **2002**, *43*, 252–258.
32. Baish, J. W.; Jain, R. K. Fractals and cancer. *Cancer Res.* **2000**, *60*, 3683–3688.
33. Bernard, F.; Bossu, J. L.; Gaillard, S. Identification of living oligodendrocyte developmental stages by fractal analysis of cell morphology. *J. Neurosci. Res.* **2001**, *65*, 439–445.
34. Rabbani, M.; Tafazzoli Shadpour, Z.; Goli Malekabadi, Z.; Janmaleki, M. Fractal Dimension Characteristics of Human Mesenchymal Stem Cell Proliferation. In *IFMBE Proceedings*; Dössel, O., Schlegel, M., Ratko, W. C., Eds.; Springer: Berlin, Germany, 2010; pp 219–221.
35. Image courtesy of Ishaan Gulrajani, is available through Wikipedia, is in the public domain free of copyright. [http://en.wikipedia.org/wiki/File:Blue\\_Mandelbrot\\_Zoom.jpg](http://en.wikipedia.org/wiki/File:Blue_Mandelbrot_Zoom.jpg) (accessed July 17, 2012).

# Editors' Biographies

## Anne Kantardjieff

Dr. Anne Kantardjieff is a Scientist in the Upstream Development group at Alexion Pharmaceuticals in Cheshire, CT. She received her Ph.D. in Chemical Engineering at the University of Minnesota, in the laboratory of Professor Wei-Shou Hu. Her graduate research was focused on the use of transcriptome analysis to characterize mammalian cell lines used for biopharmaceutical production and application of these tools to engineer improved production cell lines. She is currently focused on developing early and late stage processes for antibody and recombinant protein therapeutics aimed at treating orphan diseases.

## Prashanth Asuri

Dr. Prashanth Asuri is an Assistant Professor in the Department of Bioengineering at Santa Clara University in Santa Clara, CA. He received his Ph.D. in Chemical and Biological Engineering at Rensselaer Polytechnic Institute, where his research primarily focused on engineering protein function and stability through control of its nanoscale environment. Dr. Asuri spent two years as a postdoctoral researcher at the University of California, Berkeley, where he applied biomolecular and material approaches to obtain a better understanding of mechanisms and conditions underlying the neuronal differentiation of human pluripotent stem cells. His current research is focused on exploring the role of material properties in regulating cell fate and engineering polymer nanocomposites for applications in tissue engineering and regenerative medicine.

## Jonathan L. Coffman

Dr. Jonathan L. Coffman is an Associate Research Fellow at Pfizer in Andover, MA. He received his Ph.D. in Chemical Engineering at the University of Wisconsin Madison. He has served extensively within the Biochemical Technology (BIOT) Division of ACS, including a term as Chair of the Executive Committee and serving as Program Co-chair for the 2003 Annual Meeting. He was also Co-Chair of the Recovery of Biological Products Conference. An expert in protein purification, he holds multiple patents and has authored several high impact publications in the field. He is currently Technology and Innovation coordinator for Bioprocess R&D at Pfizer.

## Karthik Jayapal

Dr. Karthik Jayapal is a Staff Scientist in Cell Culture Process Development at Bayer Healthcare in Berkeley, CA. He received his Ph.D. in Chemical Engineering from the University of Minnesota in the laboratory of Professor Wei-Shou Hu. His research was focused on the use of genomic and proteomic tools for systems analysis of antibiotic-producing *Streptomyces* cells. He is currently focused on upstream process development for Chinese Hamster Ovary based monoclonal antibody projects for hematology, gynecology and oncology indications.

# Subject Index

## A

Affinity chromatography  
ability of arginine, suppress antibody aggregation, 71  
role of arginine and citrate as eluents, 67  
amino acid residues, 79f, 80f  
complex formation, residues involved, 78t  
conclusions, 82  
contact coefficient  
designing new protein A ligands and eluents, 82  
force field  
free energy of binding, cosolute effect, 73, 77  
human IgG, 75f  
identification of residues, protein A and Fc, 83f  
introduction, 68  
model building, IgG-protein A complex reconstruction, 71  
preferential interaction coefficient ( $\Gamma_{23}$ ), 74, 83f  
protein A and antibody Fc domain interaction, 69f  
[Protein A- Fc] complex, 77  
protein A mimetics, 81f  
reduction in yield of antibody/protein, graphical representation, 68f  
simulation details, 72  
structure of arginine, 70f  
Aqueous two-phase system (ATPS), 33  
ATPS. *See* aqueous two-phase system (ATPS)

## B

Biological replacement of petroleum biosynthesis of fuels from CO<sub>2</sub>, engineering, 10  
conclusion, 15  
driving bioproduction of *n*-butanol, 5  
biosynthetic pathways outlines, 7s  
concentration as function of time, 9f  
isobutanol production and cell growth, 6f  
*n*-butanol biosynthesis, Creating artificial driving forces, 8s

generate biofuels from atmospheric CO<sub>2</sub>, using photovoltaic cells, 11  
isobutanol synthesis, electricity-powered CO<sub>2</sub> fixation, 13f  
isobutanol synthesis in ceramic cup, 14f  
genetic manipulation of *E. coli*, biofuels production, 3  
biosynthesis of isobutanol, genes required, 5f  
*C. acetobutylicum*, biosynthesis of *n*-butanol from glucose, 4f  
*n*-butanol production, optimization, 4f  
harvesting of solar energy to production of isobutanol, estimated energy efficiency, 15t  
introduction, 1  
two step conversion of acetyl-CoA into malonyl-CoA, 11s  
various plants to biofuels, carbon flux and energy yield, 2t

## C

Cation exchange (CEX) chromatography, 51  
Chinese hamster ovary (CHO), 19  
CHO. *See* Chinese hamster ovary (CHO)  
CHO cell production culture  
conclusion, 31  
introduction, 20  
materials and methods  
antibody production process, 21  
cell culture, transfection, cloning and stress induction, 21  
cell cycle analysis, 22  
DNA constructs and luciferase assay, 21  
flow cytometry analysis and fluorescence microscopy, 22  
Western blot analysis, 21  
results and discussion  
endogenous UPR levels, monitoring, 22  
FACS analysis of dGFP, 25f  
kinetics of UPR activation, recombinant CHO cell clones, 26f

kinetics of UPR activation and production culture viability, 27/*f*  
mAb production titers, 28/*t*  
metabolite end product production, ER stress, 30  
monitor endogenous UPR activation, fluorescence reporters, 25  
real time fluorescence mean data, 29/*f*  
recombinant CHO production, 28/*f*  
relative luciferase activity, 24/*f*  
SDS-PAGE gel Western blot analysis, 23/*f*  
UPR activation and processing conditions, defining association, 29  
UPR in rCHO cells, impact of osmolarity, 30/*f*  
UPR induction, impact of ME, 29/*f*  
UPR induction levels in cells under ER stress, quantification, 22  
UPR pathway, impact of osmolarity, 31  
UPR-dependent stress inducible reporter construct, 23/*f*  
UPR pathway, schematic, 20/*f*  
Couple isobutanol and 3-methyl-1-butanol biosynthesis, 12

## D

Directed evolution  
natural products, biosynthesis, 147  
acylated CP detection, FT-ICR MS spectra, 160/*f*  
acyltransferase (AT) domains, 152  
biosynthesis of polyketide (PK), elongation process, 151/*f*  
conclusion, 161  
erythromycin and its PKS, 152/*f*  
erythromycin PKS, 153  
introduction, 147  
natural starter units and natural and non-natural extender units, summary, 149/*f*  
non-natural diacids, 157/*f*  
PKSs, process non-natural acyl-CoAs, 159  
polyketide diversification, synthetic biology approach proposed, 155/*f*  
polyketide modification, synthetic biology philosophy, 153  
polyketide natural products structure, extender unit contribution, 150/*f*  
promiscuous acyl-CoA synthetase, engineering, 154

screen acyl-CoA synthetase activity, 156/*f*  
S-N-acetylcysteamine ester and Mod6TE-based assay, 161/*f*  
structurally diverse polyketide natural products, 148/*f*  
*trans*-AT, incorporate non-natural extender units onto PKS, 158

## E

Embryonic stem cell cultures  
heterogeneous growth and differentiation, image analysis method, 165  
automated results using FracLac (24) of cell counts, 175/*f*  
Benoit fractal set, 170/*f*  
box counting method, automation, 171  
cell counts, 169  
conclusions, 178  
fractal dimension, plot, 173/*f*  
fractal dimension determination, 171  
fractal dimensions and viable cell density, 174/*f*  
fractal geometry, 170  
human ESC culture, 169  
human ESC differentiation via embryoid body formation, 169  
human H1 cell line on day 8 of differentiation, 178/*f*  
human H1 ESCs at 48 hours, 176/*f*  
human H1 ESCs on day 8 of differentiation cells, 177/*f*  
illustration of box counting method, 172/*f*  
image generation and fractal dimensions, 170  
introduction, 166  
mESC line D3 cells, standard culture data, 168/*f*  
mouse ESC culture, 168  
pluripotency and early differentiation markers, 167/*f*  
Embryonic stem cells (ESCs), 165  
ESCs. *See* Embryonic stem cells (ESCs)

## F

FACS. *See* fluorescence-activated cell sorting (FACS)

Fluorescence-activated cell sorting (FACS), 24  
Formic acid, 12  
Free radicals KadG, SodC, or NorA promoters, 12

**G**

Green approach to purification, 87  
Akta chromatograms, 108*f*  
AKTA explorer 100 system, 104  
analysis of all samples, pilot scale purification, 97*t*  
analytical methods, 93, 105  
capillary electrophoresis-sodium dodecyl sulfate (CE-SDS) analysis, 108*f*  
cleaning procedure, 91, 92*t*  
conclusions, 103  
experimental protocol, schematic view, 90*f*  
in house mAbs, analysis, 109*t*  
intact IgG protein detected after mock runs, 109*t*  
introduction, 87  
lab scale purification of mAbA, 95*f*  
lab scale purification of mAbC, results, 99*f*  
list of buffers, 104  
mAb product yields and impurity profiles of different MabSelect™ SuRe columns, comparison, 103*t*  
MabSelect™ SuRe cleaning procedure (CP), optimization, 94  
MSSCCP method, schematic outline, 98*f*  
optimized cleaning protocol of MabSelect™ SuRe column, different stages, 100*f*  
plot of protein carryover, 96*f*  
plot of protein carryover as function of CV washes, 106*f*  
plot of total protein carryover, 106*f*  
plot of total protein carryover as function of cleaning cycle, 109*f*  
protein A affinity chromatography, 89  
protein A-based mAb purification process, outline, 88*f*  
protein removal, 107*f*  
purification of mAbC, 101*f*  
purification of mAbY, analysis of all samples, 102*t*  
quantification of protein carryover and risk, 90

quantification of protein carryover without cleaning, 93  
resin selection, 91  
reusing protein A resins, economic impact, 89  
reusing resin, cost analysis, 105*f*  
standard purification procedure (mock elution), 104

## **H**

hG-CSF. *See* human granulocyte-colony stimulating factor (hG-CSF)  
High value biological products recovery aqueous two-phase processes, 33  
different volume ratio ( $V_R$ ), 36  
introduction, 34  
schematic summary, 34*f*  
thermodynamic partition coefficient ( $K_p$ ), 35  
case studies, application of ATPS double-layered Rotavirus-like particles (dLRP), primary recovery, 37  
experience with ATPS plant-based bioprocess development, 39  
recovery of natural colorants from microbial origin, 38  
simplified aqueous two-phase (ATP) extraction, 35*f*  
Human granulocyte-colony stimulating factor (hG-CSF), 33, 36, 39  
5-(Hydroxymethyl)furfural (HMF)  
analytical impact of HMF, quantitation of drug product  
DP protein concentration, quantification, 137  
drug product protein concentration issue, 137*f*  
formulation and lyophilization process, 138*f*  
improve production efficiency, improving analytical methods, 138  
product analysis, additional challenges, 138*f*  
protein lyophilized drug product, RP-HPLC trace, 139*f*  
freeze-dry mediated formation, 129  
conclusion, 140  
drug product impurity issue, 133*f*  
formation of furfural from fructose, representative schematic, 131*s*  
formation of HMF, effect of drying conditions, 135*f*

HMF impurity and authentic HMF,  
determination of  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR  
data, 140  
HPLC trace, superposition, 144/*f*  
identification of impurity, 133  
introduction, 130  
mass spectral information for  
impurities, 134/*f*  
NMR spectral data  
non-enzymatic breakdown of  
carbohydrates, by-products, 130/*f*  
UV-absorbance spectra of placebo,  
132/*f*  
toxicology  
assessment, 136  
HMF quantitation method, 136

## I

IgG and protein A, change in free energy  
of binding, 78/*t*

## L

Lasso peptides  
alanine scan, 118  
amino acid sequence of McjA, 119/*f*  
biosynthesis, 116  
biosynthesis of mature 21 amino acid  
lasso peptide MccJ25 from McjA,  
function of McjB and McjC proteins,  
117/*f*  
cells expressing orthogonally inducible  
MccJ25 production, characterization,  
123/*f*  
conclusions, 124  
essential residues determination, alanine  
scan, 121/*f*  
introduction, 113  
lasso peptide library, 118  
lasso peptide stability, 116  
lasso peptide structure, different  
representations, 114/*f*  
lasso peptides, molecular engineering,  
121  
leader peptide, sequence variation, 120  
MccJ25 production and export,  
orthogonally regulated system, 122/*f*  
microcin J25 and capistrulin, ribbon and  
bow representation, 115/*f*  
natural gene cluster mcjABCD on  
plasmid pTUC100 in *E. coli*, 116/*f*

replica plate assay, 118  
S. Newport spot-on-lawn assay, 120/*f*  
Salmonella spot-on-lawn zone, relative  
zone diameters, 124/*f*  
spot-on-lawn assay, 117  
truncation, 117  
truncation studies, 119  
Lyophilization process analysis, 135

## M

mAb. *See* monoclonal antibodies (mAb)  
MabSelect SuRe<sup>TM</sup> Campaign Changeover  
Procedure (MSSCCP), 87  
Monoclonal antibodies (mAb), 51  
MSSCCP. *See* MabSelect SuRe<sup>TM</sup>  
Campaign Changeover Procedure  
(MSSCCP)

## P

PKSs. *See* polyketide synthases (PKSs)  
Polyketide synthases (PKSs), 148

## Q

Quantification of protein carryover without  
cleaning, 93

## R

*Ralstonia eutropha*, 12

## S

Standard purification procedure (mock  
elution), 104

## T

Tie line length (TLL), 36  
Tie peptide knot, method, 113  
TLL. *See* tie line length (TLL)

## U

Unfolded protein response (UPR), 19  
UPR. *See* unfolded protein response (UPR)  
Utilizing mAb adsorption isotherms, 51  
    adsorption isotherm and static binding  
        capacity measurements, 54  
    adsorption isotherms and static binding  
        capacities, 57  
    cation-exchange chromatography as  
        post-protein A polishing step, 63  
    CEX chromatography, 52  
    conclusion, 64  
    dynamic and static capacities,  
        comparison, 63*f*  
    dynamic binding capacities at lower pH,  
        62  
    dynamic binding capacities with higher  
        pH and higher salt, 62  
    dynamic binding capacity experiment,  
        55  
    introduction, 52  
    linear gradient elution experiments, 55  
    mAb1, adsorption isotherms, 56*f*  
    mAb purification, 52  
    mAb-resin interactions  
        Fractogel® solid support, 54*f*

    materials, 53  
    solid support and chromatography, 53  
static binding capacities, re-evaluation,  
    62  
static capacity trends, 58*f*  
    dynamic binding experiments with  
        mAb1, 61*f*  
    effect of pH from linear gradient  
        elutions on recovery, 61  
    LGE parameters, comparison, 60*t*  
    linear gradient elution data, log-log  
        plot, 59*f*  
    linear gradient elution peaks, 60*f*  
    linear gradient elutions, 59

## V

Virus-like particles (VLP), 37  
VLP. *See* Virus-like particles (VLP)

## W

Western blot analysis, 21