December 2014

Novel Mechanisms In The Sorting Of Proglucagon To The Secretory Granules Of The Regulated Secretory Pathway

Leonard M. Guizzetti
*The University of Western Ontario*

Supervisor
Dr. Savita Dhanvantari
*The University of Western Ontario*

Graduate Program in Medical Biophysics

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Leonard M. Guizzetti 2014

Follow this and additional works at: [https://ir.lib.uwo.ca/etd](https://ir.lib.uwo.ca/etd)

Part of the [Molecular Biology Commons](https://ir.lib.uwo.ca/etd)

Recommended Citation
https://ir.lib.uwo.ca/etd/2536

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca.
NOVEL MECHANISMS IN THE SORTING OF PROGLUCAGON TO THE SECRETORY GRANULES OF THE REGULATED SECRETORY PATHWAY

(Thesis format: Integrated Article)

by

Leonard Michael Guizzetti

Graduate Program in Medical Biophysics, and Collaborative Graduate Program in Molecular Imaging

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Leonard Michael Guizzetti 2014
Abstract

The prohormone proglucagon encodes for multiple peptide hormones, including glucagon, glucagon-like peptide-1 (GLP-1), and GLP-2, produced through tissue-specific processing by prohormone convertase (PC) 1/3 and PC2. In alpha cells, PC2 yields glucagon, the major counter-regulatory hormone to insulin, which together, control glucose homeostasis. In contrast, GLP-1 and GLP2 are mainly produced in intestinal L-cells by PC1/3. GLP-1 stimulates insulin secretion following a meal, and therefore has opposing function to glucagon regulating glucose homeostasis; in contrast, GLP-2 enhances gut nutrient absorption. Efficient sorting of proglucagon to secretory granules is required for nutrient-regulated secretion. The aim of this thesis is to discover the molecular mechanisms by which proglucagon is targeted to secretory granules, which ensures that proglucagon is correctly processed to mature hormones, and is necessary for prompt physiologic response to nutrient status. This thesis identifies several sorting signals contained within the mature hormone domains of proglucagon that encodes its targeting information. Using quantitative immunofluorescence microscopy and co-localization analyses, the molecular mechanism that sorts proglucagon to granules is directed by the domains of glucagon and GLP-1. Despite these two hormones sharing a large degree of structural homology, it is their particular alpha-helix structures that enable the sorting of proglucagon. Further, the evidence suggests that proglucagon is first sorted to granules prior to being processed to active hormones. In alpha cells, carboxypeptidase E is required for the sorting of glucagon to granules. Together, each neuropeptide or peptide hormone carries with it a unique sorting “signature” to efficiently reach its destination, and allows alpha and L-cells to tightly regulate nutrient homeostasis.

Keywords
proglucagon, glucagon, GLP-1, peptide hormone, carboxypeptidase E, prohormone processing, protein trafficking, protein sorting, confocal microscopy, quantitative co-localization
Dedication

This thesis is dedicated to my dad, Mike Guizzetti, who passed away before he could see me fulfill his wish for me to graduate from university, and without whom, this path would not be possible.
Acknowledgements

I wish to thank Becky McGirr for teaching me many of the cell and molecular biology techniques, training me in cell culture. I would also like to thank Ian Cameron for his help and enthusiasm during his fourth-year project. I would also like to thank Tumelo Moleko, Vanessa Rota and the former members of the lab for helpful conversations and making a fun lab environment. I wish to thank the Lawson "coffee crew" for entertaining discussions and creating friendships over these past 5 years.

Thank you to my supervisor Dr. Savita Dhanvantari for her generous level of support during my Ph.D. Savita has provided me with exceptional training opportunities, and has always been very enthusiastic and encouraging in my scientific and professional development. Thank you to my advisory committee members, Drs. Jimmy Dikeakos and Dwayne Jackson for the helpful discussions of my work, their guidance and support.

To my family, thank you for your love, understanding and support to allow me to complete this degree. I may be far away, but I have never forgotten you all.

To Heather, thank you for marrying me during the last stages of my thesis. I am especially grateful for your support and love during the past five years, and reminding me that there is more to life than work. There are also coffee, Sally and our adventures together.

Everyone that I have encountered along this path, either in small or large ways, has contributed to me achieving my goal.
Table of Contents

Abstract........................................................................................................................................... ii
Dedication.......................................................................................................................................... iii
Acknowledgements............................................................................................................................ iv
Table of Contents............................................................................................................................. v
List of Tables....................................................................................................................................... viii
List of Figures....................................................................................................................................... ix
List of Appendices.............................................................................................................................. x
List of Abbreviations.......................................................................................................................... xi
Chapter 1........................................................................................................................................... 1
  1.1 Introduction................................................................................................................................. 1
  1.2 Structure, Function and Distribution of Proglucagon and the Proglucagon-Derived Peptides............................................................................................................................... 1
    1.2.1 Glucagon............................................................................................................................... 3
    1.2.2 GLP-1.................................................................................................................................. 4
    1.2.3 GLP-2.................................................................................................................................. 4
    1.2.4 Oxyntomodulin and Glicentin............................................................................................ 6
  1.3 Mechanisms of Regulated Exocytosis......................................................................................... 7
    1.3.1 Secretory Pathway................................................................................................................ 7
  1.4 Mechanisms of regulated secretion............................................................................................ 14
    1.4.1 Mechanisms of glucagon secretion from α cells................................................................. 14
    1.4.2 Mechanisms of GLP-1 and GLP-2 secretion from L-cells.................................................. 18
  1.5 Processing of Proglucagon.......................................................................................................... 21
    1.5.1 Glucagon production in pancreatic α cells.......................................................................... 21
    1.5.2 GLP-1 and GLP-2 production in intestinal L-cells............................................................... 23
    1.5.3 Proglucagon processing in the central nervous system...................................................... 24
    1.5.4 The ultra-structural localization of the proglucagon-derived peptides within granules................................................................................................................................. 25
  1.6 Mechanisms of Sorting............................................................................................................... 26
    1.6.1 Aggregation of hormones..................................................................................................... 27
    1.6.2 Dibasic sorting signals.......................................................................................................... 28
    1.6.3 Dicarboxylic sorting signals................................................................................................. 28
    1.6.4 Amphiphilic α-helices.......................................................................................................... 28
    1.6.5 Disulphide loops.................................................................................................................. 29
    1.6.6 Glycoproteins with sorting signals...................................................................................... 29
    1.6.7 Compounding proteins may increase returns...................................................................... 29
    1.6.8 Sorting Receptors................................................................................................................ 30
  1.7 Rationale and Specific Aims....................................................................................................... 33
  1.8 References.................................................................................................................................... 35
Chapter 2........................................................................................................................................... 53
  2 The sorting of proglucagon to secretory granules is mediated by carboxypeptidase E and intrinsic sorting signals.......................................................................................................... 53

v
2.1 Introduction ........................................................................................................ 53
2.2 Methods .............................................................................................................. 57
  2.2.1 Cell Culture, plasmids and transfection ...................................................... 57
  2.2.2 Depletion of Carboxypeptidase E by siRNA Knockdown ......................... 58
  2.2.3 Western Blot ............................................................................................... 58
  2.2.4 Secretion Experiments ............................................................................. 59
  2.2.5 Immunofluorescence and image acquisition ........................................... 59
  2.2.6 Image analysis using Pearson's Correlation Coefficient ........................ 60
  2.2.7 De novo Protein Modeling ......................................................................... 61
2.3 Results ................................................................................................................. 62
  2.3.1 Pearson's correlation coefficient accurately reflects degree of co-localization ........................................................................................................ 62
  2.3.2 Proglucagon is sorted to granules in the presence of Carboxypeptidase E  65
  2.3.3 Proglucagon sorting depends upon properly sorted CPE ......................... 67
  2.3.4 The effect of PC1/3 and PC2 on proglucagon sorting ............................... 69
  2.3.5 CPE affects proglucagon sorting in α cells, but not L cells ........................ 71
  2.3.6 Proglucagon contains a sorting signal within the glucagon sequence ...... 74
  2.3.7 Initial processing of proglucagon at K70R71 enhances sorting to granules 74
  2.3.8 Modelling the 3D structure of proglucagon ............................................. 76
2.4 Discussion ........................................................................................................... 77
2.5 References .......................................................................................................... 83

Chapter 3 .................................................................................................................. 87
3 Two Dipolar α-Helices Within Hormone-Encoding Regions of Proglucagon are Sorting Signals to the Regulated Secretory Pathway .............................................. 87
  3.1 Introduction ...................................................................................................... 87
3.2 Methods .............................................................................................................. 90
  3.2.1 Cell Culture and transient transfections ................................................... 94
  3.2.2 Secretion experiments .............................................................................. 94
  3.2.3 Immunoprecipitation, Western blot and secretion index ......................... 94
  3.2.4 Co-immunoprecipitation of Chromogranin A complexes ...................... 95
  3.2.5 Immunofluorescence ............................................................................... 95
  3.2.6 Image Acquisition and Analysis ................................................................. 95
  3.2.7 Secondary Structure Predictions and Biophysical Property Calculations .. 96
  3.2.8 Statistical Analyses .................................................................................. 96
3.3 Results ................................................................................................................. 97
  3.3.1 Rationale for using PC12 as a model of hormone trafficking and secretion 97
  3.3.2 Glucagon contains an α-helix sorting signal ............................................. 98
  3.3.3 GLP-1, but not GLP-2, efficiently targets Fc to secretory granules .......... 101
  3.3.4 A dipolar α-helix GLP-2 mutant sorts to secretory granules .................. 103
  3.3.5 Biophysical Properties of Glucagon, GLP-1 and GLP-2 α-helices determine sorting efficiency .................................................................................. 106
  3.3.6 MPGF, but not glicentin, sorts to secretory granules ............................... 107
  3.3.7 Oxyntomodulin sorts to secretory granules ............................................ 107
  3.3.8 The Dipolar α-helix can bind to chromogranin A in a pH- and calcium- dependent manner ...................................................................................... 108
List of Tables

3.1 Primer pairs used for cloning and mutagenesis..........................................................91
3.2 Biophysical Properties of Major Proglucagon Derived Peptides.................................105-6

4.1 Mathematical parameters used for quantitative co-localization analyses..............124
4.2 Performance of selected co-localization parameters under differing conditions of true co-localization and segregation..............................................................................135

5.1 Hypothetical novel classes of sorting signals..............................................................166
List of Figures

1.1 Post-translational processing of proglucagon to its derived peptides...............................2
1.2 Overview of key metabolic actions of insulin, glucagon and GLP-1 on glucose homeostasis........................................................................................................................................6
1.3 Mechanisms of glucagon secretion from alpha cells........................................................................15
1.4 Mechanisms of GLP-1 secretion from L cells..................................................................................20

2.1 Post-translational processing of proglucagon to its derived peptides...........................................55
2.2 Validation of the Pearson's correlation coefficient............................................................................64
2.3 CPE plays a role in the sorting of proglucagon..............................................................................66
2.5 Proglucagon is sorted to secretory granules in N2a cells expressing CPE........................................67
2.6 The E300Q mutant CPE is not sorted efficiently to secretory granules........................................68
2.7 The presence of PC1/3 or PC2 does not enhance sorting of proglucagon........................................70
2.8 CPE depletion affects regulated secretion of glucagon from alphaTC1–6 cells..........................72
2.9 CPE depletion has no effect on regulated secretion of GLP-1 from GLUTag cells..................73
2.10 Identification of putative sorting signals in proglucagon..............................................................75
2.11 Structural model prediction of proglucagon..................................................................................77

3.1 Post-translational processing of proglucagon to its derived peptides...........................................88
3.2 Schematic depiction and expression of fusion proteins.................................................................93
3.3 Glucagon contains a necessary α-helical sorting signal.................................................................99
3.4 GLP-1 is sorted to secretory granules by a sorting signal within GLP-1(7-37)..........................102
3.5 MPGF is efficiently sorted into secretory granules by the sorting signal contained within GLP-1.........................................................................................................................104
3.6 Identification of CgA as a sorting receptor....................................................................................109
3.7 Schematic representation of proglucagon sorting and processing in alpha and L cells................112

4.1 Verification of chromatic alignment...............................................................................................132
4.2 Illustration of cytofluorogram showing full, partial or no co-localization.....................................134
4.3 Effect of object overlap on the measured correlation coefficients without background noise..........................................................................................................................137
4.4 The effects of additive noise and apparent SNR on detecting degrees of object co-localization........................................................................................................................................141
4.5 Regions of interest improve parameter performance by reducing the weighting on background and noise pixels..............................................................................................................143
4.6 Performance of multiple co-localization parameters on open immunofluorescence image sets........................................................................................................................................145
4.7 Object-based co-localization of 3D rendered sub-cellular volumes of the secretory pathway and granules....................................................................................................................157
List of Appendices

1 Code Contributions to FIJI/ImageJ project..............................................................173
  1.1 Bug fixes for the FIJI Co-localization analysis plug-in.................................173
  1.2 Addition of the Spearman's Rank Correlation Coefficient Algorithm to FIJI.....174
List of Abbreviations

ACTH, adrenocorticotropic hormone
ANP, atrial natriuretic peptide
BiFC, bi-molecular fluorescence complementation
BSA, bovine serum albumin
BDNF, bone-derived neurotrophic factor
CCD, charge-coupled device
CCF, cross-correlation function
CgA, chromogranin A
CgB, chromogranin B
COP1, coat protein complex I
COPII, coat protein complex II
CMOS, complementary metal-oxide-semiconductor
CPE, carboxypeptidase E
DAG, diacylglycerol
DAPI, 2-(4-amidinophenyl)-1H -indole-6-carboxamidine
dbAMP, dibutyryl-cyclic AMP
DMEM, Dulbecco's modified Eagle medium
EGFP, enhanced green fluorescent protein
ER, endoplasmic reticulum
ERGIC, endoplasmic reticulum-Golgi intermediate complex
ERK, extracellular-signal-regulated kinase
Fc, crystalizable fragment from mouse IgG-2b
FFAR, free fatty acid receptor
FITC, fluorescein isothiocyanate
FRET, Forster resonance energy transfer
FSH, follicle stimulating hormone
GABA, gamma-aminobutyric acid
GFP, green fluorescent protein
GH, growth hormone
GIP, glucose-dependent insulinotropic factor (gastric inhibitory polypeptide)
GLP-1, glucagon-like peptide 1
GLP-2, glucagon-like peptide 2
GLUT, glucose transporter
GRPP, glicentin-related polypeptide
GSD, ground-state depletion
HBSS, Hank’s buffered salt solution
HG, high glucose
HS, horse serum
IBMX, 3-isobutyl-1-methylxanthine
ICQ, intensity correlation quotient
IgG, immunoglobulin G
IP-1, intervening peptide 1
IP-2, intervening peptide 2
ISG, immature secretory granule
K-ATP, ATP-sensitive K+-channel
LH, luteinizing hormone
LG, low glucose
MBCD, methyl-beta cyclodextrin
MOC, Mander's overlap coefficient
MPGF, major proglucagon-derived fragment
NA, numerical aperture
NPY, neuropeptide Y
OXM, oxyntomodulin
PAM, peptidylglycine alpha-amidating monooxygenase
PC, phosphatidylcholine
PC1/3, prohormone convertase 1/3
PC2, prohormone convertase 2
PCC, Pearson's correlation coefficient
PE, phosphatidylethanolamine
PCR, polymerase chain reaction
PGDP, proglucagon-derived peptide
PI, phosphatidylinositol
PI4P, phosphatidylinositol-4-phosphate
POMC, pro-opiomelanocortin
PSF, point-spread function
RFP, red fluorescent protein
SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis
SG, secretory granule
SgII, secretogranin II
SgIII, secretogranin II
SGLT, sodium-glucose co-transporter
siRNA, small interfering RNA
SLMV, synaptic-like microvesicle
SNAP, soluble NSF attachment protein
SNARE, soluble NSF attachment protein (SNAP) receptor
SNR, signal-to-noise ratio
SRC, Spearman's rank-correlation coefficient
SST, somatostatin
STED, stimulated emissison depletion
TIFF, tagged-image file format
T1DM, type 1 diabetes mellitus
T2DM, type 2 diabetes mellitus
TGN, trans-Golgi network
TRH, thyrotropin releasing hormone
TRITC, tetramethylrhodamine isothiocyanate
VAMP, vesicle-associated membrane protein (synaptobrevin)
WT, wild-type
Chapter 1

1.1 Introduction
Proper regulation of energy metabolism and glucose homeostasis are crucial for the survival of any organism. Collectively, glucose homeostasis is exquisitely regulated by the counter-regulatory actions of insulin and glucagon, the major hormones secreted from pancreatic islet beta and alpha cells, respectively. Blood glucose is the primary trigger for secretion of glucagon and insulin in the maintenance of euglycemia and the defence against deleterious excursions of glycemia. Insulin and glucagon target the liver, skeletal muscle and peripheral tissues to regulate glucose homeostasis. In order for these islet cells to be responsive to blood glucose, islet cells must first package these hormones in specialized organelles, called secretory granules. Granule storage is a prerequisite for the conversion from inactive, precursor prohormones to active hormones, and for maintaining concentrated stores of hormone such that physiologic triggers promptly counteract changes in nutrient status. The mechanism of packaging insulin in granules is understood. However, the underlying mechanism of glucagon sorting is not, and it is the aim of this thesis is to investigate this mechanism.

1.2 Structure, Function and Distribution of Proglucagon and the Proglucagon-Derived Peptides
Proglucagon is encoded by the glucagon gene (GCG), and expressed in both pancreatic alpha cells, enteroendocrine intestinal L-cells, and select neurons in in the hypothalamus and brainstem. Post-translational tissue-specific processing of proglucagon is achieved by cleavage at dibasic amino acid sequences separating hormone domains, and a single mono-basic site to produce active GLP-1 (7-37) (in L-cells only), by the differential expression and action of prohormone convertases (PCs) (Figure 1). In pancreatic alpha cells, where PC2 is expressed, proglucagon is processed to the major product glucagon (Figure 1). In contrast, PC1/3 expression in enteroendocrine L-cells and neurons produces the major products GLP-1 and GLP-2, and minor products oxyntomodulin and glicentin (Figure 1). Further details of the processing will be discussed in the section “Processing of Proglucagon.” Once secreted into circulation, the proglucagon-derived peptides
(PGDPs) transduce their actions through a family of related receptors.

Figure 1. Post-translational processing of proglucagon to its derived peptides. A schematic representation of proglucagon in which each domain has been colour-coded, and the amino acid positions of key processing events have been identified. The amino acid numbering is relative to the first N-terminal amino acid of proglucagon (lacking the signal peptide). The major hormone products produced from tissue-specific processing by their respective prohormone convertase (PC) are shown at bottom. In pancreatic islet α cells, the major products are glucagon and major proglucagon fragment (MPGF). In enteroendocrine L cells and in the brain, the major products are GLP-1 (1-37), GLP-1 (7-37), GLP-2, oxyntomodulin and glicentin.

Glucagon, GLP-1 and GLP-2 all mediate their effects through cognate G-protein-coupled receptors (1), known as the vasoactive intestinal peptide/glucagon/secretin receptor family. Mammalians have well-conserved proglucagon and cognate GPCR sequences, suggesting their physiological importance in early evolutionary nutrient metabolism and regulation (2, 3). This receptor family activates adenylate cyclase/cyclic adenosine monophosphate/protein kinase A, and may also couple to phospholipase C and store-operated Ca\(^{2+}\) signalling (4–7). The proglucagon-derived peptides have a degree of cross-reactivity for other PGDP receptors. Most importantly, glucagon and oxyntomodulin are very weak agonists of GLP-1R by about 100-fold, while GLP-1 is a weak agonist of GCG-R, approximately 10-fold weaker than glucagon (8–12). Though their sequences share homology, the PGDPs exert distinct biological effects in the regulation and counter-
regulation glucose homeostasis.

1.2.1 Glucagon

Glucagon is the primary hormone product of α cells, the major counter-regulatory hormone to insulin, maintaining euglycemia and glucose homeostasis. The prevailing belief is that glucose is secreted in response to a hypoglycemia (i.e., a fasted state or in exercise), by primarily acting upon liver hepatocytes to increase glucose output through increased glycogenolysis and gluconeogenesis (13). Normally, fine control over glucose control is achieved by the coordinated and pulsatile secretion of glucagon or insulin (14). The importance of glucagon in this balance is seen when exogenous glucagon administration potently stimulates large and transient increases in blood glucose in animals and humans (15–17). Mouse models in which proglucagon is globally deleted (GCG$^{gfp/gfp}$) (18), or glucagon signalling is abrogated either globally (GCG-R null) (19, 20) or only within liver hepatocytes (21) causes the development of a slight fasting hypoglycemia, improved glucose tolerance and normal insulin tolerance. Post-natal GCG$^{gfp/gfp}$ mice are severely hypoglycemic, and the normalization of glucose homeostasis developed in adulthood is an adaptation due to dampened insulin secretion and down-regulated hepatocyte signalling mechanisms to compensate for the loss of glucagon (22). Taken together, glucagon is an important regulator of hepatic glucose output in conditions of hypoglycemia.

In the relative or absolute absence of insulin, glucagon secretion becomes dysregulated. While a healthy individual resists hyperglycemic episodes, increased glucagon output can be considered a general defence to prolonged and severe hypoglycemic episodes. In both T1DM (23) and severe T2DM (e.g., approaching insulin-deficiency) (24), glucagon response to hypoglycemia is lost, resulting in more severe and recurrent hypoglycemia. T2DM is further characterized by hyperglucagonemia and increased basal hepatic glucose output. Fasting hyperglycemia in T2DM is further exacerbated by the inappropriate and paradoxical hypersecretion of glucagon in a hyperglycemic state (25, 26), leading to prolonged hyperglycemia. Dysregulation of glucagon secretion is central to the pathogenesis of T2DM hyperglycemia (27) although the exact mechanism underlying this dysregulation is debated, involving a mixture of abnormal intrinsic
signalling of the α cell and abnormal autonomic or paracrine control. Since regulated exocytosis of glucagon involves membrane depolarization and elevated cytosolic calcium, it is not difficult to imagine that a change in electrical excitability of the α cell may contribute to this pathogenesis (24). As T2DM progresses, impairment of K-ATP channel regulation results in the relative decrease in glucagon response. This mechanism will be discussed in further detail in the section regarding mechanisms of glucagon secretion.

### 1.2.2 GLP-1

Glucose homeostasis is also regulated by GLP-1, which exerts insulinotropic effects in a glucose-dependent manner, and opposing those of glucagon. GLP-1 and GLP-2 are co-secreted from the enteroendocrine L-cells, located mainly along the distal ileum and colon (28, 29), in a bi-phasic pattern of secretion and in response to nutrient ingestion (30–32). GLP-1 makes an essential contribution to the control of glucose homeostasis by direct stimulation of insulin secretion, as shown using both GLP-1R antagonists in humans (33), and using GLP-1R deficient mice (34). In contrast, GLP-1 indirectly inhibits glucagon secretion (35), in healthy humans at sub-insulinotropic doses (36) and is maintained in T2DM (37). However, In T2DM, both the early and late phase of GLP-1 secretion is reduced (31). Evidence from mice (38) and humans (39) suggests that glucagon suppression acts through paracrine modulation of α cells via local secretion of insulin. GLP-1 also decreases gastric motility in healthy lean and obese humans with T2DM (40–42), and decreases gastric emptying and appetite (43, 44). As T2DM is accompanied by hyperglucagonemia, this may be explained by the loss of GLP-1 secretion. It would also suggest that a healthy individual maintains tight coordination between the glucoregulatory GLP-1 and counter-regulatory glucagon.

### 1.2.3 GLP-2

GLP-2 is recognized as an intestinotrophic factor influencing multiple functions of the GI tract. Unlike GLP-1, GLP-2 does not stimulate insulin secretion (45, 46), but it does stimulate glucagon secretion (47). GLP-2 increases nutrient absorption from the gut, maintains the integrity of the intestinal epithelium (29, 48), and increase mesenteric blood flow in humans (49). This is especially important in intestinal malabsorptive disorders,
such as short bowel syndrome or distal bowel resections, whereby GLP-2 administration promotes intestinal growth and adaptation in humans (50, 51). Therefore GLP-2 indirectly contributes to post-prandial nutrient absorption and nutrient homeostasis. In T2DM, the stimulation of glucagon secretion (45) further contributes to prolonged hyperglycemia. With respect to the islet, GLP-2 protects against metabolic stress (e.g., from high-fat diet), and is essential to the adaptation of α and beta cells to obesity induced in mouse models of obesity and GLP-2R knockout (52). In this study, the obese and GLP-2R deficient mice had a disproportionate loss of beta cells and increased number of α cells, leading to ambient hyperglycemia and elevated glucagon levels. Pathological expansion of α cells is known to involve the up-regulation of pro-inflammatory cytokine IL-6 (53). IL-6 is known to mobilize serum FFA levels in mice (54) which causes oxidative stress and is toxic to beta cells (55). Such a mechanism may be operating in the GLP-2R knockout mice. Additionally, elevated IL-6 shifts the α cell phenotype toward expression of PC1/3, therefore producing intra-islet GLP-1 (56), and later co-secretion of glucagon and GLP-1. In mice, this links the action of GLP-2 to controlling systemic inflammation and cross-talk between islet cell types with the gut, in order to preserve normal glucoregulatory responses.

The diverse actions of glucagon, GLP-1 and GLP-2 operate in balance to maintain nutrient homeostasis. In a fasted state (Figure 2), glucagon is the safe-guard to hypoglycemia by stimulating increased glucose output from the liver to maintain euglycemia. Simultaneously, fasted blood glucose levels are below the stimulatory threshold needed for insulin secretion, and the lack of ingested nutrients will not stimulate secretion of GLP-1 and GLP-2. Beginning shortly after meal ingestion (~15 min), GLP-1 and GLP-2 are secreted in the first phase (Figure 2), initiating the ileal brake, modulating gut motility and increasing intestinal nutrient absorption. Therefore, GLP-1 and GLP-2 are together control the “optimal” rate of nutrient absorption. Absorbed glucose directly stimulates insulin secretion, further increasing insulin secretion by the action of GLP-1. Together, insulin, GLP-1 and GLP-2 act to suppress glucagon secretion and therefore prevent its anorectic effects in the absorptive, fed state. Perturbations of intestinal absorption or secretion, or insult to islets (e.g., by oxidative-
stress) disrupt their balance.

Figure 2. Overview of key metabolic actions of insulin, glucagon and GLP-1 on glucose homeostasis. Hypoglycemia triggers α cells to secrete glucagon, which potentiates hepatic gluconeogenesis and glycogenolysis. After a meal, ingested nutrients potentiate GLP-1 and GLP-2 secretion from L-cells. GLP-1 regulates glucose homeostasis by increasing insulin secretion in a glucose-dependent manner.

1.2.4 Oxyntomodulin and Glicentin
Oxyntomodulin (OXM) contains the glucagon sequence that is C-terminally extended by IP-1. Oxyntomodulin modulates appetite and body weight, as well as gastric acid secretion. Oxyntomodulin was discovered in 1981 (57, 58) for its hyperglycemic and glycogenolytic hepatic effects, mediated through the glucagon receptor (59, 60). In randomized clinical trials in humans, it increased energy expenditure and satiety, while decreasing food intake (61, 62), consistent with effects seen from co-administration of sub-physiologic doses of both glucagon and GLP-1 in humans (36). OXM has weak affinity to both Gcg-R and GLP-1R, when compared to glucagon and GLP-1 (9, 11). For instance, it is (~20-fold) more potent than glucagon to stimulate gastric acid secretion (61). Despite signalling through both the glucagon and GLP-1 receptors, it is believed that OXM is a GLP-1R-biased ligand, showing a preference toward phosphorylation of
extracellular-signal-related kinases (ERK)1/2 and less stimulation of in vitro cAMP assays through differential recruitment of G-α subunits when compared to GLP-1 (63). This bias is consistent with the recent findings that OXM elevates the intrinsic heart rate in mice through the GCG-R and not GLP-1R (10, 64). These properties make it attractive as a new class of dual-agonist for anti-obesity and anti-diabetic therapy as the potential hypoglycemic effects are tempered by the GLP-1R activation. Similarly, glicentin is OXM that is N-terminally extended by GRPP. It has even less well understood pharmacological effects despite also being co-secreted from intestinal L-cells. In humans, glicentin opposes those effects of OXM on gastric acid secretion (65); it has similar inhibitory effects on gut motility (66), and delays post-prandial gastric emptying (67). At present, it is believed that the molecular mechanisms of OXM and glicentin signalling are likely to be similar. The remaining PGDPs (IP-1, IP-2 and GRPP) do not appear to exert any interesting effects.

The prohormone proglucagon encodes several structurally related peptides, that have wide-ranging and contrasting physiologic effects. The α and L-cells that express proglucagon are able to liberate only the relevant constituent hormone(s) needed to respond to nutrient status. Tissue-specific processing is known to be caused by the specific cleavage by expression of either PC1/3 or PC2. In order to prepare adequate concentrations of active hormones within these cells, proglucagon is sorted into secretory granules, where conversion to glucagon or the glucagon-like peptides occurs along the late secretory pathway. Nutrient status potently stimulates secretion of the proglucagon-derived peptides so they may exert tight control over glucose metabolism and nutrient homeostasis. However, the molecular mechanism by which proglucagon is sorted from the trans-Golgi network (TGN) into granules is not understood. This thesis aims to uncover this sorting mechanism for proglucagon within α- and L-cells.

1.3 Mechanisms of Regulated Exocytosis

1.3.1 Secretory Pathway

Cells use multiple pathways to ensure exocytosis of various proteins. Beginning with the early secretory proteins, the nascent peptide chain is synthesized into the ER (68), and
packaged for delivery to the Golgi from so-called “exit sites,” loaded into vesicles defined by coat-protein (COP)-II (69). These COP-II vesicles undergo homotypic fusion in mammals, forming the ER-Golgi Intermediate Complex (ERGIC) that subsequently becomes/transits en bloc to the cis-Golgi (70). The Golgi is typically regarded as having a tripartite compartmental organization – consisting of cis, medial and trans cisternae – along with the Golgi-associated TGN. As cargo transits the stack, it may acquire post-translational modifications (e.g., sulfation, maturation of N-glycosides, addition of O-glycosides). The mammalian Golgi are morphologically similar, in which several Golgi stacks are joined laterally into an elongated Golgi ribbon (71), and sometimes the neighbouring cisternae within stacks can be joined by a tubular-mesh connections rather than existing as disconnected structures (72, 73). Neuroendocrine cells have been a recent focus to detail the 3D Golgi morphology (74). The beta cell in particular exhibits this more complex mesh-like topology (75). The exact details of how secretory proteins transit the Golgi stack are still a matter of debate.

There are two dominant models for Golgi traffic: vesicular trafficking between stable compartments, and cisternal progression/maturation. The stable compartments hypothesis describes the Golgi cisternae as biochemically distinct compartments, in which proteins progress sequentially through each cisterna, exposed to the unique enzymes and conditions of each cisterna, by means of anterograde carrier vesicles (76–78). These carrier vesicles are thought to be defined by coat protein (COP)-I (79, 80). The finding that purified COP-I vesicles contained Golgi glycosylation enzymes (81, 82), weaken this model because it implicates COP-I vesicles in both retrograde and anterograde cargo flow (83, 84). The limitations of this model compelled people to favour cisternal maturation as an alternate framework. The cisternal progression/maturation model offers a more fluid description of Golgi dynamics. The cis-cisternae is formed directly from the ERGIC complex, progressively maturing through the Golgi stack to the trans-most cisterna, and is ultimately “shed” at the TGN as secretory cargoes are carried along post-Golgi routes. As a cisterna matures, the resident enzymes are recycled to earlier positions to maintain their polarized distribution (83–86), though some groups have found a lack of some Golgi resident recycling in COP-I vesicles (86, 87). A fundamental question is whether COP-I
vesicles carry (anterograde) secretory cargo. A recent super-resolution fluorescence microscopy study showed that fused cells can exchange both resident proteins and secretory cargo over long distances in COP-I vesicles, suggesting bidirectional transportation (72, 85, 88). The cisternal maturation model is extended by considering the Ras-related protein subfamily of G-proteins, Rab proteins. The Rab network is a set of ~70 proteins spread over the various membrane-bound compartments of the cell, including the Golgi cisternae, which may provide a mechanism for both maintaining cisternal identity and signals maturation (81, 82). By analogy, Rab proteins mark distinct membranes to help define organelles and vesicles. At present, the cisternal maturation model is most consistent with the literature and is considered the most appropriate model. Once cargo arrives at the TGN for sorting, neurons and (neuro)endocrine cells possess a regulated secretory pathway for transport of specialized neuropeptide and hormone cargo.

Exit from the Golgi along the regulated secretory pathway requires formation of secretory granules that undergo several steps in order to form, mature and become competent for exocytosis. Hormones must first be sequestered into an area of the TGN that begins to protrude (extrude) from the membrane, called the granule precursor or nascent granule. Scission of the two membranes forms a nascent secretory granule, and often occurs simultaneously with cargo loading (89, 90). The granule undergoes several maturation steps, including 1) the removal of extraneous cargo proteins and membrane lipids, 2) trafficking toward the plasma membrane, 3) acidification of the lumen, 4) docking at the plasma membrane, before finally being rendered mature granules. These steps are partly dependent on each other and partly concurrent with each other; the maturation steps happen with some overlap between consecutive steps as will be discussed later with final maturation steps. Consider that a beta cell can only immediately release ~20-30 granules (i.e., readily releasable pool) (91), and that secretory granules are reduced in number (92) and secretion competency (93) following the development of T2DM. Each of these (more proximal) maturation steps are therefore very sensitive to interference, to the point that even slight perturbations could severely negatively impact the ability of the (neuro)endocrine tissue to adequately concentrate hormones into and maintain a pool of granules.
Newly forming granules are limited not only by the cargo to be packaged, but also by the specific lipid membrane components. The membrane of the TGN is enriched in phosphatidylinositol-4-phosphate (PI4P), derived from phosphatidylinositol. PI4P is a signalling lipid on the cytosolic leaflet of the TGN membrane that controls Golgi-plasma membrane traffic (94, 95). In part, PI4P is implicated in the regulation of granule biogenesis and exocytosis in chromaffin cells (96, 97), and insulin secretion from beta cells (98, 99). The precise membrane composition dictates the biophysical properties of the budding and mature granule, and also provide for spatial coordination of cargo sorting. Lipid rafts play a central role in the formation of secretory granules, of which cholesterol is a key component. Mature granule membranes contain cholesterol in the range of 33-65 mol % (100, 101). Cholesterol depletion by methyl-beta-cyclodextrin (MBCD) causes multiple defects on insulin and glucagon secretion, while insulin granules appeared morphologically smaller with less-dense cores (102), without apparent effect on cholesterol desorption from the plasma membrane (103). Cholesterol depletion caused the mis-sorting of POMC to the constitutive secretion pathway by preventing the granule targeting of carboxypeptidase E (CPE) in AtT-20 cells (101), similar to findings of diminished glucose-stimulated insulin secretion from MIN6 cells (104). Several of the key processing enzymes associate with cholesterol-sphingomyelin-rich lipid rafts, including PC1/3 (105, 106), PC2 (107, 108), CPE (101, 109) and secretogranins (100, 110). The other main component of lipid rafts are sphingolipids, including phosphatidylethanolamine (PE). Reduction of sphingosine kinase Sphk1 in INS-1 cells, which prevents the biosynthesis of downstream sphingolipids (including PE), abolished glucose-stimulated insulin secretion by 50-90% (111). Not only is cholesterol a necessary component of both constitutive and regulated secretory pathways (101, 112), sphingolipids are also critical to their formation. Additionally, they both provide membrane flexibility due to liquid-ordered phase separation in membranes (113–115).

The flexibility of the membrane is essential for granule biogenesis, a process of budding or extrusion from the TGN, ending with scission of the two compartments. Cholesterol (115) and other lipids like diacylglycerol (DAG) and phosphatidylethanolamine (PE) lend
spontaneous negative curvature to the budding granule, enabling the formation of the characteristic neck (rather than a tubular protrusion) for membrane scission (116). The next most abundant lipid in granule membranes is PE, at ~18-20 mol % (101, 110), suggesting that the granule membrane is specifically optimized for its formation. DAG has a cone-shape in an acidic environment (e.g., in the TGN) that lends itself well to the extreme negative curvature (117) and are implicated in the formation of vesicular carriers (115). Together, it is believed that these lipids promote the negative curvature of the neck to aid in granule biogenesis, and a scission complex has been implicated to be recruited to the granule neck in DAG-dependent and protein kinase D-mediated manner (118–120). Positive membrane curvature in the secretory pathway can be sensed or even induced by Bin/amphiphysin/Rvs (BAR) domain-containing proteins; the banana-shaped BAR domain, formed by a coiled-coil structure, has a convex side that associates with and can induce positive membrane curvature (121, 122), and are specifically recruited for the deformation of the TGN membrane (123). Three BAR domain-containing proteins, arfaptin-1, arfaptin-2 and PICK1, are implicated in the formation of secretory granules. PICK-1 regulates granule formation in chromaffin (124) and beta cells, such that impaired PICK-1 function results in reduced insulin secretion, glucose intolerance in mice, and is implicated as one cause for T2DM in humans (125, 126). Arfaptin-1 and arfaptin-2 are recruited to domains of the TGN marked by Arf-like 1 (Arl-1) (123) in endocrine cells, or Arf1 in non-endocrine cells (127). Arfaptins interact with DAG lipids in the budding granule (89) (presumably at the neck), to promote the formation of vesicular and tubular structures (123). Arfaptin-1 in particular mediates the fission process of granules in beta cells (89), either by recruiting a fission complex, or itself effecting the fission (128). While it is likely these same principles from beta cells that apply to α cells, granule biogenesis has yet to be studied in α cells.

Since several neuroendocrine and endocrine tissues produce secretory granules, it is expected that they would have different characteristics (e.g., size, contents, secretory response). For instance, secretory granules are quite variable in total number and size. Estimates of granule numbers range from a few hundred in lactotrophs (129), ~20,000 in bovine chromaffin cells (130), to ~7,000 in mouse α cells (131), and ~6,000 in beta cells.
Similarly, the mean granule diameter varies considerably, and thus its neuropeptide cargoes from ~250 nm in diameter in PC12 (133), 275 nm in mouse α cells (131), to ~350 nm in chromaffin cells (130), with ~600-800 nm being the largest granules belonging to pancreatic acinar zymogen (134). The reason for this variation in size is not well understood, but at least in lactotrophs, granule size correlates with increased probability of complete fusion events and full exocytosis of all cargoes (135).

Concomitant with the maturation of immature granules, the granule lumen is acidified. Vacuolar-type H+-ATPase (V-ATPase) pumps in the Golgi and granules (136) gradually acidify from pH ~6.3-6.0 in ISGs to pH 5.0-5.5 in mature granules (137–139). Critically, the production of mature endocrine hormones requires the acidification of the granule lumen (139–141), which can then activate PC1/3 and PC2 which are optimally active in the range of pH 5.5-6 (142, 143), coinciding with the conditions of the mature granule lumen. Inhibition of acidification prevents exocytic fusion, and in particular glucagon and insulin secretion (144, 145). The V-ATPase also directly interacts with the small Ca^{2+}-binding protein, calmodulin, to mediate the calcium-dependent exocytosis of insulin from beta cells (146, 147), GLP-1 from GLUTag cells (148), and presumably glucagon secretion (149, 150).

The process of transporting a maturing secretory granule toward the plasma membrane involves four discrete mechanistic steps: membrane approach/delivery, docking, priming and fusion. The extrusion process of granule formation is still debated, though there is good evidence to implicate the BAR domain-containing proteins in chromaffin cells (124), possibly by acting as a “proof-reading switch” to sense when the granule is “full.” The kinetic energy of extrusion is controversial and may result either from membrane scission alone, or from the combination of scission and extrusion due to the action of myosin Va motor (151). On the other hand, granule size is more likely to be determined during the membrane-remodelling phase of maturation, for which myosin Va interacts with a key regulator of granule maturation, Rab3D (152–154). Once ISGs are formed, they are immediately transported over long distances (i.e., several μm) toward the plasma membrane, directed along microtubules (155, 156) by the conventional microtubule-
based motor, kinesin (157, 158), until reaching the actin-rich cortex. In the cortical region, the granules are carried the last few hundred nm by myosin Va in neuroendocrine cells (e.g., chromaffin cells (159), PC12 (160)) and also in beta cell lines (161–163). Interference with cortical trafficking either by siRNA-mediated knockdown of myosin Va or a dominant-negative of myosin Va-tail protein impairs insulin secretion (162). Since delivery to the cortex takes ~3-5 sec, it is thought it precedes maturation due to membrane remodelling and cargo maturation which take take 2-4 hours. However, newly formed ISGs become competent for secretion in as little as 30 minutes (164), consistent with the timing of VAMP4 removal from immature granules in AtT20 cells (165), and the recycling of furin from PC12 granules (166).

The last step in this process before exocytosis is membrane fusion. The basic mechanism of fusion is mediated by a variety of soluble N-ethylmaleimide-sensitive factor attachment protein (SNAP) receptor (SNARE) protein complexes. Generally, fusion occurs when a “donor” or “vesicular” membrane, contains a single SNARE (denoted v-SNARE), forms four helix bundle with two additional SNARE proteins on the “target” or “recipient” membrane (denoted as t-SNARE). Each protein contains one or two SNARE motifs of ~60-70 amino acids in size, comprised of a repeating heptad of hydrophobic residues, with a central amino acid being either arginine (v/R-SNARE) or glutamine (t/Q-SNARE) (167). The specificity of these interactions is encoded by and within the SNARE motif itself (167, 168). SNARE fusion generally requires the assembly of one R-SNARE (usually from the vesicle) and three Q-SNAREs (usually from the “target” membrane) (156, 160, 169–171). The minimal fusion process is thought to occur in two steps: the initial formation of a heterodimeric Qa/Qbc complex of plasma membrane syntaxin-1 (as Qa) and SNAP-25 (as Qbc), followed by the insertion of vesicular R-SNARE (synaptobrevin-2/VAMP-2) joining the two membranes in “trans” orientation, and then “zippering” up to form a “cis” oriented complex and achieving fusion. Though this complex is sufficient for spontaneous fusion in reconstituted liposomes, it proceeds with considerably slower kinetics (t1/2 ~ 10 min) compared to the rapid exocytosis seen in neuronal synapses (t1/2 ~ few hundreds of μsec) (160, 161, 172). In vivo, several accessory proteins are important for the regulation of modulation fusion kinetics, responsible for the
divergence between synaptic vesicles and secretory granule exocytosis kinetics, reviewed in (155). Synaptotagmins are a critical accessory protein family because of their calcium-sensitivity (173). Therefore, layered into the SNARE mechanism is tissue-specific modulation for a broader complexity of regulating exocytosis.

1.4 Mechanisms of regulated secretion

1.4.1 Mechanisms of glucagon secretion from α cells

Glucagon secretion can be caused by cellular, hormonal and autonomic triggers (Figure 3). These are commonly referred to as secretagogues. The hypoglycemic status (e.g., during fasting or exercise) of the individual is arguably the most important physiologic secretion stimulus. Glucose-regulated glucagon secretion follows a U-shaped dose-response, with maximal inhibition occurring at 6-7 mM glucose, in α cell lines, isolated mouse islets (26, 174), as well as in humans (175, 176), below the glycemic threshold (7 mM) for insulin secretion (26). Chronic insult of hyperglycemia to αTC1-6 cells results in up-regulation of exocytotic machinery and enhanced basal glucagon secretion (177). By autocrine feedback, glucagon can upregulate its own biosynthesis (178), perhaps as a rapid mechanisms to replace the newly depleted granule pool. α cells directly sense glycemia via the glucose sensor, glucokinase (177). Glucose metabolism is linked to glucagon secretion, such that increases in the cellular ATP/ADP ratio, inhibiting the K-ATP channel, subsequently inhibiting glucagon secretion (179), otherwise the α cell is rendered functionally “blind” to external glycemic levels. However, the precise mechanism underlying glucose-regulated glucagon secretion remains controversial (180).

Amino acids and lipids can also stimulate glucagon secretion. It is well established that L-arginine is a potent glucagon secretagogue both in vitro and in vivo (181–183), even overcoming the inhibition present at >5 mM glucose. Leucine in particular can be both stimulatory at normal physiologic levels, or inhibitor at elevated serum leucine levels (184). Mixed meals (especially high-fat meals), will also stimulate, rather than suppress, glucagon secretion in healthy adults (174, 175). Nutrient ingestion and composition are important physiological triggers for maintenance of glucose homeostasis.
Figure 3. Mechanisms of glucagon secretion from α cells. A simplified schematic of key mechanisms involved in regulated secretion of glucagon. Stimulation by amino acids or decreased glucose triggers membrane depolarization, elevates intracellular calcium, and initiates glucagon exocytosis. Several paracrine modulators (e.g., insulin, somatostatin, GABA) can dampen or increase secretion rates, generally by coupling of GPCR to elevation of intracellular calcium.

Similar to other electrically excitable cells (e.g., chromaffin cells, beta cells), glucagon exocytosis is triggered by fired action potentials, triggering Ca\(^{2+}\) ion influx through voltage-gated calcium channels, and the resulting increase in intracellular Ca\(^{2+}\) causes membrane fusion and secretion of glucagon. A principle component to regulating action potential firing is the K-ATP channel (131, 185). The past 15 years have seen several investigations into the mechanisms of action potential general and electrical excitability of α cells. Most clearly underlying glucagon secretion are the voltage-gated calcium channels (176), and ATP-dependent K-ATP channels (for a thorough review of α cell electrophysiology, see (180, 186)). Closure of K-ATP channels (e.g., by tolbutamide) results in α cell depolarization and glucagon secretion both in vitro and from intact islets in rodents and humans (187), while membrane hyperpolarization (e.g., by diazoxide)
inhibits glucagon secretion in rodents and humans. The exact mechanism underlying K-ATP channel activity is still debated, however, since α cells are spontaneously active at low glucose, and become “silent” as glucose is elevated, a recent hypothesis suggested that the membrane voltage is maintained near the threshold of depolarization (180). Under normal conditions, excursions to high glucose should inhibit glucagon secretion, and hypoglycemic excursions should stimulate secretion. Ashcroft and Rorsman posit that under pathologic conditions of T2DM and some forms of T1DM, this sensitivity is reversed leading to paradoxical α cell behaviour. Whatever the importance of K-ATP channels on α cell behaviour, there are also K-ATP channel-independent mechanisms of stimulating glucagon secretion.

Several paracrine factors have been proposed to play a role in glucagon secretion, including insulin, somatostatin, GIP, GLP-1 and GABA. Of these, α cells respond most robustly to insulin and somatostatin. Insulin is a paracrine inhibitor of glucagon secretion, as can be expected from counter-regulatory hormones, and was found to be an accurate reciprocal relationship in T1DM humans with intravenous infusions of insulin (188). Under extreme hypoglycemia (1 mM glucose), insulin completely inhibits glucagon secretion from human islets (176); perfused rat pancreata and isolated islets (189); and mouse islets (190). Transgenic mice with a specific α-cell deletion of the insulin receptor displayed glucose intolerance, enhanced glucagon secretion under hypoglycemic clamp, and gradual increase in beta cell mass (191), confirming the importance of direct insulin action on α cell glucagon response.

Somatostatin is also a critical paracrine modulator of glucagon secretion. The main isoform of the somatostatin receptor, SST-R2, is functional in human and mouse α cells (192), that when activated, hyperpolarizes the α cell downstream of intracellular Ca²⁺ influx, though prevents exocytosis though an incompletely-defined mechanism. Mice lacking SST-R2 (193) displayed normal basal glucagon and insulin secretion. However, glucagon secretion was two-fold greater upon depolarizing stimuli, and exogenous somatostatin failed to suppress glucagon secretion. Somatostatin is a potent inhibitor of glucagon secretion in fed states. Easing this “brake” contributes significantly to non-
fasting hyperglucagonemia and hyperglycemia (194–196), consistent with pathophysiology of T2DM.

The incretins are interesting for their divergent effects on glucagon secretion. GLP-1 indirectly exerts an inhibitory effect (197–199) on glucagon by stimulating somatostatin secretion (194, 200, 201). On the other hand, GIP stimulates glucagon secretion in normoglycemic humans (202). The GLP-1R is virtually absent when detected by PCR or immunocytochemical techniques in normal rodent and human islets (203). In contrast to GLP-1, a direct effect of GIP on α cells from perfused rat pancreas was shown to stimulate glucagon secretion in conditions of hypoglycemia (201). From this, it is likely that GIP is a direct enhancer, and GLP-1 an indirect inhibitor, of glucagon secretion.

The metabolic glutamate/GABA-glutamine pathway is functionally linked to glucagon secretion. α cells pump glutamate into secretory granules via vesicular glutamate transporters (VGLUT) 1/2, co-secreted with glucagon (204, 205). In contrast, GABA and glycine are pumped into α cell synaptic-like microvesicles (SLMVs) by vesicular inhibitory amino acid transporter (VIAAT) (206, 207). The activity of VGLUT is down-regulated while VIAAT is up-regulated in response to high glucose, and vice versa (208), allowing the ambient glucose level to determine a fine balance between stimulatory and inhibitory neurotransmitters loading into granules and SLMVs. Uptake of glutamine serves as a substrate for production of glutamate and GABA, and the mechanism of uptake may mirror the case of neurons. Specifically, glutamine may be co-transported with sodium ions by diamineacyltransferase 2 (SAT2), producing a depolarizing electrogentic ion current, believed to stimulate glucagon secretion (209, 210). This is consistent with the generation of intracellular Ca\(^{2+}\) oscillations (211), which does stimulate glucagon secretion. In contrast, beta cells load both glutamate and GABA into SLMVs (212, 213), which can be secreted independently of glucose (214). At the same time, glutamate uptake depolarizes α cells and stimulates glucagon secretion, providing an autocrine and paracrine feedback mechanism (214). α cell VGLUT activity may also work in association with the vacuolar H+-ATPase, which coordinates an increasingly acidified granule lumen, similar to their operation in beta cells (204, 206, 207). Ambient
blood glucose levels can direct islet cells to load both granules and SLMVs with either excitatory and inhibitory neurotransmitters, potentially coupled with granule acidification, to amplify or dampen glucagon and insulin secretion.

1.4.2 Mechanisms of GLP-1 and GLP-2 secretion from L-cells

The “incretin effect” is so-called for amplifying insulin secretion following ingested nutrients when compared to intravenous administration (32). The incretin hormones, glucose-dependent insulinosotropic factor (GIP; previously called gastric inhibitory peptide) and GLP-1, are mainly derived from the gut. GIP is mainly secreted by K cells in the upper small intestine. In contrast, GLP-1 is secreted from L-cells in the lower small intestine and colon, also able to directly sense nutrients in the lumenal environment (215, 216). Both cell types are expressed in similar (diffuse) quantity throughout the entire gut (217), making it challenging to study directly. Because of their ability to increase glucose- and nutrient-stimulated insulin secretion, this anti-diabetic property – and in particular due to GLP-1 – remains actively pursued for T2DM therapy. Two biologically active GLP-1 (7-37) and GLP-1 (7-36 amide) peptides equally stimulate insulin secretion in humans (218, 219), though the physiological significance of two forms is not clear. The biological relevance to stimulating insulin secretion remains a key gut hormone in the maintenance of nutrient homeostasis.

The mechanisms of GLP-1 secretion from L-cells are less understood by comparison to insulin and glucagon. As a note, wherever GLP-1 secretion is discussed, the same mechanism is also assumed for GLP-2, as both hormones are secreted in equimolar concentrations, which has not been examined separately. The most reliable in vitro models remain GLUTag cells, derived from a mouse intestinal tumour expressing proglucagon-SV40 large T- antigen (220). More recently, a transgenic mouse line expressing Venus (a YFP) under the promoter of proglucagon (221) has enabled more efficient and specific primary L-cell studies, a necessary advantage for such diffuse enteroendocrine cells (28, 217). The post-prandial profile of GLP-1 secretion is biphasic, with the first peak at 15-30 minutes and the second at 1-2 h (222). Intestinal nutrients appear to be the major physiologic stimuli of L-cells, especially by sugars, fatty acids and amino acids and these mechanisms will be discussed below.
As with (neuro)endocrine cells in general, secretagogue-mediated exocytosis is coupled to electrical depolarization of the cell membrane (Figure 4), in which calcium uptake is coupled to exocytosis. Chemosensation of sweet molecules in the gut (e.g., glucose, fructose, sucrose, artificial sweeteners), activate through the coupled G protein-coupled receptor, TAS1R2-TASR3, and can be imported into the cell by facultative GLUT glucose transporters (223), and the Na⁺-glucose co-transporter, SGLT1, typical of intestinal brush-border transporters. The taste receptors are stimulated by simple sugars and common sweeteners, and couple to the G-protein α-gustducin, that upon activation, elevates intra-cellular calcium and triggers secretion. Mice with deletion of α-gustducin (-/-), TAS1R2 (-/-) or TAS1R3 (-/-) have nearly abolished glucose-stimulated GLP-1 secretion (224, 225). The electrogenic current of Na⁺ by SGLT1 is the most important glucose transport system, directly depolarizing the L-cell, thus triggering secretion (221, 226, 227). This is the key difference to the Ca²⁺-dependent action potentials in α and beta cells (228, 229). GLP-1 secretion can also be stimulated by sensation of free fatty acids and amino acids. In GLUTag cells, GPRC6A is the main amino acid sensing mechanism (230), primarily by basic amino acids. Although L-cell function was not assessed, deletion of the receptor pre-disposed mice to obesity, insulin resistance and glucose homeostasis dysregulation (231). On the other hand, fatty acids are sensed by FFAR1 (or GPR40) (232) and GPR120 (233). In humans, GLP-1 secretion is stimulated by oral administration of lipid (234). Though it is unclear which fatty acid receptor is more important, mice lacking FFAR1 had reduced response to FFA-stimulated GLP-1 secretion and serum insulin (235). Together, the fatty acid receptors also couple to α-gustducin to stimulate GLP-1 secretion (236).
Figure 4. Mechanisms of GLP-1 secretion from L-cells. A simplified schematic of key mechanisms involved in regulated secretion of GLP-1 (adapted from (237)). Intestinal lumen nutrient sensation by GPCRs, facultative glucose transport (GLUT) and electrogenic Na⁺-glucose co-transporters (SGLT) are shown. GPCR activation or inward sodium current can trigger membrane depolarization, elevation of intracellular calcium, and GLP-1 secretion. Thus, several mechanisms couple nutrient stimulation to incretin hormone exocytosis.

Only one report to date has characterized major SNARE machinery involved in GLP-1 exocytosis. GLP-1 secretion requires the interaction among VAMP2, syntaxin-1a and SNAP25 in GLUTag cells (238), consistent with their expression in L-cells and with exocytosis of both insulin and glucagon. There are several similarities between the stimulus-secretion coupling and exocytosis mechanisms among endocrine cells, and it suggests that the hormone secretion mechanisms are broadly similar among endocrine cells. Tissue-specific differences between L- and α-cells are manifest in proglucagon processing and possibly sorting mechanisms.
1.5 Processing of Proglucagon

Mammalian proglucagon is synthesized as a 180 amino acid precursor (239–241), whose domains are organized as in Figure 1. Proglucagon encodes for a plurality of peptides that are processed in a tissue-specific manner through the action of PC1/3 and PC2. The end result of differential expression of these PCs in α-cells, L-cells or neurons produces the characteristic pattern of proglucagon-derived peptides. Proglucagon is not post-translationally modified (such as glycosylated) with the exception of processing to its constituent hormones (242, 243), and C-terminal amidation that occurs to GLP-1 (discussed below).

1.5.1 Glucagon production in pancreatic α cells

The processing of proglucagon in α cells begins with cleavage at the inter-domain dibasic site, Lys70Arg71, liberating glicentin and the major proglucagon-derived fragment (MPGF) (Figure 1). Glicentin is further processed to glucagon, (along with the flanking GRPP and IP-1) (244). Proglucagon processing is accomplished by PC2, shown to be necessary and sufficient for processing to the α cell profile in αTC1-6 cells (245), and is sufficient for α-cell-like processing in AtT-20 cells (246). PC2 activity depends on a cofactor, neuroendocrine peptide 7B2, both stabilizing PC2 in its active conformation and inhibiting enzymatic activity until delivery to granules (247, 248). 7B2 also prevents PC2 from aggregating into inactive oligomers (249). PC2 is enzymatically active within mature secretory granules (137, 142). By using a pulse-chase paradigm, the time dynamics of proglucagon processing in αTC1-6 cells confirmed the sequential cleavage beginning at the inter-domain site, Lys70Arg71, producing glicentin and MPGF after 30-45 minutes chase (245, 250). These chase times are consistent with a conversion of proglucagon beginning in the TGN or immature granules. This leaves open the possibility that some amount of proglucagon is initially processed to glicentin and MPGF in the TGN or immature granules. One possible endoprotease that could substitute would be furin, found in all cell types, and active in the TGN and is removed from immature granules. This is suggested by the findings of proglucagon processing in several non-endocrine cell lines which only express furin (245, 246, 251–253), and is consistent with the expectation that PC2 would be inhibited by 7B2 in these early compartments of the secretory pathway. Notably, the remaining processing of glicentin is carried out in mature
granules at dibasic sites recognized by PC2 to remove GRPP, and in particular, the PC2-specific site to remove IP-1 (Figure 1).

Mouse models confirmed the role of 7B2-activated PC2 on α cell production of glucagon. PC2 knockout mice abolished glucagon production and severely impaired α cell proglucagon processing (254, 255), consistent with an α cell line derived from these PC2 null islets (256). These mice lacked detectable serum glucagon, had improved glucose tolerance, consistent with a lack of active glucagon. Interestingly, these α cell granules became irregular in shape and electron-translucent, suggesting PC2 expression was not necessary for granule targeting of proglucagon. Proglucagon processing was significantly decreased either upon depletion of 7B2 in an α cell line, or in 7B2-knockout mice (257), similar to the PC2-knockout phenotype. Altogether, PC2 is essential to produce glucagon.

There is recent controversy surrounding the production of bioactive GLP-1 within pancreatic α cells. Whether GLP-1 is produced at all in a healthy pancreas or terminally differentiated α cells (as described above), the relative expression is far lower than that of the gut, its principle site of production. PC2 cannot cleave the Arg77 to yield active GLP-1, and therefore if it is to be produced, it must encounter PC1/3 (discussed below). Notably, this is not the case in one recent report (258). α cell differentiation is controlled by a sequence of multiple transcription factors (reviewed in (259)) such that mature α cells repress expression of PC1/3 and thus incapable of producing bioactive GLP-1. In the face of insult or injury to the islet, α cells are known to adapt their phenotype and can even revert to a precursor α cell, allowing them some functional plasticity in response to metabolic insults. Severe beta cell chemical ablation in mice is so severe as to cause beta cells to trans-differentiate into α cells through an insulin+/glucagon+ intermediate cell (260), implying a transient ability to produce bioactive GLP-1 by PC1/3 expression. The α cells isolated from PC2-knockout islets expressed PC1/3 and produced GLP-1 (256), whose α cells were hyperplastic in the originating mice (254). Mature GLP-1 could be detected in such hyperplastic α cells resulting from the loss of glucagon signalling (by global GCG-R knock-out) (19), implying the expression of PC1/3. A time-course analysis of PC1/3 and GLP-1 production within α cells using diabetic (db/db) mice as a model of
T2DM, and NOD mice as a model of T1DM show that PC1/3 (and GLP-1 production) is prominently upregulated in T2DM (261). Obesity and T2DM systematically elevate pro-inflammatory interleukin-6, promoting α cell hyperplasia (53) and production of mature GLP-1 and PC1/3 in α cells (56). Therefore multiple models of metabolic disorders (e.g., diabetes, obesity) or chemical insults to the islet can profoundly alter the α cell phenotype and begin producing detectable quantities of mature GLP-1. However, the gut remains the major site of GLP-1 production in healthy individuals and animals.

1.5.2 GLP-1 and GLP-2 production in intestinal L-cells

The most important products of intestinal L-cells are the mature GLP-1, GLP-2, and also glicentin and oxyntomodulin (262, 263), which are increasingly being appreciated as important hormones for nutrient homeostasis. The processing profile of proglucagon within the GLUTag cell line and native L-cells (253, 264) is coordinated by PC1/3 (Figure 1). Just like pancreatic α cells, L-cells also process proglucagon at the inter-domain site, Lys70Arg71, after ~20-30 min chase (250). Glicentin is processed to oxyntomodulin after 30 minutes (250), in a molar ratio of about 3:1 of glicentin:oxyntomodulin (264). PC1/3 further processes MPGF to predominantly produce GLP-1 (7-37) by cleavage either at Arg109Arg110 or the internal Arg77 within GLP-1, as well as GLP-2 (250–252). Viral over-expression of PC1/3 into PC2-null αTC1-6 cells is sufficient to produce bioactive GLP-1 (265), suggesting that the simple change of convertase could shift the equilibrium toward GLP-1 production. Knockout of PC1/3 in mice blocks production of GLP-1 and GLP-2 (266, 267). Therefore, PC1/3 is necessary and sufficient for proglucagon processing in intestinal L-cells. Considering the cellular location of this processing, such rapid chase times suggest that initial cleavage occurs in the TGN by furin or PC1/3, or within immature granules (142). Recently, a hypermorphic PC1/3 mutant (S357G-PC1/3) showed increased specificity to the two dibasic sites flanking glucagon, liberating ~20% more glucagon than wild-type PC1/3, but did not alter its basal secretion based on conditioned media samples (268). Since this hypermorph is active at neutral pH, proglucagon is more likely to be rapidly processed prior to arriving at the ISG, and it would suggest this initial processing occurs prior to the sorting of glucagon to granules in GH4C1 cells.
GLP-1 and GLP-2 differ in the extent of their processing in vivo. GLP-1 and GLP-2 are co-secreted in equimolar amounts from the same granules (263, 269). The natural form of GLP-1 extracted from human intestine showed a further processing of the (inactive) GLP-1 (1-37) by cleavage at the monobasic Arg77 to yield bioactive GLP-1 (7-37) (262, 263). As in cell models, there is natural variation in the ratio of glycine-extended GLP-1 (7-37) and amidated GLP-1 (7-36)-amide both along the length of the intestines and between mammals (263, 269–271). Mono- and dibasic C-terminal amino acids are removed by carboxypeptidase E (CPE) (272). After carboxypeptidase activity, GLP-1 amidation is accomplished by peptidylglycine α-amidating mono-oxygenase (PAM) (273, 274). In contrast to GLP-1, removal of the C-terminal basic residues from GLP-2 is its final modification. The focus of recent GLP-1 research has shifted towards identifying enteroendocrine L-cells along the entire gut and its multiple physiologic roles.

1.5.3 Proglucagon processing in the central nervous system

The pattern of GLP-1 and GLP-2 production in the CNS is similar to that of the intestines (275), and so it is presumed that PC1/3 is the responsible convertase. Indeed, the role of PC1/3 has been extrapolated from the numerous studies of proglucagon processing using neural-derived and endocrine cell lines. However, this assumption has not been tested in vivo, partly owing to a much smaller abundance in the brain, compared to the pancreas and gut. The predominant areas of proglucagon expression in the brain are a subset of neurons in the solitary tract nucleus (276). Several sensitive peptidomics studies of the crude lysates of the brainstem (and other brain regions) in normal and transgenic mouse models such as knockouts of PC2 or PC1/3 (277, 278) or CPE mutant obese mice (279, 280) failed to detect proglucagon-derived peptides, presumably due to extremely low expression in these regions (personal communication with Dr. L. D. Fricker). Thus the actions of PC1/3 (or PC2) in the brain have only been inferred from correlation of their expression with proglucagon-expressing neurons, and this avenue of investigation remains inconclusive.

Two alternative views are parsimonious with proglucagon expression and processing in the brain. Proglucagon is certainly expressed by subsets of neurons in the brainstem and hypothalamus. By comparison with POMC- and NPY-expressing cortical neurons,
proteolytic processing to NPY, ACTH, beta-endorphine and α-MSH are significantly dependent on the action of cathepsin L (281, 282). An emerging idea is the expression of proglucagon, and secretion of GLP-1, by activated glia in the brain. A previous report identified proglucagon+ cells in the cerebral cortex and hippocampus that are thought to be glial cells in mice, and also detected GLP-1 secretion from the mouse glial cell line, BV-2 (283). An alternative endopeptidase, yapsin-1 (284) behaves similarly to PC2, with specificity to dibasic sites, and in vitro, can liberate glucagon. The mechanism of processing in glial cells remains unknown. PC1/3 is the traditional prohormone convertase for processing, though yapsin-1 and cathepsin L are recently proposed candidates. The idea that proglucagon could be expressed in two distinct regions in the brain also invites further investigation into potentially novel mechanisms of its processing in the brain.

1.5.4 The ultra-structural localization of the proglucagon-derived peptides within granules

There has been much interest in the ultra-structural characterization of α cells and localizing glucagon within the late secretory pathway. Previous findings by Orci and colleagues (285) distinctly co-localized glucagon and glicentin within the same granule in which glucagon was diffusely located within the electron dense core while glicentin is segregated to the electron-lucent halo. In those granules studied, there were equal proportions of glucagon+ granules and glicentin+ granules (which was likely GRPP), suggesting that glicentin enters mature granules where it then undergoes processing by PC2. Using anti-sera directed against mid-sequence GLP-1 or GLP-2 in α cells revealed that the staining pattern is restricted to the electron-dense core (286), consistent with the identification of MPGF in α cell granules. Similarly, glicentin, GLP-1 and GLP-2 appeared to co-localize in serial thin-section immunoelectron images of human L-cells (286). Here, GLP-1 is diffusely localized within the dense core, while GLP-2 was localized to the halo, suggesting that MPGF has undergone processing in an immature granule to allow for their topological segregation. Interestingly, there have not yet been molecular dynamic studies to determine diffusion rates within granules. None of the ultra-structural studies have shown glucagon, or any of the other PGDPs, to adopt higher-order or crystalline structures, suggesting that proglucagon is less prone to aggregation in
the secretory pathway. Previous data from our lab has also ruled out an ability for proglucagon to bind to reconstituted liposomes. This makes the existence of a receptor an attractive hypothesis for proglucagon sorting. Immunoelectron analysis of granule contents does not reveal the degree of co-storage of the various PGDPs within granules. To address the outstanding question of whether proglucagon processing precedes its initial sorting, and therefore MPGF and glicentin may enter the same or different granule populations, live cell fluorescence imaging or proteomic approaches are necessary to determine the proportion of the various PGDPs within immature granules. It remains unanswered from immunoelectron surveys whether glicentin and MPGF are always found in immature granules, or if they may be diverted to different granule sub-populations.

Processing is only one critical component to the production and secretion of hormones as they must end up in secretory granules for proper physiologic response to stimuli. Critically, sorting form the TGN toward mature granules necessarily parallels the processing events to yield the PGDPs. As proglucagon and its derived peptides transit the regulated secretory pathway, they will encounter the necessary low-pH and high calcium environments necessary for PC1/3 or PC2 activity, they will be concentrated, and this compartmentalization is required for nutrient stimulated secretion of glucagon, GLP-1 and GLP-2. To date, the molecular mechanism that determines how proglucagon is sorted to granules remains unknown. However, several clues from the study of other neuropeptides and hormones suggest that structures within proglucagon may act as a sorting signal, guiding its route toward granules.

### 1.6 Mechanisms of Sorting

In characterizing the route prohormones take along the regulated secretory pathway, an essential bottleneck that must be overcome, is the efficient sorting and condensation into secretory granules. Two models were proposed to explain this step (285). Sorting can occur either through a cargo-receptor interaction that is mediated by a sorting signal, or, cargo proteins reversibly self-aggregate prior to entry in nascent granules. Subsequent research of intracellular trafficking research has focused on identification of sorting
signals and their receptors, and characterizing cargo aggregates in the secretory pathway (reviewed in (287)). This section focuses on what is known about the specific structural motifs, or sorting signals, of peptide hormones and neuropeptides.

1.6.1 Aggregation of hormones

Condensation of granule cargo by aggregation was first described by Palade (288). The hypothesis is that cargo proteins have a weak affinity to aggregate under the ionic conditions of the secretory pathway, and that in doing so, provides an thermodynamically efficient mechanism condensation, which then enter granules as electron-dense aggregates. Two “classic” examples of aggregation are put forth as evidence for aggregation of cargo hormones. First, pro-atrial natriuretic peptide (pro-ANP) is secreted from secretory granules of atrial cardiomyocytes in response to increased blood volume. Sorting to granules in either AtT-20 or PC12 cells seems to require calcium-mediated aggregation via diacidic residues in its pro-domain (77). Mutation of Glu23,Glu24 in the sorting domain (pro-ANP 11-30) increases solubility of aggregates in vitro, and results in constitutive secretion from AtT-20 cells. Second, the somatotropin family which includes growth hormone (GH), prolactin and placental lactogen. These structurally related hormones are characterized by a four α-helix bundle, they do not undergo processing by prohormone convertases, and are stabilized by intra-molecular disulphide bonds (289). Prolactin and GH form aggregates in the Golgi that traffic as electron-dense cores through the regulated pathway (288). However, aggregation is not a one-size-fits-all mechanism for the regulated secretory pathway. The ability for cargo proteins to aggregate (either in a high calcium or low pH environment) varies substantially, and aggregation is better observed as a function of the protein rather than the secretory pathway. Human prolactin could not aggregate in 10 mM Ca²⁺ in vitro at any acidic pH, conditions that match the granule environment, negating the idea that sorting relies on this mechanism (290). Sorting-signal tagged human serum albumin traffics efficiently to granules in AtT-20 while undergoing no (or negligible) aggregation in AtT20 cells (291). While short charged segments of amino acids mediate aggregation of pro-tachykinin to sort into granules (292). Alternatively, inducing aggregation in the secretory pathway can actually reduce sorting efficiency and secretion (293, 294). Regardless of the influence of aggregation at the TGN, ample empirical evidence shows that prohormones and
neuropeptides mostly rely upon specific domains or structures that facilitate their correct targeting and secretion.

### 1.6.2 Dibasic sorting signals

Dibasic sites are well-known because of their endoproteolysis by prohormone convertases to yield mature hormones from their precursors. That these sites undergo cleavage along the regulated secretory pathway makes them attractive sorting signals. The representative example of dibasic sites influencing sorting is pro-renin (295), which contains two dibasic sites in its pro-domain that are essential for both its processing and regulated secretion. Similarly, pro-neurotensin/neuronedin N and pro-thyrotropin-releasing hormone (pro-TRH) both require processing at multiple dibasic sites to gain entry to granules (296–299). Dibasic signals are identified for pro-NPY (300), pro-gastrin (301), and pro-VGF (non-acronymic) (302).

### 1.6.3 Diacidic sorting signals

The first identified sorting signal was a di-acidic motif from pro-opiomelanocortin (POMC) (303, 304), and is also found in pro-insulin (305), pro-brain-derived neurotrophic factor (pro-BDNF) (306) and pro-enkephalin (307). For POMC, the key acidic residues for receptor interaction are Asp9,Glu14, stabilized by Leu11,Leu18, that must be constrained in an amphipathic disulphide loop conformation (303, 304). For pro-insulin the disulphide bonds constrain Glu-A17,Glu-B13 and Leu-A16,Leu-B17 necessary for receptor binding (304, 305). Therefore, this constrained di-acidic amino acid motif is a well-conserved sorting signal.

### 1.6.4 Amphipathic α-helices

The somatostain family includes somatostatin (SST) and cortistatin (CORT). Pro-somatostatin is sorted to granules via an N-terminal amphipathic α-helix in the pro-domain (308, 309). Cortistatin also contains a homologous pro-domain α-helix (309), of unknown functional significance. A final example is found in pro-cocaine and amphetamine regulated transcript (pro-CART) (310). The amphipathic nature of these helices, in which a large hydrophobic patch surrounded by charged residues can direct granule sorting (311), possibly by hydrophobic association with membrane lipids or intermolecular aggregation. Although not a hormone, PC1/3 is also targeted to granules via an
amphipathic helix (PC1/3 738-753) that can be partly attributed to both calcium-dependent aggregation of these helix domains, and membrane association (106).

1.6.5 Disulphide loops

Pro-vasopressin and pro-oxytocin are highly conserved peptides that sort to secretory granules using both oligomerization as well as multiple intra-molecular disulphide loops (312). Entry into the regulated pathway first requires cleavage of the pro-domain from the hormone (313). Once processed, these pro-domain/hormone heterodimers can aggregate via intra- or inter-molecular disulphide loops and the conformation of the hormone domain is crucial for sorting using both AtT20 and Neuro-2a cells (312, 314). In contrast, pro-somatostatin contains a mostly disordered pro-domain and two dibasic cleavage sites to generate the mature hormones, SST-14 and SST-28, characterized by a C-terminal disulphide loop. It was recently suggested that the nature of this loop can dynamically form fibrils and directly affect exocytosis, and possibly even sorting efficiency (315). Therefore, further nuances to the sorting mechanism remain to be identified.

1.6.6 Glycoproteins with sorting signals

The gonadotropins including follicle stimulating hormones (FSH) and luteinizing hormone (LH) are distinguished in that they are heterodimeric glycopeptides, consisting of an common α, and unique beta chain. Thus, the beta chain determines the identity of the hormone, its biological specificity, and the mechanism of intracellular trafficking (316). Despite being secreted from gonadotropes, FSH and LH are secreted differently: constitutive secretion of FSH and regulated secretion of LH (317). The information encoded for regulated secretion is encoded by the C-terminal heptapeptide of the LH-beta subunit, not found in the FSH-beta subunit (318, 319). The exact mechanism is not clear, and the data do not rule out a role for either increasing folding efficiency in the ER by the presence of the heptapeptide this leading to regulated secretion (320) or whether the hydrophobic nature is required aggregation, or interaction with a yet unidentified sorting receptor (321).

1.6.7 Compounding sorting signals may increase returns

Sending hormones to granules is often not accomplished via one signal alone. Instead, multiple sorting signals may synergistically combine to increase storage efficiency (302,
The advantages of this process are clear. Prohormones that are differentially processed in multiple tissue types (e.g., POMC and proglucagon) may use tissue-specific receptors, thus allowing independent routes for prohormone processing and secretion. Increasing sorting efficiency is paramount to efficient production of bioactive peptides.

What is clear is that there is not a single universal sorting signal. So far, the molecular range of identified signals can be as small as two dibasic amino acids or large loop-based structures. Aggregation may play a role in these processes, though its extent appears to be highly variable, and in many examples, is unnecessary for sorting. Still, all of these modalities may combine in any particular prohormone in order to modulate the efficiency of their sorting. Generally, sorting signals tend to be contained in the pro-domain, and not in the hormone domains. In order to complete the mechanism, protein binding partners, so-called sorting receptors, recognize these motifs to facilitate granule entry.

1.6.8 Sorting Receptors

Just as there is a variety of sorting signals, these signals are recognized by a smaller subset of receptors. The search for a universal sorting receptor is elusive, and in reality, multiple (neuro)endocrine mechanisms exist to serve the diversity of (neuro)endocrine peptides. The operational definition of a sorting receptor must meet several criteria: 1) receptors and their cargoes should co-target to the same granules; 2) receptors should bind to their ligands in a pH- or cation-dependent manner to show specificity for binding in the TGN or immature granules; and 3) a loss of the interaction or receptor should severely affect regulated secretion and cause mis-sorting the constitutive pathway. Due to the need for a receptor to selectively bind cargoes at mildly acidic conditions of granules, and then release them upon exocytosis when exposed to a neutral pH, these interactions are likely to be very dynamic, binding with both low affinity and high specificity. Sorting efficiency can be increased by the partial co-aggregation between receptor and cognate hormone(s), in a one-to-many relationship along the ionic gradient(s) of the secretory pathway. In this way, cargo is efficiently concentrated, while not creating additional demand for activation energy. Two sorting receptors used by endocrine tissues involve processing enzymes and granins. Together, these few members serve a broad range of hormones thus identified.
The first *bona fide* sorting receptor is carboxypeptidase E (CPE), which is responsible for trafficking the “POMC” proteins: POMC, pro-insulin, pro-BDNF and pro-enkephalin, as previously discussed. Inactive pro-CPE is directed to the TGN (57 kDa form) by a C-terminal trans-membrane α-helix in cholesterol-sphingolipid enriched membrane domains (323), and maturing in granules (324). The ability to bind prohormone cargo relies on Arg255 and Lys260 (304) that can bind two acidic amino acids ~12-15 Å on the hormone, stabilized by two hydrophobic residues spaced ~5-7 Å apart. Evidence for CPE as a sorting receptor is most convincing from depletion studies of transgenic mice. A strain of mice called CPE\textsuperscript{fat/fat}, harbors a mutant S202P-CPE, that is enzymatically inactive (325). CPE\textsuperscript{fat/fat} mice constitutively secrete POMC (326) and proinsulin, becoming obese, diabetic, and the granules of beta cells lack the characteristic electron-dense cores (325). Similar results for pro-insulin were observed in CPE knock-out mice (327).

PC1/3 and peptidylglycine α-amidating monoxygenase (PAM) are the other processing enzymes that act as sorting receptors and are co-targeted to granules. PC1/3 has a membrane-associated amphipathic α-helix (105, 106), while PAM is a type 1 membrane protein (274). The loss of the membrane-associated α-helix in PC1/3 mis-routes pro-renin for constitutive secretion (105, 328). Within granules, PAM acts on hormone substrates to produce C-terminal amidation of glycine residues (such as in GLP-1). In secretory granules of atrial myocytes, PAM is the sorting receptor of pro-ANP (329), despite not being a substrate of PAM. No reports on whether PAM acts as a sorting receptor within α- or L-cells exist, though it seems unlikely, since heterozygous knock-out PAM+/− mice have normal basal blood glucose and glucose tolerance at 2 months of age, and only develop mild obesity by 10 months, likely a non-specific complication of the model (329, 330). Therefore the possibility remains that PC1/3 may sort proglucagon in L-cells.

The granins are a complex family of proteins, co-operatively sorting cargo hormones and are major components of the granule matrix. Chromogranin A and B (CgA and CgB), coupled to secretorgranins II and III (SgII and SgIII) are most important for sorting. CgA
is the most abundant constituent in chromaffin and all pancreatic islet cells (331, 332), while beta and alpha cells also contain SgII and SgIII (333). At the TGN, it is believed that the granins nucleate partial aggregation at the mildly acidic pH and the calcium environment of the TGN (334–337). Importantly, the granins selectively aggregate with neuropeptides and hormones, excluding constitutively secreted cargoes (338). CgA and CgB are soluble proteins that require the aide of an adaptor protein, one of the secretogranins for granule targeting. The most profound demonstration of CgA (and also CgB) actions on hormone storage, is that CgA-/- mice have severely impaired catecholamine storage in adrenal chromaffin cells (339, 340), and is moderately impaired in CgB-/- mice (330, 341). The effects on other (neuro)endocrine tissues are rather mild (even unaffected) likely due to compensation by upregulation of other granins (342). Over-expression of CgA can enhance sorting of GH and NPY to granules (343), and CgB can direct POMC to granules and for increased secretion of ACTH (344). The secretogranins, SgII and SgIII, are adaptors of chromogranins (110, 345, 346), by directly associating with lipid rafts of granule membranes. The granin mechanism can explain the increased stoichiometry ratio of cargo to receptor, a common (yet unproven) criticism of the sorting receptor hypothesis. Co-aggregation of soluble cargo in the TGN by chromogranins produces a large mass of peptides that can then be targeted to granules by the membrane-bound secretogranins, efficiently packing large quantities of chromogranins and hormones. SgIII does bind the condensed aggregates of granin and neuropeptides (e.g., adrenomedullin), providing a high-capacity sorting mechanism (347), and SgII appears to operate in a similar fashion (110). Together, segregation of regulated cargo is retained in the area of nascent granule formation.

To illustrate an example, consider the sorting of POMC. SgIII binds the cholesterol-rich membrane at the N-terminus, while either binding CgA or CPE via the C-terminus (100, 346–348). SgIII and CPE are both closely associated in cholesterol-rich membrane domains, and can individually bind to POMC (349, 350). As a nascent granule forms, CgA is carried to granules by SgIII, while POMC is carried by CPE (and also CgA), where they will encounter each other in the lipid raft-enriched membrane. SgIII is also required to properly sort CgA in AtT-20 cells (110), but can only bind to either CgA or
CPE at one time (349). Therefore, at the nascent granule membrane, POMC can be handed off to CgA which has a high binding capacity due to aggregation, while SgIII laterally binds to either CgA (to retain POMC) or CPE (to facilitate a transfer to the SgIII-CgA complex) (349, 350). This “hand off” mechanism describes two actions happening in parallel, from the independent delivery of POMC via CPE and CgA/SgIII, to the transfer to CgA/SgIII complexes. This view is compatible with a condensing ISG actively retaining hormones in the maturing granule.

1.7 Rationale and Specific Aims

In the case of proglucagon, it is organized into several constituent hormone sequences. Each of these sequences contains at least one putative sorting signal, in addition to several dibasic sites that flank each major hormone domain. At present, neither the existence of sorting signals, nor the existence of sorting receptors involved in proglucagon trafficking, have been investigated. Based on the discussion of signals and receptors above, it is likely that such a mechanism should exists, especially considering the need for α- and L-cells to produce distinct subsets of proglucagon-derived hormones.

I hypothesize that specific sorting signals within the hormone domains of proglucagon determine its intracellular trafficking through specific interaction of a receptor to facilitate efficient storage within granules and processing to mature hormones.

Aim 1: To identify sorting receptors and sorting signals that mediate the sorting of proglucagon to secretory granules.

A) I will identify candidate sorting signals using the predicted 2D and 3D structure of proglucagon, and determine whether these structures are necessary for sorting by using site-directed mutagenesis to specifically disrupt each structure.

B) I will determine whether a known prohormone sorting receptor, carboxypeptidase E, is necessary for sorting proglucagon to granules by quantifying intracellular trafficking in the absence or presence of transfected CPE. The mechanism will be confirmed by using siRNA-mediated depletion of CPE from an α and L cell line, and measuring the degree of regulated secretion by RIA.
Aim 2: To characterize the molecular properties of sorting signals within the peptide domains of proglucagon.

A) I will determine which hormone domains contain sorting information by making specific genetic constructs and following intracellular trafficking and measuring regulated secretion.

B) To determine the biochemical nature of the sorting signal using site-directed mutagenesis.

C) I will determine the temporal sequence of proglucagon processing and sorting.

Aim 3: To develop and validate a quantitative co-localization algorithm to determine in which sub-cellular compartment a protein of interest is targeted.

I will validate the utility of several common quantitative immunofluorescence co-localization techniques by conducting computer simulation experiments. I will create a work-flow for conducting quantitative co-localization analyses for future immunofluorescence studies, then compare this method with more complex co-localization analytic methods.
1.8 References


43


incretin response after a mixed meal is associated with insulin resistance in nondiabetic men. *Diabetes Care* 24, 1640–5


47


Chapter 2

The sorting of proglucagon to secretory granules is mediated by carboxypeptidase E and intrinsic sorting signals

Statement of Author Contributions

Mr. Leonardo Guizzetti validated and conducted quantitative immunofluorescence co-localization to detect degrees of intra-cellular protein trafficking, and performed 3D protein modeling.

Ms. Rebecca McGirr conducted siRNA transfections and fluorescence microscopy.

Statement of Copyright

This chapter is reformatted from its original publication as an article in the Journal of Endocrinology. The citation for the original work is:
The * denotes co-first authorship of the original manuscript.

2.1 Introduction

Ingested nutrients stimulate the secretion of glucoregulatory hormones from both the pancreas and the gut to exert control over glucose homeostasis. Proglucagon is a polypeptide prohormone expressed in pancreatic α cells, intestinal L cells, and a select subset of neurons in the brainstem. It is processed in a tissue-specific manner to produce distinct sets of constituent hormones. Expression of prohormone convertase (PC)-2 in α cells produces glucagon, the major pancreatic hormone product (1, 2). Glucagon is secreted in response to hypoglycemia and is the major counter-regulatory hormone to insulin, activating hepatic glycogenolysis and gluconeogenesis to maintain euglycemia
In contrast, PC1/3 in intestinal L cells and neurons produces oxyntomodulin, GLP-1 and GLP-2 (4–7). These nutrients are secreted in response to nutrient ingestion, and each has a distinct effect to control glucose homeostasis within a nominal range (8). GLP-1 enhances post-prandial insulin secretion, whereas GLP-2 is an intestinotrophic hormone that increases intestinal blood flow and nutrient assimilation (9). Oxyntomodulin is an anorectic hormone, decreasing food intake and increasing satiety and energy expenditure (10). Therefore, nutrient status potently stimulates secretion of the proglucagon-derived peptides so they may exert tight control over glucose metabolism and nutrient homeostasis.

In order for endocrine and neuroendocrine tissues to robustly respond to changes in nutrient status, neuropeptides and hormones are directed along the regulated secretory pathway. This biosynthetic pathway is distinct from unregulated, constitutive secretion, and maintains specialized storage organelles called secretory granules. Two hypotheses have been proposed for the targeting of hormones to granules. In the “sorting-by-retention” model, hormones condense into large aggregates in the trans-Golgi network (TGN) that are retained in nascent secretory granules, while the maturing granule selectively removes other proteins. Alternatively, prohormones in the “sorting-for-entry” model specifically bind to either membrane-bound sorting receptors or membrane lipids, by a structural sorting signal. Once prohormones enter immature granules, they will undergo endoproteolysis to yield bioactive hormones, which are then condensed and stored in the granules as an electron-dense core, characteristic of their appearance by electron microscopy. Secretion can then be stimulated in a Ca\(^{2+}\)-dependent manner by an appropriate secretagogue (e.g., nutrient ingestion). Based on the “sorting-for-entry” model, we can begin to explore the molecular mechanism of the sorting signals of proglucagon.

The known sorting signals of prohormones range in their molecular complexity. These include 1) dibasic amino acid sequences, found in pro-neuropeptide Y (11); 2) amphipathic α-helices (12) such as those in pro-somatostatin (13), pro-CART (14), VGF (non-acronymic) (15), pro-renin (16), pro-gastrin (17), and pro-neurotensin (18); and 3)
acidic residues found on disulphide loops found within POMC (19) and proinsulin (20). An α-helix predominates the structures of glucagon (21) and GLP-1 (22) based on their known X-ray crystal structures, and GLP-2 (23) by the NMR solution structure, suggesting that proglucagon contains significant helical content within its hormone-encoding domains. Proglucagon also contains five proteolytic dibasic sites located at hormone domain junctions (Figure 1). The α-helix of glucagon contains an embedded dibasic site, a motif that targets VGF to granules (15). Lastly, different sorting motifs may synergistically enhance granule targeting (24). Therefore proglucagon trafficking may rely on information encoded in several putative sorting signals, located in the α-helices of mature glucagon and glucagon-like peptide domains, and multiple dibasic cleavage sites.

Figure 1. Post-translational processing of proglucagon to its derived peptides. A schematic representation of proglucagon in which each domain has been colour-coded, and the amino acid positions of key processing events have been identified. The amino acid numbering is relative to the first N-terminal amino acid of proglucagon (lacking the signal peptide). The major hormone products produced from tissue-specific processing by their respective prohormone convertase (PC) are shown at bottom. In pancreatic islet α cells, the major products are glucagon and major proglucagon fragment (MPGF). In enteroendocrine L cells and in the brain, the major products are GLP-1 (1-37), GLP-1 (7-37), GLP-2, oxyntomodulin and glicentin.
In order for hormones to be targeted to granules, a sorting receptor must recognize and bind the sorting motif. Carboxypeptidase E (CPE) is a granule-resident enzyme that cleaves C-terminal basic amino acids. It is known to be a sorting receptor for both pro-opiomelanocortin (POMC) and pro-insulin by interaction of two acidic residues of the sorting signal, with complementary basic residues on the receptor surface (19, 20, 25). Individually, PC1/3, PC2 and CPE are co-targeted to granules with hormones. PC1/3 co-targets pro-renin to granules in order to produce renin in AtT-20 cells (26). Deletion of the C-terminal tail of PC1/3 that tethers to lipid rafts of granule membrane mis-routes pro-renin for constitutive secretion (27, 28). However, no such evidence exists to suggest that PC2 directly impairs granule targeting, though this hypothesis was not specifically investigated. Even in α cells isolated from PC2 null islets (2) and a derivative α cell line (29) show no defect of regulated secretion (aside from the obvious defect in processing). A strain of mice called CPE$^{fat/fat}$ harbors a mutant S202P-CPE, that is enzymatically inactive (30). CPE$^{fat/fat}$ mice constitutively secrete POMC (25) and proinsulin, becoming diabetic, and the granules of beta cells lack the characteristic electron-dense cores (30). Similar results for pro-insulin were observed in CPE knock-out mice (31). Because these enzymes are co-targeted to granules with their substrate hormones, and CPE has been shown to specifically interact with the POMC sorting motif, the possibility exists that CPE, PC1/3 and PC2 are candidate receptors for proglucagon sorting.

In the present study, we investigated the requirement of CPE as a sorting receptor for proglucagon. We hypothesized that in the absence of CPE, sorting of proglucagon to granules would be greatly diminished. Furthermore, we also investigated whether PC1/3 or PC2 co-expression would have an effect on sorting of proglucagon. We found that CPE alone plays a role in trafficking proglucagon to granules and that this may be specific to α cells. We also investigated the hypothesis that proglucagon contains sorting signals by focusing on the glucagon α-helix and its embedded dibasic site. By finding that granule targeting was impaired by specific mutation of these two structures, we provide the first characterization of sorting signals contained within the glucagon α-helix of proglucagon.

In order for hormones to be targeted to granules, a sorting receptor must recognize and bind the sorting motif. Carboxypeptidase E (CPE) is a granule-resident enzyme that cleaves C-terminal basic amino acids. It is known to be a sorting receptor for both pro-opiomelanocortin (POMC) and pro-insulin by interaction of two acidic residues of the sorting signal, with complementary basic residues on the receptor surface (19, 20, 25). Individually, PC1/3, PC2 and CPE are co-targeted to granules with hormones. PC1/3 co-targets pro-renin to granules in order to produce renin in AtT-20 cells (26). Deletion of the C-terminal tail of PC1/3 that tethers to lipid rafts of granule membrane mis-routes pro-renin for constitutive secretion (27, 28). However, no such evidence exists to suggest that PC2 directly impairs granule targeting, though this hypothesis was not specifically investigated. Even in α cells isolated from PC2 null islets (2) and a derivative α cell line (29) show no defect of regulated secretion (aside from the obvious defect in processing). A strain of mice called CPE$^{fat/fat}$ harbors a mutant S202P-CPE, that is enzymatically inactive (30). CPE$^{fat/fat}$ mice constitutively secrete POMC (25) and proinsulin, becoming diabetic, and the granules of beta cells lack the characteristic electron-dense cores (30). Similar results for pro-insulin were observed in CPE knock-out mice (31). Because these enzymes are co-targeted to granules with their substrate hormones, and CPE has been shown to specifically interact with the POMC sorting motif, the possibility exists that CPE, PC1/3 and PC2 are candidate receptors for proglucagon sorting.

In the present study, we investigated the requirement of CPE as a sorting receptor for proglucagon. We hypothesized that in the absence of CPE, sorting of proglucagon to granules would be greatly diminished. Furthermore, we also investigated whether PC1/3 or PC2 co-expression would have an effect on sorting of proglucagon. We found that CPE alone plays a role in trafficking proglucagon to granules and that this may be specific to α cells. We also investigated the hypothesis that proglucagon contains sorting signals by focusing on the glucagon α-helix and its embedded dibasic site. By finding that granule targeting was impaired by specific mutation of these two structures, we provide the first characterization of sorting signals contained within the glucagon α-helix of proglucagon.
2.2 Methods

2.2.1 Cell Culture, plasmids and transfection

Wild-type Neuro 2a (N2a wt) cells and Neuro 2a cells stably transfected with mouse CPE (N2a-CPE, clone 17) were obtained from Dr. Y. P. Loh (Bethesda, MD, USA). Cells were maintained in DMEM containing 10% FBS and stable transfectants were maintained in media containing 400 µg/ml G418. N2a wt cells were also stably transfected with the enzymatically inactive form of CPE, E300Q (32) using Lipofectamine 2000 (Life Technologies). Stable transfectants were selected in 800 µg/ml G418, pooled and maintained in 400 µg/ml G418. To examine the role of CPE in sorting proglucagon, N2a wt, N2a-CPE and N2a-E300Q cells were transfected with hamster pre-proglucagon (in pcDNA 3.1; a kind gift from Dr. D. F. Steiner, Chicago, IL, USA). To examine the roles of other prohormone convertases, both N2a wt and N2a-CPE cells were transiently transfected with plasmids encoding mouse PC1/3 and PC2 (kind gifts from Dr. N. G. Seidah, Montréal, QC, Canada). To determine whether proglucagon co-localizes to granules of N2a-wt cells, we constructed a proglucagon-EGFP plasmid. Pre-proglucagon cDNA was PCR amplified to add the KpnI (proG-EGFP forward primer, 5’-GGT ACC ATG AAG AAC ATT TAC ATT GTG G-3’) and BamHI (proG-EGFP reverse primer, 5’-GGA TCC GTT TTC TTG TCA GTG ATT TTG GT-3’) restriction sites (underlined) in frame with EGFP (bolded nucleotides). Using PCR amplification and restriction digest, pre-proglucagon was cloned into the pEGFP-N1 backbone (Clontech Labs, Takara Bio) to add a C-terminal EGFP tag. To determine possible sorting signals, the sequence of proglucagon was mutated independently at three sites: at the processing site, R70K71; at the dibasic site within glucagon, R17R18; and at two leucines that were postulated to flank the helix structure within glucagon, L14 and L26. The processing site mutant, K71Q, was a kind gift from Dr. D. F. Steiner (Chicago, IL, USA). We generated the mutation at the dibasic site, R18Q (forward primer sequence 5’-AAA TAC CTG GAC TCC CGC TCC GCT TTC TTG TCA GTG ATT TTG GT-3’), and the double leucine-to-proline mutation was done in two steps; L14P was made using forward primer 1, 5’-TAC AGC AAA TAC CGG GAC TCC CGC CGA GCC GCC-3’, and L26P was subsequently generated using forward primer 2, 5’-CAA GAT TTT GTG CAG TGG CCG ATG AAC ACC ACC-3’. Bold sequences in primers indicate site of mutation. All site-directed mutagenesis reactions
were carried out using the QuikChange™ II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA) and results were confirmed by sequencing at the London Regional Genomics Facility, University of Western Ontario.

2.2.2 Depletion of Carboxypeptidase E by siRNA Knockdown
Specific knock-down of CPE in either alphaTC1-6 or GLUTag cells was achieved by synthesizing siRNA using the Silencer siRNA Construction Kit™ (Life Technologies, Carlsbad, CA, USA). The kit included a negative control, Ambion Silencer Negative Control #1, a sequence with no known targets in the mouse genome. Initial targets were selected from the mouse CPE mRNA sequence (GenBank™ accession #NM_013128.1) as generated by Ambion's online algorithm, and subsequently screened using a BLAST search to identify seven candidate sequences with no homology to known mouse mRNAs. All seven candidate siRNA sequences were synthesized, screened in both cell lines and validated for the extent of CPE depletion by western blot using the CPE 4-5 antibody that detects the N-terminal domain of CPE (a gift from Dr. Y. P. Loh, Bethesda, MD, USA) ([Figure 7A](#)). A single target was chosen, CPE73, spanning nucleotides 1408-1428 (sense strand: 5'-C CCT GCT GTT GGG GTG GAC TT), for its significant and reproducible ability to decrease CPE levels in both cell lines ([Figure 7A](#)). CPE73 was used in subsequent depletion experiments, in which 200 nM concentrated CPE73 or control siRNA were transfected using Oligofectamine (Life Technologies) as per manufacturers instructions for 72 h prior to secretion experiments or fixation for immunofluorescence microscopy.

2.2.3 Western Blot
For western blot analysis, all cells were grown in six-well plates, lysed and protein extracted. Protein concentration was quantified by densitometry and 20 µg of total protein was separated on 12% SDS-PAGE gels as previously described (33). Proteins were transferred onto nitrocellulose membranes and probed using specific antisera against CPE (CPE 4-5), or against PC1/3 or PC2 (kind gifts from Dr. N. G. Seidah, Montréal, QC, Canada). Bands were visualized using Super Signal West Pico chemiluminescent substrate (Thermo-Fisher Scientific, Toronto, ON, Canada) and bands were quantified by densitometry as described (33). Secreted CPE was detected by media
collection and lyophilized. The concentrated residue was resuspended in sample buffer and loaded entirely into one well. As a result of concentration, these samples migrated more slowly through the gel and had a greater apparent molecular weight ($M_r$) of 60 kDa.

### 2.2.4 Secretion Experiments

AlphaTC1–6 cells (a kind gift from Dr. C. B. Verchere, Vancouver, BC, Canada) were maintained in DMEM containing 15% horse serum and 2.5% FBS. GLUTag cells (a kind gift from Dr. D. J. Drucker, Toronto, ON, Canada) were grown in low-glucose DMEM containing 10% FBS. For stimulated secretion experiments, αTC1–6 cells were seeded in 24-well plates in replicates of six and pre-incubated in Hank's buffered salt solution (HBSS; 138 mM NaCl, 5.33 KCl, 4.00 NaHCO₃, 1.26 CaCl₂, 0.50 MgCl₂, 0.44 KH₂PO₄, 0.41 MgSO₄, 0.30 Na₂HPO₄) containing 25 mM glucose for 1 h. Cells were then incubated in HBSS containing 1 mM glucose and 10 µm each of forskolin and 3-isobutyl-1-methylxanthine (IBMX) for 1 h. GLUTag cells were incubated in low-glucose (5.5 mM) DMEM +0.5% FBS without (basal) or with (stimulated) 10 µm each of forskolin and IBMX for 2 h. After all secretion experiments, media were collected and trifluoroacetic acid (TFA) was added to 0.1%. Cells were rinsed twice in HBSS, and scraped in 1 mL homogenisation buffer (1 M HCl, 1 M formic acid, 1% (v/v) TFA, 1% (w/v) NaCl). The cells were sonicated in one 12 s burst, centrifuged and the supernatant was collected, and both media and cell extracts were passed through a Sep-Pak C18 reverse-phase cartridge to elute proglucagon-derived peptides as previously described (4, 5). Glucagon and GLP-1 content were assessed by RIA using the glucagon and GLP-1 RIA kits (Linco/EMD Millipore, Bedford, MA, USA). The antibody provided in the glucagon RIA kit specifically detects the C-terminal end of glucagon and does not cross-react with any other proglucagon-derived peptides.

### 2.2.5 Immunofluorescence and image acquisition

For immunofluorescence, cells were grown on glass coverslips, as described (33). The primary antibodies used were directed against glucagon (rabbit; 1:1000; Bachem/Peninsula Laboratories, Torrance, CA, USA), and either the Golgi marker p115 (mouse; 1:50; Transduction Laboratories, San Jose, CA, USA), or the secretory granule marker chromogranin A (rabbit; CgA; 1:50; Abcam). Additionally, primary antibodies
against insulin (guinea pig; 1:1000), GFP (rabbit; 1:250; A. v. peptide, Clontech Laboratories, Takara Bio, Mountain View, CA, USA), or transcription factor PDX-1 (rabbit; 1:50) were used in the validation of the Pearson's correlation coefficient image analysis (described below). Coverslips were blocked in 10% goat serum and 1% bovine serum albumin in PBS. The anti-glucagon antibody was raised against the entire sequence of glucagon, and thus recognises unprocessed proglucagon, glicentin, oxyntomodulin and glucagon. To examine CPE immunoreactivity, slides were incubated with the CPE 7–6 antibody (raised in rabbit; kind gift from Dr. Y. P. Loh, Bethesda, MD, USA), which recognises the C-terminus of CPE. Alexa 488-IgG (Invitrogen) was used to visualise the glucagon or CPE antibody, and Alexa 594-IgG for the p115 or CgA antibody.

Cells were visualized using an Olympus IX81 widefield fluorescence microscope and images were acquired using In Vivo software. Ten optical sections were acquired at 0.2 µm steps to cover the z-axis field of the cells, using a 60x oil immersion objective lens. Image stacks were then processed by the blind (3D) deconvolution algorithm provided in Image-Pro Plus software (Media Cybernetics Inc., Rockville, MD, USA).

2.2.6 Image analysis using Pearson's Correlation Coefficient

Image analyses were conducted using FIJI (FIJI is just ImageJ) (version 1.46h) (34), a distribution of ImageJ (NIH, Bethesda, MD, USA). Co-localization analyses were conducted using the Co-localization 2 plugin in FIJI. During the course of image analysis, it became apparent that there were bugs in the quantification of co-localization. I fixed the bugs and contributed the code to the FIJI/ImageJ project. Details of the bug fixes can be found in Appendix 1. Regions of interest were drawn around single- or multi-cell bodies from pseudo-coloured red and green fluorescent images to exclude as much background as possible. The segmented region was defined by a binary mask and applied to both image sets. Using raw, unprocessed images presents the possibility that any identified co-localization could instead be a false-positive. To ensure false-positives were not included in subsequent analyses, each red/green image pair was first screened using the Coste's significance threshold test (35). Costes' algorithm scrambled one image to test for spurious co-localization and simultaneously determines a threshold intensity for each colour, above which pixels are considered to be statistically correlated. Paired
red and green pixels were then used to quantify co-localization using the Pearson's correlation coefficient (PCC) ($r$), according to:

$$r = \frac{\sum_i ((S_{C1j} - S_{1av}) \times (S_{C2j} - S_{2av}))}{\sqrt{\sum_i ((S1_i - S_{1av})^2) \times \sum_i (S2_i - S_{2av})^2}}$$

Here, the $i$-th and $j$-th pixels represent the respective regions of interest, $S_{C1}$ and $S_{C2}$, of co-localized pixel intensities in each colour channel, and all signal intensities, regardless of co-localization, are denoted by $S1$ and $S2$ or their averages, $S_{1av}$ and $S_{2av}$. Importantly, the numerator is dependent on an unbiased estimation of co-localized pixels as chosen by the Costes' algorithm, while the denominator uses all pixels within respective regions of interest to establish a signal-to-background ratio. This measurement also has the desirable properties of being invariant to the region of interest shape, and each pair of images is normalized to their own mean fluorescence intensities. Calculated correlation coefficients were treated as a single experimental dataset and differences were tested using a one-factor ANOVA omnibus test, and Tukey's HSD post-hoc test using a significance threshold of $\alpha = 0.05$. Group sample sizes ($n$) varied between 10-25, with adequate power to detect the effect sizes shown here. The Pearson's correlation coefficient method was validated in INS-1 832/13 cells using markers known for a low degree of co-localization (insulin and PDX-1) and a high degree of co-localization (insulin and CgA).

### 2.2.7 De novo Protein Modeling

At present no crystal or solution structure of the full proglucagon molecule exists. To determine whether such a model may provide mechanistic insight to the sorting of proglucagon, the full-length proglucagon polypeptide sequence was submitted to the Robetta 3D modelling web server for de novo 3D protein structure prediction (36). Robetta uses the Ginzu model (37) to find match sub-domains against known homologous structures that have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb/). Proglucagon was parsed into two sub-domains, each modeled using the incretin hormone, glucose-dependent insulinotropic peptide (GIP; PDB:2B4N and PDB:2OBU). The N-terminal domain (proglucagon 1-77; PDB:2B4N) had a position-specific iterative BLAST (PSI-BLAST) (38) confidence score of 5.40; the C-terminal domain (proglucagon 78-158) (PDB:2OBU) had a HH-Search (39) confidence
of 2.33. The Robetta server assembled both sub-domain models for final folding and side-chain packing, to produce five full-chain 3D structures.

2.3 Results

2.3.1 Pearson's correlation coefficient accurately reflects degree of co-localization

If proglucagon sorting to granules is sensitive to changes in CPE expression, we should be able to detect a change in sub-cellular distribution at steady state using quantitative co-localization. We used Pearson's correlation coefficient (PCC) to quantify the degree of co-localization. The degree of co-localization correlates with sorting efficiency to granules, and therefore we used co-localization as a measure of sorting efficiency of proglucagon. The PCC measurement has previously been used to describe the co-localization of the neuropeptide secretogranin II to granules in both PC12 and chromaffin cells (40). Further, the PCC is thought to be an accurate method of co-localization because it correlates normalized fluorescence intensities on a pixel-by-pixel basis (41). (A more detailed discussion on the quantitative techniques used in co-localization and common misunderstandings of PCC are addressed in Chapter 4.) Here we describe the qualitative degree of co-localization alongside quantitative measurements of Pearson's correlation coefficient. Because the explanation of qualitative images is subjective by nature, Zinchuk and colleagues have proposed a method to translate qualitative descriptions to quantitative ranges¹ of co-localization parameters (42). Using these mappings adds consistent quantitative meanings to these subjective qualifiers. Before using the PCC to measure proglucagon sorting efficiency, we first validated the method by measuring two proteins that are known to have either a low or high degree of co-localization with each other.

To verify the degree of co-localization between two sets of proteins in their native cells, we used the insulin-producing rat INS-1 832/13 beta cell line. These cells efficiently store a large amount of insulin in secretory granules, which is co-stored with a granule matrix

¹ As described by Zinchuk, the “fuzzy” ranges for Pearson's correlation coefficient (auto-threshold) map as follows. **Very weak**: -1.0 ~ -0.27; **Weak**: -0.26 ~ 0.09; **Moderate**: 0.10 ~ 0.48; **Strong**: 0.49 ~ 0.84 and **Very strong**: 0.85 ~ 1.0.
protein, chromogranin A (CgA). Therefore we expect a high degree of steady-state co-localization between insulin and CgA in the Golgi and granules. Both insulin and CgA immunoreactivity were detected in a para-nuclear structure, consistent with the Golgi, and in punctate structures along the cell processes, consistent with secretory granules (Figure 2A). PCC analysis on these images showed a very strong co-localization (Figure 2B), as expected. In contrast, the beta-cell specific pdx-1 transcription factor is active in the nucleus, and should have a low degree of co-localization with insulin. The subcellular distribution of insulin immunoreactivity was the same, while nearly all pdx-1 was localized to the nucleus (Figure 2A). As expected, insulin and pdx-1 were not co-localized with each other, but the PCC analysis showed that it was just barely “moderately” co-localized by using the numeric mappings (PCC = 0.49, on the border between weak and moderate co-localization) (Figure 2A). Correspondingly, there is very low overlap after inspection of the pdx-1 immunofluorescence signal (Figure 2B). Insulin was very significantly co-localized to granules (p<0.001) than it was with pdx-1 in the nucleus. This high degree of insulin and CgA co-localization by Pearson's has also recently been demonstrated in rat insulinoma INS-1 cells (43). Therefore we confirm the specificity of Pearson's correlation coefficient to measure co-localization.
Figure 2. Validation of Pearson's correlation coefficient. (A) INS-1 832/13 cells were immunodetected using antibodies against insulin (green) and pdx-1 (red) or chromogranin A (CgA, red). Insulin and CgA are located in the Golgi and secretory granules (punctate signal located in cell processes), while pdx-1 was localized to the nucleus. Representative images are shown. (B) Mean PCC of pdx-1 and insulin (light grey bar) and CgA and insulin (dark grey bar). As expected, there is weak correlation between insulin and pdx-1, and very strong co-localization between CgA and insulin. * $p<0.01$. Values are mean ± SEM ($n=10$).
2.3.2 Proglucagon is sorted to granules in the presence of Carboxypeptidase E

We next investigated the role of CPE on the sorting efficiency of proglucagon. CPE immunoreactivity was not detected by western blot in wild-type N2a cells (N2a-wt), and could only be detected after stable transfection of CPE (N2a-CPE) (Figure 3A). In the absence of CPE (N2a-wt cells), proglucagon immunoreactivity was mostly localized to the cis/medial-Golgi, as indicated by p115 co-immunostaining (Figure 3B), and there was a strong co-localization between proglucagon and p115 by PCC analysis (Figure 3C). In contrast, in N2a-CPE cells which over-express CPE, proglucagon immunostaining shifted toward being stored in punctate structures in the cell processes at steady state, consistent with storage in granules (Figure 3B). PCC analysis showed moderate co-localization in the presence of CPE through a significant decrease in co-localization (p<0.01) (Figure 3C). Therefore, the presence of CPE efficiently increased proglucagon sorting to granules, as reflected by a loss of co-localization with the early Golgi marker, p115, and punctate appearance of proglucagon in granules.

In order to validate that proglucagon is co-localized to granules, we transfected N2a-CPE with proglucagon-EGFP (enhanced green fluorescent protein). Cells were co-immunostained for EGFP to detect proglucagon, and CgA to detect granules (as in INS-1 cells) (Figure 4). As expected, proglucagon immunoreactivity very strongly co-localized (r = 0.84 ± 0.01) with CgA-positive punctate structures in the cell processes (Figure 4). Therefore, N2a-CPE cells efficiently sort proglucagon to granules.
Figure 3. CPE plays a role in the sorting of proglucagon. (A) Western blot analysis of cell extracts from Neuro2A cells lacking CPE expression, and after transfection of wt or E300Q mutant CPE. (B) Full-length proglucagon was transfected into N2a cells, either alone (– CPE), or co-transfected with CPE or E300Q mutant CPE. Cells were immuno-
stained using antibodies against glucagon (green) and the Golgi protein p115 (red). Representative images are shown. Scale bar represents 10 µm. (C) FIJI/ImageJ software was used to determine the fluorescence intensity co-variance (PCC) of proglucagon and p115. Proglucagon showed a significant decrease in mean correlation with p115 when co-expressed with CPE, and a significant increase when co-expressed with the E300Q mutant CPE. Values are means ± S.E.M. (n=10–25). *p<0.01 vs – CPE; # p<0.01 vs both cells with and without CPE.

2.3.3 Proglucagon sorting depends upon properly sorted CPE

In an effort to resolve whether the effect of CPE on proglucagon sorting was influenced by its enzymatic activity or its ability to sort to granules, we made stable transfectants of N2a-wt cells with the inactive Glu300Gln mutant of CPE (Figure 3A). If CPE acted as a sorting receptor for proglucagon, we hypothesized that proglucagon sorting efficiency should correlate with CPE trafficking. In E300Q CPE cells (N2a-E300Q), proglucagon immunoreactivity was localized to the Golgi, as in the cells lacking CPE (Figure 3B). Co-localization of proglucagon with p115 was very strong (Figure 3C), as reflected by a significantly increased PCC (p<0.01) when compared to that of N2a-wt, indicating a

---

**Figure 4.** Proglucagon is sorted to secretory granules in N2a cells expressing CPE. N2a cells were transfected with CPE and proglucagon-EGFP, and immunostained using antibodies directed against EGFP (green) and CgA (red), a granule resident protein. Mean Pearson's correlation coefficient was significant (PCC=0.84 ± 0.01, n=6) and showed very strong co-localization.
greater abundance of proglucagon in the Golgi. Furthermore, co-localization analysis of proglucagon with p115 in N2a-wt cells was significantly greater than in N2a-CPE ($p<0.01$) (Figure 3C), suggesting that the presence of CPE significantly enhanced sorting efficiency of proglucagon. This was interesting because it suggested that the E300Q mutation impaired the post-Golgi sorting of proglucagon. In support of this hypothesis, E300Q-CPE was robustly detected in the Golgi (Figure 5A) and could not be detected in media (Figure 5B). On the other hand, CPE was detected along the tips of cell processes (Figure 5A) could be detected in media (Figure 5B), consistent with storage in granules. This suggests that E300Q CPE exit from the Golgi was also impaired, and is not properly sorted in N2a cells.

**Figure 5.** The E300Q mutant CPE is not sorted efficiently to secretory granules. (A) In cells expressing wild-type CPE, punctate staining along the tips of the cell processes indicates presence in secretory granules. In cells expressing E300Q, CPE
immunofluorescence largely coincides with that of p115. (B) Wild-type CPE is secreted from N2a cells where a portion is cleaved from the granule membrane resulting in a smaller, soluble form of secreted CPE. The E300Q mutant CPE is not present in the media, further indicating that is it not transiting through the secretory pathway in N2a cells. The apparent increase in molecular weight of the secreted form is due increased salt content of the lyophilized sample resulting in slower migration through the gel, as stated in Methods.

2.3.4 The effect of PC1/3 and PC2 on proglucagon sorting

Since proglucagon is co-targeted to granules with either PC1/3 (in L cells) or PC2 (in α cells), we also tested the hypothesis that PC1/3 or PC2 were sufficient for proglucagon sorting. We were also interested in whether the co-expression of CPE with either PC1/3 or PC2 could enhance CPE-mediated proglucagon sorting. N2a-wt cells were a convenient system for this experiment because they lack endogenous PC1/3 and PC2 (Figure 6A). We transfected N2a-wt cells with either PC1/3 or PC2 alone, or in combination with wild-type CPE. When PC1/3 was transfected alone, it was detected as a single 84 kDa pro-form, while co-expression with CPE resulted in both an 84 kDa form and an active 66 kDa form (Figure 6A). The expression pattern of PC2 remained unchanged (Figure 6A). Proglucagon immunoreactivity was mostly localized to the Golgi when expressed either with PC1/3 or PC2 alone (Figure 6B), and the Pearson's analysis showed that there were no significant differences from the N2a-wt cells lacking CPE (Figure 6C). The co-localization of proglucagon and p115 was unaffected by PC2, with or without CPE (Figure 6B,C). Proglucagon was localized in the Golgi similarly to the absence of CPE (Figure 6B), and Pearson's analysis showed similarly poor co-localization (Figure 6C). Addition of CPE mostly restored proglucagon sorting, as small amounts were seen in granules (Figure 6B) and Pearson's analysis showed improved co-localization ($p<0.05$) compared to PC1/3 alone, though it was not restored to the same level as CPE alone. These results are consistent with CPE enhancing proglucagon sorting alone.
Figure 6. The presence of PC1/3 or PC2 does not enhance sorting of proglucagon. (A) Western blot analysis of PC1/3 (upper panels) and PC2 (lower panels) expression after before and after transfection into N2a cells. Wild-type N2a cells have no detectable expression of either PC. In cells lacking CPE, PC1/3 immunoreactivity is present as a
single 84 kDa band, while in the presence of CPE, it is mostly present as a 66 kDa band. PC2 is present as a 66 kDa band with or without CPE. (B) Full-length wild-type proglucagon was transfected into N2a cells, either with PC1/3 or PC2 alone or in combination with CPE. Cells were processed using antibodies against glucagon (green) and the Golgi protein p115 (red). Representative images are shown. Scale bar represents 10 µm. (C) Mean correlation of proglucagon and p115 fluorescence in cells expressing PC1/3 or PC2, alone or in combination with CPE, was not significantly different from cells not expressing CPE. Values are means ± S.E.M. (n=10–25). *p<0.01 compared with − CPE, # p<0.01 compared with + PC1/3 alone.

2.3.5 CPE affects proglucagon sorting in α cells, but not L cells
To determine if the effect of CPE of proglucagon sorting is relevant to α or L cell physiology, we conducted knockdown studies using alphaTC1-6 cells, an α cell line derived from mouse glucagonoma, and GLUTag cells, a mouse L cell line derived from mouse enteroendocrine tumours. Out of a screen for CPE-specific siRNAs, we selected CPE73 for its reliable and reproducible ability to deplete CPE (see Methods; Figure 7A). To confirm whether CPE had a physiologic effect, we conducted secretion assays on alphaTC1-6 cells for glucagon, and GLUTag cells for mature GLP-1. Using CPE-73 siRNA, we significantly reduced CPE expression by 61±3% in alphaTC1-6 cells (n=3, p<0.05) (Figure 7B) and by 74±7% in GLUTag cells (n=3, p<0.05) (Figure 8A). Knockdown of CPE in alphaTC1-6 significantly increased basal glucagon release (n=6, p<0.05) with a corresponding decrease in cell content of glucagon (n=6, p<0.05) (Figure 7C,D). The transfection efficiency of alphaTC1-6 cells, as assessed by EGFP transfection, was 36.6 ± 3.7% (calculated from 10 images, each with ~50 cells). However, stimulated secretion by forskolin/IBMX was lost with reduced CPE (n=6, p<0.01) (Figure 7C,D). In contrast, no differences in regulated secretion of GLP-1 were detected in GLUTag cells using forskolin/IBMX stimulation (n=6, p<0.001) (Figure 8B). CPE appears to specifically affect glucagon secretion from α cells, while L cells remain unaffected.
Figure 7. CPE depletion affects regulated secretion of glucagon from alphaTC1–6 cells. (A) Screen of siRNA directed against CPE. Seven siRNA were synthesized against non-homologous targets of mouse CPE, transfected into both cell lines, and screened for the extent of depletion by western blot using the N-terminally directed CPE4-5 antibody. A siRNA directed against GAPDH and a mock transfection without siRNA were used as controls. Equal protein lysates were separated on SDS-PAGE as determined by BCA.
assay. The greatest degree of depletion was achieved using siRNA-73, and so this was used for subsequent experiments. (B) Western blot analysis shows siRNA-mediated knockdown of CPE expression in alphatC1–6 cells. (C and D) AlphaTC1–6 cells were treated with transfection agent alone (mock), or 200 nM of control siRNA or siRNA-73 against CPE. (C) There was a significant increase in secretion in response to 10 mM forskolin/IBMX treatment in mock- and control- transfeected cells. In cells transfected with siRNA CPE, basal secretion was significantly higher than in control cells, and there was no response to stimulation with forskolin/IBMX. * p<0.001 compared with control transfected basal secretion. Values are means ± S.E.M. (n=6). (D) Glucagon cell content decreased in cells transfected with siRNA-73 compared with control siRNA. * p<0.05, ** p<0.01 compared with control siRNA values. Results are representative of four independent experiments.

Figure 8. CPE depletion has no effect on regulated secretion of GLP-1 from GLUTag cells. (A) Western blot analysis shows siRNA-mediated knockdown of CPE expression in GLUTag cells. (B) GLUTag cells were treated with transfection agent alone (mock) or 200 nM of either control siRNA or siRNA-73 against CPE. There was robust stimulated secretion in response to 10 mM forskolin/IBMX treatment in all groups. Values are means ± S.E.M. (n=6). *** p<0.001 compared with basal secretion. Results are representative of four independent experiments.
2.3.6 Proglucagon contains a sorting signal within the glucagon sequence

Sequence analysis of proglucagon revealed that glucagon contains a substantial α-helical structure, which contains two embedded basic arginines (R18R19). If these structures act as sorting signals, then sorting should be impaired by their specific mutation. We generated independent mutations to disrupt either the dibasic site by an arginine to glutamine mutation (R18Q), or the α-helix by dual leucine-to-proline kinks (L14P,L26P), and expressed these mutant proglucagons in the presence of CPE. Each mutant was more abundant in the Golgi (Figure 9A). Compared to wild-type proglucagon, both the L14P,L26P mutant \( (p<0.01) \) and R18Q dibasic mutant \( (p<0.05) \) were strongly and more significantly co-localized with p115, and similar to each other in their degree of co-localization (Figure 9B). These data suggest that the specific α-helix of glucagon encodes the information for proper sorting of proglucagon.

2.3.7 Initial processing of proglucagon at K70R71 enhances sorting to granules

The precise order of when proglucagon undergoes cleavage at the K70R71 site and initial sorting to granules is not known. It is possible that this cleavage event makes the sorting signal accessible for more efficient sorting, or that sorting occurs first, and proglucagon is subsequently processed. While PC1/3 and PC2 are the physiologically important convertases in the production of proglucagon-derived peptides, neither are expressed in Neuro2a cells and therefore would not cleave at this site. A related enzyme, furin, is expressed in all cells (44), and has been shown to process proglucagon at the K70R71 inter-domain site (45). As furin is trafficked from the TGN to immature granules within the secretory pathway, cleavage of proglucagon by furin in these compartments may facilitate entry to granules. To examine this hypothesis, we blocked furin-mediated cleavage by mutation of the K70R71 site to R71Q. The R71Q mutant was mainly localized in the Golgi (Figure 9A), and Pearson's analysis showed that co-localization with p115 (Figure 9B) was significantly greater \( (p<0.05) \) than that of wild-type proglucagon (Figure 9B). This suggests that processing of proglucagon at the K70R71 enhances sorting efficiency, and supports the view that the initial processing event occurs before the sorting event.
Figure 9. Identification of putative sorting signals in proglucagon. N2a cells expressing CPE were transfected with wt proglucagon, or the following mutants: K70R71Q, which disrupts a furin cleavage site; R18Q, which disrupts a dibasic site within glucagon; or L14P/L26P, which reduces the α-helical content of glucagon. (A) wild-type proglucagon shows strong immunoreactivity in the tips of the cell processes, and very weak signal in the Golgi. By contrast, there is strong Golgi signal that co-localized with p115 in cells expressing the K70R71Q mutant proglucagon, as well as signal in the cell processes. Representative images are shown. Scale bar represents 10 µm. The R18Q and L14P/L26P mutants show some Golgi co-localization and post-Golgi signal. (B) Mean correlation of proglucagon and p115 fluorescence in cells expressing wt proglucagon (wt) or indicated mutants. All mutant constructs showed significantly higher mean correlation with p115 compared with wt proglucagon, indicating some disruption in sorting. * p<0.05, ** p<0.01 compared with wild-type.

2.3.8 Modelling the 3D structure of proglucagon
We sought to produce a 3D conformational structure of proglucagon in order to see if any new information could be provided regarding the mechanism of contained sorting signals. We submitted the full proglucagon sequence for de novo structure prediction to the Robetta software suite (Seattle, WA, USA; http://robetta.bakerlab.org/), and five structures were predicted. The model with the greatest confidence score is shown in Figure 10. Though the five structures contain very different spatial conformations, each predicted structure contains significant α-helical content, three of which are found within the sequences of glucagon (proglucagon 33-61), GLP-1 (proglucagon 72-108), and GLP-2 (proglucagon 126-158), and the K70R71 inter-domain site are indicated in Figure 10. Since these helices appear to be randomly oriented within a disordered structure, joined by two flexible hinges of IP-1 and IP-2, this is consistent with a previous partial model of proglucagon (45). This model is unlike the previously reported structure (45) in which glucagon, GLP-1 and GLP-2 are depicted as rigid α-helices arranged in a trimeric structure. Furthermore, our protein modelling data predict that the α-helices of GLP-1 and GLP-2 may also serve as sorting signals. If CPE directly binds to proglucagon, our model the canonical “POMC signal” is formed from the flexibility of the molecule, or it may be that proglucagon contains a novel class of sorting signal that interacts with CPE.
Figure 10. Structural model prediction of proglucagon. The Robetta modelling web server was used to generate a *de novo* 3D protein structure prediction for entire sequence of proglucagon. The results are presented in cartoon format, with N- and C-terminal indicated, and the inter-domain cleavage site K70R71 labeled. The sequence of glucagon is coloured green, GLP-1 (1–37) yellow and GLP-2 red. The structure is largely disordered, with α-helices of glucagon, GLP-1 and GLP-2 oriented randomly within the structure. The K70R71 cleavage site appears to be located in an unstructured region, which may provide flexibility and accessibility to furin.

2.4 Discussion

This is the first report to investigate the mechanisms of the sorting of proglucagon to the secretory granules of the regulated secretory pathway. Proglucagon processing largely occurs in granules, where, distinct sets of constituent hormones in both α and L cells are produced by tissue-specific PC enzymes. The mechanisms of granule exocytosis that result in the nutrient-regulated secretion of glucagon and GLP-1 have also been investigated. The question that remains is, what are the mechanisms underlying how proglucagon is sorted to granules so that it may be processed and secreted. Proglucagon must eventually enter granules, where the mildly acidic and millimolar calcium environment of the granule lumen are required for optimally active prohormone.
convertases to produce glucagon, GLP-1 and GLP-2. Furthermore, efficient storage in granules is necessary for α and L cells to remain sensitive to nutrient status and thus trigger a robust secretory response. The initial sequence of whether sorting precedes processing (or vice versa), are not known. As proglucagon is co-targeted to granules with its processing enzymes, we investigated whether these enzymes play a role in the underlying mechanism. Here we have identified sorting signals that directly affect the efficiency of sorting to granules. We have also identified a role for CPE as a sorting receptor to mediate proglucagon sorting to granules in α cells and not L cells.

For this study, we chose the model Neuro-2a mouse neuroblastoma cell line because it correctly targets hormones to the regulated secretory pathway (46), and because we obtained a clone that lacked endogenous CPE (47), that made for an excellent control to examine the role of CPE as a sorting receptor. Conveniently, the Neuro-2a wild-type cells also lack detectable endogenous PC1/3 and PC2, which would prevent endogenous processing of proglucagon and allow their individual roles to be examined in the sorting of proglucagon. In order to quantify the degree of sorting, we chose to examine co-localization by calculating Pearson's correlation coefficient. The use of Pearson's correlation coefficient (PCC) is readily applicable to this task (48) and is thought to be an accurate quantitative parameter for diffraction-limited microscopy (41). It has been used to characterize the sorting of neuropeptide SgII in PC12 cells (40). From its mathematical formulation, Pearson's is robust to varying protein densities and uniform noise. A common alternative parameter, Mander's coefficient, was created in order to solve a perceived sensitivity problem (41), and ironically, is far too sensitive in practice as to approach near perfect co-localization where non-exists (49). Other methods of co-localization (e.g., overlap coefficient) discard meaningful fluorescence signal, or are not specific. A greater degree of correlation results in a more positive PCC value and two proteins are said to be co-localized. This should not be confused with measurements of the fraction of proteins which are co-localized, which is only accurately measured by sub-diffraction limiting techniques (e.g., super-resolution). When used with background subtraction, and compared to appropriate biological controls, as used in this report, Pearson's correlation is a useful quantitative technique. We measured co-localization
between proglucagon and p115, a marker of the cis- and medial-Golgi, by immunofluorescence microscopy. We used p115 instead of a granule marker, because in these cells, CgA gave a strong Golgi signal which masked changes to proglucagon localization. The Pearson's method successfully discriminated between insulin and pdx-1, proteins known to be co-localized or segregated, and confirmed that proglucagon was co-localized in CgA⁺ compartments in our Neuro-2a cells. Therefore, our co-localization analyses can accurately distinguish co-localized proteins within a sub-cellular compartment.

Since it has been shown that CPE acts as a specific sorting receptor for a wide range of hormones, including POMC (25) and proinsulin (20), this was the most promising candidate for a sorting receptor for proglucagon. When CPE was absent, the punctate granular immunofluorescence pattern was reduced, a pattern also seen with POMC in these N2a-wt cells (47). Since α and L cells both express CPE, we sought to confirm these results in cells that normally synthesize proglucagon. When the siRNA knockdowns were conducted, two types of controls were included, a mock transfected control without siRNA, and a scrambled siRNA non-specific to any known target in the mouse genome. It was felt that loading controls were not necessary because the mock and siRNA control transfections are internal controls. In the absence of CPE in alphaTC1-6 cells, regulated secretion of glucagon was lost. Consistent with our previous findings in which long-term culture of alphaTC1-6 cells in high glucose (25 mM) upregulated granule exocytotic machinery, these cells secrete glucagon in the presence of CPE (33). Our results now suggest that CPE is needed for the sorting of glucagon to granules, and that CPE plays an α cell specific role for sorting proglucagon to granules, as GLP-1 secretion was unaffected in the absence of CPE. There is relatively less CPE content in α cells when compared to L cells (50, 51), so it is theoretically possible that alphaTC1-6 cells were more sensitive to knockdown studies than were GLUTag cells by the property of containing less CPE. Following knockdown, GLUTag cells could have retained a minimum effective threshold of CPE to still allow for the sorting of proglucagon to granules and would then appear unaffected by knockdown. In both cell types, CPE removes the C-terminal basic residues from both proglucagon and its intermediates to
produce mature hormones. This may also suggest that α cells are more sensitive to the location of CPE activity on PGDPs compared to L cells, and there could exist a CPE-independent mechanism in L cells.

In an effort to test if the enzymatic activity of CPE was required for sorting proglucagon, we used the E300Q-CPE mutant, which had been previously shown to have an intact C-terminal sorting signal (47, 52) and target to granules of NIT3 cells derived from CPE<sup>fαt/fαt</sup> mouse beta cells (32). However, the enzymatically inactive E300Q-CPE failed to sort proglucagon to granules. This was interesting because the immunofluorescence staining of both proglucagon and E300Q-CPE localized to the Golgi, and proglucagon was strongly blocked from sorting to granules. In our hands, the E300Q-CPE was not secreted from Neuro-2a cells, while proglucagon strongly co-localized to the Golgi, contradicting this previous study (32). The strong immunofluorescence signal in the Golgi suggests that E300Q-CPE is not degraded. Instead, it is possible that the impaired trafficking of E300Q-CPE is specific to N2a cells. Similar cell-specific findings were found with the S202P-CPE (<i>fat</i>) mutant, in which a significant fraction escapes proteosomal degradation and sorts to granules in NIT-3 cells (53), yet appears completely degraded in pituitary cells (30). Nevertheless, we could not distinguish whether proglucagon sorting is independent of the CPE enzymatic activity. This situation is similar to the impaired sorting of PC1/3 that prevented pro-renin from entering granules (16).

CPE is activated in the acidic environment of the granules (pH ~5.5-6.0), and although the membrane bound form maintains some enzymatic activity, the soluble form of CPE is the major active form (54–56). Activation of CPE involves a conformational change of its enzymatic pocket, which can be specifically blocked by catalytic site-directed inhibitors (e.g., bromoacetyl-D-arginine) (55). The conformational state of CPE is already established to be a major determinant that determines whether the POMC sorting motif binds to complementary motifs on CPE-sorted prohormones (57, 58), independent on its enzymatic activity (55), and consistent with the characterization that the E300Q-CPE maintained sorting function (32). Therefore, our results suggest the possibility that proglucagon is sorted by binding to a novel interaction surface of CPE distinct from that
used by POMC.

We generated several putative 3D structures of proglucagon in order to resolve whether it is compatible with the known sorting receptor motif of CPE. On the surface of CPE are two exposed basic residues, Arg255 and Lys260, that can bind two acidic amino acids ~12-15 Å apart on POMC, proinsulin, pro-enkephalin and pro-BDNF, stabilized by two hydrophobic residues spaced ~5-7 Å apart. Our model of proglucagon reveals that the mature hormone domains contain predominant α-helices, consistent with the known structures of purified recombinant hormones, yet the 3D organization of the molecule is largely disordered. Given the flexibility of the molecule, it is conceivable that a CPE binding motif could be in the disordered GRPP or IP-2, which contain several acidic amino acids. Alternatively, a coiled-coil conformation comprising glucagon and GLP-1, is another compatible CPE binding motif. Though hard to predict, the inherent flexibility of some intrinsically disordered proteins show binding diversity for multiple, specific ligands. Nevertheless, the specific nature of the interacting domains between proglucagon and CPE still need to be determined as it may represent novel sorting motif on the surface of CPE.

In addition to prohormone maturation, processing enzymes can influence trafficking of prohormones to granules. Both PC1/3 and PC2 are targeted to granules with proglucagon, and it is not yet known whether they are involved in the sorting of proglucagon. PC1/3 has been implicated in sorting of pro-renin, so it may aid proglucagon sorting. One report of a human neuroendocrine cell line (BON-1) found that induction of PC1/3 up-regulated expression of both CgA and SgII, two granins that function together as sorting machinery, providing evidence of co-operation among independent sorting receptors. Neither the pro-form of PC1/3 or the active form had any effect on proglucagon sorting efficiency beyond the effect seen with CPE. In contrast, there is no evidence that PC2 plays a role in hormone sorting or secretion, but we decided to investigate it given that it producing glucagon within granules. In the extreme, PC2 knockout mice are expectedly deficient in glucagon. These α cell granules had abnormal morphology, yet they appear to maintain regulated secretion of proglucagon.
and processing intermediates (2), indicating that sorting of proglucagon was unaffected in the absence of PC2. Consistent with this report, we found that PC2 had no effect on the sorting of proglucagon, so it would seem that PC2 has no bearing on proglucagon sorting. Our data shows that CPE plays a role in sorting proglucagon in Neuro-2a and α cells, while L cells may possess an alternate sorting mechanism.

Adding to this complexity is the fact that proglucagon undergoes several steps of processing to yield mature hormones. It is thought that the first step of processing occurs at the inter-domain Lys70Arg71 cleavage site to yield glicentin and MPGF. These two peptides are consistently detected when proglucagon is transfected into a non-endocrine cell line, leading to the hypothesis that this cleavage occurs in the TGN. This processing event in non-endocrine cells that lack a regulated secretory pathway would be mediated by furin, since it is expressed in all cell types (4, 45). By pulse-chase studies, proglucagon is first processed into glicentin and MPGF in α cells and many endocrine cell lines by ~30 minutes of chase (1, 45, 62). However, this technique lacks sufficient resolution to identify the compartment in which proglucagon is processed, and therefore this initial cleavage could occur in the TGN or within immature secretory granules. Both glicentin and MPGF must then sort into secretory granules, where PC1/3 and PC2 will produce mature hormone. Therefore, the R71Q mutant used in our study represents an important block early in the sequence of post-translational processing, which we found significantly reduced sorting efficiency. This is interesting because it suggests that processing at this site increases sorting efficiency, and should happen prior to the sorting event. Another way to address the temporal order of the initial sorting and processing events would be to identify whether glicentin and MPGF enter the same, or distinct granule populations. Due to a lack of specific antibodies to monitor every single proglucagon-derived peptide, and adequate systems to monitor proglucagon processing and trafficking in vivo, it is unclear which granule populations contain glicentin and MPGF. As a result, this alternative approach cannot distinguish whether initial processing precedes sorting, and if so, is this initial processing is required to enhance sorting efficiency. While there are several examples whereby different hormones enter the same granule population, whether this is the dominant sorting mode for proglucagon is not
known. In the case of pro-TRH, processing intermediates are known to enter distinct sub-populations of granules (63). Our results raise the possibility that glicentin and MPGF contain independent sorting signals since glucagon and GLP-1 are liberated from these distinct intermediates.

An important finding of this paper is the first documentation of sorting signals contained within proglucagon. On the basis of sorting signals identified in other prohormones (64), we focused on two likely signals within proglucagon, 1) the glucagon α-helix, and 2) the dibasic site within the glucagon α-helix (Arg17Arg18). Since mutations to either structure similarly reduced the sorting efficiency of proglucagon, it is likely that sorting information is uniquely encoded by the entire glucagon α-helix. That the two homologous α-helices in GLP-1 and GLP-2 apparently cannot, or do not, compensate for this mutation, may be a result of lacking an embedded dibasic site. Note that the information encoded in the α-helices of GLP-1 and GLP-2 is investigated in greater detail in Chapter 3 (65). Efficient sorting to granules is important to guarantee both proper processing to mature hormones, and that α and L cells have sufficient stores of active hormones to remain responsive to changes in nutrient status. Such a robust response is needed if glycemic control by glucagon and GLP-1 are to be tightly controlled. The nature of the glucagon sorting signal, and whether GLP-1 and GLP-2 encode similar information is examined in Chapter 3.

2.5 References

canine L cells expressing endogenous prohormone convertase 1/3 and prohormone convertase 2. Endocrinology 140, 4800–8
microscopy. J. Microsc. 185, 21–36
Chapter 3

Two Dipolar α-Helices Within Hormone-Encoding Regions of Proglucagon are Sorting Signals to the Regulated Secretory Pathway

Author Contributions

Mr. Leonard Guizzetti conducted cell assays, plasmid construction, fluorescence microscopy and quantitative co-localization analysis.

Ms. Rebecca McGirr assisted with plasmid construction.

Statement of Copyright

This chapter is reformatted from its original publication as an article in the Journal of Biological Chemistry. The citation for the original work is:
The official Journal of Biological Chemistry policy regarding permissions for authors may be accessed online (http://www.jbc.org/site/misc/Copyright_Permission.xhtml).

3.1 Introduction

Proglucagon is an endocrine prohormone that is expressed in pancreatic α cells, intestinal L cells, and select neurons of the hypothalamus and brainstem. It is the precursor for the peptide hormones glucagon, glucagon-like peptide (GLP)-1 and GLP-2. Glucagon is the main glucose counter-regulatory hormone, principally stimulating hepatic gluconeogenesis and glycogenolysis to maintain euglycemia (1). Conversely, GLP-1 and GLP-2 are secreted from intestinal L cells in response to nutrient ingestion; GLP-1 stimulates insulin secretion in a glucose-dependent manner, and GLP-2 increases intestinal blood flow and nutrient absorption (2). Oxyntomodulin, which contains the sequence of glucagon plus a 6 amino acid C terminal extension (Figure 1) is also
postprandially secreted from L cells and acts as a potent appetite suppressant (3). Therefore, all three hormones exert distinct metabolic actions to maintain nutrient homeostasis.

Figure 1. Post-translational processing of proglucagon to its derived peptides. A schematic representation of proglucagon in which each domain has been colour-coded, and the amino acid positions of key processing events have been identified. The amino acid numbering is relative to the first N-terminal amino acid of proglucagon (lacking the signal peptide). The major hormone products produced from tissue-specific processing by their respective prohormone convertase (PC) are shown at bottom. In pancreatic islet α cells, the major products are glucagon and major proglucagon fragment (MPGF). In enteroendocrine L cells and in the brain, the major products are GLP-1 (1-37), GLP-1 (7-37), GLP-2, oxyntomodulin and glicentin.

The post-translational processing of proglucagon by prohormone convertases (PCs) follows a strict temporal sequence, in which an initial cleavage at K70R71 yields two fragments, glicentin and major proglucagon fragment (MPGF) in both α cells and L cells (Figure 1) (4, 5). Pancreatic α cells produce glucagon through cleavage of glicentin by PC2 (4, 6), and PC1/3-mediated processing yields glicentin, oxyntomodulin, GLP-1 and GLP-2 within L cells and neurons (7–9). There is evidence that proglucagon processing in α cells is altered under conditions of beta cell injury such that bioactive GLP-1 is produced (10). Each of these peptide hormones must be stored in dense-core secretory
granules, a compartment unique to endocrine and neuroendocrine cells, for nutrient-regulated secretion. It is well documented by pulse-chase and immunoelectron microscopy studies that the final stages of processing occur in the secretory granules (5, 11, 12), and therefore, the sorting of proglucagon to secretory granules is essential for the production of its bioactive peptide hormones. However, it is not known if the initial cleavage of proglucagon to glicentin and MPGF occurs before or after sorting to granules. One component of a sorting mechanism that appears to be common to a number of prohormones is a sorting signal that is contained within the prohormone sequence. If the initial processing of proglucagon precedes sorting to granules, then a sorting signal must be present in each of glicentin and MPGF, leading to the intriguing possibility that proglucagon contains at least two sorting signals that are spatially segregated.

Several types of prohormone sorting signals have been described that mediate specific interactions with membrane-bound sorting receptors or co-target with processing enzymes. Proinsulin (13, 14) undergoes aggregation mediated by hydrophobic residues. A disulfide-bonded loop exposes two acidic amino acid residues comprise a sorting signal within pro-opiomelanocortin (POMC) (15, 16), proinsulin (14), proenkephalin (17) and pro-brain-derived neurotrophic factor (pro-BDNF) (18), and interacts with the sorting receptor carboxypeptidase E (CPE). Paired basic amino acids that are cleavage sites for PCs serve as sorting signals in pro-neuropeptide Y (pro-NPY) (19), pro-renin (20), progastrin (21), proneurotensin (22) and pro-VGF (non-acronymic) (23), suggesting that these prohormones are co-targeted with their processing enzymes. Finally, amphipathic α-helix regions/domains are required for the sorting of prosomatostatin (24) and pro-cocaine and amphetamine regulated transcript (pro-CART) (25). Any or all of these sorting signals may exist within a single prohormone and may synergize to increase sorting efficiency (23, 26, 27).

Of these various known sorting signals, proglucagon contains two predicted types: significant α-helical content within glucagon, GLP-1 and GLP-2, as documented by their known crystal or NMR structures (28–30), and a dibasic amino acid sequence within the α-helix of glucagon (R17R18; proglucagon 49-50). Interestingly, unlike other
prohormones, the α-helices lie within ordered hormone-encoding regions, and not in a prodomain (31). Additionally, these hormone domains are evolutionarily conserved, particularly regarding their biophysical characteristics (32). We have previously identified R17R18 and the α-helix within glucagon as putative sorting signals, and our results also suggested that processing of proglucagon to glicentin and MPGF precedes sorting (33). Therefore, in the present study, we investigated the possibility that proglucagon contains multiple sorting signals in the different hormone domains. To this end, we have extensively characterized the role of each predicted α-helix within proglucagon in sorting to the regulated secretory pathway in the well characterized neuroendocrine PC12 cell line. Our study reveals that two non-amphipathic α-helix domains within the sequences of glucagon and GLP-1 are necessary and sufficient to target proglucagon to granules. We also combine these results to a model of proglucagon processing and sorting in α- and L-cells.

3.2 Methods

Plasmid construction and reagents

Fusion proteins were constructed using proglucagon-derived peptide sequences attached to the 3’ end of the cDNA encoding the CH2/CH3 domains of mouse IgG-2b (termed Fc), preceded by the pro-renin signal peptide, as previously described (Figure 2) (24), (a kind gift from Dr. T. Reudelhuber, Montreal, QC, Canada). Proglucagon-derived DNA sequences were amplified from Syrian hamster pre-proglucagon cDNA (a kind gift from Dr. D. Steiner, Chicago, IL, USA; GenBank™ accession J00059.1). All primers were purchased from Sigma Aldrich (Oakville, Ontario, Canada), and the specific primers used in this study for PCR amplification or site-directed mutagenesis can be found in Table 1. All fusion constructs were constructed in pcDNA3.1 (Life Technologies, Burlington, Ontario, Canada).
Table 1. Primer pairs used for cloning and mutagenesis

<table>
<thead>
<tr>
<th>Fusion Construct</th>
<th>Oligonucleotide Sequence Pair (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forward: 5'-AAGCTTGGCATGGATCAATTCCGAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Fc (for construction of Fc alone and Fc-Glucagon)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forward: 5'-GACCATCTCCCGGTCTCCGGTTTCTGAGCAGT</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Fc stop mutation</td>
<td>Forward: 5'-CCGTGGATCCACTCAGTCCAGTGATG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Fc-Glucagon frame shift</td>
<td>Forward: 5'-GACCATCTCCCGGTCTCCGGGTAGCCTTGG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucagon *</td>
<td>Forward: 5'-GAATTCCTCACAGGGAACA</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu26Pro-Glucagon amplification</td>
<td>Forward: 5'-GCTAGCATGGTCAATCCGGATGG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg18Gln-Glucagon mutagenesis</td>
<td>Forward: 5'-AATACCTCTGAGTCTCCGGCAAGGCAAGATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu14Pro-Glucagon mutagenesis</td>
<td>Forward: 5'-TACAGCAAAATACCCGGGAGTCCGGAGTGG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Fc (for construction with GLP-1, GLP-2, Glicentin, MPGF)</td>
<td>Forward: 5'-CTCAGCAAAATACCCGGGAGTCCGGAGTGG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP-1 (1-37) §</td>
<td>Forward: 5'-GGATCCAGGAGGAGGAGAGGAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP-1 (7-37) §</td>
<td>Forward: 5'-GGATCCAGGAGGAGGAGGAGGAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP1 E27 mutagenesis</td>
<td>Forward: 5'-GGCGAGGCTCGGAAGAGAGGAAGGGAATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP-1 Stop codon mutagenesis</td>
<td>Forward: 5'-GGTGAAAGGCAAGAGGAATTTGGAATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP-2 §</td>
<td>Forward: 5'-GGATCCAGGAGGAGGAGGAGGAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>GLP-2 Stop codon mutagenesis</td>
<td>Forward: 5'-GGTGAAAGGCAAGAGGAATTTGGAATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp3Gln-GLP-2 mutagenesis</td>
<td>Forward: 5'-GGATCCAGGAGGAGGAGGAGGAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp8Lys,Glu9Lys,Asn11Lys, Asp15Lys-GLP-2 mutagenesis</td>
<td>Forward: 5'-GGTGAAAGGCAAGAGGAATTTGGAATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Glicentin §</td>
<td>Forward: 5'-GGATCCAGGAGGAGGAGGAGGAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>MPGF §</td>
<td>Forward: 5'-GGATCCAGGAGGAGGAGGAGGAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxyntomodulin §</td>
<td>Forward: 5'-GGTGAAAGGCAAGAGGAATTTGGAATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Underlined sequence indicates the restriction sites used for cloning. Bolded sequences indicate site-directed mutations. A conservative mutation was made within GLP-1 to remove an internal EcoRI restriction site. Bolded and underlined text indicates a stop codon. Constructs used a flexible linker of either eight amino acids (denoted by §; sequence: KLGTELGS) or a ten amino acid linker (denoted by *; sequence: GSTQSSVVEF).
The cDNA sequence of Fc was selectively amplified using the Fc primers (Table 1) and ligated to the pcDNA3.1 backbone, between HindIII and BamHI restriction sites. To construct an Fc expression plasmid, an in-frame stop codon was mutated between the coding region and the HindIII restriction site, using the Fc stop primers (Table 1). Mutagenesis reactions were performed using the QuickChange II Site-Directed Mutagenesis kit (Agilent Technologies, Mississauga, Ontario, Canada) according to the manufacturer’s protocol.

The Fc-wild-type glucagon fusion construct was generated ('Fc-WT Glucagon'; Figure 2), in which glucagon cDNA was amplified using specific primers for glucagon (Table 1), and ligated between the EcoRI and NotI restriction sites, joined by a 10 amino acid linker (Table 1). To determine possible sorting signals, the sequence of glucagon was mutated in two ways (Figure 2): two leucines, L14 and L26, which are postulated to stabilize the α-helix, were mutated to L14P,L26P ('Fc-LP glucagon') by specific amplification and mutation using the respective L14P and L26P glucagon primers (Table 1); and the dibasic sequence R18R19 was changed to R18R19Q ('Fc-RQ glucagon') using the R18Q mutagenesis primers (Table 1).

Subsequent proglucagon-derived peptide constructs used a similar Fc expression system, in which Fc was ligated into the NheI and HindIII restriction sites. Expression constructs were generated for the following peptides: GLP-1 (1-37), GLP-1 (7-37), GLP-2, oxyntomodulin, glicentin and MPGF (referred to as Fc-GLP-1 (1-37), Fc-GLP-1 (7-37), Fc-GLP-2, Fc-OXM, Fc-Glicentin and Fc-MPGF, respectively; Figure 1 & 2), and ligated into the BamHI and EcoRI restriction sites. An internal EcoRI cut site was silently mutated using the GLP-1 E27 primers (Table 1). From Fc-GLP-2, we generated site-directed point mutations, specifically changing five acidic amino acids to either neutral, D3Q, or basic, D8K, E9K, N11K, D15K (referred to as Fc-Dipolar GLP-2; Figure 1 & 2, Table 1). These mutations were chosen in order to mimic the dipolar nature of the glucagon α-helix, which share less than 40% homology, and also to keep the α-helix intact. Lastly, each of these constructs was terminated by an in-frame stop codon, introduced either by site-directed mutagenesis or PCR amplification. All results were
confirmed by sequencing at the London Regional Genomics Facility, University of Western Ontario.

**Figure 2.** *Schematic depiction and expression of fusion proteins.* A, Fusion proteins were comprised of a signal peptide (SP), mouse IgG-2b heavy chain (Fc), and a proglucagon-derived peptide (PGDP). A control was comprised of Fc alone immediately followed by a stop codon. The listed peptides were individually fused in frame with Fc to generate fusion proteins. Numbers indicate amino acid positions relative to the PGDP sequence. Amino acid mutations and position are indicated within schematic illustration. *Dashed boxes* indicate the groups of related peptides. Fc-LP glucagon decreases helical content of glucagon, while Fc-RQ glucagon changes the dibasic sequence. Fc-Dipolar GLP-2 mutations introduce a positively charged surface within the α-helix to mimic the charge distribution of glucagon/GLP-1. B, Expression of fusion proteins in PC12 cells by Western blot. Far right lane shows expression of Fc-WT glucagon in alphaTC1-6 cells.
3.2.1 Cell Culture and transient transfections
Wild-type PC12 cells (a kind gift from Dr. W. J. Rushlow, University of Western Ontario, London, Ontario, Canada), were maintained in high-glucose (25 mM) DMEM (Life Technologies), supplemented with 15% horse serum (Life Technologies) and 2.5% FBS (Life Technologies). AlphaTC1-6 cells (a kind gift from Dr. C. B. Verchere, University of British Columbia, Vancouver, British Columbia, Canada) were cultured as previously described (34). Cells were transfected using Lipofectamine 2000 (Life Technologies). To prepare cells for microscopy, cells were grown on glass cover-slips coated with rat-tail type I collagen (100 μg/mL; Sigma) at a density of $4 \cdot 10^5$ cells/cm$^2$ the day prior to transfection. For secretion assays, cells were grown in poly-D-lysine coated 6-well tissue culture dishes (Coming, Mississauga, Ontario, Canada). Cells were allowed to grow for 48 h following transfection.

3.2.2 Secretion experiments
On the day of the experiment, media were changed to high-glucose DMEM supplemented with 1% dialyzed FBS. After preincubation, cells were incubated for 3 h in the same medium ("3 h basal") followed by 15 min incubations without ("-K") and with ("+K") 55 mM KCl to stimulate granule exocytosis (35). Cells were quickly rinsed in Hanks' buffered salt solution (HBSS) between incubations. All media (1 mL per sample) were collected on ice, with fresh protease inhibitor cocktail (Roche, Mississauga, Ontario, Canada), 2 μg/mL aprotinin, 55 mM Tris, 1 mM EDTA, pH 7.4 for immunoprecipitation, and cell lysates were collected and protein concentration was quantified as previously described (34).

3.2.3 Immunoprecipitation, Western blot and secretion index
The media and cell lysates were applied to 50 μL of Protein A-Sepharose (GE Health Care, Uppsala, Sweden), incubated at 4°C overnight with rotation, after which beads were recovered, and protein was eluted by heating to 70°C for 10 min. The immunoprecipititated proteins were separated on 10% NuPAGE pre-cast gels (Life Technologies) or SDS-PAGE and transferred to nitrocellulose membranes. Fc-immunoreactive bands were visualized by incubating membranes with goat anti-mouse
IgG HRP-conjugated antibody (1:5000 concentration; Life Technologies) followed by SuperSignal Chemiluminescent substrate (Thermo-Fisher Scientific, Toronto, Ontario, Canada). Bands were quantified by densitometry as described previously (26). Secretion indexes were expressed as a ratio of stimulated to basal secretion, normalized to total protein (22), and were used for statistical analysis.

3.2.4 Co-immunoprecipitation of Chromogranin A complexes
To perform co-immunoprecipitation experiments, the Fc-construct transfected PC12 cells were lysed in either a neutral or pH 5.5 buffer (20 mM sodium acetate, 0.1 M KCl, 1% Triton X-100), cell debris was cleared by centrifugation (22,000x g for 10 min at 4°C), and supernatant was taken for protein quantification. Protein concentration was quantified using a colorimetric BCA assay (Thermo-Fisher), and results read out on a multi-well plate reader (Bio-Rad). Immunoprecipitations were done for the Fc epitope using Sepharose A beads (as described), using a total of 100 ug whole cell lysate, diluted to a final volume of 1 mL (plus a protease inhibitor cocktail) to facilitate over-night incubation (16 h, 4°C) on a tube rotator. The following morning, Sepharose beads were pelleted (5,000x g, 30 sec), washed in the same ice-cold buffer used for immunoprecipitation. The protein complexes were eluted by heating (70°C, 10 min) and immediately chilled on ice. Samples were then immuno-blotted for CgA to detect possible co-immunoprecipitated complexes.

3.2.5 Immunofluorescence
Cells were processed for immunofluorescence as previously described (26). Slides were incubated with antibodies against the secretory granule marker, chromogranin A (CgA) (1:100; Abcam, Cambridge, MA, USA), or the synaptic-like microvesicle marker, synaptophysin (1:250; Abcam). AlexaFluor488 IgG (Life Technologies) was used to visualize the reporter, Fc, and AlexaFluor594 IgG for the CgA or synaptophysin antibody. Coverslips were mounted using a ProLong Gold Anti-fade mounting medium (Life Technologies).

3.2.6 Image Acquisition and Analysis
Immunofluorescence images were acquired using a Zeiss LSM 510 Duo Vario confocal
microscope (Zeiss Canada Inc., Toronto, Ontario, Canada) and a 63x 1.4 NA Plan-Apochromat oil differential interference contrast objective lens using the Zen 2009 software (Zeiss Canada Inc.). Three cover slips per transfection were imaged for analysis. Image analysis was conducted using FIJI version 1.46h (36), a distribution of ImageJ (NIH, Bethesda, MD, USA), using the Co-localization 2 plug-in within FIJI. Regions of interest were manually drawn around distinct single- or multi-cell bodies, positive for Fc and either chromogranin A or synaptophysin. Co-localization of these pixels from each pseudo-coloured image were used to calculate Pearson’s correlation coefficient, as previously described (33). To generate a 3D rendering of the spatial localization of Fc-WT glucagon and Fc-Dipolar GLP-2, the Imaris software package (Bitplane AG, Zurich, Switzerland) was used. The 3D voxel information was used to assign 0.25-0.30 μm spheres to computed point sources of light in each channel. Only the co-localized spots are shown, as determined by spatial overlap within a maximum distance of 0.30 μm. Correlation coefficients from each experiment were treated as one experimental data set (n=30-35).

3.2.7 Secondary Structure Predictions and Biophysical Property Calculations
Secondary structure predictions (see Table 2) were carried out with the PSI-PRED algorithm (version 3.1) (37). Percent helical content was calculated as the ratio of total α-helical residues to the peptide length. The corresponding pI was calculated using the ExPASy bioinformatics portal (38) and the mean hydrophobic moment was calculated using the method of Eisenberg et al. (39). Hydrophobic cluster analysis was carried out by the method of Gaboriaud et al. (40).

3.2.8 Statistical Analyses
Differences were assessed using a one-way ANOVA with Tukey’s HSD post-hoc test. Statistical significance was accepted at the level of p<0.05, and the results are expressed as the mean ± standard error. Statistical analyses were performed using GraphPad Prism version 5.02 (GraphPad Software Inc, La Jolla, CA, USA).
3.3 Results

3.3.1 Rationale for using PC12 as a model of hormone trafficking and secretion

While in our previous publication we observed abundant localization of proglucagon in secretory granules in Neuro2a cells (33), these cells did not respond to any secretagogue (K\(^+\), Ba\(^{2+}\), dbcAMP, IBMX alone or in combination) in our hands. Since showing regulated secretion of our fusion constructs was a necessary part of our study, we sought another model of a cell type with a regulated secretory pathway. We avoided the use of α or L cell lines so as not to confound our results with endogenous proglucagon and derived peptides due to the multi-step nature of proglucagon processing. We chose the PC12 neuroendocrine cell line, because: 1) they have a very well-characterized regulated secretory pathway; 2) they express CPE, which we have shown to be a sorting receptor for proglucagon in α cells; and 3) they lack significant PC1/3 and PC2 activity, thus allowing us to assay individual proglucagon-derived peptides for sorting independently of processing. PC12 cells were used to characterize the sorting of multiple classes of neuropeptides and hormones that agree with the mechanism in their native tissues, including: proinsulin and pro-enkephalin (41), pro-BDNF (18, 42), pro-CART (25), pro-neurotensin (43) POMC (17), and pro-renin (44). In fact, the regulated secretory pathway in PC12 cells better characterized than in either of the accepted models of proglucagon processing, alphaTC1-6 and GLUTag cells, and they have many key proteins in common. PC12 and alphaTC1-6 cells express chromogranin A, human α cells additionally express chromogranin B, and secretogranins III (34, 45, 46), while L cells express CgB, and secretogranins II, III and V (47), thus showing similarities in sorting machinery. The exocytosis machinery is also similar, with both PC12 and alphaTC1-6 cells expressing the SNARE proteins syntaxin-1a, VAMP2, SNAP25 (34, 48) and the SNARE-associated proteins Munc13-1 and Munc18-1 (49, 50), while AP-1 and AP-3 are expressed by PC12 and mouse α cells (51). More recently, it has been shown that the GLUTag L cell model also expresses SNAP25, VAMP-1, -2 and -3, syntaxin-1a and Munc18-1 (52). The literature therefore strongly supports the use of the PC12 cell line as a model for the sorting of proglucagon to the regulated secretory pathway in α and L cells.
3.3.2 Glucagon contains an α-helix sorting signal

In order to identify sorting signals contained within glucagon, we transfected PC12 cells with either Fc alone or the Fc-WT Glucagon fusion constructs described in Figure 2A. Expression of fusion constructs was confirmed by Western blot (Figure 2B). First, we determined the extent of regulated secretion of fusion proteins from PC12 cells using 55 mM K⁺ as a secretagogue. The KCl secretagogue causes depolarization of the plasma membrane, triggering a rapid, calcium-dependent fusion of secretory granules with the plasma membrane, resulting in exocytosis of granule cargo. A lack of response to secretagogue stimulation (i.e., secretion index equal to unity) indicates constitutive secretion, whereas a significantly elevated secretion index indicates the ability of glucagon to direct Fc into secretory granules of the regulated secretory pathway (53). Second, we examined the extent to which the Fc fusion proteins were sorted to secretory granules by quantitative co-localization with the secretory granule marker, chromogranin A (CgA). Taken together, these experiments specifically determined the nature of sorting signals within glucagon that direct it to granules.

As expected, the fragment of the mouse IgG heavy chain, Fc, was secreted in a constitutive manner as shown by the lack of K⁺-stimulated release, and a secretion index of 1 (Figure 3A, B). In contrast, fusion of Fc to WT glucagon resulted in regulated secretion, as indicated by a robust secretory response to 55 mM K⁺ (Figure 3A) and a secretion index that was significantly elevated (p<0.05) compared to Fc alone (Figure 3B). Therefore, WT glucagon contains a signal that is sufficient to sort Fc to granules. We then investigated the structural nature of the sorting signal within glucagon by mutating the α-helix (Fc-LP Glucagon) and the role of the R17R18 motif in sorting (Fc-RQ Glucagon). The secretion of Fc-LP glucagon secretion did not increase upon secretagogue stimulation (Figure 3B), and the secretion index was not significantly different from Fc alone (Figure 3B). In contrast, Fc-RQ glucagon showed similar regulated secretion to WT glucagon (Figure 3A), and significantly greater secretion index (p<0.05) compared to Fc alone (Figure 3B). These results suggest that the α-helix within glucagon, and not the dibasic site, may serve as a sorting signal to direct proglucagon into granules.
Figure 3. Glucagon contains a necessary α-helical sorting signal. PC12 cells were transfected with either Fc alone, or Fc-Glucagon fusion constructs. A, Western blot analysis of regulated secretion. PC12 cells were incubated for 15 min without (-K; constitutive), then with (+K; stimulated) 55 mM K⁺. Media were immunoprecipitated for Fc prior to Western blot analysis. Representative blots are shown. (n=6) B, Secretion indexes from (A). The dashed line indicates a secretion index of unity. Values are means ± S.E.M. (n=6). ** P<0.01, * P<0.05, vs Fc alone. C, Subcellular localization of Fc-glucagon fusion proteins. Arrows indicate co-localization of Fc with CgA. Arrowheads denote Golgi localization. Scale bar: 10 μm. D, Pearson's correlation coefficient (PCC) for co-localization between Fc and CgA. Values are means ± S.E.M. (n = 30-35). *** P<0.001, * P < 0.05, vs Fc alone, # P<0.001 vs Fc-GLP-2.
To identify the subcellular distribution of the fusion proteins, we conducted immunofluorescence confocal microscopy to visualize Fc immunoreactivity and the extent of co-localization with the secretory granule marker, CgA. Co-localization was quantified as the fluorescence intensity co-variance between Fc and CgA immunofluorescence, using Pearson’s correlation coefficient (PCC), described previously in (33). Fc alone had a para-nuclear staining pattern characteristic of Golgi localization (arrowhead, Figure 3C). The corresponding measured fluorescence correlation of Fc and CgA (Figure 3D) appears high, but likely reflects the fact that both Fc and CgA are co-trafficked through the Golgi under steady-state conditions, rather than localization of Fc in granules. In contrast, Fc-WT glucagon expression was localized in CgA-positive granules along the cell periphery and toward the tips of the cell processes (arrow, Figure 3C), a pattern that indicates localization in secretory granules (33). Pearson’s correlation of Fc-WT glucagon with CgA was significantly greater than Fc alone \( (p<0.01; \text{Figure } 3D) \), thus demonstrating the sorting of Fc-WT glucagon to secretory granules. When the \( \alpha \)-helix of glucagon was disrupted in Fc-LP glucagon, Fc immunoreactivity was predominantly localized to the Golgi, and the corresponding Pearson’s correlation was not significantly different from Fc alone (Figure 3D). Lastly, Fc-RQ glucagon was localized within CgA-positive secretory granules in a punctate pattern similar to that of Fc-WT glucagon (arrow, Figure 3C). Pearson’s correlation of Fc-RQ glucagon with CgA was significantly greater than Fc alone \( (p<0.001) \) (Figure 3D) and not significantly different from Fc-WT glucagon. Taken together, our results indicate that the \( \alpha \)-helix within glucagon is a necessary and sufficient sorting signal, while the dibasic R17R18 motif is not required for sorting.

In order to show that the sorting of the Fc constructs is not an artifact of the cell type, we repeated secretion and immunofluorescence experiments using Fc-WT glucagon in alphaTC1-6 cells, a glucagon-secreting cell line (34). Fc-WT glucagon exhibited a similar degree of stimulated secretion with 15 mM arginine \( (SI = 2.3 \pm 0.1 \text{ vs } 1.1 \pm 0.2 \text{ for Fc alone, } p<0.01) \) to that seen in PC12 cells stimulated with 55 mM K\(^+\). Similar values were also observed in alphaTC1-6 cells with the localization of Fc-WT glucagon in CgA-positive granules \( (\text{PCC} = 0.80 \pm 0.03 \text{ vs } 0.51 \pm 0.03 \text{ for Fc alone, } p<0.001) \) as in PC12
cells. As shown in Figure 2B, the level of expression of Fc-WT glucagon in alphaTC1-6 cells was within the range observed for PC12 cells. These results validate the use of PC12 cells and Fc-PGDP constructs to identify sorting signals in proglucagon.

3.3.3 GLP-1, but not GLP-2, efficiently targets Fc to secretory granules

Since our previous work indicated that the processing of proglucagon to glicentin and MPGF may precede entry into granules (33), we tested the possibility that proglucagon may contain sorting signals within its other constituent peptides. Fc fusion proteins of the glucagon-like peptides, GLP-1 and GLP-2, were constructed and expressed in PC12 cells. Both GLP-1 (1-37) and GLP-1 (7-37) were included so as to test the role of the N-terminal 6 amino acids of full-length GLP-1. Both the Fc-GLP-1 (1-37) \((p<0.001)\) and Fc-GLP-1 (7-37) \((p<0.001)\) exhibited robust \(K^+\)-stimulated secretion compared to the constitutively secreted Fc reporter (Figure 4A,B). Surprisingly, Fc-GLP-2 did not exhibit regulated secretion (Fig 4A,B). Immunofluorescence microscopy showed that both forms of Fc-GLP-1 directed Fc to granules, as evidenced by a punctate fluorescence pattern along the cell periphery and toward the tips of cell processes (arrow and inset, Figure 4C). There was significant correlation between CgA and Fc fluorescence for Fc-GLP-1 (1-37) \((p<0.001)\) and Fc-GLP-1 (7-37) \((p<0.001)\) (Figure 4D) Fc-GLP-2 showed a stronger para-nuclear localization, and interestingly, was also present in punctate vesicles which appeared to be distinct from those that were immuno-positive for CgA (arrowhead and inset, Figure 4C). There was no significant difference in co-localization with CgA and Fc compared with Fc alone (Figure 4D), consistent with the lack of \(K^+\)-stimulated secretion. Therefore, our data suggest that GLP-1 (7-37) contains sufficient information for granule sorting. However, GLP-2 is not sorted efficiently into dense-core secretory granules, and may instead be routed to another vesicle compartment in PC12 cells.
Figure 4. GLP-1 is sorted to secretory granules by a sorting signal within GLP-1(7-37). PC12 cells were transfected with Fc-GLP-1 or Fc-GLP-2 fusion constructs. A, Western blot analysis of regulated secretion. PC12 cells were incubated for 15 min without (-K; constitutive), then with (+K; stimulated) 55 mM K+. Media were immunoprecipitated for Fc prior to Western blot analysis. Representative blots are shown. (n=6) B, Secretion indexes from (A). The dashed line indicates a secretion index of unity. Values are means ± S.E.M. (n=6). *** P<0.001 vs Fc alone, # P<0.001 vs Fc-GLP-2. C, Subcellular localization of Fc-GLP-1, Fc-GLP-2 and Fc-Dipolar GLP-2 fusion proteins. Arrows indicate co-localization of Fc with CgA. Scale bar: 10 µm. D, Pearson's correlation coefficient (PCC) for co-localization between Fc and CgA. Values are means ± S.E.M. (n = 30-35). *** P<0.001 vs Fc alone.
3.3.4 A dipolar α-helix GLP-2 mutant sorts to secretory granules

Since GLP-2 shares only 38% and 32% homology with glucagon and GLP-1, respectively, it is possible that the sequence context contributes to the differences in the sorting of glucagon and GLP-1 compared with GLP-2. We therefore introduced point mutations in GLP-2 that would mimic the charge distribution within the sequence of glucagon by changing four acidic amino acids in the α-helix to basic lysines (see Methods, Table 1), termed “Fc-Dipolar GLP-2” (Table 2). In contrast to Fc-GLP-2, Fc-Dipolar GLP-2 showed a robust response to 55 mM K⁺ (Figure 4A), and the secretion index was similar to that of Fc-WT glucagon and significantly greater than Fc alone \((p<0.05)\) and Fc-GLP-2 \((p<0.05)\) (Figure 4B). Immunofluorescence microscopy showed that Fc-Dipolar GLP-2 was localized to CgA-positive granules (Figure 4C). The extent of co-localization between Fc-Dipolar GLP-2 and CgA was significantly greater than Fc alone \((p<0.001)\) and WT GLP-2 \((p<0.001)\) (Figure 4D). Therefore, altering the charge distribution of the α-helix of GLP-2 was sufficient to direct Fc to secretory granules.
Figure 5. MPGF is efficiently sorted into secretory granules by the sorting signal contained within GLP-1. PC12 cells were transfected with Fc-Glicentin, Fc-MPGF or Fc-OXM fusion constructs. A, Western blot analysis of regulated secretion. PC12 cells were incubated for 15 min without (-K; constitutive), then with (+K; stimulated) 55 mM K⁺. Media were immunoprecipitated for Fc prior to Western blot analysis. Representative blots are shown. (n=6) B, Secretion indexes from (A). The dashed line indicates a secretion index of unity. Values are means ± S.E.M. (n=6). ** P< 0.01 vs Fc alone. C, Subcellular localization of Fc-Glicentin, Fc-MPGF and Fc-OXM fusion proteins. Arrows indicate co-localization of Fc with CgA. Arrowheads denote Golgi localization. Scale bar: 10 μm. D, Pearson's correlation coefficient (PCC) for co-localization between Fc and CgA. Values are means ± S.E.M. (n = 30-35). ** P< 0.01 vs Fc alone.
Table 2. *Biophysical Properties of Major Proglucagon Derived Peptides.*

| Fusion Peptide | Peptide Sequence | 1) Mutation  
2) pI  
3) Mean Hydrophobic Moment | Helical Cluster Projection |
|----------------|-----------------|---------------------------|---------------------------|
| Fc-WT Glucagon | HSQGFTSDSKYLDSDLRAQDFVQWLMT  
(72% α-helix content) | 1) WT  
2) 5.76  
3) 0.55 | ![Image](image1)
| Fc-LP Glucagon | HSQGFTSDSKYPDSSLRAQDFVQWLPMT  
(45% α-helix content) | 1) L14P, L26P  
2) 6.39  
3) 0.47 | ![Image](image2)
| Fc-RQ Glucagon | HSQGFTSDSKYLDSRALQDFVQWLMT  
(72% α-helix content) | 1) R18Q  
2) 4.43  
3) 0.58 | ![Image](image3)
| Fc-GLP-1 (7-37) | HAEGFTSDVSYLEGQAKEFIAWLVKGRG  
(68% α-helix content) | 1) WT  
2) 4.68  
3) 0.69 | ![Image](image4)
| Fc-GLP-2 | HADGSFSDEMTILDSLARDWNLIQTKITD  
(67% α-helix content) | 1) WT  
2) 4.23  
3) 0.66 | ![Image](image5)
Amino acid sequences of the peptides used in this study are shown, with the wild-type or mutant sequence indicated. Underlined portions of sequence correspond to α-helical content of the peptides. Stars indicate proline (hydrophobic), diamonds indicate glycine (uncharged, hydrophobic), open squares indicate threonine (uncharged, polar), and dotted squares indicate serine (uncharged, polar). Enclosed amino acids represent hydrophobic patches.

### 3.3.5 Biophysical Properties of Glucagon, GLP-1 and GLP-2 α-helices determine sorting efficiency

Despite the fact that the amino acid sequences of glucagon, GLP-1(7-37) and GLP-2(1-33) are all highly conserved and contain a predominantly α-helical structure, our results clearly show that the α-helix alone is not sufficient to target PGDPs to granules. We determined the biophysical nature of the helices by calculating the hydrophobicity and charge distribution for the helical portion of each peptide. The hydrophobic clusters within wild-type glucagon (Table 2 & Figure 3) were disrupted within Fc-LP glucagon and remained intact in Fc-RQ glucagon, indicating that the leucines are important in the formation of larger hydrophobic clusters. Therefore, the signal within glucagon must consist of an intact α-helix. We then conducted hydrophobic cluster analysis of glucagon, GLP-1 and GLP-2, and did not observe any differences in either size or location of hydrophobic clusters (Table 2) between these highly-conserved sequences (54). These α-helices are flanked by highly conserved N- and C-terminal tails, indicating that these α-helices are in a similar peptide context. The mean hydrophobic moments for the α-helix regions of glucagon, GLP-1 and GLP-2 were similar, reflecting the degree of amphiphilicity of these helices. However, there were significant differences in net charge of the α-helices. The calculated pIs for the glucagon and GLP-1 α-helices were greater than that of GLP-2 (Table 2), suggesting the net charge (electrical polarization), rather than hydrophobicity, is a more important determinant of proglucagon sorting (Table 2). Lastly, based on the charged amino acid distribution, glucagon and GLP-1 have a net
polarization along the length of their helices, while GLP-2 has a more uniform negative charge distribution. By introducing a dipolar mutation to GLP-2, the charge distribution resembled that of glucagon/GLP-1, thus reconstructing a net polarization within GLP-2. Our results demonstrate that efficient targeting of glucagon (Figure 3), GLP-1 (Figure 4) and the GLP-2 dipolar mutant (Figure 4) to granules is determined by dipolar α-helices, that contain distinct positive and negative patches to polarize the length of the helix are sufficient to target glucagon and GLP-1 to secretory granules.

3.3.6 MPGF, but not glicentin, sorts to secretory granules
It has been documented that initial processing of proglucagon occurs at K70R71 early in the secretory pathway (5), possibly in the Golgi, to yield glicentin and MPGF (Figure 1). In this scenario, the processing of proglucagon to glicentin and MPGF may precede sorting to granules. We therefore examined the sorting behavior of glicentin and MPGF, with the hypothesis that processing at K70R71 would occur prior to sorting. Surprizingly, however, secretion of Fc-Glicentin was not stimulated by 55 mM K⁺ (Figure 5A) and its secretion index was similar to Fc alone (Figure 5B), indicating that glicentin was not sorted to the regulated secretory pathway. In contrast, secretion of Fc-MPGF was significantly stimulated by 55 mM K⁺ (p<0.001) (Figure 5A, B). These results were corroborated by analyses of subcellular localization. In contrast, Fc-Glicentin showed very little co-localization with CgA (arrow and inset, Figure 5C). Quantification of Pearson's correlation coefficients showed Fc-MPGF had a significantly higher value than Fc-Glicentin (p<0.01) and Fc alone (p<0.01) (Figure 5D). Therefore, our data demonstrate that MPGF, but not glicentin, is sorted to granules, thus implying that proglucagon must be sorted to granules prior to being cleaved to glicentin and MPGF. This is an intriguing finding because both glicentin and MPGF contain sorting signals (glucagon and GLP-1, respectively), yet they are sorted quite differently. These results suggest that 1) the sorting signal within the sequence of GLP-1 is sufficient to direct MPGF to secretory granules, and 2) the sorting signal within glucagon is masked by the N-terminal GRPP (Figure 1).

3.3.7 Oxyntomodulin sorts to secretory granules
In order to determine if GRPP is masking the sorting signal within glucagon, we
generated Fc-OXM (Figure 2). Secretion of Fc-OXM was significantly stimulated by 55 mM K$^+$ ($p<0.01$) (Figure 5A, B), in contrast to Fc-Glicentin. Immunofluorescence microscopy of Fc-OXM showed co-localization with CgA-positive granules (Figure 5C), and quantification of Pearson's correlation coefficient showed that Fc-OXM had a significantly greater value than Fc alone ($p<0.01$; Figure 5C, D). Therefore, the sorting signal within glucagon is sufficient to direct oxyntomodulin to granules. These results are consistent with the hypothesis that GRPP masks the glucagon sorting signal in the context of glicentin (Figure 1), thus providing a mechanism by which glicentin is not sorted to granules.

3.3.8 The Dipolar α-helix can bind to chromogranin A in a pH- and calcium-dependent manner

Since CgA is the most abundant protein constituent of PC12 secretory granules, I tested whether CgA could specifically bind the dipolar α-helix. To test for a specific reaction in conditions that mimic the lumen of mature secretory granule, pH 5.5 and 10 mM calcium, two sets of pull-downs were examined, for a total of four conditions. PC12 cells transfected with a panel of Fc constructs were immunoprecipitated, using either a neutral or pH 5.5 sodium acetate buffer, with or without 10 mM added calcium. The only condition that could pull down CgA by immunoprecipitating for Fc, was the pH 5.5 and 10 mM Ca$^{2+}$ condition (Figure 6). CgA was not pulled down using untransfected PC12 cell lysate, nor with Fc alone, or Fc-LP glucagon. In contrast, CgA bound most strongly to Fc-WT glucagon, Fc-GLP-1 (7-37) and Fc-MPGF, with minimal detected binding to Fc-WT GLP-2 and Fc-Glicentin. These results suggest a specific interaction between CgA and the dipolar α-helix sorting signal.
**Figure 6.** Identification of CgA as a sorting receptor. Immunoprecipitation for an Fc-fusion protein, and immunoblot for CgA. Western blot image is from co-immunoprecipitation in a pH 5.5 sodium acetate buffer with 10 mM Ca\(^{2+}\). No co-immunoprecipitated complexes were detected under neutral pH, or without 10 mM Ca\(^{2+}\) (total of 4 conditions tested). At right is a CgA immunoblot from 10 µg lysate of untransfected PC12 cells. + symbol indicates relative quantity of immunoprecipitate.

### 3.4 Discussion

Highly efficient sorting of proglucagon is required for the maturation of the proglucagon-derived peptides, and subsequent storage within secretory granules. Proglucagon is a unique prohormone from the perspective of its structural organization. Several prohormones, such as pro-thyrotropin-releasing hormone (pro-TRH) and pro-gonadotropin-releasing hormone (pro-GnRH), have structured prodomains, while the active hormone domain(s) are completely disordered (31). In contrast, proglucagon exhibits disordered prodomains (GRPP, IP-1 and IP-2), with mostly ordered hormone domains, as our previous work has shown (33). Additionally, the sequences of glucagon, GLP-1 and GLP-2 are highly conserved with respect to their charge distribution (32).

With this information in hand, we wished to characterize how proglucagon is targeted for regulated secretion by identifying the relevant sorting signals encoded within the ordered hormone domains of proglucagon. We constructed fusion proteins linking each PGDP to a reporter, Fc. Our results demonstrate that both glucagon and GLP-1 contain dipolar α-helices in which charged residues are distributed around hydrophobic patches, and that these helices direct sorting to granules. In contrast, GLP-2, which contains an α-helix that is not polarized, is very inefficiently sorted. Surprisingly, the sorting of glicentin, which contains the sequence of glucagon and therefore the identified dipolar α-helix, was inefficient, while MPGFR maintained its sorting efficiency. Oxyntomodulin was sorted efficiently to secretory granules, thus demonstrating that the N-terminal sequence of
glicentin masked the sorting signal contained within the α-helix of glucagon. We conclude that proglucagon contains two sufficient sorting signals contained within the sequences of glucagon and GLP-1, in the form of a dipolar α-helix, and that the α-helix of glucagon is masked after proglucagon is processed to glicentin.

In our previous studies of proglucagon trafficking using Neuro2a cells, our index of sorting efficiency was the co-localization between proglucagon and the cis/medial-Golgi marker, p115 (33). The high correlation value of R18Q-proglucagon led us to conclude that the dibasic R17R18 sequence within glucagon could contribute to sorting. In the present study, we calculated co-localization of the Fc constructs with the granule-resident protein CgA. Here, a high correlation reflected more efficient co-localization in granules, indicating that the R17R18 sequence may not be a factor in the sorting of proglucagon to granules, or that it may be cell type-specific. However, it is important to note that the sorting of the α-helix mutant of glucagon was calculated to be inefficient in both systems, indicating that the α-helix within glucagon is a primary sorting signal for proglucagon regardless of the cell type.

Previously identified α-helical sorting signals indicates that their amphipathic nature directs sorting of prohormones and their processing enzymes to the regulated secretory pathway. Prohormones containing granule-targeting amphipathic helices include pro-somatostatin (24) and pro-CART (25). The sorting signals of the prohormone processing enzymes PC1/3 (53, 55), PC2 (56), and CPE (57) are also amphipathic α-helices. Our previous work showed that reducing the α-helical content within glucagon reduced proglucagon sorting efficiency in Neuro2a cells (33). We now show that proglucagon contains two sorting signals in the form of non-amphipathic α-helices with a unique arrangement of hydrophobic patches and charged residues. Dikeakos et al. addressed sorting determinants by using synthetic α-helices, finding that the tested amphipathic helices with a charged face, or a non-amphipathic helix with a substantial hydrophobic patch and segregated charged residues, were efficiently sorted to granules (27). They inferred two important features of helical sorting signals: segregation of charged residues from hydrophobic patches is essential; and the degree of hydrophobicity correlates well
with sorting efficiency. While synthetic α-helices were sorted with as few as 5 charged residues (27), our data showed that as few as three charged residues within the sequences of glucagon and GLP-1 can direct sorting. The pIs of the helices within glucagon and GLP-1 are more similar than GLP-2 to the granule lumen environment (pH 5.5), possibly aiding their targeting to granules. Our hydrophobic cluster analyses show a dipolar charge distribution segregated from hydrophobic patches within glucagon and GLP-1, which were able to sort to granules. In contrast, the helix within GLP-2 has slightly different characteristics; while the nature of the hydrophobic patches are identical to those in glucagon and GLP-1, the charge distribution is not dipolar, consisting of only negatively charged residues along the helix. This difference resulted in very inefficient sorting, suggesting that charge distribution is more important than hydrophobicity for the non-amphipathic α-helices of proglucagon. We may now estimate the minimal hydrophobic domain required for sorting, in which a large, contiguous hydrophobic face (27), can be reduced to two dis-contiguous patches of 3 to 4 residues on opposing faces of the helix. This inference is supported by the recent finding that the pro-CART helix contains a smaller hydrophobic face relative to synthetic helices (25). This underscores the importance of the α-helix as a platform for sorting signal construction in general, and in proglucagon, the sorting information is encoded by the dipolar distribution of electric charge in relation to hydrophobic patches along the helix surface.

The differences in sorting efficiency between glicentin and MPGF suggest a context-dependent regulation of sorting when considering that glicentin does not efficiently sort to granules, despite containing a sorting signal within the sequence of glucagon. After proglucagon is initially processed at the inter-domain cleavage site, K70R71, glucagon is flanked by the sequences of GRPP and IP-1 (Figure 1). We used the PSI-PRED server (37) to analyze the predicted secondary structure of glicentin, and it revealed that IP-1 is disordered when not joining glucagon to GLP-1, a characteristic of omega loops (58). The N-terminal GRPP domain is also highly disordered, and enriched in acidic residues. Our previous model of proglucagon shows this disordered region masks the basic N-terminal residues of the glucagon helix (33), and the present study confirms this masking by showing that removal of the GRPP domain results in the targeting of Fc to granules.
Conformational masking has been demonstrated in moesin, in which an α-helical domain regulates the degree of unmasking between its N- and C-terminal ligand-binding domains (59); and in the prohormone protachykinin, in which the negatively charged pro-region masks the positively charged product, calcitonin gene related peptide (60). We now show that glicentin experiences a similar conformational masking by GRPP. On the other hand, MPGF experiences no such masking since IP-2, which links the helices of GLP-1 to GLP-2, appears to be partially helical (5, 33), thus maintaining the availability of the GLP-1 helix for efficient targeting to granules.

Figure 7. Schematic representation of proglucagon sorting and processing in α and L cells. Proglucagon (red bars) is synthesized in the ER and transported through the Golgi to the trans-Golgi network (TGN). Our data support the hypothesis that proglucagon is first sorted to immature secretory granules (ISGs) via dipolar α-helices within glucagon and GLP-1, and then cleaved to glicentin and MPGF (squares), possibly by furin. Within mature secretory granules (SGs), the prohormone processing enzyme PC2 processes glicentin to glucagon in α cells, while PC1/3 cleaves glicentin and MPGF to yield oxyntomodulin, GLP-1 (7-37) and GLP-2 in L cells.
This asymmetry between glicentin and MPGF trafficking presents interesting implications for the temporal relationship between proglucagon processing and sorting. It is well documented that proglucagon processing begins with the early cleavage event at the interdomain site, K70R71 (5, 9); in alphaTC1-6 cells, glicentin and MPGF were detected at 30-45 min via a pulse-chase paradigm (4). Our previous work has shown that mutation of K70R71 reduced the efficiency of proglucagon sorting in Neuro2a cells (33), and this result led us to conclude that processing may occur before sorting. However, the present study does not support this conclusion. If proglucagon processing occurs prior to sorting, our present model would predict that glicentin would be sorted inefficiently, which would impact the production of glucagon in α cells. Therefore, we now propose that the K70R71 site simply acts as another sorting signal, and together with the α-helices of glucagon and GLP-1, targets intact proglucagon to granules, whereupon processing to glicentin and MPGF occurs, as illustrated in Figure 7. Our model also suggests that the two α-helical sorting signals are functionally redundant, perhaps reflecting an evolutionary selection towards a high sorting efficiency for proglucagon. That the sequences of glucagon and GLP-1 are highly conserved (32) lends evidence to this reasoning. This is in contrast to pro-TRH, where PC1/3-mediated processing early in the secretory pathway is required for efficient sorting to distinct sub-populations of granules (61, 62).

Identification of sorting signals within proglucagon gives rise to the question of potential binding partners, or sorting receptors. While the results of Chapter 2 indicated a role for CPE in α cells, I also tested the possibility that CgA could act as a receptor because it is the most abundant soluble granule protein in both α and PC12 cells. CgA could only be co-immunoprecipitated using Fc-constructs at mildly acidic and millimolar calcium conditions that mimic the mature granule chemical environment, and that the interaction was specific to the dipolar α-helix and the quantity of complex that was pulled down correlated well with co-localization analyses. Therefore it is promising that CgA may be a sorting receptor for glucagon in α cells. The absence of the processing enzymes PC1/3 and PC2 from PC12 cells supports our previous findings (33) that neither enzyme plays a role in the sorting of full-length proglucagon. The amphipathic α-helices identified within
prohormone processing enzymes, PC1/3 (26, 63), PC2 (56), and CPE (16), are known to associate with the cholesterol-rich domains of granule membranes. However, we could not demonstrate binding of purified proglucagon to liposomes (*data not shown*), and therefore hypothesize that proglucagon may bind to granule proteins. It is possible that granins bind prohormones, such as POMC (64). We have some evidence that proglucagon sorting involves interaction with CPE in α cells (33). Studies investigating the roles of other granin proteins in sorting proglucagon are currently underway.

In conclusion, we have shown that proglucagon contains two dipolar non-amphipathic α-helices with relatively small hydrophobic faces that act as sorting signals for entry into secretory granules of endocrine cells. Our data support a mechanism by which proglucagon is sorted to granules prior to the initial cleavage event that results in the production of glicentin and MPGF (*Figure 7*). That these sorting domains lie within the ordered domains of encoded proglucagon-derived peptides, and not in a disordered prodomain that characterizes many other prohormones, highlights the unique sorting “signature” of proglucagon, and further emphasizes the disparate nature of sorting signals that lie within prohormones and other proteins destined for the secretory granules of the regulated secretory pathway.

### 3.5 References

44. Chidgey, M. a, and Harrison, T. M. (1990) Renin is sorted to the regulated secretory pathway in transfected PC12 cells by a mechanism which does not require expression of the pro-peptide. Eur. J. Biochem. 190, 139–44
Chapter 4

Practical Co-localization Analysis for Quantitative Fluorescence Microscopy

Statement of Author Contribution

This review is wholly the work of Mr. Guizzetti.

4.1 Introduction

In the previous two chapters, I have used quantitative fluorescence microscopy to objectively examine the sub-cellular trafficking of proglucagon in situ. The conclusions of these chapters have heavily relied upon co-localization analyses of these images to draw conclusions about the varying degrees of sorting efficiency along the secretory pathway. Image quantification allows the researcher to extract quantifiable data, sometimes from biologically variable samples or using limited image data, that can be then be used for hypothesis testing to reach accurate and precise conclusions. Increasingly more cell biology studies are using quantitative fluorescence microscopy techniques to examine the function and the associations between two labelled molecules within specific intracellular compartments, most often proteins. Co-localization analysis remains the most common image analysis tool for confocal immunofluorescence microscopy in cell biology research. While super-resolution studies are still less common than those using the more accessible confocal microscope, it is useful to discuss the practical application and interpretation of co-localization in diffraction-limited image sets. Despite the literature already devoted to the use of co-localization, there are still new studies published that claim to analyze co-localization, or fail to make any analysis and instead switch a qualitative descriptions, and in general, these studies are not rigorous in their analytic methodology. To complicate matters, researchers will sometimes alternate between their intended meaning of the word “co-localization.” The usage changes from qualitative descriptions of overlapping/coincident fluorescence signals, or if calculating a co-localization, still choose to use statistics that are less than ideal for the application, or
they may struggle with interpretation of the result and return to vague qualitative descriptions of results.

4.1.1 Principles of Digital Image Analysis

Image quantification is fundamentally a trade-off between adequate resolution of the specimen such that the details of interest are resolvable, and acquiring high signal-to-noise (SNR) images so that the digital image accurately represents the truth of the fluorescent labels within the cell. High resolution images are a challenge, as the fluorescence intensity is dispersed over multiple pixels, potentially reducing the SNR. Resolution may suffer when fluorophores are not very bright, is too diffuse or when rapid photobleaching is a concern. The SNR of the optical system can vary greatly, but care should be placed in acquiring high SNR images to make quantitation reliable and accurate. Here I will briefly cover the essentials of digital image formation, and good practices for analysis.

Resolution in optical imaging is a function of the light blurring due to the point-spread function of the optical system, as determined by the wavelengths being imaged, and the specific properties of the optics being used in the microscope (e.g., magnification, numerical aperture, photomultiplier, digital sensor, detector electronics). Confocal microscopes employ a pinhole in the light path just prior to the detector, rejecting much of the out-of-focus light, and can greatly reduce the width of the point-spread function. In practice, to achieve an increase in resolution, the fluorophore must be both abundant and bright enough to allow for a narrowed pinhole size, or else risk photobleaching by increasing pixel dwell time. The width of the PSF determines the lateral resolution of the microscope, and the resolution limit can be described as \( r_{xy} = \frac{0.61 \lambda}{NA_{objective}} \), where \( \lambda \) is the emitted wavelength, and NA is the numerical aperture of the objective. For fluorescence microscopy, the height of the PSF determines the axial resolution, described as \( r_{ax} = \frac{2 n \lambda}{NA_{objective}^2} \), where \( n \) is the refractive index of the mounting media (usually around 1.4-1.5). In fluorescence imaging (wide-field or confocal) the limiting factor for resolution is the numerical aperture of the objective lens, and typically, the resolution limits are \(~200-250\) nm laterally, and \(~550-700\) nm axially. These limitations should be
considered for the structures that need to be resolved for protein localization and intracellular trafficking.

In order to convert from analog light into a digital image, a fundamental hardware component is the photomultiplier, converting the incident photons into a digital signal. As an image is acquired, the laser beam is raster scanned or swept across the sample, and the emitted light is collected, amplified by the photomultiplier, converted to a digital signal and is registered as a single pixel. Therefore, each complete image is formed by the sequential acquisition of each pixel. It is good practice while doing optical imaging to increase SNR by monitoring the live exposure histogram of an acquired image in real-time. By adjusting emitted light intensity (e.g., by increasing laser power, electronic gain, pinhole aperture), the PMT will detect more incident photons, translating to a greater discrete pixel intensity and utilizing as much of the image dynamic range as possible. This permits more accurate and precise quantification because the collected image data will better represent the specimen. Care should be taken such that no pixels are over-saturated due to over-exposure, as this will lose higher intensity pixel information. Conversely, the minimum pixel intensity should be just above the adjusted black value of the PMT, to ensure the emitted photons are properly represented. The quality of downstream image processing is only as good as the final image produced by the entire digital microscope system, and digital images are more sensitive to imperfections than compared to what our eyes perceive.

Image processing are “off-line” techniques used to correct for some deficit in the collected images. These techniques may improve image SNR by removing background (segmentation), filtering out noise, or using deconvolution. Some techniques can be to reduce size or dimensionality of the data, such as drawing regions-of-interest, or making 3D projection images, while others can be to interpolate new information such as 3D volume rendering and reconstruction. This is not an exhaustive list, but gives the reader ideas of some image processing techniques. The researcher should be aware that raw images are data, and must be treated as such. When performing any image processing technique, it is essential that they be made on duplicates of the original images so that the raw data can always be accessed, and it necessitates the creation of a digital backup. Any
of the techniques performed must also be stated in the publication's methods section for transparency and reproducibility.

There are also some practical tips to keep in mind in how computers represent the intensity of each pixel. A pixel is represented by a computer as a number of bits (usually 8-, 12- or 16-bits), called a bit-depth. The bit-depth represents the range of possible values that a pixel can have. For example, an 8-bit image can take $2^8$ possible values (256 shades of grey), and is the dynamic range of all pixel values in image. When working with digital images, the bit-depth must never be reduced or truncated. Doing so compresses the dynamic range of the image, and “clips”/truncates the high intensity pixel data. This high-intensity detail is likely to contain the signal-of-interest, and would represent a degradation of image and data quality. As a practical example, a 16-bit image (65,536 levels of grey) contains pixel intensities in the range of 10-40,000, and when converted to 8-bit, clips all pixels at or above 256 to the same intensity, resulting in a severe loss of image detail and apparent over-saturation of the clipped 8-bit image. Similarly, histogram compression results in an altered image contrast between pixels. Acquired images are often converted to tagged-image file format (TIFF) files for analysis because all pixel information is maintained in its raw state and does not allow for “lossy” compression schemes, unlike JPG and other formats, that discard information to produce a smaller image file. Image processing should be done on the raw image data so that the researcher is assured the results are representative of the image, and not an artifact of the image file format. Several commercial and free software packages conduct image conversion and analysis. The techniques that are applied to one image (of one image channel) must be applied to all image channels for consistency of data quality. This is important for conducting co-localization studies, which require two (or more) image channels.

### 4.2 Quantitative Multi-Channel Image Analysis and Co-localization

For consistency, it will be useful to first define what is meant by co-localization, and what it can tell the user regarding their species of interest (usually proteins). Ideally, the researcher wants an image analysis tool that can determine whether two proteins are
physically interacting with each other, and would turn to co-localization as a quantitative measure of interaction. However, diffraction-limited microscopy does not offer sufficient resolution to quantify such a physical interaction due to the nature of image acquisition. Instead, suitable microscopy techniques to validate inter-molecular interaction must utilize the physical phenomenon of Förster resonance energy transfer (FRET) (detecting interaction with spatial resolution of <10 nm), or bimolecular fluorescence complementation (BiFC) (producing an active fluorophore upon protein complementation) (for reviews see (1, 2)). Additionally, the increasing adoption of super-resolution microscopy is better suited for the quantification of the percent of interacting proteins. The much greater spatial resolution affords the development of new algorithms and more sophisticated tests of interaction, such as object-based or proximity-based methods (3–6), or adopting techniques from statistical physics to account for several proteins of interest, for example (7).

Co-localization can report on the degree of spatial co-distribution. Quantitative co-localization accounts for two components: simple overlap (sometimes called co-occurrence), and the tendency to co-distribute (or co-segregate). Therefore, fluorescence co-localization is best suited for determining the tendency to associate with the same cellular or molecular structure, as opposed to measuring the association of freely diffusing and interacting molecules. Simple overlap alone is insufficient to determine co-localization, and various overlap coefficients have been used as the basis for evaluating co-localization. Worse yet, “judging by eye,” for the degree of yellow signal when red and green signals overlap, is highly subjective and offers meaningless interpretation. This can best be demonstrated by the optical illusion in which the perceived colour contrast of one colour varies within the context of its background, in other words, seeing colours that are really not there (e.g., Munker illusion1) (8). Incorporation of spatial association adds robustness to the co-localization methods, in that two (or more) bodies that are co-localized should be observed in the same location and should have concordant spatial locations. In studies of intracellular trafficking of cargo proteins, the most common approach is to fluorescently label one cellular compartment, such as the cis- or trans-

---

1 This optical illusion can be generated using FIJI by going to File > Open Samples > Spirals, and observing that the blue and green spirals are actually the same colour.
Golgi (e.g., (9–11)), or specific vesicular compartments, such as secretory granules (e.g., (10, 12, 13)), or lysosomal/endosomal compartment (e.g., (14–16)). In the context of examining protein intracellular trafficking through the secretory pathway, accurate quantification can be used to examine steady-state distribution of protein localization when compared to transient co-localization in the Golgi, or it can be used to assess efficiency of storage within a terminal destination within secretory granules. Live cell imaging can also extend this analysis to reveal kinetics of movement through distinct sub-cellular compartments, a question that is increasingly more relevant with confocal microscopy (17) and super-resolution imaging (14). Accurate co-localization quantification can lead to new insights for cell biological processes, such as the preference to secrete insulin from younger granules of pancreatic beta-cells (17). However, it is not justified by co-localization alone that just because two proteins exist in the same compartment that they should necessarily be interacting, just as it is unreasonable to expect that co-localization can always quantify the stoichiometric ratios of interacting proteins. It is important to understand the capabilities and limitations of such quantification techniques.

Here I present a framework for conducting a quantitative co-localization study to measure protein co-localization and intracellular trafficking. To better inform the user on how to interpret various commonly used correlation parameters, I have generated simulated image sets containing point source objects, in order to provide a cell-free context for understanding parameter interpretation and evaluating their performance. The simulated image sets are also useful to control the statistical distribution of pixel intensity and spatial location. I will make several assumptions about the reader in order facilitate the discussion. Firstly, the reader should be familiar with good confocal imaging practices (a good primer on common pitfalls in confocal microscopy is found at ((18))). Images are captured in which the “signal” is well above the “background” and “noise” due to cellular and substrate auto-fluorescence (e.g., ((19, 20))), controlling uneven illumination (e.g., (21)), and the fluorophores are sufficiently separated in their excitation/emission spectra to prevent fluorophore cross-talk and bleed-through. Furthermore, image sets must be acquired in uncompressed image formats. A good review on the digital aspects of image files and microscopy can be found by (22).
Both commercial (e.g., Imaris, Image Pro, Axiovision, Colocalizer Pro) and free software packages (e.g., FIJI/ImageJ) that measure co-localization include several different algorithms from which the user is expected to know the correct method to use. Especially in earlier co-localization studies, researchers often presented several computed measurements alongside their immunofluorescence images, not deciding on a single “best” algorithm for their data. Ideally, the choice of co-localization parameter should be an objective measurement of true co-localization of 2 (or more) labelled molecules. The co-localization statistics should be 1) sensitive to the effect of one molecule co-distributing with the other; 2) insensitive to background noise; 3) insensitive to relative changes in fluorescence intensity channels (e.g., scale independent); and because no two cells are alike, 4) the statistics should be insensitive to changes in overall cell morphology, such that they are shape independent. I will discuss these four points to help interpret the multitude of co-localization measurements.

4.2.1 Mathematical parameters used for co-localization analyses

This section discusses the parameters used for quantitative co-localization and their mathematical definitions. The most commonly used parameters for these studies are Manders' overlap coefficient and Pearson's correlation coefficient, and some less common correlation parameters (listed in Table 1). If the reader is familiar with the mathematical definitions of the parameters, they can skip ahead to the next section for a discussion of the merits of these parameters.

Table 1. Mathematical parameters used for quantitative co-localization analyses.

<table>
<thead>
<tr>
<th>Coefficient Name</th>
<th>Measured quantity</th>
<th>Equation</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlap</td>
<td>Simple overlap</td>
<td>1</td>
<td>(23)</td>
</tr>
<tr>
<td>Manders' overlap (MOC)</td>
<td>Intensity-weighted Overlap</td>
<td>2,3</td>
<td>(23, 24)</td>
</tr>
<tr>
<td>Pearson's correlation (PCC)</td>
<td>Correlation between channel intensity</td>
<td>4,5</td>
<td>(25)</td>
</tr>
<tr>
<td>Spearman's rank-correlation (SRC)</td>
<td>Rank-ordered correlation between channel intensity</td>
<td>4</td>
<td>(26)</td>
</tr>
<tr>
<td>Kendall's tau</td>
<td>Rank-ordered correlation between channel intensity</td>
<td>6,7</td>
<td>(27, 28)</td>
</tr>
</tbody>
</table>
Li’s intensity correlation quotient (ICQ)        Binary/sign-ordered correlation between channel intensity  8 (29, 30)

The simplest (and least reliable) parameter to calculate co-localization would be to measure the amount of overlap between any pair of signals (23, 24). The greater the overlap, the greater the coefficient. The overlap coefficient is the ratio of all pixel product intensities, normalized by the sums of each channel absolute pixel intensities.

\[
OCC = \frac{\sum_{i=1}^{N} S_{1,i} S_{2,i}}{\sum_{i=1}^{N} |S_{1,i}| \sum_{i=1}^{N} |S_{2,i}|}; OCC \in [0,1] \tag{Equation 1}
\]

Where \( S_{1,i} \) and \( S_{2,i} \) represent the \( i \)-th pixel value in the image. An improvement to this method was developed by Manders, and defined the spatial overlap of one channel on the other as weighted by their absolute intensities, and normalized by the sum of pixel intensities of a given channel.

\[
MOC = \frac{\sum_{i=1}^{N} S_{1,i} S_{2,i}}{\sum_{i=1}^{N} (S_{1,i})^2 \sum_{i=1}^{N} (S_{2,i})^2}; MOC \in [0,1] \tag{Equation 2}
\]

Where \( S_{1,i} \) and \( S_{2,i} \) represent the \( i \)-th pixel value in the image. However, this method is still an “overlap” coefficient. A later improvement to the Manders' overlap coefficients allow for the definition of a channel-specific threshold value, such that pixels above their respective thresholds are considered “co-localized.”

\[
M_1 = \frac{\sum_{i=1}^{N} S_{1,i,\text{coloc}}}{\sum_{i=1}^{N} S_{1,i}}; M_2 = \frac{\sum_{i=1}^{N} S_{2,i,\text{coloc}}}{\sum_{i=1}^{N} S_{2,i}}; M_1, M_2 \in [0,1] \tag{Equation 3}
\]

Where \( S_{1,i} \) and \( S_{2,i} \) represent the \( i \)-th pixel value in the image, and the \( S_{1,i,\text{coloc}} \) and \( S_{2,i,\text{coloc}} \) are constrained such that \( S_{2,i} > 0 \) and \( S_{1,i} > 0 \), respectively. Confusingly, the use of a threshold for MOC is sometimes called “object-based,” but does not actually take
into account any 3D rendered volumes whatsoever.

The first introduction of the Pearson's correlation coefficient (PCC) was by (24). The PCC is well known statistical measure of linear correlation and dependence between two variables (25). In the context of immunofluorescence images, it is the sum of products of mean-subtracted pixel intensities, scaled by the individual channels' sum of squared differences. In other words, the correlation is computed on the paired pixel values.

\[ r_p = \frac{\sum_{i=1}^{N} (S_{1,i} - S_{1,av})(S_{2,i} - S_{2,av})}{\sqrt{\sum_{i=1}^{N} (S_{1,i} - S_{1,av})^2} \cdot \sqrt{\sum_{i=1}^{N} (S_{2,i} - S_{2,av})^2}}; \quad PCC \in [-1, 1] \]  

\[ \text{Equation 4} \]

Where \( S_{1,i} \) and \( S_{2,i} \) represent the \( i \)-th pixel value in the image, and \( S_{1,av} \) and \( S_{2,av} \) are the average pixel intensities in each channel. A threshold condition can be applied to the Pearson's correlation coefficient, such that \( S_{1,i} \) and \( S_{2,i} \) in the numerator are constrained to be above a channel specific threshold.

\[ r_p = \frac{\sum_{i=1}^{N} (S_{1,i,\text{thresh}} - S_{1,av})(S_{2,i,\text{thresh}} - S_{2,av})}{\sqrt{\sum_{i=1}^{N} (S_{1,i} - S_{1,av})^2} \cdot \sqrt{\sum_{i=1}^{N} (S_{2,i} - S_{2,av})^2}}; \quad PCC \in [-1, 1] \]

\[ \text{Equation 5} \]

In general, adding a threshold imposes the condition on the pixels that in order for a pair of pixels to be considered, both pixel intensities must be greater than their respective channel's threshold value. This has the effect of dividing the scatter plot of paired pixel intensities into quadrants, with co-localized pixels residing in the upper-right quadrant (above both channel thresholds).

In the general case, Spearman's rank-correlation coefficient (26) is equivalent to the Pearson's correlation coefficient (Equation 5), except that the pixel intensities are replaced by the pixel rank within each image channel. Ties in rank are resolved by calculating the average rank of the tied positions. For example, four pixels each have the same intensity, and they would be assigned the rank orders of 3, 4, 5, and 6. The tie is broken by computing the rank average, 4.5.
Kendall's tau (sometimes called \textit{tau-a}) \textup{(27)} is similar to the Spearman calculation, in that each image's pixel values are first assigned to their rank order, and makes no assumption of sampling distribution.

\[ t_a = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}; t_a \in [-1,1] \]  \hspace{1cm} \text{(Equation 6)}

Where \( n \) is the number of observed pairs (e.g., the number of pixels in each image), the denominator represents the total number of pair combinations; and \( n_c \) and \( n_d \) correspond to the number of concordant and discordant pairs, respectively. Any pair of pixels \( (x_i,y_i) \) and \( (x_j,y_j) \) is said to be\textit{ concordant} if the ranks for both comparisons agree (\( x_i > x_j \) and \( y_i > y_j \), \textit{or} \( x_i < x_j \) and \( y_i < y_j \)). If the ranks do not agree \textit{(i.e.,} \( x_i > x_j \) and \( y_i < y_j \), \textit{or} \( x_i < x_j \) and \( y_i > y_j \)) then the observed pair is\textit{ discordant}. However, this formulation makes the implicit assumption that no ties exist in rank order. Therefore, in the case of ties, they are first resolved as in SRC (assigning the average rank-order) and the \( b \)-variation of tau \textup{(28) pp. 34-48} is calculated instead.

\[ t_b = \frac{n_c - n_d}{\sqrt{\left[ \frac{n(n-1)}{2} - \sum \frac{t_i(t_i-1)}{2} \right] \left[ \frac{n(n-1)}{2} - \sum \frac{u_i(u_i-1)}{2} \right]}}; t_b \in [-1,1] \]  \hspace{1cm} \text{(Equation 7)}

Where the numerator is the calculated as \textit{tau-a}, and \( t_i \) and \( u_i \) are the number of observations tied at a particular rank of \( S_1 \) and \( S_2 \), respectively. At present, FIJI's Coloc 2 plug-in computes \textit{tau-b}.

Kendall's tau improves on Spearman's rank-correlation coefficient with an easier interpretation when the null hypothesis that two images are independent of each other is rejected. In the context of an image, no pixel can be considered to be objectively more important in rank than another, so if two images tested for rank-correlation, Kendall's tau
reports the degree of similarity between two images conditioned on the assumption that they are independent. The possible advantage to using Kendall's tau over Spearman's rank-correlation coefficient is that Kendall's tau correlation can be assumed normal in real applications under the null hypothesis that all possible rank-ordered pairs occur with equal likelihood (27). Since Kendall's tau calculation of rank-pairs is exhaustive, the complete distribution can be quickly calculated (27), using a computer to verify the Normality assumption, allowing the computation of approximate confidence intervals. These qualities make it a strong candidate to consider alongside more accepted parameters of co-localization for co-localization analysis, and should also be applicable in super-resolution image studies.

The original version of this test are based on interpretation of scatter plots in the intensity correlation analysis (ICA) (30, 31). The ICA method is based on the underlying assumption that pixel intensity in one image is symmetrically distributed (not necessarily from a Normal distribution), such that, for an image which as random (or mixed) staining, the sum of mean differences is zero ($\sum S_i - S_{av} = 0$). By extension, the two channel version assumes that the joint probability distribution of pixel pairs is also symmetrically distributed, such that $\sum (S_1,i - S_{1,av})|S_2,i - S_{2,av}| \approx 0$. When individual pixel intensities are plotted against the product $(S_1,i - S_{1,av})(S_2,i - S_{2,av})$, a scatter plot is produced, that can be visually examined to judge correlation of immunostaining. A scatter plot that is biased to the right of vertical line at $(S_1,i - S_{1,av})(S_2,i - S_{2,av}) = 0$ is interpreted as correlation, while a scatter plot mainly along this vertical line represents independent staining, and a bias to the left represents segregation. However, because this interpretation is subjective, a it required the computation of Li's intensity correlation quotient (ICQ), in order to allow for a hypothesis testing using the sign-test.

$$ICQ = \frac{\sum \left| (S_1,i - S_{1,av})(S_2,i - S_{2,av}) \right| > 0 \cdot \left| (S_1,i - S_{1,av})(S_2,i - S_{2,av}) \right| > 0.5}{\sum \left| (S_1,i - S_{1,av})(S_2,i - S_{2,av}) \right|} \in [-0.5, 0.5]$$  \hspace{1cm} (Equation 8)

The ICQ computes the sum of the product of concordant pixels (e.g., those that
correlate), over the sum of all products of pixel pairs (e.g. those that correlation and those that don't). In a sense, Li's ICQ is a binary correlation, and just as the other parameters presented here, does not measure co-localization. The subtraction of 0.5 is to centre the range about zero, with ICQ values less than zero representing segregation, and greater than zero representing correlation, and thus co-localization. At present, Li's ICQ is only implemented in FIJI.

4.3 Methods
Several parameters have been proposed to measure co-localization, the two most common of which are Pearson's correlation coefficient (Table 1, Eq. 4,5) and Manders' overlap coefficient (Table 1, Eq. 2,3) for quantitative fluorescence microscopy, as determined by PubMed searches reporting co-localization. For reasons detailed below, this review is focused on the PCC, and less common but potentially useful rank-based parameters – SRC, Kendall's tau and Li's ICQ. In the interest of practising quantitative co-localization, I recommend against using the overlap and Manders' coefficients (Table 1, Eq. 1-3). The overlap coefficient makes no attempt to normalize the channel intensities, and is simply too sensitive to variations in noise, relative channel fluorescence intensity, and channel mean to be of any practical use (32). Therefore it is equivalent to creating a simple overlap, and the researcher is left to “measure by eye.” Similarly, inspection of the Manders' overlap coefficient formulation, and experimental results from Adler agree that it is best "to abandon the MOC and the related k1 and k2 pair of coefficients" (32) in favour of Pearson's correlation and rank/sign-based correlations or another of the parameters discussed herein. Their findings are also in agreement with the findings of others (11, 33) and the authors' experience, with real and simulated data showing poor performance with the MOC and complete insensitivity to the degree of correlation. For these reasons, MOC will not be considered further here, in favour of more robust parameters. The inclusion of Li's ICQ and the rank-based correlation coefficients, SRC and Kendall's tau, are interesting but have not been considered extensively in the literature. Li's ICQ (Eq. 8) was first introduced a decade ago, and reduces the parameter to a binary comparison of pixel values. It has the potential to be as robust as the PCC to noise, background and differences in relative channel intensity. Similarly, the SRC and Kendall's tau are rank-based, and those may also offer similar
advantages. These three parameters, along with the classic PCC and Costes' automated threshold version of PCC (34) are compared in detail.

4.3.1 Image simulations and Co-localization Analysis

Simulated greyscale images were generated using a custom written script in the Python programming language (Python Software Foundation. Python Language Reference, version 3.2. http://www.python.org) using the Python Image Library (Pillow, version 2.5; http://github.com/python-pillow/Pillow). Each image set contains a pair of pseudo-coloured images to represent two fluorescence acquisition channels. All simulations contained randomly located point objects, which were subsequently convolved with a Gaussian kernel, to simulate the effect of point-spread function blurring of confocal microscopy image formation. To control for the percent of object co-localization, both simulated image channels shared a subset of the randomly generated objects for partial co-localization, or the same set of objects for perfect co-localization. Objects with 0% co-localization (segregated objects), had unique sets of random object locations. To simulate image noise, randomly generated Gaussian values were subsequently added to image sets. The ratio of mean object intensity to mean background noise was held constant to maintain signal-to-noise. For display purposes, the generated grey-scale images were then pseudo-coloured, for display purposes. Simulations were performed for 1,000 repetitions, and the calculated parameter values were averaged. Co-localization was measured using the Coloc 2 plugin of the FIJI/ImageJ software package (http://fiji.sc/) (35), and verified using a custom written Python script for selected co-localization statistics. Calculated correlation values for co-localization were averaged over all simulations and are presented as the mean value, and analyzed using the R software package (version 3.1; http://www.r-project.org/).

4.3.2 Image Acquisition of Multifluorescent Beads and Chromatic Aberration Calibration

Multi-fluorescent beads (Thermo-Fisher) 5.0 µm in diameter were imaged using a Zeiss LSM 510 laser-scanning confocal microscopy under 60x objective magnification. Microscopy was performed at the Biotron imaging facility at the University of Western Ontario. Images were imported into FIJI for measurement of each colour channel's x- and y-axial line profile. Overlay of red, green and blue line profiles show that the microscope
is calibrated for lateral chromatic alignment, and no chromatic defect is observed.

4.4 Results

4.4.1 Chromatic calibration of the microscope

Before collecting any images for co-localization analysis, the confocal microscope must be chromatically calibrated. Improper calibration leads to colour fringing artifacts and will already degrade the detection ability of co-localization from a systematic bias, most especially pronounced when trying to analyze punctate structures that correspond to secretory granules (as is commonly done). The easiest method is to use multi-fluorescent beads, which are both inexpensive and commercially available. (Core imaging facilities may even have these beads readily available for this purpose.) The fluorescent coatings of the beads are chosen to be compatible with the most common filter sets used in immunofluorescence imaging, typically blue (e.g., DAPI), green (e.g., FITC, AlexaFluor488), orange (e.g., Cy3, TRITC) and red (e.g., Cy5, AlexaFluor647). Using the same light path that will be used for capturing immunofluorescence images, acquire a few images of fluorescent beads in each channel that will be used for imaging. By drawing a line profile through the beads for each of the x- and y-axes, and then overlaying the profiles, any offset of the bead centre/edges can be measured, as in Figure 1. Ideally, this pinhole alignment need only be done once prior to beginning image acquisition on sample specimens. If there is an offset (in any direction) and images have already been acquired using the uncalibrated microscope, proper alignment can be restored by digitally translating one channel with respect to another, restoring alignment post-acquisition. In reality, there will always be some residual optical aberration left in the system, but chromatic and optical alignment is necessary to properly detect co-localization.
**Figure 1.** Verification of chromatic alignment. (A) A z-stack of 5.0 μm diameter multi-fluorescent beads were acquired for each colour channel. Line profiles drawn through the bead centre in (B) x- and (C) y-directions confirm that the edges of the red and green channels align, and the centre of the blue channel matches the centre of the bead.

### 4.4.2 Cytofluorograms, correlations, and co-localization

Co-localization should be thought of as a co-distribution of two labelled molecules, either on the plasma membrane, intracellular membranes or within the same intracellular compartment. In biology, co-localization is rarely perfect. Potential issues can arise if a protein exists in dynamic equilibrium either between different states of assembly/recruitment with a complex, or trafficking through different compartments, and experimental considerations such as incomplete molecule labelling, for example by non-optimal antibody affinity for its epitope or antibody cross-reactivity. How then to quantify such a correlation?

For all descriptive statistics and inferential analysis, *always* “plot the data” – image analysis is no different! A special type of 2D histogram/scatter plot often used for immunofluorescence images, called a cytofluorogram, plots the pixel intensities of one
channel against the other channel. These graphs can be viewed during real-time acquisition in some commercial packages (e.g., those offered by Zeiss or Leica) or post-acquisition in many software packages (e.g., FIJI). In this scatter plot, each data point is made up of spatially matched pixel intensity values, one from each image. Essentially, a cytofluorogram represents a 2D histogram of pixel matched values across the image set, with a variation of this to also show the intensity or height of each intensity pair as a heat map. If the researcher uses Pearson's correlation coefficient, it is exactly equal to the slope of the regression line through the cytofluorogram. Assessment of image data in this graph is superior to examining a simple overlay, because the user can immediately see the quality of image data, how much of the signal is not correlated or contains noise, and an initial assessment of co-localization.

Let's consider three conditions of overlap in Figure 2. First, a set of simulated objects that (by design) perfectly overlap and co-distribution (Figure 2A). Evaluating this cytofluorogram of perfect co-localization (Figure 2A'), I expect every yellow pixel to coincide with a magenta pixel, and their pixel intensity changes concordant with each other. As expected, all calculated parameters of co-localization, shown in Table 2, agree with the theoretical maximum value. Partial co-localization of simulated objects was achieved by translating the objects in one image by 1 pixel in both the x- and y-directions (Figure 2B). This results in a partial overlap of all objects, but should otherwise still be considered co-localized. It is instructive to notice that this scenario is also representative of uncorrected chromatic aberration, in which colour fringing effects will degrade correlation values. The cytofluorogram (Figure 2B') still shows positive correlation, indicating co-localization, albeit with added noise contributed by the non-overlapping regions. The corresponding correlation parameters in Table 2 all show good agreement with the perfect co-localization case, reflecting the notion of very good, but imperfect, co-localization.
Figure 2. Illustration of cytofluorogram showing full, partial or no co-localization. Computer generated images with synthetic point objects following a 2D discrete Gaussian blur (with variance 2 px) to simulate idealized point-spread blurring in the microscope's light path. Each image contains either 100 (A and B) or 200 (C) random objects, and are psuedo-coloured yellow and magenta (with their colour overlay). In the associated 2D cytofluorogram, regression lines are drawn in red. Conditions represent are (A) perfect co-localization between objects, (B) partial co-localization with noise due to chromatic aberration, and (C) perfect segregation (no co-localization). Pixels displaying only fluorescence in one channel lie along the x- or y-axes. See Table 1 for a list of co-localization parameters for each of these three scenarios.
Table 2. Performance of selected co-localization parameters under differing conditions of true co-localization and segregation.

<table>
<thead>
<tr>
<th>Condition</th>
<th>PCC</th>
<th>PCC (with threshold)</th>
<th>SRC</th>
<th>Kendall's tau</th>
<th>Li's ICQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical range</td>
<td>[-1,+1]</td>
<td>[-1,+1]</td>
<td>[-1,+1]</td>
<td>[-1,+1]</td>
<td>[-0.5,+0.5]</td>
</tr>
<tr>
<td>Perfect co-localization</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>0.5*</td>
</tr>
<tr>
<td>Partial co-localization (1 px shift)</td>
<td>0.88</td>
<td>0.85</td>
<td>0.93</td>
<td>0.88</td>
<td>0.46</td>
</tr>
<tr>
<td>Minor co-localization (2 px shift)</td>
<td>0.61</td>
<td>0.52</td>
<td>0.85</td>
<td>0.76</td>
<td>0.42</td>
</tr>
<tr>
<td>Perfect segregation</td>
<td>-5.0·10^{-2}</td>
<td>-7.6·10^{-4}</td>
<td>-0.23</td>
<td>-0.21</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Note: These values correspond to images shown in Figure 2. Values are presented as the mean over n=1,000 simulations. * values are presented as exact, and conform with their respective theoretical maximum.

Lastly, I consider perfectly segregated objects (Figure 2C). The exact same distribution of point objects have been horizontally translated so as to not overlap at all, and the cytofluorogram shows all magenta and yellow objects clustered along the axes (Figure 2C'). The slope of the regression line is -0.05 (approximately equal to 0), indicating good agreement with object segregation. With respect to the PCC, these values are approximately equal to 0, though in both conditions of partial co-localization and segregation, the un-thresholded Pearson's value over-estimates the degree of correlation. Also some authors have misinterpreted the Pearson's formula concerning perfect segregation as yielding a value of -1. However, the proper interpretation for no association (perfect segregation) would be a value of 0, indicating no linear co-dependence, while a value of -1 represents perfectly negatively correlated data. It is very rare for immunofluorescence images to contain truly negatively correlated labelled molecules, and therefore negative Pearson's correlation values much below 0 should be inspected more closely as it may be symptomatic of imaging artifacts (such as uneven illumination). In real immunofluorescence images, the “lobes” that extend along the axes of the cytofluorogram represent uncorrelated background signal (i.e., noise and auto-fluorescence). Lastly, Li's intensity correlation quotient (ICQ) (Table 1, Eq. 8) performed
the poorest of these parameters to distinguish co-localized from segregated objects, given its theoretical range is -0.5 to 0.5. While the intuition behind the ICQ is easy to comprehend, using this parameter may result in inefficient detection of co-localization differences, consistent with the findings of (32, 36). In the case of even less co-localization (Table 1), but not complete segregation, the values continue to trend toward no association, with the rank-based methods (Spearman's, Kendall's tau) maintaining power to detect co-localization from small offsets. Overall, the correlation parameters of Pearson, Spearman and Kendall preserve much of their dynamic range for quantification of co-localization.

4.4.3 Li's ICQ is insensitive to known co-localization

Previous attempts to estimate sensitivity of a co-localization parameter to various “ground truth” degrees of co-localization fixed the percent of co-localized point objects (23, 31). However, the observed effects are exaggerated when these tests are arranged as a regular grid for MOC, SRC and PCC, because they are susceptible to slight changes on a change of a few co-localized objects. To produce a more robust estimate when the set of co-localized objects is fixed (“ground truth”), point objects were randomly generated as in Figure 2, to remove any effect of objects arranged in a grid, while accounting for a small amount of random overlap, representing spurious co-localization. For this set of simulations, a fixed percentage of object co-localization was imposed, ranging from 0% to 100% (Figure 3). Li's ICQ displays the least sensitivity to co-localization, and at 0% object co-localization, yields a value slightly higher than reported in Table 1 (0.37 vs 0.32). This difference is not due to the number of objects (both have 200 random objects), but is instead showing that when a group of objects is random dispersed in the same image area (as in Figure 3), there may be minimal spurious overlapping, increasing the parameter value, but when objects are segregated in the image, the parameter is a little more sensitive to this separation.
Figure 3. Effect of object overlap on the measured correlation coefficients without background noise. 200 random point objects were generated, and Gaussian blurred to simulate optical blurring by the microscope's point-spread function. A predefined percentage of object overlap (from 0-100%) was imposed, such that "ground truth" object co-localization was fixed. For each level of object co-localization, 1,000 simulations were run, and the mean parameter values plotted. Connecting lines are shown as an aid for the eye.

4.4.4 Background and noise

Image background is from the specimen caused by several factors, the most common of which is auto-fluorescence (from cells or substrate coatings), mounting medium, the choice of glass, and other sources that uniformly offset (or add to) pixel intensity. Prior to the imaging session, the photomultiplier amplifier offset should be adjusted to a true black value. Diffuse fluorescence signal arising from the specimen is not typically considered valuable for co-localization, and should be accounted for in analysis by setting an appropriately high threshold value for co-localization. On the other hand, noise creates variance in the pixel intensity, and obscures the “real” value. Noise is caused by electronic sources in the detector electronics, as well as fluorophore cross-talk and bleed-through. Methods that specifically reduce Poisson noise (shot/detector noise) from
Electronic sources in confocal imaging were developed with good performance (37). Another common noise removal technique is median filtering (low-pass filtering) (38), which is scalable to 2D and 3D, but the tradeoff is a loss of resolution and only reduction, not elimination, of noise. The concept of signal thresholds was quickly introduced with co-localization measurements, and different methods exist. Coste's (34) proposed an automated iterative method to select a threshold using orthogonal regression, iterating through progressively lower threshold values for each axes of the cytofluorogram until the PCC (below threshold) becomes negative (it is implemented in some software packages, such as FIJI's Coloc 2 plugin). In most cases, researchers are interested in positive correlations, and this auto-threshold algorithm is appropriate in that case. An alternative approach (39) produces threshold by weighting the effect of each pixel to its contribution to both MOC and PCC, and is also applicable in the case when negative association is of interest. Thresholding methods are most effective for removing (constant intensity) background and only partially reduce the effect of noise.

4.4.5 Higher order image structure in images may negatively skew co-localization

For an ideal co-localization parameter, it would be beneficial if the value were linearly proportional to the degree of co-localization. Over the entire range of percent co-localization, the SRC, Kendall's tau, and PCC (without threshold) have an approximately linear response. The Pearson's using the Coste's auto-threshold shows some instability at <5% of co-localization, and this is because an acceptable threshold value could not be determined. This instability is possible (though not guaranteed) when the PCC has values very close to its extremes (-1 or 1) (40). However, the PCC (with auto-threshold) is stable and is the most sensitive across the majority of this range. Coincidentally, these three parameters also estimate the percent of overlap, though there is no mathematical basis for this in the general case because they are not guaranteed to have a linear response, even in this ideal simulation. (Here, this approximation is because of a uniform background level and equal object intensity.) While simulation results of Figures 3 and Table 1 agree when there is partial or perfect co-localization, it reveals a negative bias in the SRC and Kendall's tau for images that have regular, or higher ordered structures of interest, resulting a lower reported parameter that the case where no co-localization is
expected in overlapping regions of interest. In these cases, cross-correlation or auto-correlation methods (see Discussion) may offer a performance advantage. In biomedical application, randomly oriented and localized sub-cellular structures would be expected to have values closer to the 0% overlap values than the more negative extremes, arguing for the need to have an appropriate negative co-localization control to define the lower limit of the detection range for a particular biological system using a particular co-localization parameter.

4.4.6 Effects of gain and bias
The PCC, and its derivatives, SRC and Li's ICQ, are all robust (or insensitive) to effects of image gain and bias. This can be seen by substitution of the pixel value $S_{1,i} = a \cdot S_{1,i}$ in the case of gain, or $S_{1,i} = S_{1,i} + a$ in the case of bias, into the formula for Pearson's and observing that the effect cancels out. Likewise, these substitutions do not change the relative rank-order or sign-test, and therefore the SRC, Kendall's tau and Li's ICQ will not be affected. This finding has not been reported with respect to co-localization analysis. Therefore, the PCC, SRC Kendall's tau or ICQ are favourable to fluorescence microscopy in which digital offset is usually adjusted in each imaging session to set the black background level, and gain may need to be adjusted slightly for differences in antibody staining between markers.

4.4.7 Effect of noise
An ideal co-localization should be robust to random background noise and still detect co-localization by correlation. To determine how these parameters performed, I varying the type and amplitude of noise, and the signal-to-noise ratio when using point objects (as in Figure 4), calculated by ratio of the mean noise level to the mean object intensity. For confocal fluorescence images, having a minimum SNR of 20 is suggested for reliable object detection and quantification. However, confocal images often suffer from less SNR due to limiting the collected light of the pinhole. A high SNR image is produced in direct relation to the effective resolution of objects within the image, as well as intensity measurements, for which these correlation parameters rely on. To evaluate performance using low SNR conditions, image sets were generated with either severe ($SNR=2$) (Figure 4A) or moderate ($SNR=5$) (Figure 4B) degradation, in which it is either very
difficult or minimally sufficient to discern objects within the image (41). Under three scenarios of known object co-localization (0%, 50% and 100%), randomly generated noise was added to the images prior to computing each co-localization parameter, and the results are presented in Table 2. In the first, low SNR scenario (Figure 4A), random Gaussian noise is added to represent noise, relative to the true mean signal (specified by the coefficient of variation). When objects are randomly dispersed, and no co-localization is expected, the PCC, SRC, Kendall's Tau and Li's ICQ correctly report no correlation. As the degree of co-localization is increased to 100%, all parameters generally report increased correlation or association, as expected. However, the dynamic range of correlation as reported by SRC, Kendall's tau and Li's ICQ, under all noise conditions, is far too small to accurately reflect any co-localization, and therefore all three are susceptible to extremes of noise. With respect to PCC, it is similar susceptible to high noise, yet detected co-localization given with less than 5% noise, though it under-reports the degree of co-localization. When the Costes' auto-threshold is used for the PCC, performance at low noise is substantially improved over non-thresholded PCC, but quickly destabilizes at greater levels of noise. The cause of this instability is a result of the auto-threshold algorithm over-fitting, and finding an unacceptably high threshold value, and therefore removing “true” data, as seen in the case of 15% noise. Therefore, these parameters all detect a lack of correlation/association even with unacceptably high noise, yet the SNR determines overall co-localization sensitivity, with the PCC most robust to this noise at low noise levels.
Figure 4. The effects of additive noise and apparent SNR on detecting degrees of object co-localization. Each panel shows the average parameter value of 1,000 simulated image sets, each with 250 randomly located point objects, with pre-defined SNR, noise and degree of object co-localization. The SNR is fixed either at (A) 2, or at (B) 5, with a representative image of 15% noise. A dashed horizontal line at zero value is added for reference of no association/correlation. As noise levels increase, the response range is compressed (difference from 0% to 100% co-localization).

As expected, co-localization parameters for the higher SNR images, under the same conditions of noise, have appreciably improved performance (Figure 4B), and the response range of all parameters has also scaled with SNR. Importantly, the scale for the degree of correlation in the case of 0% co-localization remains similar under all noise conditions and levels of SNR, with the exception of Li's ICQ. The ICQ range and shifted to closer match the idealized case of object co-localization with no background or noise, and again, maintains the smallest response range. Li's ICQ is promising in ideal conditions of low-noise and high SNR, especially when compared to the more common PCC. Unsurprisingly, the parameters perform best under lower noise conditions, especially with high SNR. However, even under low SNR, both PCC maintain the
greatest response range, similar to that of the ideal case with no noise (Figure 3). The
instability of Costes' auto-threshold is apparent at low SNR and high noise conditions
(PCC (threshold) at $SNR = 2$, 15% noise, Figure 4A) compared to low SNR and low
noise (SNR=2 and 5% noise, Figure 4A), and is a result of over-fitting the orthogonal
regression to find a threshold value due to the noise. Greater noise and lower SNR causes
a general compression in the response range to different degrees of true co-localization.
Conversely, greater overall SNR improves parameter sensitivity, and with respect to the
PCC, is important for the stability of the auto-threshold stability. Overall, the PCC is the
most robust to noise conditions, and with high SNR image data, offers the greatest
response range to detect co-localization differences.

4.4.8 Effect of ROI selection
When co-localization analyses are conducted, it is considered good practice to draw
regions-of-interest around the whole cell area when individual cells are well-
distinguished, if for nothing else but to remove excess background and noisy pixels.
Often times ROIs are drawn around sub-cellular compartments as defined by a
fluorescent marker. To determine the effect of ROI selection on the calculated statistics, I
simulated point objects that were generated with Gaussian noise (5% coefficient of
variation, mean of 100) in a situation where SNR was a not limiting factor, and with
known degrees of object co-localization (from 0% to 100%) (Figure 5). Gaussian noise
was chosen because of its symmetry, and would be expected to cancel out over a large
number of pixels. Objects were randomly distributed in the centre of the image (original
image dimensions were 512x512 pixel). When an ROI was selected around this centre
region (225x225 pixel image dimension), the noise was removed from calculation, and
the calculated statistics are consistent with the detection range reported in Figure 4 and
Table 1. However, not using an ROI let a large area of uncorrelated noise on the image
margins. The calculated parameters became inflated across all degrees of co-localization,
compressing the response range, as a direct result of the extra noise heavily weighting the
parameter calculations. Similar results were obtained with Poisson distributed noise to
mimic detector “shot” noise (results not shown).
Figure 5. Regions of interest improve parameter performance by reducing the weighting on background and noise pixels. Images were generated with Gaussian noise with constant average background. Parameters were calculated for whole images (“no ROI” group) and using the bounding area of point objects (“using ROI” group), with a fixed degree of known object co-localization. The response range of parameters is expanded when the ROI was used compared to not using an ROI.

4.4.9 Bench-marking common co-localization parameters using open data sets

In order to evaluate these parameters in real fluorescence image sets, it would be instructive to see how they perform using cells that are labelled for two structures that do not co-localize, and in cells that display a wide variation in morphology. The morphological variation is an important aspect because they will always be different, even in clonal cell lines, and the argument could be made that the relative areas of the two labels could have an influence on the calculated statistics. To this end, I used a freely available image set in which the same cells, under experimental conditions, undergo
extreme variation in their morphology, while stained for truly segregated markers. I used image set BBBC020 from the Broad Bioimage Benchmark Collection (42), in which murine bone-marrow derived macrophages were treated with a drug to investigate their ability to spread. These cells were immunostained for CD11b/APC (a cell membrane protein) and counter-stained with DAPI, and there is no real co-localization between these two structures. All computed parameters for this data set are shown in Figure 6. Even though these images were taken with a wide-field fluorescence microscope, it is clear that the PCC and Li's ICQ parameters do not falsely identify co-localization. The SRC and Kendall's tau perform relatively worse, with a broader distribution of values, when compared to simulated data, would suggest at least partial co-localization. Also evident is that the Costes' auto-threshold appears to transform the Pearson distribution by creating a heavy left (lower) tail. Here, I suggest that for biological data, a negative value should be have the same interpretation as a zero value: the two labels are not co-localized – in this case, a cell surface marker is definitely not in the nucleus.
Figure 6. Performance of multiple co-localization parameters on open immunofluorescence image sets. An open image set of wide-field immunofluorescence images of bone marrow-derived macrophages \((n=25)\), immunostained for cell membrane protein, CD31b/APC, and counter-stained with DAPI, serving as a control for non-co-localizing labels. Images were collected during a time-lapse experiment to study macrophage spreading, and the heterogeneity of cell morphology serves as a good co-localization false-positive benchmark. Details of the study can be found at (43) and image set “BBBC020” was obtained from (42).

4.5 Discussion

The term co-localization is used to mean many different things, including the degree of association between two proteins; a quantifiable measure of protein interaction; are two proteins in the same place; and the more rudimentary meaning of visual overlap between two fluorescent signals. I take co-localization to mean any two (or more) fluorescently labelled objects that are highly correlated and co-distribute within immunofluorescence images. Ultimately, the cell biologist wants to extract quantitative meaning from the images they acquire. Even with perfect images, the biological question must be matched to the quantitative techniques. Co-localization analysis is a routine method for immunofluorescence microscopy, and cell biologists should strive to rigorously analyze their images beyond the “eyeball” metrics. However, these methods all have their
limitations. They cannot answer the question of what percentage of one protein interacts with a second protein. Many examples of reliable protein associations can be found (e.g., secretory protein trafficking (10, 44)), but potential limitations arise from transient co-localization, or restricted localization within a sub-cellular compartment, or systematic errors to labelling strategy (e.g., (45)). These parameters are useful to answer questions of whether two (or more) fluorescence patterns are similar within the cell, as a measure of the degree of statistical association (but not chemical interaction).

An immediate quality check of immunofluorescence images can be observed from the cytofluorogram and PCC parameter. The section of the cytofluorogram that is correlated and contains co-localized pixels is usually described as the upper-right quadrant (those methods discussed here) or less commonly, by a radial arc centred along the diagonal of perfect co-localization (46). These areas are then quantified (e.g., by linear correlation) to determine the degree of co-localization and represents the biologically significant data. The cytofluorogram gives a good idea of how many uncorrelated pixels exist in each channel relative to what may contribute to correlated pixels. Since uncorrelated pixels and noise cannot be avoided in even the most perfect immunofluorescence images, it is critical to have a threshold value for each image channel. It is usually a good idea to restrict the ROI for analysis to entire cell, or if the labelled molecules exist in one structure (e.g., the nucleus), then drawing the ROI around the nucleus will improve the quality of measurement. Of course, one must always compare to both a negative (known segregated) and positive (known highly co-localized) biological control, to establish the limits of detection within the particular biological and optical system.

There are several misconceptions regarding the use of correlation values as indications of co-localization that should be addressed. Regarding the PCC, some believe that it is not appropriate for immunofluorescence co-localization applications because the acquired pixel data is rarely (if ever) normally distributed (32). Pearson's correlation coefficient does not assume normality because it is defined based on the moments of the distribution, and therefore only requires finite variance and co-variance for the joint distribution. Both finite variance and finite co-variance are guaranteed since images are acquired with finite spatial resolution (determined by the CCD and pixel dimensions) and finite pixel
dynamic range. In the condition where the joint probability distribution (of paired pixel intensities) is bi-variate Normal then the PCC is an exhaustive parameter to describes any linear dependence relationship. Second, the supposition that the PCC is not an acceptable parameter for hypothesis testing (31) follows from the presumption that the normality condition is not met. However, the PCC has a long pedigree of being resilient to even severe departures from normality (47), and is considered a reasonable statistic for hypothesis testing with specific application to co-localization analysis (33). While PCC is by is by no means perfect in every application (e.g., (48)), it offers many advantages over other correlation parameters discussed here with respect to less complex cases of correlation. Li's ICQ was reported to be a better indication of co-localization (31), and it does make the implicit assumption that pixel intensities be symmetrically distributed. The parameter was chosen to simplify hypothesis testing by using the sign-test. However, this is not a reasonable assumption especially when images contain noise. Secondly, the choice of the underlying distribution of the sign-test for Li's ICQ is the Normal approximation to the binomial distribution, but in all fluorescence microscopy applications, pixels do not satisfy the binomial conditions of independence (between pixels) and the probability of “success” (correlation), because pixels are more likely to be similar to their immediate neighbours due to the blurring caused by the point-spread function of the imaging system. This is a property of both the point spread function in image formation (one pixel intensity is smoothed out over many), and because the cellular location of a labelled molecule may be restricted to protein complexes, membranes or compartments. The SRC and even more, the Kendall's tau are promising statistics for co-localization analysis, and are currently implemented in FIJI.

4.5.1 Choice of appropriate controls

In terms of control experiments, it is first necessary to validate the probes that will be use for co-localization experiments, and then prepare to take good quality images. A set of slides should be sequentially imaged in all filter channels, where the slides are: 1) unlabelled, to assess autofluorescence of the sample, 2) labelled with only one of the set of probes (one slide per probe) to determine the performance of each probe, and 3) if necessary, slides in which primary antibodies are omitted to test specificity of the secondary antibody labelling. Once validated, the optical system should be checked for
proper chromatic calibration (e.g., using fluorescent beads), and optical alignment to ensure the best quality image formation. If noise correction is anticipated, a replication based noise correction calculation (37) is a method of reducing shot noise specifically from confocal images, and requires a set of four images and two sequential scans to produce a correction factor for co-localization analysis. Other methods to consider to reduce background and noise: software deconvolution, thresholding, sequential confocal image acquisition using line scanning or frame averaging.

The issue of biological controls is more flexible, and must be considered based on the particular system being examined. Negative controls should be collected in which two proteins are known to not interact or associate (except possibly by random diffusion). For example, if two proteins are expected to be restricted only to the nucleus, the choice could be as simple as a nuclear-localized GFP and DAPI, to characterize random intensity correlations within the ROI containing only the nucleus. On the other hand, if protein trafficking is being monitored, one marker should be a stable marker of the compartment(s) of interest (e.g., calnexin for the ER, or a BGALT1 for the TGN); the second probe may be a protein that either does not traffic through the compartment but has a similar sub-cellular distribution pattern (e.g., endosomal and vesicular compartments), or does traffic through the compartment but does not interact within that compartment with a protein of interest. If using soluble dyes or GFP-derived proteins of acidic compartments, take precaution that they actually label the correct compartments (see for example, (45)). Positive controls are meant to mimic the correlation of interest, and serves as a maximum of the detectable co-localization. Such a control may be two proteins that are known to interact within the compartment of interest, they may be artificial interactions in the same compartment (e.g., adding high-affinity tags to recombinant proteins). A positive control can also be done by simultaneously labelling the same protein with limited amounts of two secondary antibodies, each labelled with a different fluorophore. The latter method guarantees both correlated and co-distributed labelling of a protein of interest, but could suffer from the loss of sub-cellular context, and doesn't account for performance of a second labelled molecule. Having appropriate biological controls is necessary to establish a dynamic range of independent and co-localized protein detection within the parameters of the imaging and biological systems,
and can then be used for statistical comparison. Lacking credible evaluation of known associated and independent proteins will be very difficult to interpret true from spurious co-localization.

4.5.2 Hypothesis testing of correlation parameters

Once all the images have been collected, the quantification techniques complete, the cell biologist will next want to test their working hypothesis. Often the questions being asked are related to “does this mutant protein affect protein trafficking?” or “does a drug treatment alter protein localization?” or “does depletion or over-expression of one protein affect the distribution of two other proteins?” These are all very fine questions to ask, and with co-localization, they will be answered by hypothesis testing, usually turning to Student's t-test or an ANOVA. Under the null hypothesis, none of these treatments will have any measurable effect, and one must compute a test statistic (e.g., t-statistic, F-statistic) that is conditioned on the null distribution to determine whether the observed value is more extreme than is likely to have occurred by random chance within an acceptable significance level (by convention, alpha is usually set to 5%).

The difficulty arises when considering what the joint probability distribution of two non-co-localizing images looks like, and consequently, the shape of the test-statistic under the null hypothesis. This is not a trivial matter because if the null distribution deviates too far from normal, a standard t-test could have a much higher false-positive rate, and would lead the researcher to conclude a significant finding where none existed. Having a significant finding first assumes that the underlying null distribution is accurate (or close enough to accurate) and the finding was not likely to occur by chance. Conversely, failing to find a significant result could mean that the finding was likely to occur by chance given the null distribution, or that the null distribution is not accurate, and the conclusion is irrelevant because it was predicated on an invalid statistic. Most of the quantitative co-localization studies neglect to address this specific point – what does a null distribution look like for any two non-correlation and non-co-localized image set? Presumably it will depend on a number of parameters, especially on the particular cell system, localization of the proteins, fluorophore labelling and the optical system used for imaging. It is also difficult, if not impossible, to predict this distribution in a general case a priori. How then
can I expect the statistical conclusions being drawn to be valid?

There is still no answer to the question of what a general null distribution looks like, but some researchers have made efforts in this way to find approximate distributions. Costes’ significance test proposes to work with two colour channels (34). Briefly, one image channel is divided into discrete blocks, scrambling the image, and calculating the PCC with the other, original image, and repeated (usually 1,000 times). The expectation is that a scrambled PCC will be less correlated that the original image, and the resulting scrambled PCC values form the null probability distribution due only to random correlation, and the likelihood (p-value) can be computed. This scrambling technique, a form of bootstrapping, is not specific to PCC, and can be applied to any of the pixel- or voxel-based intensity correlation measures. A drawback of both this approach and its improvements (49, 50) have been compared (33), finding the null distribution is too broad and leading to false-positive rates greater than the acceptable significance threshold. A stronger implementation of this randomization technique not considered by McDonald was extended to multiple image labels, and is applicable to pixel-, voxel-, object- and intensity-based correlation techniques, but only considered the case of MCC (5). Through analysis of simulated and real image data (33), they find that the PCC is an acceptable parameter to use for hypothesis testing when using a one- or two-sample t-test, and expect similar performance for more complex experimental designs (e.g requiring ANOVA or regression). Therefore, for most cell biological application of quantitative co-localization, the standard statistical tests apply.

Another common pitfall is the issue of sample size, and often not reporting any power analysis calculation to determine the minimum sample size. The process of calculating a sample size is outside the scope of this paper (an excellent text on the subject can be found by (27, 28)), the maximum anticipated effect size should be carefully considered using a pilot experiment, and whether the researcher is interested in a one-tailed or two-tailed hypothesis. For instance, many cell biological applications are interested in a one-tailed hypothesis: the two labelled proteins are well segregated or they are co-localized (to some degree). A two-tailed hypothesis would be the anticipate greater or lesser co-localization depending on the treatment. A well powered analysis requires consideration
of how the proteins of interest are localized in the cell. In the case of proteins that traffic together, more images will be needed for a well-powered study than if they are segregated under some condition.

4.6 Conducting Quantitative Co-localization: A suggested workflow

I suggest a workflow for co-localization analysis of immunofluorescence images start with ensuring that the microscope be optically calibrated, especially for chromatic alignment, and samples labelled with spectrally distinct fluorophores to minimize cross-talk and bleed-through. While imaging, most software allow real-time monitoring of the histogram which should be checked to ensure that the image pixels are within the dynamic range of the camera and that no pixels are saturated from over-exposure, and at a sufficient pixel dwell time to ensure an adequate signal-to-noise ratio. Antibodies should be validated such that the labelling is specific to the protein of interest and localized within the known sub-cellular compartment(s) from literature. Alternatively, recombinant fluorescent proteins also need validation for immuno-detection and biological validation demonstrating that the fluorescent protein tag does not interfere with its usual function or sub-cellular trafficking. Following the characterization of the imaging system, antibodies and fluorophores, the researcher must then choose of appropriate biological controls, as discussed earlier. For 3D imaging, optical z-stack images should be acquired for sufficient axial resolution of the sub-cellular structure (e.g., 0.2 µm step size), and for each imaging channel, the pinhole diameter must be matched for isometric resolution across each acquisition channel. Deconvolution for 3D confocal image stacks is recommended for imaging intracellular trafficking in fine-structure organelles. Deconvolution requires a z-stack acquisition of the PSF by imaging single fluorescent beads that are much smaller than the optical resolution of the microscope using the identical light path settings of the microscope. This PSF should be used for deconvolution of microscope images. These steps are all important to ensure equal resolution and good signal-to-noise ratios in collected images to produce good quality image data. All image sets should also be saved in the microscope's native image format to preserve the experimental meta-data (containing optical path information, microscope settings, etc.), and when exported for analysis, should be saved in their native
microscope software image format to preserve raw images, and exported to uncompressed tagged-image file format (TIFF) images for analysis.

Next, images for the negative and positive controls should be acquired. If possible, the live cytofluorogram at each stage of image acquisition should be viewed to monitor the relative amount of background in each channel, and can alert you to image quality issues during acquisition. Whether imaging individual or discrete clusters of cells, scan area should be increased to maximize pixel resolution and results in scanning the laser over a smaller sample area, reducing sample photobleaching and decreasing acquisition time. Once control images have been acquired, analysis can be done off-line. It is best to draw ROIs around single or distinct clusters of few cells for analysis, removing irrelevant background pixels. (Depending on the software package used, calculating co-localization coefficients can be performed using ROI selections, or these selections can be converted to a binary mask.) To maximize efficiency several co-localization coefficients can be calculated simultaneously, firstly the PCC with Costes' auto-threshold, and optionally the SRC and Kendall's tau. By examining the measured degrees of segregation (negative control) and association (positive control), the researcher now has a defined “dynamic range” of the particular experimental system. The Costes' auto-threshold can also be examined using the cytofluorogram to see if it is reasonable, in that a positive threshold value is reported, and it should be less than the mean signal intensity for each channel. If it is not reasonable, it may be symptomatic of poor image SNR, or not using an ROI and the background pixels prevent a reasonable threshold from being detected. If the control images look good and the co-localization parameters appear reasonable, images can be acquired for experimental treatments using the same guidelines, and correlation parameters calculated. If the co-localization results by PCC are not clear, then consider the rank-correlation coefficients. If these results are still not clear, then the researcher should explore using the more advanced cross-correlation functions (described below).

4.7 Advanced methods of 3D co-localization and extensions to super-resolution imaging

The co-localization methods described above work in simple cases, and in two colours only, making no assumptions about the underlying biological processes involved in
protein co-localization. Further improvements to co-localization analysis have been developed by including more information or using more sophisticated analysis techniques. Ideally, the quantitative co-localization method should be sensitive to co-localization of one channel with another (and extended to multiple fluorophores), insensitive to noise and background, and be able to detect non-linear interaction patterns, such as protein complex assembly. Efforts made in this direction have been to propose new co-localization parameters, and increasingly, moving toward super-resolution imaging.

With careful collection of z-stacks in microscopy, it is possible to achieve even greater resolution that the Rayleigh criteria by using optical deconvolution. Though more often used for wide-field fluorescence, 3D deconvolution can also be performed on confocal stacks, with the caveat that the volume be somewhat over-sampled (z-step of 0.2 µm) and voxel sizes matched in each channel to achieve equal resolution in each channel. Deconvolution is less often applied to confocal microscopy image sets, possibly because resolution is deemed “good enough” for the proteins or structures of interest. A comparison was performed using noise filtering and 3D confocal deconvolution techniques for multi-channel imaging (51) in which the both filtering and deconvolution improved co-localization parameter sensitivity. In their comparison, Pearson's correlation and Li's ICQ performed best. However, when resolution becomes a limiting factor (often by increased noise), noise removal techniques are necessary to simultaneously increase the signal-to-noise ratio while possibly offering improved resolution, with resolution approaching the diffraction limit (41). However, when noise reduction is applied, confocal 3D deconvolution performs better than filtering methods in both simulated and real image sets (38, 52–54), and deconvolution should be performed on confocal stacks prior to co-localization analysis. Since 3D deconvolution is a computationally expensive task, this was an impediment, especially a decade ago when computational power of desktop computers was more limited. However, a recent computational technique uses ImageJ and low-cost computer hardware to run a “blind” (Lucy-Richardson) deconvolution in real-time, removing this barrier (55). Such “blind” deconvolution methods are advantageous because they do not require any measurement of a point-spread function, but rather iteratively fit an ideal point-spread function to the images. To
provide better quality data for co-localization analysis deconvolution should be applied to improve image resolution and SNR.

Super-resolution techniques are now capable of dual-colour nanoscopic imaging technique. Fluorescence techniques using photoactivatable fluorescent proteins (photoactivatable localization microscopy, PALM; and stochastic optical reconstruction microscopy, STORM) use statistical reconstruction of time-resolved photoactivatable fluorophores to create super-resolution images. These images often create fusion proteins using photoactivatable or photoswitchable GFP (and derivatives thereof) to label specific intra-cellular structures or molecules. However, GFP-derived fluorophore brightness is known to be highly pH-sensitive, and a common problem of imaging within mature granules (pH 5-5.5) is the partial or full quenching of fluorophore signal. Creating more stable PALM/STORM fluorophores in acidic conditions is an active area of development (56). Presently there are limited numbers of fluorophores with a pKa compatible for imaging within acidic compartments, though it has received little attention for granule imaging despite its apparent promise to imaging insulin secretory granules (57).

An alternative super-resolution approach called stimulated emission depletion (STED) uses one laser to excitation a diffraction-limited spot, while simultaneously using a second laser to force stimulated emission of fluorophores within a sub-region, switching the fluorophores “off”, resulting in better lateral resolution (58). STED achieves lateral resolution of about 15 nm in situ (59), and has the advantage of imaging live samples in real-time with approximately 50 nm lateral resolution (60). STED systems are now compatible with multi-colour imaging (61, 62), and can be combined with TIRF or multi-photon imaging. Recently, deconvolution techniques have also been considered for super-resolution fluorescence imaging, as suggested by (60). Developing STED deconvolution algorithms is still actively being investigated (63), providing near-real time deconvolution speed (on the order of 1-10 seconds). The scaled-gradient-projection deconvolution method is not just limited to STED, but any fluorescence microscopy method when the point-spread function is supplied. This method not only increases STED image resolution, but when applied to confocal images, brings the lateral resolution to ~90 nm, below the diffraction limit. Since the lasers used in confocal and STED microscopy are
fundamentally diffraction-limited, STED imaging can still suffer from poor axial resolution. Further improvements have been made using “compact” STED systems, increasing the axial resolution to ~100 nm (62, 64, 65), bringing STED into the realm of 3D nanoscopic imaging. STED is now a viable nanoscopic imaging technique for quantitative co-localization. For example, three-colour imaging can be used to determine object-based co-localization (61) or using conventional PCC analysis to determine biased interactions between hexokinase and distinct isoforms of mitochondrial voltage-dependent anion channel (66). The natural increase in resolution lends itself to developing more sophisticated object-based co-localization methods, especially for the application to monitoring protein trafficking along the secretory pathway (e.g., within secretory granules). These nanoscopic imaging techniques offer substantial increases in spatial resolution, and although they still cannot directly answer whether proteins are directly interacting, the increased resolution does offer greater insight to the dynamics of the secretory pathway and greater certainty to image within individual granules.

4.7.1 Novel and advanced co-localization parameters

A novel rank-based correlation coefficient, a dual-channel rank-weighted correlation coefficient (RWC), has the clear advantage of being sensitive to both spatial correlation and intensity fluctuation (11). The intuition is similar to the PCC, in that a correlation is measured, but every pixel value is also weighted by rank. The biological benchmark for this parameter was to measure intracellular co-localization of the temperate-sensitive VSV glycoprotein (tsO45-GFP), a model secretory cargo protein (67, 68). By using a pulse-chase paradigm, a shift from the non-permissive incubation temperature 42°C to a permissive 32°C releases a wave of protein from the ER, and by fixing cells at intervals they measured co-localization from the ER to the trans-Golgi (11). Not only was this the first quantitative analysis of co-localization for this cargo protein, their findings of intracellular trafficking kinetics agree with biochemical kinetic measurements, and show that the parameter performs well to detect small co-localization differences in a spatially compact structure of the Golgi.

Novel co-localization parameters have been proposed to incorporate structural information from 3D image stacks. Wang and colleagues (69) reported two parameters
called the co-localization intensity and binary coefficients (CIC, CBC). This method has
similar intuition to the correlation parameters discussed above, weighting similarities of
pixel intensity from neighbouring focal planes to remove spurious cross-over that may
otherwise contribute to false positive co-localization. This method is more robust to
spurious co-localization in noisy image sets, as tested using a model of nerve myelantion.
An interesting aspect of their technique is that they specifically account for the axial
signal overlap across neighbouring focal planes, and have tested the method within the
limits of confocal axial resolution. It will be interesting to see how this method performs
with super-resolution techniques that can improve axial resolution by ~2-10-fold.
Currently no commercial software packages perform this calculation.

A powerful technique for taking advantage of 3D information from confocal stacks is
when sub-cellular compartments can be segmented and reconstructed into 3D volumes.
At present, commercial (Image Pro, Imaris) packages and non-commerical packages both
offer this ability. True “object-based” methods use these rendered volumes to detect co-
localization by several means. First, compartments such as vesicles can be reconstructed
from image stacks based on an estimated size. Typically, object-based co-localization
occurs when the centroid of one object is near enough to another object's centroid, or it is
within the reconstructed volume of another object (see for example, (5, 48)). This
advanced method builds a model of point-based object detection for investigating
punctate objects (be it labelled proteins or vesicular compartments) with a bootstrapped
statistical framework for making inferences, and is easily extended to 3D and time-lapse
microscopy. This is still an area of active research to examine in such phenomena as virus
entry into cells (70), observing protein machinery maintaining cell polarity in S. pombe
using spherical geometry (71), and dynamic vesicle identification (3). In the case of
intracellular protein trafficking, I have previously validated object-based co-localization
of two secretory granule-targeted proteins, glucagon and a modified, dipolar GLP-2, to
introduce a granule-targeting sorting signal (Figure 7). The co-localization of these two
proteins were confirmed by PCC (72). The 3D renderings were reconstructed using each
protein, and a granule resident protein, chromogranin A, and the percentage of
overlapping objects agrees with the degree of co-localization as measured by PCC.
Object-based methods have the advantage of being able to exclude objects of
heterogeneous size which may skew co-localization analysis (e.g., the trans-Golgi), and focus instead on the secretory granules.

**Figure 7.** Object-based co-localization of 3D rendered sub-cellular volumes of the secretory pathway and granules. Objects of approximately 250-300 μm in diameter were segmented and rendered in Imaris Bitplane from 3D confocal image stacks (0.2 μm z-step). Rendered “spots” of from chromogranin A granules (in red) were co-localized with either Fc-tagged (A) glucagon or (B) dipolar GLP-2, in green. The percent of co-localized objects (~72%) (overlapping volumes) agrees with the high degree of co-localization as reported by PCC. These images were reproduced from (72).

Two different classes of co-localization methods have been proposed that are fundamentally cross-correlation functions. The cross-correlation function (CCF) of van Steensel (73), used to track glucocorticoid and mineralocorticoid receptor trafficking within the nucleus, is essentially the PCC iteratively calculated following horizontal translations of one colour channel relative to the other (usually 20 pixels in either direction). By plotting the resulting PCC values as a function of the horizontal shift, the curve shape is expected to better indicate true co-localization from spurious or segregated correlations, because only co-localization objects would maximally co-localization when the correlated pixels are maximally aligned. The CCF is useful when considering small, isotropic particles (74) and its detection power is fundamentally limited by the resolution of the system. A natural extension of this cross-correlation approach was later proposed by the name of a protein-proximity index (PPI) (75). This technique is meant for two
colour image sets, and considers the effects of how the cross-correlation (both x- and y-shifts of channel one with respect to channel two), and fitting a two-component Gaussian decay curve to the cross-correlation peak. When a fast-decaying component is present, this differentiates “true” co-localization from the slow-decaying curve that represents auto-fluorescence. This technique also takes advantage of increased resolution of STED microscopy, explaining its greater sensitivity to detecting co-localization. The idea of cross-correlation has been extended to dynamic single-particle tracking of confocal image stacks by comparing image cross-correlation with each particle's 3D trajectories to spatio-temporally match fluorescent probes (76). These cross-correlation and auto-correlation functions are particularly suited to monitoring dynamic processes over time, when intracellular trafficking or particle movement cannot be predicted a priori, but only after re-alignment. They also have the advantage of being sensitive to repetitive structural elements, something that may be useful when monitoring trafficking through tubular structures or the Golgi cisternae.

Two novel methods take advantage of nearest-neighbour distance metrics and spatial inhomegeneity to detect particle clustering as a marker of dynamic protein-protein interactions. A recent approach (77, 78) whose method of co-localization is designed for super-resolution images, termed spatial association/apposition analysis (SAA). Each particle object is segmented and evaluated for co-clustering probability with other nearby particles. First demonstrated to reveal receptor clustering on the cell surface (77), it has now been applied to monitor intracellular trafficking by vesicular transport and is extended to multiple interacting particles. The spatial inhomogeneity assumption is that, if objects are not clustered, they should not exhibit a bias in nearest-neighbour distance or radial clustering to other objects. This technique has been validated for GSD microscopy, and should also be applicable to any single-particle super-resolution microscopy method. A mathematically related technique was reported to observe individual RNA polymerase II complexes, previously thought to exist as clusters of transcriptional factories within the nucleus (79). The imaging technique is novel, adapting light-sheet and super-resolution microscopy to achieve much higher SNR images than conventional confocal microscopy, while spatio-temporal tracking of particles is used to detect spatial clustering as a measure of protein co-localization. This technique can also be be extended to measure
stoichiometric interactions within particle clusters. The combination of single particle tracking methods using super resolution imaging techniques are advantageous for delineating stoichiometric ratios of protein clusters, and profiling the dynamic spatio-temporal co-localization of proteins.

Lastly, a promising family of techniques that originate from the machine learning and image segmentation literature are Bayesian approaches to co-localization. The foundations of Bayesian analysis as applied to analysis of super-resolution images are best explored for image feature extraction (for review see (80)), though these methods are computationally more complex than many of the alternative method described here. Some progress in this direction have been to use Bayesian post-processing techniques of confocal and super-resolution images to localize proteins with similar precision to super-resolution imaging alone (81,82). Recently super-resolution images were used for Bayesian localization of proteins, to detect spatial clusters of fluorescently tagged receptors in the cell membrane (83). This technique is promising to apply when investigating multiple protein-protein interactions, such as the sorting events that occur in the regulated secretory pathway, as receptors are localized to newly forming granules with their cargo. As computational power increases, Bayesian analysis methods will be very interesting to follow for co-localization studies.

4.8 Conclusion

Extracting quantitative information from optical microscopy requires care in the preparation of the biological specimen but also knowledge of the image processing techniques that are applied. Diffraction limited microscopy in particular suffers from blurring due to the point-spread function and variable image SNR. These aspects can be accounted for by appropriate image processing techniques, but it should be understood that these alter the “raw” image data. Trade-offs involved in image processing and image acquisition are illustrated by a discussion of quantitative immunofluorescence confocal microscopy for co-localization analyses, which represent the most common quantitative technique applied in confocal imaging. I suggest a workflow to acquiring 3D confocal image stacks for co-localization analysis and consider the most commonly used correlation parameters. As super-resolution and single particle imaging methods continue
to be adopted, extracting quantitative information will require the development of novel techniques. More sophisticated methods can also be applied to answer questions related to stoichiometry of the interaction between labelled molecules. As the resolution limits are pushed further down, the sheer amount of image data and associated analysis techniques will become the limiting factors for quantitative microscopy, and these data will continue to be mined to extract more quantitative spatial and temporal information.

4.9 References

Zymogen Granule Exocytosis is Dependent on Endosomal Trafficking via the Constitutive-Like Secretory Pathway. *J. Biol. Chem.* **8**


restoration and time-lapse confocal microscopy to MAP kinases in mitochondria. *PLoS One* 6, e19031


Chapter 5

5.1 Discussion and Future Directions

A central question in α- and L-cell biology is how proglucagon is sorted into granules in order to undergo processing to its constituent hormones and be effectively concentrated for later exocytosis. The mechanisms of differential processing by PC1/3 and PC2 are well understood, but the specific location of where the convertase enzyme encounters its substrate, proglucagon, is not yet resolved. It is recognized that many (pro)hormones and neuropeptides contain sorting signals that are encountered and recognized by granule-targeted receptors as granules are formed. Here I have identified that proglucagon contains several sorting signals, and interestingly, two are contained within the mature hormone domains of glucagon and GLP-1 sequences. The molecular structure represents a novel class of α-helical sorting signals: an electrically polarized (dipolar) α-helix with discrete positively and negatively charged surfaces. This is distinct from the previously identified amphipathic α-helices which rely on a large hydrophobic face. I have also characterized some of the possible molecular interactions in both α- and L-cells that could serve as receptor-mediated binding of proglucagon. I have documented that sorting of proglucagon is cell-specific, requiring CPE, and is consistent with several reports that characterize CPE as a receptor for multiple hormones. I also have preliminary evidence for CgA serving a similar receptor role (in PC12 cells), binding to both the glucagon- and GLP-1 sorting signals in a pH-sensitive manner. Therefore, I have identified distinct sorting signals within proglucagon and propose two receptors for interaction such that α-cells can efficiently and adequately sort proglucagon to granules for processing to mature glucagon.

This work has raised new and unresolved questions concerning the biology of proglucagon and the field of hormone biology. Since the identification of dipolar α-helices as sorting signals, this molecular motif may be represented and unrecognized in other neuropeptides and hormones, proving that existing mechanisms for sorting are more adaptive than previously thought, or supporting the existence of yet unidentified receptors. On the basis of predicted protein secondary structure, dipolar helices may be
functional signals within pro-adrenomedullin (ADML), pro-pituitary adenylate cyclase-activating polypeptide (PACA), glucose-dependent insulinotropic polypeptide (GIP), pancreatic polypeptide (PAHO) and peptide-YY (PYY). Additionally, reinterpretation of the amphipathic α-helix data of pro-cocaine and amphetamine-related transcript (1) would also support the notion of a dipolar α-helix. As a general mechanism for dipolar α-helices, they are likely to undergo electrostatic association with charged granule targeted proteins (e.g., granin association is calcium- and pH-dependent) or charged lipid head groups of the granule membrane. A pH-dependent or cation-dependent interaction would support this view as suggested by initial experiments with glucagon and GLP-1. While no single universal mechanism for sorting prohormones exists, and there is a real possibility that new classes of signals have yet to be identified. An interesting possibility is that a β-sheet structure could be such a motif. The beta-sheet structure may form large-scale aggregates, or target to granules by association with zwitterionic membrane lipids (e.g., phosphatidylcholine) or by association with cholesterol. One example is found in hepcidin, a key metabolic regulator of iron homeostasis. Both the pro-form and mature hepcidin are small peptides (60 amino acids in the pro-form, and 20 or 25 in mature form), and the mature hepcidin structure has been resolved (see Table 1). Then there is the curious xenin-25, a regulated secretory protein found within granules of many endocrine tissues. It is unique in that it is likely the shortest pro-peptide at 35 amino acids in length, and is cleaved to mature xenin-25 (2). Since it constitutes the N-terminal domain of coatamer sub-unit A, a cytosolic coat protein, it is unknown how pro-xenin enters the regulated secretory pathway. Yet, xenin is still a secreted gut hormone (3) and the beta sheet structure may constitute a novel sorting signal (Table 1). Further characterization of how neuropeptides and hormones will likely reveal new classes of motifs and interactions, and will contribute to the general understanding of sorting along the regulated secretory pathway.
Table 1. Hypothetical novel classes of sorting signals.

<table>
<thead>
<tr>
<th>Putative sorting signal(s)</th>
<th>Example Protein (Gene) name</th>
<th>3D Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-sheet or hairpin</td>
<td>Hepcidin (HEPC)</td>
<td><img src="image" alt="Protein Data Bank ID: 1M4E" /></td>
</tr>
<tr>
<td>B-sheet</td>
<td>Xenin-25 (COPA)</td>
<td><img src="image" alt="Results were generated by de novo prediction." /></td>
</tr>
</tbody>
</table>

Concerning new questions specifically related to the biology of proglucagon, there are some major questions to resolve: 1) what is the location where proglucagon is both initially sorted?; 2) where precisely is proglucagon initially cleaved into glicentin and MPGF?; 3) what is the nature of the distribution of the processed PGDPs within the granule population?; and 4) what is the nature of interaction between proglucagon and CPE and CgA? It is well known that secretory granules are functionally transported/diverted to distinct pools that are operationally distinguished based on
whether they are docked at the plasma membrane or nearby, and whether they are immediately releasable or if there is a lag prior to exocytosis. This difference is largely explained by the type of effector proteins on the cytosolic leaflet of the granule (in particular Rab3, Rab27 and SNARE proteins and SNARE accessory proteins) and the particular secretory state of the cell. For example, young granules spatially segregate near the membrane over older granules in bovine chromaffin cells, appearing docked at the membrane, while older granules are more mobile (4, 5). A related and under-considered question concerns the existence and identity of distinct sub-populations of newly formed secretory granules. In other words, are (neuro)endocrine peptides further able to differentially load cargo, thereby producing distinct subs-populations of granules? Many reports, principally by immuno-electron microscopy, report natural variation the spatial location and density of cargo proteins in the granule, distributed among the dense core and the translucent halo. However, whether this process is a by-product either of natural variations in granule biogenesis, or a result of divergent biogenesis mechanisms, remains a mostly unexplored question (6). Immunoelectron studies of the distribution of GLP-1 and GLP-2; glicentin and MPGF; and glucagon and MPGF; are all suggestive that the PGDPs are co-stored within the same granule, and thus enter the same population of granules. However, this has not been definitively addressed, and the data at hand only suggest a correlation. Recent evidence has been identified in GLUTag cells whereby the classical techniques of labelling granules by means of GFP-tagged hormones (e.g., GH, BDNF and NPY), were each shown to co-distribute to varying degrees with GLP-1 (7). Though it was not the focus of that study, they found that GFP-tagged NPY and BDNF strongly localized to GLP-1-containing granules, while GFP-tagged GH was completely segregated from GLP-1. Super-resolution microscopy is an excellent method to examine the distribution of GLP-1 and GLP-2, or glucagon and MPGF, within granules.

Evidence in support of distinct sub-populations of granules have only been reported for pituitary-derived neuroendocrine cells. In AtT-20 cells (corticotrophs), exogenous expression of syncollin enters granules containing ATCH, but not VAMP-2 (8), the granule membrane v-SNARE required for exocytosis and in agreement with similar findings of unequal cargo distributions in AtT-20 cells (9). A parallel observation was
reported in multiple beta cell lines, in which the loading of CgB and insulin was heterogeneously distributed, some granules containing only insulin, a mix of insulin and CgB, or CgB alone (in a percent ratio of 66:27:7) (10). Though they demonstrated different profiles of secretion through various types and doses of secretagogue stimulation, the explanation could be due to truly distinct sub-populations of granules (as the authors suggested), or a coincidental result of a single population of granules that is being dynamically regulated for distinct kinetics of exocytosis, as has been well-documented from PC12 and chromaffin cells (11–13). This remains a possible additional layer of complexity to the granule biogenesis process and the functional significance of a diverse granule population.

To understand when and where proglucagon is sorted and processed, advanced visualization techniques and proteomics could be used to distinguish the sub-compartments of the TGN from immature and mature SGs. This would necessarily add temporal resolution to the examination of the dynamic sorting process. While the specificity of antibodies to detect all species of PGDPs without cross-reactivity is major challenge, both sensitive and specific detection methods are required for monitoring proglucagon undergoing processing. LC-MS-based proteomic detection (14, 15) certainly meets these requirements. However, proteomics methods destructively sample the contents of cells, requiring pooling and a loss of temporal and spatial resolution. The challenge is to not only distinguish proglucagon from the PGDPs (i.e., glucagon and glicentin), but also to delineate the multiple compartments of the secretory pathway. The latter is somewhat easier to accomplish by immuno-staining proteins that either reside in ISGs but do not traffic to mature granules. Some examples include furin (16), mannose-6-phosphate receptor (M6PR), syntaxin 6, clathrin (17), carboxypeptidase D (18), synaptotagmin IV, VAMP4 (19), and Golgi-associated, γ-ear-containing, ADP-ribosylation factor-binding protein (GGA) (20), adding non-specific compartment labelling. More specific labelling approaches have been proposed on the basis of using fluorescent proteins as molecular timers.

A molecular timer is any molecule that can develop or change its fluorescence properties
over time. Early uses of a fluorescent timer applied to intracellular trafficking used a pH-sensitive GFP, pHluorin. As the pHluorin is exposed to increasingly acidic environments, the GFP emission profile dims from physiologic pH and is quenched by the acidic environment of mature granules (21). Applications to intracellular trafficking are particularly convincing when using fluorescence TIRF imaging to monitor live granule exocytosis, the fluorophore immediately recovering fluorescence as soon as the fusion pore forms to neutralize the acidic granular pH. The disadvantage of this technique is the progressive signal loss as granules mature, and no clear distinction of immature granules. A related idea takes advantage of the slowly-maturing RFP variant that changes emission from green to red, has been used to monitor chromaffin granule ageing and secretion (4) when fused to a granule-targeted protein. One such protein, syncollin, is targeted to granules of endocrine (22) and neuroendocrine cells (23), and unlike its natural inhibitory action of granule exocytosis, does not affect exocytosis kinetics as a fusion (24). A major draw back to this timer is that the transition time is ~ 16 h (4, 25), which would preclude the temporal resolution of nascent and maturing granules, but is ideal to monitor age-coded mature granules. Therefore, a faster timer was engineered from the mKusabira-GO fluorophore, undergoing a similar green to orange transition (5), which has a linear transition over 1-10 hours, making it ideal for monitoring nascent and maturing granules. An interesting third approach to timing makes use of a transgenic mouse expressing recombinant SNAP-tagged insulin for monitoring young granules in beta cell (26). The SNAP domain is derived from the human O6-methylguanine-DNA methyltransferase, and when exposed to a benzylguanine-containing substrates, catalyzes a reaction that is used to transfer a fluorescent dye (e.g., TMR-Star, BG-505) to the SNAP tag (27). As used in the SOFIA mouse (26), the SNAP-tag approach offers the temporal resolution, and was used in a pulse-chase paradigm to preferentially label older and younger granules with different fluorescent dyes. While they showed preferential exocytosis of “young” over “old” insulin granules, the technique is extensible to specific “gating” of granules with defined age, and can be applied to specifically monitor immature granules.

To directly address the question of where the initial cleavage event of proglucagon occurs, fluorescent tagging strategies could be leveraged to enrich a population of
immature granules combined with downstream analysis by proteomics techniques. A challenge to investigate proglucagon has always been the availability and performance of antibodies against the multitude of PGDPs. To circumvent this issue, proteomics are ideally suited for accurate quantitation and fingerprinting of proglucagon and PGDPs. The approach described by (26) could be modified to only label the immature granules (granules only 15-30 minutes old) by first “blocking” a granule-targeted SNAP protein with the non-fluorescent BTP, followed by a short pulse of a rhodamine-derived TMR-Star. Cell membranes can be gently lysed in hypotonic solution leaving in tact the granule population, and subjected to fluorescent-activated organelle sorting (28). Double-labelling approaches to specifically label the immature SGs, either by a a fluorescent timer or specifically tagged ISG protein could be used as a positive control or for increased specificity by dual-fluorescence-activated sorting. By preserving the labelled immature granules, they can be enriched for LC-MS/MS proteomic analysis, or “printed” onto a thin layer of mica for imaging by atomic force microscopy, electron microscopy or optical imaging to ensure the quality and enrichment of immature granules. In this way, the contents of ISGs will reveal the relative abundance of proglucagon to the PGDPs. If this experiment is conducted in native cell lines, and compared with cells that either lack PC1/3 and PC2 (e.g., PC12), or in native cell lines in which PC1/3 and PC2 have been depleted or inhibited, this approach also answers whether initial cleavage occurs in the TGN or ISG. If the initial cleavage occurs in both compartments, this technique can also assign the degree of cleavage to each compartment. While more elaborate schemes can be imagined, this approach is advantageous because it combines the sensitivity of MS analysis with the specificity of granules that have been defined by their age. One could also imagine a time course comparison to observe the changes in proteomic profile as granules age to confirm the conversion of proglucagon to its constituent hormones.

From Chapter 2, I concluded that the initial processing of proglucagon was necessary to enhance sorting, on the basis that a block in this event resulted less efficient sorting. I refined this model in Chapter 3 by more closely examining the nature of the sorting signals within the proglucagon-derived peptides. Based on the data at hand, I concluded that the proglucagon cleavage event must occur after it is sorted to granules, on both the
nature of the signals and the marked difference in trafficking between glicentin and MPGF, where glicentin partially entered synaptic-like micro-vesicles. I developed and validated a more rigorous approach to quantitative co-localization in Chapter 3, and discuss how it compares to other common methods in Chapter 4, as well as considering more innovative quantitative co-localization techniques. A more rigorous and quantifiable technique to directly address proglucagon sorting can take advantage of one of the advanced co-localization techniques described in Chapter 4, and complementary proteomics-based approaches as previously described. Single-particle tracking analyses can be used to visualize the spatial distribution of glicentin and MPGF when combined with high-temporal resolution labelling techniques (such as fluorescent “timers” and SNAP-tag labelling), bringing sufficient resolution to monitor sorting and processing events in the secretory pathway.

5.2 References

10. Giordano, T., Brigatti, C., Podini, P., Bonifacio, E., Meldolesi, J., and Malosio, M. L. (2008) Beta cell chromogranin B is partially segregated in distinct granules and can be released separately from


Appendix

Appendix 1. Code contributions to the FIJI/ImageJ project

In order to explore the computational application of co-localization, I sought a software package that both implemented Pearson's correlation coefficient and was intended for scientific applications. Several commercial and non-commercial software packages incorporate Pearson's correlation coefficient and multiple other co-localization algorithms. Among them, I chose to use FIJI/ImageJ, because of its popularity in biological sciences; it is free (in cost and license); open-source; has a responsive developer community; and runs on platforms. Within FIJI, the Colocalization analysis plug-in (called “Coloc 2”) measures co-localization either on whole images, masked images or selected regions-of-interest (ROIs). I first tested the validity of the calculation by using “toy” image sets, such that the Pearson's correlation coefficient could be independently verified by manual and Excel spreadsheet calculations. From these tests, it became apparent that the Pearson's correlation coefficient calculations were numerically flawed, and resulted (at times) in substantially under- or over-inflated correlation values on test image sets. Obviously, this was problematic, and in reaching out the core FIJI developers, I discovered that this bug had not yet been reported. Therefore, it is possible that had Coloc 2 been used for co-localization, incorrect correlation values may have been reported in scientific literature. At the same time as fixing this bug, I had also added another algorithm to compute Spearman's rank correlation coefficient. The Spearman's rank correlation might be useful for future co-localization studies using FIJI.

Bug fixes for the FIJI Colocalization analysis plug-in

The existing bugs inside of the Colocalization analysis plug-in could be divided into two categories: a numerical error, and a logical error. The latter, logical error, failed to properly handle ROI selections. In theory, any selected ROI would have its co-localization parameters calculated for only those pixels within the ROI. In this bug, a selected ROI correctly returned the pixels contained therein, yet the image statistics of each channel mean intensity and sum of squared differences were computed from the entire image, ignoring the selected ROI, causing the parameter to be miscalculated. Second, the numerical error was the use of incorrect arithmetic in translating the mathematical equation for Pearson's correlation into computer arithmetic, resulting again in miscalculation. A by-product of this numerical error
was that there was a mismatch between the software calculation of an integer and a floating-point value, that results in rounding error. This code was sent to the developer mailing list and was incorporated in the code base by the core developers (at approximately the same time as the Spearman rank correlation coefficient below).

Addition of the Spearman's Rank Correlation Coefficient Algorithm to FIJI

The Spearman's rank correlation coefficient is the rank-ordered equivalent of the Pearson's correlation coefficient. Specifically, when both continuous variables used for Pearson's correlation are converted to their rank-order, the Pearson's \( \rho \) is equivalent to the Spearman's \( \rho \). An additional component of this correlation value is that ties in rank must be resolved according to the average of their positions. For a sample of \( i \) values (e.g., a pair of images each with \( i \) pixels), the Spearman's \( \rho \) is calculated as follows:

\[
\rho = \frac{\sum_{i} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i} (x_i - \bar{x})^2 \sum_{i} (y_i - \bar{y})^2}}
\]

where subscript \( i \) is the rank of each pixel, \( x \) and \( y \) represent each image channel, and overbars represent mean pixel rank. Similar to Pearson's \( \rho \), Spearman's \( \rho \) is also continuous on the domain \([-1,1]\).

An attractive feature of using this technique is that Spearman's \( \rho \) is more sensitive to non-linear statistical co-relationships. Despite similar advantages to Pearson's correlation, Spearman's rank correlation has the potential to discriminate sub-cellular compartments by nature of its rank-ordered computation. However, this has not yet been critically applied to fluorescence microscopy for the purposes of protein co-localization in real image sets.

The revision history for the Spearman's rank correlation coefficient is publicly available from GitHub (commit# 5e134f48; found online at http://github.com/fiji/fiji/commit/5e134f487709fc37e5afdf7e72b0f8d92605b4168). The substantive algorithm code has been re-produced below, from the `SpearmanRankCorrelation.java` source file:

```java
public static <T extends RealType<T>> double calculateSpearmanRank(TwinCursor<T> cursor) {
    // Step 0: Count the pixels first.
```
int n = 0;
while (cursor.hasNext()) {
    n++;
    cursor.fwd();
}
cursor.reset();

data = new double[n][2];
ch1raw = new double[n];
ch2raw = new double[n];
ch1ranks = new double[n];
ch2ranks = new double[n];

for (int i = 0; i < n; i++) {
    cursor.fwd();
    T type1 = cursor.getChannel1();
    T type2 = cursor.getChannel2();
    data[i][0] = type1.getRealDouble();
    data[i][1] = type2.getRealDouble();
    ch1raw[i] = data[i][0];
    ch2raw[i] = data[i][1];
}

/**
 * Here's the concept. Rank-transform the data, then run
 * the Pearson correlation on the transformed data.
 *
 * 1) We will sort the dataset by one column, extract the
 *    column values and rank them, and replace the data by
 *    the ranks.
 * 2) Repeat the process now with the remaining column.
 * 3) Calculate the coefficient from the individual rank
 *    columns, the t-statistic and the df's of the test.
 */

// Step 1: Sort the raw data, by column #2 (arbitrary choice).
Arrays.sort(data, new Comparator<double[]>() {
    public int compare(double[] row1, double[] row2) {
        return Double.compare(row1[1], row2[1]);
    }
});

for (int i = 0; i < n; i++) {
    ch2raw[i] = data[i][1];
}

// Rank the data then replace them into the dataset.
ch2ranks = rankValues(ch2raw);
for (int i = 0; i < n; i++) {
    data[i][1] = ch2ranks[i];
}

// Step 2: Repeat step 1 with the other data column.
Arrays.sort(data, new Comparator<double[]>() {
    public int compare(double[] row1, double[] row2) {
        return Double.compare(row1[0], row2[0]);
    }
});
for (int i = 0; i < n; i++) {
    ch1raw[i] = data[i][0];
}

ch1ranks = rankValues(ch1raw);
for (int i = 0; i < n; i++) {
    data[i][0] = ch1ranks[i];
    ch2ranks[i] = data[i][1];
}

// Step 3: Compute statistics.
rhoValue = calculateRho(ch1ranks, ch2ranks);
tStatisticSpearman = getTStatistic(rhoValue, n);
dfSpearman = getSpearmanDF(n);
return rhoValue;
}

public static double getTStatistic(double rho, int n) {
    double rho_squared = rhoValue * rhoValue;
    return (rhoValue * Math.sqrt((n - 2) / (1 - rho_squared)));
}

public static double[] rankValues(double[] sortedVals) {
    int len = sortedVals.length;
    int start = 0;
    int end = 0;
    double[] newranks = new double[len];
    double avg = 0, ranksum = 0;
    boolean ties_found = false;

    // first assign ranks, ascending from 1
    for (int i=0; i<len; i++) {
        newranks[i] = i+1;
    }

    // check for tied values
    for (int i=0; i<len; i++) {
        start = i;
        end = i;
        // Advance values while you haven't exceeded the final value in the
        // ranked data,
        // and until we break a tie in values.
        while ((++end < len) && (sortedVals[start] == sortedVals[end])) {
            ties_found = true;
        }

        // Check if we advanced our end position
        if ((end-start != 1) && (ties_found)) {
            // Compute arithmetic average of rank according to Spearman's
            // method:
            // average = sum of ranks / number of ranks
            avg = 0;
            ranksum = 0;
            for (int j=start; j<end; j++) {
                ranksum += newranks[j];
            }
        }
    }
    return newranks;
}
125 avg = ranksum / (end-start);

// Assign averages to the tied ranks
for (int x=start; x<end; x++) {
    newranks[x] = avg;
}

ties_found = false;

//reset i
i=end-1;

return newranks;

public static double calculateRho(double[] x, double[] y) {
    // Define some variables.
    double rho;
    int len = x.length; // the lengths should be the same for each array
    double mean_x = 0.0, mean_y = 0.0;
    double sum_x = 0.0, sum_y = 0.0;
    double sd_x = 0.0, sd_y = 0.0, sd_xy = 0.0;
    double ssd_x = 0.0, ssd_y = 0.0;
    double denominator = 0.0;
    int i = 0;

    // Calculate the mean of each rank set
    for (i = 0; i < len; i++) {
        sum_x += x[i];
        sum_y += y[i];
    }
    mean_x = sum_x / len;
    mean_y = sum_y / len;

    for (i = 0; i < len; i++) {
        // Calculate the Sum of Differences (numerator)
        sd_x = x[i] - mean_x;
        sd_y = y[i] - mean_y;
        sd_xy += sd_x * sd_y;

        // Calculate the Sum of Squared Difference
        ssd_x += (x[i] - mean_x) * (x[i] - mean_x);
        ssd_y += (y[i] - mean_y) * (y[i] - mean_y);
    }

    /** Calculate rho
    * We calculate it this way rather than the alternative in the case
    * (1 - [ (6 sum d^2) / (n(n^2 - 1)) ] which is a simplification when
    * there are no ties in rank transformed data.
    */
    denominator = Math.sqrt(ssd_x * ssd_y);
    rho = sd_xy / denominator;
    return rho;
}
Curriculum Vitae

Name: Leonard Michael Guizzetti

Post-secondary Education and Degrees:
B.Sc. (Medical Biophysics, Collaborative Graduate Program in Molecular Imaging), 2005-2009
Western University
London, ON, Canada

Ph.D. (Medical Biophysics), 2009-2014
Western University
London, ON, Canada

Honours and Awards:
University of Western Ontario Entrance Scholarship, 2005

American Society for Cell Biology Travel Award, 2011

Collaborative Graduate Program in Molecular Imaging Travel Award, 2011 – 2013

American Society for Cell Biology Travel Award, 2013

Graduate Thesis Research Award, 2013

Society of Graduate Students Travel Award, 2013–2014

Related Work Experience:
Teaching Assistant
Biology 1001A/1002B
University of Western Ontario, 2012–2013

Teaching Assistant Training Program (Certificate)
University of Western Ontario, 2013

Teaching Assistant
Biology 2244A/B
University of Western Ontario, 2013–2014

Course Content Co-creator
Statistics 1024A/B
University of Western Ontario, 2014

Publications:

Journal Articles:

**Refereed Abstracts:**


**Refereed Oral Presentations:**


**Refereed Poster Presentations:**


