7. Appendix

Abstract

The completion of the sequencing of the human genome and the development of global high throughput approaches have changed and enriched biological experimentation. The most important task is nor the elucidation of the function of all encoded proteins. One of the limitations that prevent the application of biochemical methods to high-throughput experimentation is the inability to express and purify large numbers of proteins in high-throughput format.

To develop a high-throughput method to purify proteins from *E. coli*, a 96-well format compatible protein purification process was developed for His_6-tagged proteins under denaturing conditions. Under non-denaturing conditions the choice of the purification tag can have a significant impact on the purification success of individual proteins. To develop a general method for the parallel purification of many proteins, four different purifications tags were evaluated using a test set of 32 sequence-verified human cDNAs of varying sizes and activities. The basic purification process was adapted to the four different chemistries and all 128 constructs were purified and characterized with respect to yield and purity. This analysis revealed that the GST-tag and the MBP-tag were equally effective and purified 28/32 proteins. The purification methods were shown to be compatible with the functional integrity of the proteins.

In order to evaluate the developed methods on a larger test set, 771 and 428 different proteins were purified with the His_6-tag under denaturing conditions and the GST-tag under non-denaturing conditions respectively. In this experiment 67% of all His_6-tagged proteins were be purified under non-denaturing and 49% of the GST-tagged proteins were purified under non-denaturing conditions.

The biochemical and biophysical parameters of the purified proteins were analyzed to identify protein properties that are predictive of protein purification success from bacteria. We found that the length of the single longest hydrophobic stretch in the primary structure of His_6-tagged proteins determines the likelihood of successful protein purification.

Zusammenfassung

Die vollstaendige Sequenzierung des menschlichen Genoms und die Entwicklung von globalen experimentellen Ansätzen haben die biologische Forschung verändert und bereichert. Die funktionelle Charakterisierung aller Proteine ist zur Zeit eine der dringensten Aufgaben. Leider sind biochemischer Methoden für funktionelle Studien im Verfahren mit hohem Durchsatz zur Zeit nicht nutzbar, da die hierzu erforderlichen Proteine nicht im Horchdurchsatzverfahren aufgereinigt werden koennen.

Um eine allgemein anwendbare Methode fuer die parallele Aufreinigung von vielen verschiedenen Proteinen unter nicht-denaturierenden Bedingungen zu entwickeln, wurde zunächst ein Reaktionsplatten-kompatibler Prozess fuer Proteinaufreinigungen von Bakterien erarbeitet und anschliessend wurden vier verschiedene Affinitätsankerproteine im Kontext von 32 Test-Proteinen evaluiert. Alle 128 Fusionsproteine wurden im Reaktionsplattenformat aufgereinigt und in Bezug auf Reinheit und Ausbeute charakterisiert. Diese Analyse ergab, das der Glutathione-S-Transferase-Anker ('GST-tag') und der Maltose-Bindungs-Protein-Anker ('MBP-tag') gleich effektiv waren und die Aufreinigung von 28/32 beziehungsweise 28/31 Proteinen ermoeglichten. Es wurde gezeigt, dass beide Aufreinigugsmethoden funktionelle Proteine produzierten.

Um die entwickelten Methoden mittels eines größeren Proteinsatzes zu evaluieren, wurden 771 beziehungsweise 428 verschiedene Proteine als His_6-Fusionsproteine beziehungsweise als GST-Fusionsproteine aufgereinigt. In diesem Experiment konnten 67% aller His_6-Fusionsproteine unter denaturierenden Bedingungen und 49% aller GST-Fusionsproteine unter nicht-denaturierenden Bedingungen aufgereinigt werden.

Letztlich wurden biochemische und biophysikalische Daten aller Proteine untersucht um Eigenschaften zu identifizieren, die es ermöglichen den Aufreinigungserfolg einzelner Proteine vorherzusagen. Unter anderem wurde festgestellt, dass die Länge, des längsten durchgehenden Abschnittes der Primärstruktur mit einer durchschnittlichen Hydrophobizität >0.85 (nach Goldman, Engelman, Steiz) mit dem Aufreinigungserfolg von His_6-Fusionsproteinen korreliert.

Abbreviations

2DE	2 Dimensional Electrophoresis
2D-LC	2 Dimensional Liquid Chromatography
2-HA	2 Hybrid Assay
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myelogenous Leukemia
APC	Anaphase Promoting Complex
Appr>p	Adenosine Diphosphate Ribose 1”-2” Cyclic Phosphate
B	Bead Bound Fraction
BAC	Bacterial Artificial Chromosome
b-gal	beta-galactosidase
CAP	Cellulose Associated Protein
CBP	Calmodulin Binding Peptide
CDK	Cyclin Dependent Kinase
cDNA	Complementary DNA
CDS	Coding Sequence
DDR	DNA Damage Response
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme Linked Immunosorbant Assay
Elu	Eluate
EST	Expressed Sequence Tag
FLEX	Full Length Expression
Gal-1-P	Galactose 1 Phosphate
Gal4-AD	Gal4 Activation Domain
GB1	B1 Domain of Protein G
GFP	Green Fluorescent Protein
GO	Gene Ontology
GST	Glutathione-S-Transferase
HA	Hemaglutinin

HGP	Human Genome Project
Hrs	Hours
HT	High Throughput
IHF	Integration Host Factor
In	Input (Fraction)
INK4	Inhibitor of Cyclin Dependent Kinase 4
Int	Integrase
IPTG	Isopropylthio-ß-D-galactoside
kDa	Kilodalton
LB	Luria-Bertani
M	Moles
MALDI	Matrix-Assisted Laser Desorption
MBP	Maltose Binding Protein
Md	Maryland
min	Minutes
ml	Microliter
mM	Millimolar
MML	Myelomonocytic Leukemia
mRNA	Messenger RNA
MS	Mass Spectrometry
n/a	Not Applicable
Ni-NTA	Nickel
NMR	Nuclear Magnetic Resonance
NP-40	Nonidet P-40
O/N	Over Night
OD	Optical Density
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PTM	Post Translational Modification
QC	Quality Control

RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
S	Supernatant
SAGA	Spt-Ada-Gcn5-acetyltransferase transcriptional activator–histone acetyltransferase complex
SDS-PAGE	Sodium-Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis
sec	Seconds
SNP	Single Nucleotide Polymorphism
T	Total (starting material)
TB	Terrific Broth
TM	Transmembrane protein(s)
Trx	Thioredoxin
UTR	Untranslated Region
W	Wash (Fraction)
Xis	Excisionase

Publications

Braun P, Hu Y, Weng H, Schick M, Korn B, LaBaer J; Different Parameters Dictate Protein Expression Success of His_6- and GST-tagged Proteins in Bacteria, manuscript in preparation

Braun P, LaBaer J; Protein Production Systems for Proteomics, Trends in Biotech, invited review - manuscript submitted

Podgrabinska S, **Braun P**, Velasco P, Kloos B, Pepper MS, Jackson DG, Skobe M; Molecular characterization of lymphatic endothelial cells, *Proc Natl Acad Sci U S A.* 2002 Dec 10;99(25):16069-74

Braun P, Hu Y, Shen B, Halleck A, Koundinya M, Harlow E, LaBaer J; Proteome Scale Purification of Human Proteins from Bacteria *Proc Natl Acad Sci U S A.* 2002 Mar 5; 99 (5): 2654-9

Brizuela L, **Braun P**, LaBaer, J; FLEXGene repository: from sequenced genomes to gene repositories for high-throughput functional biology and proteomics. *Mol. Biochem. Par.,* 2001 Dec;118(2):155-165

Eisenhardt D, Fiala A, **Braun P**, Rosenboom H, Kress H, Ebert PR, Menzel R: Cloning of a catalytic subunit of cAMP-dependent protein kinase from the honeybee (Apis mellifera) and its localization in the brain. *Insect Mol Biol.* 2001 Apr;10(2):173-81.

Classon M, Salama S, Gorka C, Mulloy R, **Braun P**, Harlow E; Combinatorial roles for pRb, p107, and p130 in E2F-mediated cell cycle control. *Proc Natl Acad Sci U S A.* 2000 Sep 26;97(20):10820-5.

Vidal M, **Braun P,** Chen E, Boeke JD, Harlow E**;** Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. *Proc Natl Acad Sci U S A* 1996 Sep 17;93(19):10321-6

Curriculum Vitae

Part I: General Information

Name: Pascal Braun

Home Address: 7 Revere St. Apartment #3
Boston, MA 02130

Email: pbraun@hms.harvard.edu Phone**:** + 1 (617) 983 1243

Marital Status: single

Citizenship: German

Education:

Year	Degree	Institution
1997 – present	Graduate Studies	Institute of Proteomics – Harvard Medical School Advisor: Ed Harlow
1996	Diploma in Biochemistry	Free University of Berlin
1991	Abitur (High School degree)	Staedtisches Gymnasium, Luisenschule Essen

Experience/Techniques:
Laboratory: Cell culture of mammalian and insect cell lines; protein expression in bacteria, human cells and in the baculovirus system; protein purification using ion exchange chromatography and protein affinity tags; immunoprecipitation of endogenous and overexpressed proteins, western blot analysis, enzyme assays, immunfluorescence, recombinational cloning, enzymatic cloning techniques, PCR, hybridization, TECAN Gemini, TECAN GenMate, high throughput experimentation

B. Report of Teaching

Training and supervision of undergraduate students	2003 - present
Cold Spring Harbor Course Proteomics Teaching Assistant	2002
Training and supervision of technical and laboratory assistants	1999- 2002

Part III: Bibliography

A. Publications: see above

B. Awards and Fellowships:

Merck Pilot Research Program Grant	2003
Boehringer Ingelheim Fonds (Fellowship)	1997 - 2000
Society for Biological Chemistry – GBCh (Training Award)	1995

C. Selected Presentations

ABRF (invited speaker)	2003
Proteome Society (selected speaker)	2002
IBC Conference - Bioinformatics (invited speaker)	2000
Cold Spring Harbor - Cancer Genetics and Tumor Suppressor Genes (selected speaker)	2000

Acknowledgements

This work would not have been possible without the support of many remarkable individuals.

Zunaechst moechte ich meiner Mutter, meinem Bruder, meiner Familie und meinen Freunden fuer ihre kontinuierliche Unterstützung in guten und dunklen Zeiten danken.

Secondly, I would like to thank Ed Harlow for his continuous support and challenging discussions and the truly friendly interaction, which helped me to grow professionally and personally, and made the time of the doctorate an enjoyable growth experience.

I would like to thank Josh LaBaer for his enthusiastic assistance and backing and the provision of many great opportunities.

Furthermore I would like to appreciate the colleagues and friends in the Harlow/ Dyson and in the HIP labs, who helped with advice, support and a pleasant and enjoyable working atmosphere.

I would like to appreciate the Boehringer Ingelheim Fonds, who provided financial support and Monika Beutelspacher und Herrmann Froehlich, who build a great forum to make friends and meet interesting colleagues.

A special thank goes to Marc Vidal who was an important source of inspirations.

And most importantly, I would like to say ‘Thank You’ to Martina Schinke for being herself and being in my life.
Du machst mich glücklich, Schatz.

Thank you all! **Vielen Dank euch Allen!**