LITERATURE REVIEW:
PROTEIN AGGREGATION

EXPERIMENTAL WORK:
SOLUBLE EXPRESSION OF HUMAN CILIARY NEUROTROPHIC FACTOR IN *Escherichia coli*

Jaakko Itkonen
Centre for Drug Research/ Division of Pharmaceutical Biosciences
Faculty of Pharmacy
University of Helsinki

August 2014

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos Institution – Department
Faculty of Pharmacy	Division of Pharmaceutical Biosciences/Centre for Drug Research

Tekijä Författare – Author
Jaakko Matias Itkonen

Työn nimi Arbetets titel – Title
Protein aggregation / Soluble expression of human Ciliary neurotrophic factor in Escherichia coli

Oppiaine Läroämne – Subject
Biopharmacy

Työn laji Arbetets art – Level	Aika Datum – Month and year	Sivumäärä Sidoantal – Number of pages
Master's Thesis	08/2012 - 08/2014	50 + 36 + Appendix

Tiivistelmä Referat – Abstract

Proteins are endogenous molecules that carry out most biological functions in vivo. They are called as the biological workhorses. Proteins are made up of polypeptide chains that usually fold in the three dimensional space to adopt a native stable conformation. Stability of proteins is dependent on the interplay of environmental factors (pH, temperature, ionic strength). For most proteins, the biological function closely relates to the structural attributes of the protein. Misfolding or unfolding of proteins often result in aggregation. Protein aggregation in vivo is known to cause debilitating and fatal diseases such as Alzheimer's, Huntington's, Parkinson's and age related macular degeneration (AMD). Instability (physical and chemical) of proteins in vitro is believed to result in aggregation. This is a huge concern for the biopharmaceutical industry as it not only limits the effectiveness of the manufacturing process but also poses a great risk of fatality in vivo due to the immunogenic nature of the aggregates.
Mechanisms of protein aggregation are complex and not well understood. Regulatory requirements for patient safety in biopharmaceutical products require characterization and analysis of aggregates in protein drug formulations. This review provides an overview of protein aggregation in general and highlights the different analytical methods used to characterize protein aggregates in biopharmaceuticals.

Neurotrophic factors influence survival, differentiation, proliferation and death of neuronal cells within the central nervous system. Human ciliary neurotrophic factor (hCNTF) has neuroprotective properties and is also known to influence energy balance. Consequently, hCNTF has potential therapeutic applications in neurodegenerative, obesity and diabetes related disorders. Clinical and biological applications of CNTF necessitate a recombinant expression system to produce large amounts of functional protein. Previous studies have reported that recombinant expression of CNTF in Escherichia coli (E. coli) was limited by low yields and the need to refold the protein from inclusion bodies.
In this report, we describe a strategy to effectively screen fusion constructs and expression conditions for soluble hCNTF production in E. coli. Most conditions tested with the codon optimized hCNTF sequence in fusion with soluble tags resulted in soluble expression of the protein. The construct 6-His-CNTF showed soluble expression in all the conditions tested. Our results suggest that codon optimization of the hCNTF sequence is sufficient for soluble expression in E. coli. The recombinant hCNTF was found to bind to CNTFRα with an EC50 = 36 nM.

Avainsanat – Nyckelord – Keywords
biopharmaceuticals, excipients, stability, formulation, hCNTF, codon optimization, factorial screening, soluble expression, E. coli

Säilytyspaikka – Förvaringställe – Where deposited
Division of Pharmaceutical biosciences

Muita tietoja – Övriga uppgifter – Additional information
Supervisors: Sanjay Sarkhel Ph.D, Arto Urtti Pharm.D, prof.

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty	Laitos Institution – Department
Farmasian tiedekunta	Farmaseuttisten biotieteiden osasto/ Lääketutkimuksen keskus

Tekijä Författare – Author
Jaakko Matias Itkonen

Työn nimi Arbetets titel – Title
Protein aggregation / Soluble expression of human Ciliary neurotrophic factor in Escherichia coli

Oppiaine Läroämne – Subject
Biofarmasia

Työn laji Arbetets art – Level	Aika Datum – Month and year	Sivumäärä Sidoantal – Number of pages
Pro Gradu	08/2012 - 08/2014	50 + 36 + Liitteet

Tiivistelmä Referat – Abstract

Proteiinit ovat endogeenisiä molekyylejä, jotka suorittavat lähes kaikki solun toiminnot ja ovat solun normaalille toiminnalle välttämättömiä. Proteiinit muodostuvat polypeptidiketjuista, jotka laskostuessaan normaalisti muodostavat stabiilin ja toimintakykyisen kolmiulotteisen rakenteen. Proteiinien stabiilius on riippuvainen ympäristön tekijöistä (pH, lämpötila, ionivahvuus, jne.) ja niiden keskinäisistä vuorovaikutuksista. Useimmilla proteiineilla niiden biologinen toiminta on läheisesti kytköksissä niiden rakenteeseen. Proteiinien vääränlainen laskostuminen tai laskoksen aukeaminen johtaa usein proteiinien ryvästymiseen eli aggregaatioon. In vivo proteiinien aggregaation tiedetään olevan osallisena aiheuttajana vaikeissa ja tappavissa sairauksissa, kuten Alzheimerin ja Parkinsonin taudissa, sekä verkkokalvon ikärappeumassa. In vitro, proteiinien fysikaalisen ja kemiallisen epästabiiliuden uskotaan johtavan aggregaatioon. Tämä on suuri huolenaihe varsinkin biofarmaseuttiselle teollisuudelle, sillä se rajoittaa proteiinilääkkeiden valmistuprosessin tehokkuutta. Tärkeämpänä syynä on kuitenkin proteiiniaggregaattien vahva immunogeeninen luonne, joka voi aiheuttaa vakavia seurauksia potilaissa.
Proteiinien aggregaation mekanismit ovat monimutkaisia ja huonosti ymmärrettyjä. Biologisten lääkkeiden potilasturvallisuutta koskevat säännökset vaativat proteiinilääkeformulaatioilta proteiiniaggregaattien analyysiä ja karakterisointia. Tämä katsaus antaa yleiskuvan proteiiniaggregaatiosta ja keskittyy biologisten lääkkeiden proteaggregaattien analyysissa ja karakterisoinnissa käytettyihin metelmiin.

Neurotrofiset tekijät vaikuttavat keskushermoston neuronien eloonjäämiseen, erilaistumiseen, leviämiseen ja säädeltyyn kuolemaan. Ihmisen siliaarisella hermokasvutekijällä (human Ciliary neurotrophic factor, hCNTF) on neuroneja suojaava ominaisuus ja sen tunnetaan vaikuttavan myös energiatasapainoon. Tästä syystä CNTF:llä onkin potentiaalista terapeuttista käyttöä neurodegeneratiivisten tautien ja aineenvaihduntatautien hoidossa. Tälläisia tauteja ovat esimerkiksi liikalihavuus ja diabetes. CNTF:n kliininen ja biologinen käyttö vaatii rekombinanttien ilmennysmenetelmien käyttöä, jotta voidaan tuottaa suuria määriä toiminnallista proteiinia. Aikaisemmissa tutkimuksissa on raportoitu rekombinantin CNTF:n Eschrecchia coli (E. coli) -soluissa tapahtuva ilmentäminen on rajoittunut alhaisen saannon vuoksi ja vaatii usein proteiinin uudelleenlaskostusta inkluusiojyväsistä.
Tässä raportissa kuvaamme menetelmän, jolla voi seuloa tehokkaasti eri konstrukteja ja ilmentämisolosuhteita liukoisen hCNTF:n tuottamiseen E. coli-soluissa. Liukoisen proteiinien ilmentymistä havaittiin useimmilla liukoisen fuusioproteiinin ja kodonioptimoidulla hCNTF-sekvenssin yhdistelmillä ja tutkituissa ilmentämisolosuhteissa. Fuusiokonstruktin 6-His-CNTF havaittiin ilmentyvän liukoisena kaikissa tutkituissa olosuhteissa. Havaintomme viittaavat hCNTF:n sekvenssin kodonioptimoinnin olevan riittävä aikaansaamaan hCNTF:n ilmentyminen liukoisena E. coli-soluissa. Rekombinantin hCNTF:n havaittiin sitoutuvan CNTFRα:aan EC50-arvolla 36 nM.

Avainsanat – Nyckelord – Keywords
Biologiset lääkkeet, apuaineet, stabiilisuus, formulaatio, hCNTF, kodonioptimointi, tekijöihin perustuva seulonta, liukoinen ilmennys, E. coli

Säilytyspaikka – Förvaringställe – Where deposited
Farmaseuttisten biotieteiden osasto

Muita tietoja – Övriga uppgifter – Additional information

Ohjaajat: Sanjay Sarkhel, FT, Arto Urtti, FaT, professori

ABBREVIATIONS

3D	Three-dimensional
AF4	Asymmetrical flow field flow fractionation
ALS	Amyotrophic lateral sclerosis
AMD	Age-related macular degeneration
Anti-TNF	anti-tumor necrosis factor
AUC	Analytical ultracentrifugation
bCNTF	Biotinylated CNTF
BSA	Bovine serum albumin
CD	Circular dichroism spectroscopy
CHO	Chinese hamster ovary cell
CNTF	Ciliary neurotrophic factor
CNTFRα	CNTF receptor subunit α
DLS	Dynamic light scattering
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	DL-dithiothreitol
E. coli	*Escherichia coli*
EMA	European Medicines Agency
ESI-MS	Electrospray-ionization mass spectrometry
FDA	Food and Drug Administration
FFF	Field flow fractionation
fp	Forward primer
GdHCl	Guanidine hydrochloride
hCNTF	human CNTF
hCNTFRα	human CNFTRα
HD	Huntington's disease
HEK	Human embryonic kidney cells
hGH	human growth hormone

HPV	Human papilloma virus
HRP	Horseradish peroxidase
HSA	Human serum albumin
IDP	Intrinsically disordered proteins
IL-6	Interleukin-6
IL-6Rα	Interleukin-6 receptor α chain
IMAC	Immobilized metal ion affinity chromatography
IPTG	isopropyl-1-thio-β-D-galactopyranoside
KGF-2	Keratinocyte growth factor 2
LB	Lysogeny broth / Luria-Bertani broth
LIF	Leukemia inhibitory factor
LIFRβ	Leukemia inhibitory factor receptor β
LOD	Limit of detection
LOQ	Limit of quantitation
mAb	Monoclonal antibody
mRNA	Messenger ribonucleic acid
NBE	New biological entity
Ni-IDA	Nickel iminoacetic acid
Ni-NTA	Nickel nitriloacetic acid
NMR	Nuclear magnetic resonance
OPG-Fc	Osteoprotegerin fused with the Fc-portion of an immunoglobulin
OPPF-UK	Oxford Protein Production Facility UK
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCS	Photon correlation spectroscopy
PEG-GCSF	Polyethylene glycosylated granulocyte colony stimulating factor
PEG-MGDF	Polyethylene glycosylated megakaryocyte growth and development factor
PIC	Protease inhibitor cocktail
PPB	Power *Prime* Broth™
PTM	Post-translational modification
QC	Quality control

QELS	Quasi-elastic light scattering
r_h	Hydrodynamic radius
RNA	Ribonucleic acid
rp	Reverse primer
sCNTFRα	Soluble CNTFRα
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE	Sedimentation equilibrium
SEC	Size-exclusion chromatography
SV	Sedimentation velocity
TAE	Tris-Acetic acid-EDTA buffer
TBONEX	Overnight Express™ Instant TB Media
TBS-T	Tris-buffered saline with Tween
TEM	Transmission electron microscopy
T_m	Melting temperature
TMB	3,3’,5,5’-tetramethylbenzidine
tRNA	Transfer-ribonucleic acid
TSE	Transmissible spongiform encephalopathy
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

PROTEIN AGGREGATION

Jaakko Itkonen
Centre for Drug Research/ Division of Pharmaceutical Biosciences
Faculty of Pharmacy
University of Helsinki
September 2012 – August 2014

1 INTRODUCTION

The last decade has seen a vast increase in the development of biomolecules as therapeutic agents to treat a wide array of diseases. Various types of bio-drugs (proteins, antibodies, nucleotides) investigated in the recent past have entered the late-stage clinical studies but almost all that have reached the market, are protein based drugs (Walsh 2010). Recombinant insulin for treatment of diabetes was the first biopharmaceutical product to be launched in 1982 (Goeddel *et al.* 1979, Rader 2013).

It was estimated in 2009 that the sales of recombinant therapeutic proteins and monoclonal antibodies (mAb) was around 100 billion dollars. This yielded the biopharmaceutical market one-sixth the total volume of the 600 billion dollar pharmaceutical industry (Walsh 2010). Promising data stood out in 2012; seven out of the fifteen top-selling pharmaceuticals by revenue were biologicals (FiercePharma 2012). Over 40 % of new pharmaceuticals currently being developed are reported to be biopharmaceuticals (Rader 2013). Over two hundred mAbs and a hundred proteins were reported to be in clinical trials (Sheridan 2010). However, the development of protein drugs has been limited by low approval rates; for the period between 2006 and 2010, only twenty-five new biological entities (NBE) were released in the US and EU market (Walsh 2010) and only four in 2013 (Kling 2014). Current estimates for protein therapeutics project a growth rate between 7 and 15% annually for coming years (Walsh 2010).

Therapeutic proteins are endogenous (or engineered proteins closely resembling endogenous proteins) and are therefore expected to have better specificity and safety profile as compared to the conventional small molecule drugs (Crommelin *et al.* 2003, Leader *et al.* 2008). The major challenges for developing protein drugs include cost, complex manufacturing, relative instability, inadequate pharmacokinetic properties and formulation in conventional dosage forms, such as tablets or capsules (Leader *et al.* 2008, Swami and Shahiwala 2013).

For most proteins to elicit biological functions, they need to fold and adopt a stable conformation, commonly referred to as the native state (Dobson 2004, Tyedmers *et al*. 2010). Environmental conditions *in vivo* are controlled by homeostasis and remain relatively stable. This helps proteins maintain their conformational integrity *in vivo*. However, *in vitro* changes in protein environment such as pH, temperature and ionic strength are often of a higher magnitude that might reversibly or irreversibly perturb the higher-order protein structure and result in denaturation (partial or complete unfolding from the native state). In most cases, unfolding of protein enhances intermolecular interactions between polypeptide chains leading to higher order aggregates. In some instances, associated tertiary or quaternary structures are considered as aggregates but for the purpose of this review, we refer to protein aggregation as a phenomenon that results in clustering of proteins and formation of higher order oligomerization with diminished or complete loss of biological function. Conformational stability of proteins is of paramount importance for biological functionality (Manning *et al*. 2010).

Although the ability of proteins to form higher order aggregates was initially associated with *in vivo* diseases, it is now acknowledged that this ability is an inherent characteristic of proteins (Chiti *et al*. 2001). Protein aggregates have varying properties. They are either soluble or insoluble (Philo 2006), vary in the number of monomers, size, life-time, shape and structure (Wang 2005). Apart from severely limiting the efficacy of production and manufacturing processes, protein aggregates are known to elicit potentially life threatening immunologic responses *in vivo* (Rosenberg 2006). Thus, from both bioprocessing and regulatory perspective, protein aggregation presents a huge challenge. Mechanisms of protein aggregation are poorly understood and new analytical tools are being developed for their characterization.

2 PROTEIN STRUCTURE AND FOLDING

Proteins are composed of polypeptide chains; they follow a sequence of events to acquire the biologically active native structural state *via* a process called ‘protein folding’ (Branden and Tooze 1998). Attainment and maintenance of the native structural state is critical for biological activity *in vivo*. Polypeptide chains are composed of twenty different amino acids linked to each other *via* peptide bonds. The chain forms a repeating backbone with alternating side-chains of the amino acid residues (Richardson 1981, Branden and Tooze 1998). The order of amino acids (sequence) as well as the length of the chain (number of amino acids) is determined by the genetic code. The sequence of the amino acids in a polypeptide chain defines the primary structure of the protein.

As the polypeptide chain gets synthesized, it folds and forms ordered secondary structural elements. These are attributed to multiple non-covalent interactions such as hydrogen bonds (between polar residues), ionic bonds (between oppositely charged residues) and hydrophobic interactions (between non-polar residues). Secondary structural elements present in proteins are α-helices, β-sheets, turns and random coils. Multiple secondary structural elements are arranged into motifs that pack into compact structural domains (Branden and Tooze 1998). The three-dimensional (3D) arrangement of motifs within the folding space is referred to as the tertiary structure. It is generally believed that the driving force for protein folding is a consequence of the relatively hydrophobic residues trying to shield themselves from the aqueous environment (Wang 1999, Levy and Onuchic 2006). This so-called ‘hydrophobic collapse’ is postulated to result in the burial of hydrophobic residues within the core of protein structure. Polar and charged residues usually decorate the surface of the protein, which is solvent exposed.

The interplay between covalent (disulphide bonds between cysteine residues) and non-covalent interactions stabilizes the tertiary structure of the protein (Branden and Tooze 1998). In multimeric proteins, the biologically functional unit is composed of a quaternary structure formed by oligomerization of the monomeric components.

The paradigm of structure-activity relationship is based on the fact that the biological function of a protein is a consequence of the three dimensional structure. A partial or complete loss of structural integrity might result in loss of activity *in vivo*. Improper folding of proteins is known to cause pathological diseases (Dobson 2004). However, it is noteworthy to mention that many biologically active proteins do not possess a structural fold in the native state. These proteins are commonly referred to as 'intrinsically disordered proteins' (IDPs). IDPs have in recent times generated a lot of interest for their physiological role (Dyson 2011). This review however, limits its scope to only natively folded proteins and not the IDPs.

3 *IN VIVO* PROTEIN AGGREGATION

Many diseases of the central nervous system, such as amyotrophic lateral sclerosis (ALS), prion diseases and amyloidoses (Parkinson's and Alzheimer's) involve protein misfolding and aggregation (Stefani and Dobson, 2003). For example, ALS is associated with aggregation of Cu/Zn superoxide dismutase (Stathopulos *et al.* 2003); transmissible spongiform encephalopathies (TSEs) are associated with the prion, a protein which is misfolded and aggregates *in vivo* (Aguzzi and Calella 2009); Parkinson's disease is associated with the aggregation of the α-synuclein protein (Nath *et al.* 2011) and in Alzheimer's disease, the microtubule associated protein Tau when hyper-phosphorylated results in the formation of neurofibrillary tangles (Kfoury *et al*. 2012). It is not clear whether the manifestation of pathological diseases is a direct consequence of the aggregation of physiological proteins; nonetheless, aggregation presents a common feature in the pathologies of the diseased state. The exact mechanisms of protein aggregation and their effects on the organism at the cellular level are not well understood and are subjects of intense research (Dobson 2004). Some evidences point to the fact that intracellular protein aggregates might impair cellular functions *via* interactions between exposed hydrophobic regions of protein aggregates and cellular proteins or the cell membrane (Stefani and Dobson 2003, Tyedmers *et al.* 2010).

Newly synthesized polypeptide chains experience a high intracellular concentration of proteins, commonly referred to as 'molecular crowding'. High intracellular concentrations might tend to favor aggregation (Dobson 2004, Barral *et al.* 2004) and be harmful to the cellular machinery. To mitigate such adversities, cellular mechanisms have evolved to either course correct misfolded/unfolded proteins or degrade them (Stefani and Dobson 2003).

Molecular chaperones, such as heat shock proteins, function by either assisting the correct folding of the newly synthesized polypeptide chain or by blocking incorrect folding pathways. Some chaperones are known to be able to rescue proteins from aggregates (Barral *et al.* 2004, Tyedmers *et al*. 2010). The formation and sequestration of intracellular protein aggregate deposits is considered to be a next-in-line response in protecting the cellular machinery from harmful effects of the aggregates. (Tyedmers *et al.* 2010). Aggregation of proteins *in vivo* is often a part of the cellular response to an imbalance in protein homeostasis – cell's own 'quality control' process. We appreciate and understand the importance of protein aggregation *in vivo* but to limit the scope of this review, we discuss here *in vitro* protein aggregation that has more direct relevance in the context of biopharmaceutical development.

4 AGGREGATION OF THERAPEUTIC PROTEINS

Therapeutic proteins are produced recombinantly in expression systems and the environmental conditions that the protein experiences in its journey from synthesis to purification are very different from those *in vivo*. The protein is exposed to environmental stresses both physical and chemical in nature that can affect the stability of the native state. Both the physical and chemical factors can perturb the tertiary structure of the protein and result in unfolding of the polypeptide chain. Unfolding of proteins is believed to be one major cause of aggregation. The physical instability is brought about by changes in conditions and external factors such as temperature, shear stresses, pressure etc. (Wang 2005). Chemically induced changes and degradation of the protein structure are caused by factors such as oxidation, deamidation, acylation and peptide bond hydrolysis (Crommelin *et al*. 2003).

The propensity of a protein to undergo aggregation is dependent on the inherent properties of the protein, namely its sequence and structure (Pawar *et al.* 2005, Tartaglia *et al.* 2008). This hypothesis is supported by the fact that in certain cases, even point mutations are known to significantly influence the aggregation propensity of a protein (Fink 1998, Chiti *et al.* 2002, Stathopulos *et al.* 2003). Also, it has been observed that β-sheet structured proteins are more prone to aggregation than α-helical proteins. Many protein aggregates have diminished α-helix content and an increased β-sheet content as compared to the native structure (Wang 2005, Laurence and Middaugh 2010). However, not all proteins with high β-sheet content have the propensity to aggregate. Conformational changes causing transition from α-helices to β-sheets might enhance the propensity to aggregate (Chiti *et al.* 2002). Conversely, in certain cases, transitions from β-sheets to α-helices might result in increased stability (Villegas *et al.* 2000).

Strategies to prevent protein aggregation in formulations present a significant challenge for biopharmaceutical development. The approach in general, is to find conditions that would stabilize the native state or destabilize the unfolded state. This can be achieved by adding excipients, chemical modifications to the protein or by recoding the genetic sequence to remove aggregation prone regions in the protein.

4.1 Pathways and mechanisms of protein aggregation

Protein aggregates may consist of native protein monomers or completely unfolded proteins, although a growing body of evidence supports the idea that intermediates between native and completely unfolded structures serve as precursors to protein aggregates (Fink 1998, Dobson 2004). These intermediates can be envisioned as partially unfolded proteins with exposed aggregation-prone regions that can interact with corresponding regions of other proteins.

Protein folding and aggregation can take place by different pathways (Figure 1). Aggregate formation is not exclusive and can occur *via* different mechanisms simultaneously (Wang 2005; Philo and Arakawa 2009). The presence of aggregates in a therapeutic protein sample and importantly, the amount and type of aggregates present might change with time depending on the kinetics of association and dissociation. It has been observed that for slow processes, it might take several hours or even days to equilibrate after a perturbation in protein conditions (Philo 2003). On the other hand, some aggregates are transient and have a relatively short life-time (Philo 2006). The types of aggregates formed depend on the mechanism and pathway of aggregation (Tyedmers *et al.* 2010). Understanding the phenomenon of aggregation is thus of fundamental importance to develop rationale for its inhibition; a major challenge in biopharmaceutical development.

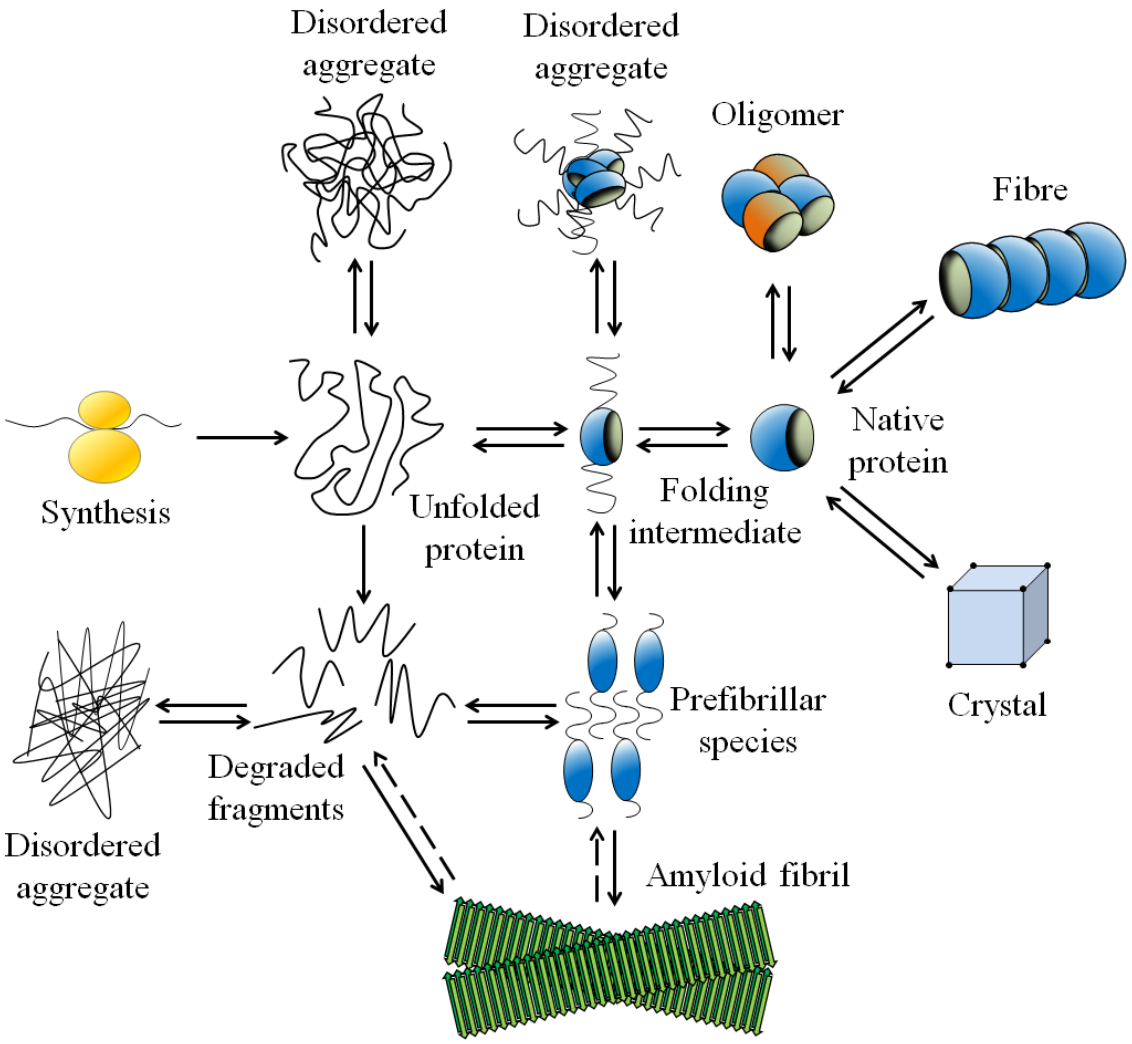


Figure 1: Protein folding, degradation and aggregation pathways.
Adapted, with permission, from Dobson © (2003) Macmillan Publishers Ltd.

Protein aggregates are often categorized based on properties of reversibility and irreversibility. However, irreversibility doesn't necessarily mean a permanent change; perturbations in environmental conditions (e.g., pH, temperature) may alter an irreversible aggregate into a reversible one (Philo 2006). In the same logic, terminologies such as soluble or insoluble aggregates might be misleading. The nomenclature used in the field of protein aggregation varies widely and terminologies are often used interchangeably by research groups. Efforts are being directed to harmonize the use of terms regarding protein aggregates (Narhi *et al*. 2012).

Factors leading to protein aggregation include but are not limited to: changes in protein concentration, pH and temperature and different processing steps involving mechanical stress caused for instance by stirring, shaking, pumping and freeze/thaw cycles (Wang 2005, Mahler *et al.* 2009). To which extent each factor affects the rate of formation and the type of aggregates formed differs. These factors and their effects on protein aggregation are discussed in Section 4.4.

4.1.1 Covalent aggregates

Structural modifications induced by chemical changes such as deamidation, transamidation (Strickley and Anderson 1997) and oxidation of amino acid residues such as histidine (Khossravi *et al*. 2000) arginine, lysine, methionine, and proline (Stadtman 1993). Changes in chemical structure of a protein alter the electronic charges that can potentially influence the attractive or repulsive forces between residues of protein monomers and/or expose create novel regions capable of interacting with other surfaces and forming aggregates.

Covalent disulphide bonds between cysteine residues stabilize protein structure. However, presence of unpaired cysteine residues also makes them susceptible to form intermolecular disulphide bonds between polypeptide chains and aggregate (Yoshioka 1993, Trivedi *et al*. 2009, Brych *et al.* 2010). Non-native disulphide bonds are believed to stabilize intermediate states of aggregation. Tyrosine residues are also known to undergo oxidation to form covalent dityrosine bonds (Malencik and Anderson 2003). Covalent aggregates are generally irreversible in nature.

4.1.2 Non-covalent aggregates

Native protein monomers may associate with each other by virtue of weak non-covalent interactions such as, hydrogen bonds, ionic bonds and hydrophobic interactions. Self-complementary regions of a protein are therefore prone to aggregation (Figure 2a) Aggregation via this pathway (intermolecular association) is a function of the protein concentration; higher concentrations likely result in more encounters between the monomers that enhance association. The size of such aggregates also tends to increase over time (Philo and Arakawa 2009). Insulin monomers are known to associate and form hexamers inside the pancreas (Xu *et al*. 2012, Bryant *et al*. 1993). The hexamers subsequently dissociate into dimers in the blood stream and finally into monomers (Dobson and Steiner 1998). The different oligomeric states of insulin do not affect the bioactivity or immunogenicity of insulin as long as dissociation of the multimer takes place. However, the pharmacokinetic profile varies for the different oligomers and this has been utilized to engineer new insulin-products with desired pharmacokinetics and half-life (Crasto *et al*. 2009, Danne and Bolinder 2012).

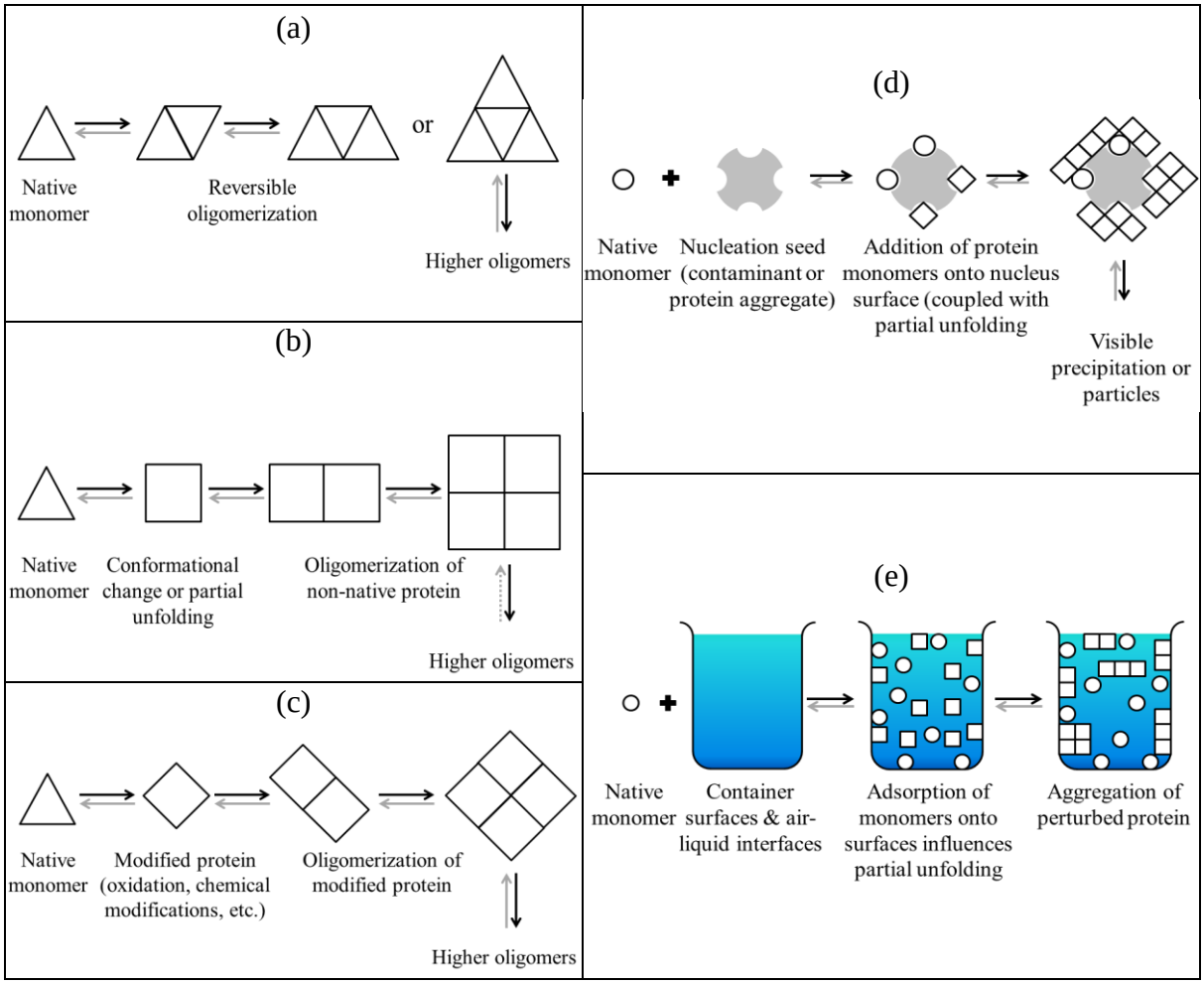


Figure 1: General mechanisms of protein aggregation.
Adapted, with permission, from Philo and Arakawa © (2009) Bentham Science Publishers Ltd.

In a related mechanism, although the native monomer does not self-associate but partial unfolding or a conformational change from the native state might expose regions of the protein that are prone to associate (Figure 2b). The perturbation in structure can be caused by external stress, commonly encountered in the different manufacturing and processing steps. Exposed regions might also be susceptible to chemical changes enhancing aggregation (Figure 2c) (Wang 2005).

Protein aggregation can be triggered by a contaminant or an impurity present in the sample (Figure 2d). Usually, such contaminants are derived from manufacturing and storage processes. Examples include, steel particles (Tyagi *et al.* 2009) tungsten, rubber (Sharma 2007), silica (Chi *et al*. 2005), particles from syringe filters (Liu *et al*. 2012) and silicone oil droplets (Majumdar *et al*. 2011). Such foreign particles usually act as nucleation seeds (Philo 2009), for aggregation. The nucleus grows in size as more protein molecules associate. Such a process of aggregation is often termed as heterogeneous nucleation and reinforces the importance of avoiding foreign particles/contaminants in protein formulations. Protein aggregates formed *via* other mechanisms may also act as nuclei to promote aggregation by this mechanism (Andrews *et al*. 2008, Kiese *et al*. 2010) but referred to as homogenous nucleation. Aggregation *via* this mechanism typically results in visible particles or precipitates as the large aggregates associate (Speed *et al*. 1997). Characteristic to this mechanism also is that at an early phase aggregates are virtually undetectable, but then rather suddenly large aggregates are present and accumulate (Chi *et al*. 2003, Philo and Arakawa 2009). The lag-phase associated with the nuclei based aggregation is due to an energy barrier; the initial rate of aggregation around the nucleation seed is slow but after a critical size has been reached, aggregation takes place rapidly. Formulation strategies are targeted towards influencing the lag phase to minimize aggregation e.g., alteration of the formulation's viscosity to decrease diffusion. Apart from foreign particles and previously formed aggregates acting as nucleation seeds in aggregation process, the interfaces and surfaces with which the proteins have contact also affect aggregation (Figure 2e). For instance, polystyrene (Smith *et* al. 2007), air-water interface (Bee *et al*. 2011), stainless steel surfaces (Bee *et al*. 2010) and ice-water interface (Kueltzo *et al*. 2008) have shown to induce aggregation in proteins. The contacts between protein and surfaces / interfaces are mediated by hydrophobic and electrostatic interactions that might result in binding of the protein followed by conformational changes of the native structure (Philo and Arakawa 2009). The conformational change increases the protein's propensity to aggregate.

4.2 External factors affecting protein aggregation

Environmental factors and conditions have a key role in protein stability. Identification and understanding their mechanism of action in protein stability is crucial towards developing rational approaches for inhibition of aggregation. Although each factor influences the stability of the protein, it is the interplay of the different factors that ultimately affect the conformational state of the protein. Some common factors that affect protein stability and aggregation propensity are discussed below. Some of the most common factors that influence protein aggregation and methods to inhibit them are discussed below.

4.2.1 Protein concentration

Concentration of protein is a significant factor in its aggregation. Higher protein concentrations enhance intermolecular protein - protein interactions that can potentially initiate the process of aggregation (Saluja and Kalonia 2008). It appears that for certain mechanisms, a minimal threshold concentration might be necessary for the initiation of aggregation (Wang 2005). Conversely, cases where high concentration of protein results in associates that are less prone to aggregation (Saluja and Kalonia 2008), high protein concentrations might not enhance aggregation propensity.

Protein therapeutics is most commonly formulated as solutions intended for parenteral administration (e.g., subcutaneous injections) since, other more conventional (oral) dosage forms are unsuitable due to enzymatic degradation and limited permeability across the gastro-intestinal epithelium (Frokjaer and Otzen 2005). Therapeutic formulations are often necessary to be of high concentrations to achieve high doses and at the same time limit the injection volumes (Shire *et al*. 2004, Frokjaer and Otzen 2005). Thus, even though having a low concentration of protein in formulations seems to be an easy solution (Shiraki *et al*. 2002) to mitigate aggregation related issues, practical requirements for clinical use limits its applicability.

4.2.2 Solution pH

Proteins are stable (retain native conformation) within a narrow range of pH, characteristic to each individual protein. Outside the optimal pH range, proteins might undergo transitions from native conformation, unfold and eventually aggregate (Chi *et al*. 2003). Changes in pH redistribute the charges in protein that can affect the attractive or repulsive interactions between molecules. During bioprocessing, a therapeutic protein may encounter changes in solution pH that can adversely affect its stability (Wang 2005). For instance, aggregation of partially unfolded monoclonal IgG was observed after a brief exposure to low pH (Filipe *et al*. 2012).

4.2.3 Temperature

Proteins retain stability below a critical temperature that is characteristic to each protein. Above the critical temperature, proteins might undergo structural unfolding making them susceptible to aggregation (Speed *et al*. 1997, Mahler *et al*. 2009). For example, thermal stress has been observed to result in the formation of small soluble monoclonal IgG_1 aggregates (Hawe *et al*. 2009).

An increase in temperature results in higher reaction rates and the frequency of intermolecular collisions increase with temperature. Consequently, proteins are known to aggregate at higher temperatures (Weiss IV *et al.* 2008, Mahler *et al.* 2009). Conversely, low temperatures can also affect the aggregation behavior of proteins, a phenomenon known as ‘cold denaturation’.

To avoid aggregation caused by increase in temperature during manufacturing and processing, care must be taken to avoid working temperatures above the melting temperature (T_m). T_m is defined as the temperature at which 50 % population of the protein is unfolded and is typically between 40° and 80° C for most proteins (Wang 1999). It is preferable to have operational temperatures well below T_m, typically between 2 - 8° C (Mahler *et al*. 2009). In certain cases, amino acids as excipients have shown to prevent temperature-induced aggregation (Wang 1999, Shiraki *et al.* 2002).

4.2.4 Foreign particles and container materials

Contaminants and particles present in a protein solution may serve as nucleation seed and induce aggregation. Binding of proteins to particles and surfaces can be reversible or irreversible and is mediated by forces such as, electrostatic and hydrophobic interactions (Bee *et al*. 2010, Bee *et al*. 2011). Adsorption of proteins onto different surfaces may also denature and unfold proteins (partially or completely) to cause aggregation (Manning *et* al. 2010, Bee *et al*. 2011). For instance, monoclonal IgG_2 absorbs onto surfaces like Teflon™ during the process of freeze thawing (Kueltzo *et al*. 2008).

In certain cases, the container material, its closure and stoppers is not a direct cause of protein aggregation but a brief exposure to other interfaces, surfaces and / or particles during processing steps, storage and delivery might be sufficient to initiate aggregation (Bee *et al*. 2011, Liu *et al*. 2012). For example, a leachable such as, Fe ions from steel, can act as a nucleation seed and also cause aggregation by oxidation (Lam *et al.* 1997). The phenomenon of leaching seems to be influenced by the properties of the solution. Stainless steel surface was reported to induce aggregation in a mAb and increase the amount of carbonylated degradation products (Bee *et al*. 2010). Shedding of stainless steel nanoparticles from piston pump contributed to heterogeneous nucleation of an IgG (Tyagi *et al*. 2009). Silicone oil, a coating frequently present on the surface of prefilled syringes commonly used with protein therapeutics, is also known to induce aggregation in many proteins (Jones *et al*. 2005, Majumdar *et al*. 2011). Similarly, microparticles and nanoparticles shed from syringe filters have been reported to stimulate particulate formation and aggregation of keratinocyte growth factor 2 (KGF-2) especially, under conditions of agitation (Liu *et al*. 2012). This observation assumes immense significance and concern as the very filters that are used to remove aggregates actually contribute to the process of aggregation.

Modifications to formulation and manufacturing processes or materials associated with equipment and containers are usually seen as methods to inhibit or slow aggregation caused by surfaces or foreign particles (Jones *et al.* 2005, Bee *et al*. 2011). For instance, the use of BD-42 (BD Technologies) coating inside prefilled syringes instead of silicone oil has been proposed for silicone-sensitive proteins (Majumdar *et al*. 2011). The use of surfactants has also been successfully utilized; polysorbate 20 suppressed aggregation of IgG_1 caused by stainless steel microparticles (Bee *et al*. 2010). Removal of foreign particulates suppressed heterogeneous nucleation and delayed aggregate formation in the formulation of recombinant human platelet-activating factor acetylhydrolase (Chi *et al*. 2005).

4.3 Detection and characterization of protein aggregates

Presence of protein aggregates in therapeutic protein formulations can be a serious threat to the efficacy and safety of the product. Protein aggregates are known to elicit immunogenic response that can be fatal. Regulatory guidelines regarding particles in dosage forms intended for injection in general have been made by the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA). The U.S. pharmacopoeia sets limit for acceptable amounts of visible and sub-visible particles in parenteral injections. However, acceptable levels of sub-visible and soluble particles (small sized protein aggregates, dimers etc.) are poorly defined and general consensus is still being evolved in this area of active research (Cromwell *et al.* 2006, Carpenter *et al.* 2009, Mahler *et al.* 2009, Singh *et al.* 2010).

In adherence to regulatory guidelines and guarantee the safe use of protein drugs, detection and characterization of protein aggregates is an absolute necessity. Apart from safeguarding regulatory interests, such efforts also help detect and investigate mechanisms of aggregation at each processing step and consequently in devising strategies to inhibit aggregation.

The methods used to characterize protein aggregates are wide and diverse; from simple visual inspection to sophisticated analytical methods that require expertise (den Engelsman *et al*. 2011). No single method is sufficient as the nature and size of aggregates influence the choice of analytical technique (Mahler *et al*. 2009, den Engelsman *et al.* 2011). Furthermore, each method is able to assess only a certain aspect of the protein aggregate, for instance aggregate size, morphology, aggregate concentration etc. The transient nature of the aggregates contributes to the complexity of characterization as the life-time of aggregates ranges from milliseconds to even days in the life-cycle of a protein therapeutic. The complexity is further compounded by the fact that the analytical method could itself interfere in the aggregation process and create artifacts (Philo 2006, Mahler *et al*. 2009). Current approaches utilize orthogonal methods to characterize protein aggregates (Philo 2006; den Engelsman *et al.* 2011). The use of orthogonal methods for similar sized particles might be useful in comparing different methods and optimization (Mahler *et al*. 2009).

Analytical tools used to detect and characterize protein aggregates can be broadly categorized as chromatographic, electrophoretic, light scattering, microscopic and spectroscopic methods (den Engelsman *et al*. 2011). The methods differ from each other in detection principles, detection size range and time needed for the analysis. Some methods are better suited for quality control (QC) purposes and others for more extensive characterization purposes. Some of the most commonly utilized methods for detection and characterization of protein aggregates are discussed in following sections.

4.3.1 Size exclusion chromatography

Size exclusion chromatography (SEC), also known as gel filtration, is one of the most popular analytical methods in quantifying and sizing protein aggregates (Mahler *et al*. 2009). Particles and aggregates of different sizes are separated on a column matrix based on their differential ability to enter and remain inside small pores and cavities of the matrix, dictated by the size of the particles or molecules (den Englesmann *et al*. 2011). Large aggregates cannot enter the pores to the same extent as small aggregates; consequently, their presence is limited to the outer volume of the column and they traverse shorter distance as compared to the smaller sized particles / molecules. The larger sized particles / molecular entities are thus eluted out first from the column. Medium-sized molecular entities can enter the pores to a limited extent and are retained inside the column longer than large aggregates but are still eluted out before smaller entities. The elution time / volume is monitored by measuring absorbance. The method allows sizing and quantitation of protein aggregates based on peak areas (Mahler *et al*. 2009). However, the accuracy of SEC in quantitation of aggregates can give rise to artifacts due to adsorption of proteins to the column matrix (Gabrielson *et al.* 2007).

Advantages in using SEC for protein aggregate analysis include relatively fast analysis and high sample throughput due to automation and high sensitivity (Carpenter *et al*. 2010, Zölls *et al*. 2012). Typically, small sample volumes are needed for SEC (den Engelsmann *et al*. 2011).

SEC can however be limited in its application primarily due to requirements in sample preparation. Larger insoluble aggregates cannot be applied to SEC and need to be filtered out during sample preparation (Mahler *et al*. 2009). High molecular weight oligomeric species might accumulate on the column or elute in the void volume (Zölls *et al*. 2012) and thus be overlooked in the analysis. These larger particles can also degrade the column over time and affect its performance adversely (den Engelsmann *et al*. 2011). Protein shape affects the accuracy of SEC analysis and correlations between elution volume / time with molecular weight might result in erroneous interpretation (Mahler *et al*. 2009). Dynamic ranges of SEC columns are also rather limited (Mahler *et al*. 2009); a column that results in a good separation between a monomer and a dimer may fail to resolve between a trimer and larger aggregates.

Buffer conditions during SEC might dissociate some aggregates and thus influence adversely the characterization of protein sample (Ahrer *et al*. 2003, Carpenter *et al*. 2009, den Engelsmann *et al*. 2011). Analyte properties may also be altered due to interactions with the matrix (Zölls *et al*. 2012). Adsorption of protein monomers and aggregates onto the column matrix is also a major concern (Carpenter *et al* 2010). Formation of protein aggregates during the process of SEC has also been observed (Philo 2006). Due to these limitations, SEC is generally used in conjecture with other orthogonal methods to validate the quality of protein therapeutics (Philo 2006, Carpenter *et al*. 2010).

4.3.2 Dynamic light scattering

Dynamic light scattering (DLS), also known as quasi-elastic light scattering (QELS) and photon correlation spectroscopy (PCS), was discovered in the 1950's, which measures the light-scattering intensity fluctuations caused by Brownian motion of particles in solution (Gun'ko *et al.* 2003, Li *et al.* 2011). This is influenced by hydrodynamic radius (r_h) of the particles, a characteristic feature dependent on the particle's mass and shape (Philo 2009). The method is non-destructive and requires minimal amounts of sample and thus suits very well early stage diagnostics on protein aggregation with limited sample availability (den Engelsman *et al*. 2011). Also, the components of the sample do not need to be separated. DLS is capable to detect both reversible and irreversible aggregates (Philo 2009).

Although DLS presents a versatile analytical tool to characterize protein aggregates, it does have certain limitations. It does not provide quantitative but only qualitative estimates (den Engelsman *et al*. 2011). Non-spherical particles are treated as spheres in estimations of hydrodynamic radii (Philo 2009). In principal, the size of an aggregate can be measured with DLS but when characterizing a polydisperse solution containing aggregates of varying sizes (which is often the case), the absolute sizes reported are inaccurate and lack precision (Ahrer *et al.* 2003, Philo 2009). The method is also inherently rather poor in resolution and not well suited for characterization of small oligomers as the signal intensity (proportional to the sixth power of diameter as due to the Rayleigh law) is weighted in favor of larger molecular species (Li *et al*. 2011). DLS also suffers from inability to distinguish between the identities of similar sized particles (Philo 2006). Ideally, DLS is best suited for rapid screening of large aggregates and particles in low concentration solutions (Philo 2009). DLS has been utilized in the detection of bovine serum albumin (BSA) and human serum albumin (HSA) aggregates (Jachimska *et al*. 2008), soluble aggregates of recombinant human factor VIII (Grillo *et al*. 2001), and aggregates in human IgG samples (Ahrer *et al*. 2003) in biopharmaceutical formulations.

4.3.3 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is an analytical method based on differing sedimentation characteristics between molecules of varying sizes during high-speed centrifugation (Philo 2009). The sedimentation follows Stokes' law and the sedimentation velocity of any particle depends on sedimentation coefficient, a parameter dependent on molecular mass, size and shape of the particle, viscosity of the fluid and the density difference between the particle and the fluid (Philo 2009, Zölls *et al*. 2012). Two different methods are most often followed: sedimentation velocity (SV), for measuring protein size and shape (Liu *et al*. 2006, Mahler *et al*. 2009) and sedimentation equilibrium (SE), for determination of protein quaternary structure (Philo 2003) and mass (Mahler *et al*. 2009). SV is more commonly utilized in protein aggregate studies and the term is often used interchangeably with AUC (Philo 2009).

In SV, the sample(s) are spun at very high speeds for 2 - 4 hours to separate the particles based on differing sedimentation velocities (Philo 2003). The data is analyzed with different data analysis programs (Berkowitz 2006, Liu *et al*. 2006). In SE, the sample(s) are conversely spun at low speeds and the centrifugal force pushes the particles outwards. Increase in concentration is opposed by diffusion of the particles resulting in an equilibrium distribution inside the centrifugal cell based on particulate mass (Philo 2003).

AUC is advantageous for investigating protein aggregation as no sample preparation other than dilution is often necessary and the solution protein formulation can often be used as a sample directly (Berkowitz 2006, Arthur *et al*. 2009). Multiple samples (e.g., protein in different formulation buffers) can be analyzed in parallel (Berkowitz 2006). A wide range of particle masses and sizes can be analyzed using AUC by altering the centrifugation velocity. This approach is useful for polydisperse solutions. Both, SV and SE are based on physical principles and do not require protein standards for calibration. AUC allows comparison of data acquired years apart and this makes it an attractive method for studies on long-term quality assessment and stability of protein samples (Philo 2003).

The ability to detect aggregates present in low concentrations is critical and failure to do so might mean that some aggregates remain unaccounted for in a sample. This might result in potential artifacts in data analysis (Gabrielson *et al*. 2009). Both, the limit of detection (LOD) and limit of quantitation (LOQ), are relatively high for SE, especially for certain types of protein aggregates (Philo 2009). Due to the relatively long experimental run time, AUC might fail to detect aggregates of transient nature (Philo 2006). AUC also suffers of rather poor sensitivity towards smaller aggregates and also of poor reproducibility as the phenomenon of sedimentation is a consequence of interplay between several factors that cause considerable variation (Philo 2009). For instance, misalignment of the centrifugal cell was found to be a major contributor to variability in the analysis of an antibody (Arthur *et al*. 2009). Sedimentation of excipients in protein formulation are known to produce gradients of density and viscosity that might affect the sedimentation of protein aggregates significantly and hinder their detection (Philo 2009). For example, 5% sorbitol was shown to mask the presence of low concentration of mAb aggregates during SV due to co-sedimentation of sorbitol (Gabrielson *et al*. 2009). The quality of the centerpieces and cell holders can also affect the precision and accuracy of the measurements and requires regular inspection of the integrity of such parts (Pekar and Sukumar 2007).

Despite certain limitations, AUC remains a powerful tool for detection and characterization of protein aggregates. It can be used orthogonally to validate and complement other analytical methods (Liu *et al*. 2006, den Engelsmann *et al*. 2011) and also in development of more accurate SEC methodology (Berkowitz 2006).

4.3.4 Asymmetrical flow field flow fractionation

Different field flow fractionation (FFF) methods have been developed for analytical purposes (Mahler *et al*. 2009). Of these, asymmetrical flow field flow fractionation (AF4) is most commonly utilized to study protein aggregation due to its ability to separate protein aggregates. As in SEC and SV-AUC, the separation of protein aggregates in AF4 is based on differences in hydrodynamic radii. Separation occurs in an asymmetrical thin channel into which the sample is injected and the molecular species are transported along by a laminarily flowing mobile phase (Fraunhofer and Winter 2004). A perpendicular cross-flow to the laminar flow directs the analytes towards a semipermeable membrane that is impermeable to analytes. The smaller particles as compared to the larger ones, diffuse readily back to the laminar flow and are eluted from the flow channel before the larger particles. The eluted particles are characterized using various detectors (den Engelsmann *et al*. 2011).

AF4 lacks a stationary phase and this mitigates problems caused by interactions of the stationary phase with aggregates. Also, shear-sensitive analytes can be analyzed (Fraunhofer and Winter 2004). The sample volumes needed are typically low (Zölls *et al*. 2012) and protein formulations can be applied directly without much sample preparation (Liu *et al*. 2006). Even though protein molecules' propensity to aggregate can be influenced due to interactions with the membrane, this can often be mitigated by reducing the cross-flow and by the choice of a low absorption membrane (Liu *et al*. 2006) or by the addition of surfactants (Fraunhofer and Winter 2004).

One of the main drawbacks of AF4 is the need for tedious method development to obtain good separations and results as the methodology is still evolving and not developed like other methods such as, SEC (den Engelsmann *et al*. 2011). Since the separation is dependent on multiple factors, the method development process needs extensive optimization (Zölls *et al*. 2012). Changes in concentrations during the analysis might also have an impact on the estimation of aggregates present (Mahler *et al*. 2009).

FFF methods are best suited for analysis of large aggregates (Hawe *et al*. 2012) that would otherwise be difficult to detect by AUC due to sedimentation and by SEC due to removal by filtration during sample preparation (Liu *et al*. 2006). A wide range of particle size (1 nm - 100 µm) can be analyzed by AF4 by tuning the rate of cross-flow (Fraunhofer and Winter 2004). The dynamic size range makes AF4 a robust analytical tool in detection and characterization of protein aggregates. A stressed IgG formulation containing a large proportion of submicron aggregates was used for AF4 method development that successfully detected submicron aggregates of another IgG and etanercept, a fusion protein used as an anti-tumor necrosis factor (anti-TNF) (Hawe *et al*. 2012).

4.3.5 Spectroscopy and other methods

Apart from the methods discussed above to study protein aggregation, several other methods exist and are being developed to address emerging challenges in studying protein aggregation. For example, early detection of antibody A and human calcitonin aggregates was reported with fluorescence microscopy after staining with Nile Red – a dye that binds to hydrophobic regions in proteins and protein aggregates, and also proposed as a method for early detection of changes in protein formulations as sample preparation is unnecessary (Demeule *et al*. 2007). Trastuzumab aggregates caused by the addition of 5% dextrose were detected and characterized with fluorescence spectroscopy and transmission electron microscopy (TEM) (Demeule *et al*. 2009). Application of mass spectrometry in protein aggregation studies has proved to be challenging due to instrument bias caused by ionization. However, electrospray ionization mass spectrometry (ESI MS); a method that utilizes a ‘soft’ ionization step, has been reported in characterizing aggregates of recombinant human antithrombin III (Wang *et al*. 2012). In addition, the use of commonly used methods such as sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), capillary electrophoresis-SDS, circular dichroism (CD) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy has also been reported for studies on protein aggregation (Mahler *et al*. 2009, den Engelsmann *et al*. 2011, Zölls *et al*. 2012).

4.4 Control of protein aggregation

Therapeutic proteins undergo several environmental and stress factors throughout their life cycle: expression in cell cultures to purification, formulation and *in vivo* administration (Manning *et al*. 2010). The most commonly encountered environmental factors are changes in protein concentration, pH, temperature, ionic strength as well as shear stresses, freeze drying related stress and exposure to different interfaces (Wang 2005, Cromwell *et al*. 2006). Environmental factors influence protein stability and proneness to aggregation; although a single processing step may not result in an aggregate *per se*, it can result in a precursor for aggregation that is more likely to aggregate in further processing steps. Apart from these factors, higher order aggregates can be formed even in the native conformational state of proteins (Chi *et al*. 2003).

Generally, aggregation of therapeutic proteins is viewed as a major challenge *in vitro* for production, formulation and storage. However, aggregation might also potentially take place *in vivo* after administration (Frokjaer and Otzen 2005, Mahler *et al.* 2010a). For instance, when administered subcutaneously, the molecules are initially localized to a relatively small area resulting in high concentrations. In combination with changes in environmental factors, this could potentially cause aggregation.

The following sections highlight the issue of protein aggregation at different stages in the life cycle of protein therapeutics and discuss means to effectively mitigate this challenge.

4.4.1 *In silico* assessment of protein aggregation

Computational methods, such as *in silico* molecular simulation (Bratko *et al*. 2007) are being used and developed to predict the propensity for aggregation of a given protein sequence. Models and algorithms derived from computational methods identify sequences and motifs that are prone to aggregation (Tartaglia *et al.* 2008, Agrawal *et al.* 2011, Roughton *et al.* 2013). Even though the computational models are simplistic (Bratko *et al*. 2007, Cellmer *et al.* 2007) and cannot predict *de novo* aggregation of a given protein, they are useful in understanding mechanisms, precursors and inducers of aggregation. Recent efforts are being directed to utilize computational tools at an early stage of drug development to identify promising protein molecules (Agrawal *et al*. 2011). Proteins with desirable conformational stability and aggregation propensity can be screened relatively rapidly using *in silico* methods (Agrawal *et al*. 2011). Such an approach has the potential to reduce the number of experiments and aid in the design of formulation screening with preferable excipients (Cellmer *et al*. 2007, Roughton *et al.* 2013).

4.4.2 Protein expression

Prokaryotic and eukaryotic expression systems are utilized to produce recombinant therapeutic proteins. Eukaryotic expression systems are mostly suited for producing therapeutic proteins of interest since they offer post-translational modifications. However the bacterial expression system remains an attractive option due to rapid growth, easy culture conditions and overall cost-effectiveness (Sørensen and Mortensen 2005).

Bacterial expression systems, such as *E. coli*, are however, limited by their inability to carry out post-translational modifications or disulphide bond formation. This tends to hinder proper folding of the polypeptide chain and results in the formation of non-functional proteinaceous granules (inclusion bodies) inside the bacterial cytoplasm (Villaverde and Carrió 2003, Singh and Panda 2005). Recovery of the recombinant protein in its native conformational state from inclusion bodies requires solubilization under denaturing conditions and refolding (Sørensen and Mortensen 2005).

To achieve soluble expression of recombinant proteins, several approaches have shown promise. Preconditioning of cells with a complex growth media (Hoffmann *et al*. 2004); tuning of expression conditions such as, temperature (Vasina & Baneyx 1997, Hammarström *et al.* 2002, Vera *et al.* 2007), induction conditions (Bentley and Kompala 1990, Galloway *et al.* 2003, Studier 2005, Tu *et al.* 2009, San-Miguel *et al.* 2013), expression rate (Galloway *et al.* 2003, Vera *et al.* 2007); co-expression of molecular chaperones – proteins that aid folding – (de Marco *et al.* 2004, Kim *et al.* 2005, de Marco *et al.* 2007) and fusion with soluble protein / peptide tags (Lichty *et al.* 2005, Waugh 2005, Esposito and Chatterjee 2006, Bird 2011). Refolding from inclusion bodies can be achieved by solubilizing them in a denaturant (GdHCl) and subsequent gradual removal of chaotropic reagents (Singh and Panda 2005). Arginine is known to assist refolding of recombinant proteins from inclusion bodies (Arakawa and Tsumoto 2003). Immobilized molecular chaperones *in vitro* have been successful in refolding from inclusion bodies (Rudolph and Lilie 1996, Altamirano *et al.* 1997, Gao *et al.* 2003, Jhamb *et al*. 2008). Engineered strains of *E. coli* have also been utilized to express soluble proteins (Miroux and Walker 1996, Sørensen and Mortensen 2005, Berrow *et al.* 2006).

Soluble expression of proteins is an extensive field of research and a detailed discussion on this topic is beyond the scope of this review. Several excellent treatises (Jonasson *et al.* 2002, Sørensen and Mortensen 2005, Peti and Page 2007, Sahdev *et al.* 2008) exist on this subject for the interested reader.

4.4.3 Purification

A protein therapeutic must be absolutely pure and free from impurities for both efficacy and safety concerns. After the expression of the recombinant protein, the protein is purified by chromatographic methods (Cromwell *et al.* 2006). Such a process generally involves several steps, each utilizing differences in physicochemical properties of the protein of interest from others.

Although the aim of purification is to remove impurities such as, viruses, other host organism proteins and protein aggregates from the protein of interest, protein aggregation can occur during the process of purification (Wang *et al*. 2005). Varying conditions of pH, concentration and ionic strength could potentially induce aggregation. For instance, during affinity chromatography with protein A, conditions for elution at low pH might induce aggregation. This was indeed observed in the case of IgM (Phillips *et al*. 2001) and Fc fusion protein (Shukla *et al.* 2007). However, interactions between the protein and protein A might also be a cause of aggregation.

Optimization of the purification methodology with different buffer systems and process parameters can help minimize protein aggregation during purification (Cromwell *et al*. 2006). A decrease in temperature diminished aggregation of a Fc fusion protein during protein A affinity chromatography (Shukla *et al*. 2007). Additives such as arginine, when included in the elution buffer, have shown to enhance recovery of mAbs with a lower percentage of aggregates during protein A affinity chromatography; a similar effect has also been reported for arginine derivates acetyl-arginine and agmatine (Ejima *et al*. 2005) .

4.4.5 Filtration

Filtration of protein solutions are carried out to remove particulate material and pathogenic microbes (Cromwell *et al.* 2006). It is also standard laboratory technique to concentrate and buffer exchange protein solutions (Liu *et al*. 2012).

Although the process of filtration is intended to remove foreign particles that very often act as nuclei for aggregation, particles shed from the filters and membranes of filtration units may induce aggregation (Liu *et al*. 2012). This is a serious matter of concern since protein solutions are routinely filtered during various processes in the lifetime of a protein therapeutic. Local concentration of proteins is also considered to be higher at the membrane surface than in the bulk solution; a condition that might result in aggregation (Cromwell *et al*. 2006). To mitigate such issues, additives such as surfactants are added to prevent adsorption to the filter surfaces (Mahler *et al*. 2010b). The choice of a suitable filter is also a necessity; the choice is influenced by the properties of protein of interest.

4.4.6 Shaking, shearing and agitation

Liquid formulations of protein therapeutics are an obvious choice for ease of application; however, the solution state of proteins expose them to air-water interfaces and container surfaces. The air-water interface is considered to be hydrophobic where proteins potentially might accumulate, denature and expose buried hydrophobic regions (Gidalevitz *et al*. 1999). This could lead to aggregation during shipping, handling and storage (Treuheit *et al*. 2002). Agitation caused by shaking and/or shearing increases the area of such an interface further aggregation. For example, recombinant murine growth hormone was observed to aggregate on agitation (Fradkin *et al*. 2011).

Addition of surfactants is a common strategy to counter problems caused by air-water interfaces (Chi *et al*. 2003). Surfactants are amphiphilic; they possess both hydrophilic and lipophilic properties. They align themselves on water interfaces in a manner that minimizes the contact of the hydrophobic regions with water. Two key mechanisms have been proposed for surfactant action. In the first mechanism, direct binding of the surfactant to protein provides steric hindrance for association of polypeptide chains. Surfactants might also bind to folding intermediates and prevent incorrect folding or prevent association of the partially unfolded states (Bam *et al.* 1998, Kreilgaard *et al.* 1998, Chi *et al.* 2003). In the second mechanism, surfactants compete with proteins for adsorption onto interfaces and thereby inhibit proteins from partitioning onto those interfaces. For instance, propylene glycol and polysorbate 20 increased the stability of agitated ciliary neurotrophic factor (CNTF) (Arakawa *et al.* 2003); Tween 20 protected recombinant human growth hormone (hGH) from agitation-induced aggregation (Bam *et al.* 1998). Aggregation of agitated polyethylene glycosylated megakaryocyte growth and development factor (PEG-MGDF), polyethylene glycosylated granulocyte colony stimulating factor (PEG-GCSF) and osteoprotegerin fused with the Fc portion of an immunoglobulin (OPG-Fc) was also inhibited by the addition of polysorbate 20 (Treuheit *et al*. 2002). However, at certain concentrations, polysorbate 20 has shown to increase aggregation of IgG1 antibody (Kiese *et al*. 2008). Furthermore, optimization of container head-space also seems to diminish aggregation on agitation (Kiese *et al.* 2008) and has been observed to prevent the formation of visible particles of an IgG A antibody in shaken vials (Brych *et al.* 2010).

4.4.7 Drying

Therapeutic proteins are preferably formulated as aqueous solutions. These liquid formulations must be maintained at low temperatures to ensure stability and thus long-term storage. However, in some instances, formulation of an aqueous protein solution might not be feasible or an attractive option due to reasons such as, lack of proper cold chain during shipping and storage. In such situations, proteins are dried to improve their stability (Manning *et al*. 2010). Protein solutions can be dried using methods such as, freeze-drying (lyophilization), spray-drying and vacuum drying (Abdul-Fattah *et al*. 2007, Manning *et al*. 2010).

All the methods of drying remove water from protein solution. During this process, the protein can be exposed to different stress factors. The water content of a dried protein product is often less than 10 % compared to a fully hydrated protein and due to the heavy loss in the native hydration layer, the protein structure might undergo sufficient perturbation and increased proneness to aggregation (Wang 2005, Hamada *et al*. 2009). Changes in temperature commonly take place during drying processes and cause thermal stress on the proteins (Manning *et al*. 2010). Freeze-drying is known to cause aggregation in proteins such as, IgGs and hGH (Sarciaux *et al*. 1999, Salnikova *et al*. 2008). Spray-drying is also known to cause aggregation in IgGs and mAbs (Maury *et al*. 2005, Andya *et al*. 1999).

Excipients are occasionally used to stabilize proteins in the dried state. It is believed that the excipients function as ‘water substitutes’; they are thought to mimick the water structure by potentially forming hydrogen bonds with the protein. This probably helps preserve the secondary structure and inhibit protein unfolding and aggregation (Maury *et al*. 2005, Abdul-Fattah *et al*. 2007, Salnikova *et al*. 2008, Manning *et al*. 2010). Compounds commonly utilized to stabilize proteins during drying processes include amino acids (arginine and proline), saccharides (sucrose), sorbitol, trehalose and surfactants (Andya *et al.* 1999, Chi *et al.* 2003, Webb *et al.* 2003, Maury *et al.* 2005, Abdul-Fattah *et al.* 2007, Salnikova *et al.* 2008). Process modifications, such as addition of a heat treatment step – annealing – to the lyophilization cycle of a human interferon-γ suppressed aggregation (Webb *et al*. 2003).

4.4.8 Freezing and thawing

Therapeutic proteins are usually stored at low temperatures to inhibit degradation and extend the shelf-life. Freezing can be a processing step although exposure to freezing temperatures can also occur accidentally, for instance during shipping (Kreilgaard *et al*. 1998). This often presents a problem as freezing and thawing are complex processes that can induce multiple stresses. Cold denaturation (spontaneous unfolding of protein due to cold temperature), increased protein and solute concentrations, changes in pH and ice-induced denaturation are some of the key reasons for freeze thaw induced aggregation in proteins (Bhatnagar *et al*. 2007, Kueltzo *et al*. 2008, Manning *et al*. 2010). For example, insoluble aggregates have been observed after freeze-thawing CNTF (Chang *et al*. 1996). Freeze-thawing resulted in the formation of subvisible aggregates of monoclonal IgG_1 ranging from 1 µm to 25 µm along with smaller aggregates (Hawe *et al*. 2009). Aggregation propensity of a model IgG_2 increased with freeze-thawing (Kueltzo *et al*. 2008).

Excipients have been used to control the rate of freezing in an attempt to minimize freezing related aggregation (Wang 2005, Cao *et al*. 2003, Kueltzo *et al*. 2008, Bee *et al.* 2011). Diverse excipients such as, sugars, polyols, inorganic salts and amino acids are used to mitigate aggregation from the freezing process (Carpenter and Crowe 1998, Hamada *et al*. 2009). Surfactants are also commonly used (Chang *et* al. 1996, Kreilgaard *et* al. 1998, Chi *et al.* 2003). In some cases, excipients might be rendered ineffective at low temperatures. Long-term storage at -30° C resulted in the crystallization of sorbitol that adversely affected its protective capability and resulted in aggregation of proteins (Piedmonte *et al*. 2007). Similar findings have also been reported for trehalose (Singh *et al*. 2011).

4.5 Protein aggregates, immunogenicity and regulatory aspects

Protein aggregates are believed to present novel epitopes (multiple-epitope character) that is recognized by the immune system as non-endogenous. The non-self molecular epitopes presented by the aggregates are commonly encountered in pathogens and thus evoke strong immune responses (Hermeling *et al*. 2004, Cordoba-Rodriguez 2008, Sauerborn *et al.* 2010). Of course, not all protein aggregates are immunogenic. Factors that influence the immune response include origin of the protein aggregates (endogenous vs. exogenous), presence of contaminants and the immunogenic state of the host (Rosenberg 2006). The intensity of immune responses is also dependent on the level of tolerance towards the aggregates and/or their endogenous counterparts.

The immune system reacts to aggregates by forming antibodies targeted against them. This results in a loss of therapeutic efficacy, altered pharmacokinetics that necessitates a new dosing regimen or in the neutralization of endogenous proteins (Hermeling *et al*. 2004, Schellekens 2005, Rosenberg 2006, Carpenter *et al.* 2009). The direst reactions involve hypersensitivity and particularly in the case of endogenous proteins serving essential biological functions being neutralized, the consequences might be fatal.

It is currently not well understood on the nature and sizes of aggregates that are responsible for triggering immune responses (Rosenberg 2006, Mahler *et al*. 2009, Sauerborn *et al.* 2010). It is however believed that the immune system recognizes the aggregates more readily than the native parent proteins (den Engelsman *et al*. 2011) and that molecular size and solubility of the aggregates are key triggering factors (Rosenberg 2006). The potential for inducing immune responses is usually higher for protein aggregates that exceed 100 kDa in molecular weight; size is not the only determining factor as monomers of a high molecular weight protein are not necessarily more immunogenic than their smaller counterparts. Thus, the multimeric character of protein aggregates seems to play a significant role in recognition and stimulation of the immune response. It also seems that insoluble aggregates may be more potent in triggering immune responses. As our current understanding of immune responses mediated by protein aggregates is still rather poor, it is very important to screen for the presence of protein aggregates and find means to prevent the formation of aggregates in protein biopharmaceuticals. However, for patients who are immunocompromised, presence of visible aggregates might be of less concern and a lesser strict criterion of acceptance applicable (Carpenter *et al.* 2009, den Engelsman *et al.* 2011).

Regulatory guidelines exist on acceptable limits and preferred detection methods for visible and subvisible particles in injectable biopharmaceuticals for the US and European markets[1]. As these regulations are mostly quantitative, no specific regulations for protein aggregates in particular exist for the moment (Cordoba-Rodriguez 2008). The regulatory guidelines are constantly being updated to incorporate new information and data as they emerge from investigations. For example, in its present form, the guidelines do not adequately cover the full size-range of protein aggregates. As a consequence, there is a possibility that smaller aggregates are routinely ignored. Manufacturers are aware of this fact and on their part make efforts to develop and utilize complementary methods to detect and characterize most protein aggregates (Cordoba-Rodriguez 2008, Carpenter *et al.* 2009).

[1] United States Pharmacopoeia <788>, European Pharmacopoeia 2.9.19 and 2.9.20

5 CONCLUSIONS

In this review, we have highlighted the challenges associated with protein aggregation towards the development of protein therapeutics. The phenomenon of aggregation is complex and currently not well understood. Apart from loss of efficacy, protein aggregates are potential risk factors of fatality due to immunogenic responses. The safety concerns related to the use of protein products has lead to intense research and setting up of regulatory guidelines for detection and characterization of aggregates. The size and nature of protein aggregates in relation to the immunogenic response they elicit is a matter of intense research.

Faced with the reality that protein aggregates are unavoidable, their presence and nature needs to be detected and characterized. No single analytical tool offers aggregate detection of all size ranges and every method has its set of limitations. Generally, it is appreciated that different methods are used orthogonally and results from one is compared and validated against another. Some methods are more applicable for rapid detection whereas others more suited for detailed characterization.

The field of research on protein aggregation is evolving and new methods are being developed to effectively characterize aggregation. Such efforts will not only help in addressing safety concerns but also in understanding mechanisms of aggregate formation and formulation of inhibition strategies.

6 REFERENCES

Abdul-Fattah AM, Truong-Le V, Yee L, Nguyen L, Kalonia DS, Cicerone MT and Pikal MJ: Drying-induced variations in physico-chemical properties of amorphous pharmaceuticals and their impact on stability (I): Stability of a monoclonal antibody. J Pharm Sci 96: 1983-2008 (2007).

Agrawal NJ, Kumar S, Wang X, Helk B, Singh SK and Trout BL: Aggregation in protein-based biotherapeutics: computational studies and tools to identify aggregation-prone regions. J Pharm Sci 100: 5081-5095 (2011).

Aguzzi A and Calella AM: Prions: Protein aggregation and infectious diseases. Physiol Rev 89: 1105-1152 (2009).

Ahrer K, Buchacher A, Iberer G, Josic D and Jungbauer A: Analysis of aggregates of human immunoglobulin G using size-exclusion chromatography, static and dynamic light scattering. J Chromatogr A 1009: 89-96 (2003).

Altamirano MM, Golbik R, Zahn R, Buckle AM and Fersht AR: Refolding chromatography with immobilized mini-chaperones. Proc Natl Acad Sci 94: 3576-3578 (1997).

Andrews JM, Weiss IV WF and Roberts CJ: Nucleation, growth, and activation energies for seeded and unseeded aggregation of α-chymotrypsinogen A. Biochem 47: 2397-2403 (2008).

Andya JD, Maa Y-F, Constantino HR, Nguyen P-A, Dasovich N, Sweeney TD, Hsu CC and Shire SJ: The effect of formulation excipients on protein stability and aerosol performance of spray-dried powders of a recombinant humanized anti-IgE monoclonal antibody. Pharm Res 16: 350-358 (1999).

Arakawa T, Dix DB and Chang BS: The effects of protein stabilizers on aggregation induced by multiple-stresses. Yakugaku Zasshi 123: 957-961 (2003).

Arakawa T and Tsumoto K: The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. Biochem Biophys Res Co 304: 148-152 (2003).

Arthur KK, Gabrielson JP, Kendrick BS and Stoner MR: Detection of protein aggregates by sedimentation velocity analytical ultracentrifugation (SV-AUC): Sources of variability and their relative importance. J Pharm Sci 98: 3522-3539 (2009).

Bam NB, Cleland JL, Yang J, Manning MC, Carpenter JF, Kelley RF and Raldoph TW: Tween protects recombinant human growth hormone against agitation-induced damage via hydrophobic interactions. J Pharm Sci 87: 1554-1559 (1998).

Barral JM, Broadley SA, Schaffar G and Hartl FU: Roles of molecular chaperones in protein misfolding diseases. Semin Cell Dev Biol 15: 17-29 (2004).

Bee JS, Davis M, Freund E, Carpenter JF and Randolph TW: Aggregation of a monoclonal antibody induced by adsorption to stainless steel. Biotechnol Bioeng 105: 121-129 (2010).

Bee JS, Randolph TW, Carpenter JF, Bishop SM and Dimitrova MN: Effects of surfaces and leachables on the stability of biopharmaceuticals. J Pharm Sci 100: 4158-4170 (2011).

Bentley WE and Kompala DS: Optimal induction of protein synthesis in recombinant bacterial cultures. Ann N Y Acad Sci 589: 121-138 (1990).

Berkowitz SA: Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. The AAPS Journal 8: E590-E605 (2006).

Berrow NS, Büssow K, Coutard B, Diprose J, Ekberg M, Folkers GE, Levy N, Lieu V, Owens RJ, Peleg Y, Pinaglia C, Quevillon-Cheruel S, Salim L, Scheich C, Vincentelli R and Busso D: Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. Acta Crystallogr D 62: 1218-1226 (2006).

Bhatnagar BS, Bogner RH and Pikal MJ: Protein stability during freezing: Separation of stresses and mechanisms of protein stabilization. Pharm Dev Technol 12: 505-523 (2007).

Bird LE: High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. Methods 55: 29-37 (2011).

Branden C and Tooze J: Introduction to protein structure. 2nd edition.Garland Publishing, Inc., New York 1998.

Bratko D, Cellmer T, Prausnitz JM and Blanch HW: Molecular simulation of protein aggregation. Biotechnol Bioeng 96: 1-8 (2007).

Bryant C, Spencer DB, Miller A, Bakaysa DL, McCune KS, Maple SR, Pekar AH and Brems DN: Acid stabilization of insulin. Biochem 32: 8075-8082 (1993).

Brych SR, Gokarn YR, Hultgren H, Stevenson RJ, Rajan R and Matsumura M: Characterization of antibody aggregation: Role of buried, unpaired cysteines in particle formation. J Pharm Sci 99: 764-781 (2010).

Cao E, Chen Y, Cui Z and Foster PR: Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. Biotechnol Bioeng 82: 684-690 (2003).

Carpenter JF and Crowe JH: The mechanism of cryoprotection of proteins by solutes. Cryobiology 25: 244-255 (1988).

Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJA, Middaugh CR, Winter G, Fan Y-X, Kirshner S, Verthelyi D, Kozlowski S, Clouse KA, Swann PG, Rosenberg A and Cherney B: Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. J Pharm Sci 98: 1201-1205 (2009).

Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJ, Middaugh CR and Winter G: Potential inaccurate quantitation and sizing of protein aggregates by size exclusion chromatography: Essential need to use orthogonal methods to assure the quality of therapeutic protein products. J Pharm Sci 99: 2200-2208 (2010).

Cellmer T, Bratko D, Prausnitz JM and Blanch HW: Protein aggregation *in silico*. Trends Biotechnol 25: 254-261 (2007).

Chang BS, Kendrick BS and Carpenter JF: Surface-induced denaturation of proteins during freezing and its inhibition by surfactants. J Pharm Sci 85: 1325-1330 (1996).

Chi EY, Weickmann J, Carpenter JF, Manning MC and Randolph TW: Heterogenous nucleation-controlled particulate formation of recombinant Human platelet-activating factor acetylhydrolase in pharmaceutical formulation. J Pharm Sci 94: 256-274 (2005).

Chi EY, Krishnan S, Randolph TW and Carpenter JF: Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. Pharm Res 9: 1325-1336 (2003).

Chiti F, Bucciantini M, Capanni C, Taddei N, Dobson CM and Stefani M: Solution conditions can promote formation of either amyloid protofilaments or mature fibrils from the HypF N-terminal domain. Protein Sci 10: 2541-2574 (2001).

Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G and Dobson CM: Kinetic partitioning of protein folding and aggregation. Nat Struct Biol 9: 137-143 (2002).

Cordoba-Rodriguez RV: Aggregates in MAbs and recombinant therapeutic proteins: A regulatory perspective. Biopharm Int 21: 44-53 (2008).

Crasto W, Jarvis J, Khunti K and Davies MJ: New insulins and insulin regimens: a review of their role in improving glycaemic control in patients with diabetes. Postgrad Med J 85: 257-267 (2009).

Crommelin DJA, Storm G, Verrijk R, de Leebe L, Jiskoot W and Hennink WE: Shifting paradigms: biopharmaceuticals versus low molecular weight drugs. Int J Pharm 266: 3-16 (2003).

Cromwell EM, Hilario E and Jacobson F: Protein aggregation and bioprocessing. The AAPS Journal 8: E572-E579 (2006).

Danne T and Bolinder J: New insulins and insulin therapy. Int J Clin Pract 66: 30-34 (2012).

Demeule B, Gurny R and Arvinte T: Detection and characterization of protein aggregates by fluorescence microscopy. Int J Pharm 329: 37-45 (2007).

Demeule B, Palais C, Machaidze, Gurny R and Arvinte T: New methods allowing the detection of protein aggregates. mAbs 1: 142-150 (2009).

Dobson CM: Principles of protein folding, misfolding and aggregation. Semin Cell Dev Biol 15: 3-16 (2004).

Dobson G and Steiner D. The role of assembly in insulin's biosynthesis. Curr Opin Struc Biol 8: 189-194 (1998).

Dyson HJ: Expanding the proteasome: disordered and alternatively folded proteins. Q Rev Biophys 44: 467-518 (2011).

den Engelsman J, Garidel P, Smulders R, Koll H, Smith B, Bassarab S, Seidl A, Hainzl O and Jiskoot W: Strategies for the assessment of protein aggregates in pharmaceutical biotech product development. Pharm Res 28: 920-923 (2011).

Ejima D, Yumioka R, Tsumoto K and Arakawa T: Effective elution of antibodies by arginine and arginine derivatives in affinity column chromatography. Anal Biochem 345: 250-257 (2005).

Esposito D and Chatterjee DK: Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotech 17: 353-358 (2006).

Filipe V, Kükrer B, Hawe A and Jiskoot W: Transient molten globules and metastable aggregates induced by brief exposure of a monoclonal IgG to low pH. J Pharm Sci 101: 2327-2339 (2012).

Fink AL: Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des 3: R9-R23 (1998).

Fradkin AH, Carpenter JF and Randolph TW: Glass particles as an adjuvant: A model for adverse immunogenicity of therapeutic proteins. J Pharm Sci 100: 4953-4964 (2011).

Fraunhofer W and Winter G: The use of asymmetrical flow field-flow fractionation in pharmaceutics and biopharmaceutics. Eur J Pharm Biopharm 58: 369-383 (2004).

Frokjaer S and Otzen DE: Protein drug stability: A formulation challenge. Nature Rev 4: 298-306 (2005).

Gabrielson JP, Brader ML, Pekar AH, Mathis KB, Winter G, Carpenter JF and Randolph TW: Quantitation of aggregate levels in a recombinant humanized monoclonal antibody formulation by size-exclusion chromatography, asymmetrical flow field flow fractionation, and sedimentation velocity. J Pharm Sci 96: 268-279 (2007).

Gabrielson JP, Arthur KK, Kendrick BS, Randolph TW and Stoner MR: Common excipients impair detection of protein aggregates during sedimentation velocity analytical ultracentrifugation. J Pharm Sci 98: 50-62 (2009).

Galloway CA, Sowden MP and Smith HC: Increasing the yield of soluble recombinant protein expressed in *E. coli* by induction during late log phase. Biotechniques 34: 524-530 (2003).

Gao YG, Guan YX, Yao SJ and Cho MG: On-column refolding of recombinant human interferon-gamma with an immobilized chaperone fragment. Biotechnol Prog 19: 915-920 (2003).

Gidalevitz D, Huang Z and Rice SA: Protein folding at the air-water interface studied with x-ray reflectivity. Proc Natl Acad Sci USA 96: 2608-2611 (1999).

Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K and Riggs AD: Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Biochem 76: 106-110 (1979).

Grillo AO, Edwards K-LT, Kashi RS, Shipley KM, Hu L, Besman MJ and Middaugh CR: Conformational origin of the aggregation of recombinant human factor VIII. Biochem 40: 586-595 (2001).

Gun'ko VM, Klyueva AV, Levchuk YN and Leboda R: Photon correlation spectroscopy investigations of proteins. Adv Colloid Interfac 105: 201-328 (2003).

Hamada H, Arakawa T and Shiraki K: Effect of additives on protein aggregation. Curr Pharm Biotechnol 10: 400-407 (2009).

Hammarström M, Hellgren N, van den Berg S, Berglund H and Härd T: Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. Protein Sci 11: 313-321 (2002).

Hawe A, Kasper JC, Friess W and Jiskoot W: Structural properties of monoclonal antibody aggregates induced by freeze-thawing and thermal stress. Eur J Pharm Sci 38: 79-87 (2009).

Hawe A, Romeijn S, Filipe Vasco and Jiskoot W: Asymmetrical flow field-flow fractionation method for the analysis of submicron protein aggregates. J Pharm Sci 101: 4129-4139 (2012).

Hermeling S, Crommelin DJA, Schellekens H and Jiskoot W: Structure-immunogenicity relationships of therapeutic proteins. Pharm Res 21: 897-903 (2004).

Hoffmann F, van den Heuvel J, Zidek N and Rinas U: Minimizing inclusion body formation during recombinant protein production in *Escherichia coli* at bench and pilot plant scale. Enzyme Microb Tech 34: 235-241 (2004).

Jachimska B, Wasilewska M and Adamczyk Z: Characterization of globular protein solutions by dynamic light scattering, electrophoretic mobility, and viscosity measurements. Langmuir 24: 6866-6872 (2008).

Jhamb K, Jawed A and Sahoo DK: Immobilized chaperones: A productive alternative to refolding of bacterial inclusion body proteins. Process Biochem 43: 587-597 (2008).

Jonasson P, Liljeqvist S, Nygren P-Å and Ståhl S: Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli.* Biotechnol Appl Biochem 35: 91-105 (2002).

Jones LS, Kaufmann A and Middaugh CR: Silicone oil induced aggregation of proteins. J Pharm Sci 94: 918-927 (2005).

Kfoury N, Holmes BB, Jiang H, Holtzman DM and Diamond MI: Trans-cellular propagation of Tau aggregation by fibrillar species. J Biol Chem 287: 19440-19451 (2012).

Khossravi M, Shire SJ and Borchardt RT: Evidence for the involvement of histidine A(12) in the aggregation and precipitation of human relaxin induced by metal-catalyzed oxidation. Biochem 39: 5876-5885 (2000).

Kiese S, Pappenberger A, Friess W and Mahler H-C: Shaken, not stirred: Mechanical stress testing of an IgG1 antibody. J Pharm Sci 97: 4347-4366 (2008).

Kiese S, Pappenberger A, Friess W and Mahler H-C: Equilibrium studies of protein aggregates and homogenous nucleation in protein formulation. J Pharm Sci 99: 632-344 (2010).

Kim S-G, Kweon D-H, Lee D-H, Park Y-C and Seo J-H. Coexpression of folding accessory proteins for production of active cyclodextrin glycosylt ransferase of *Bacillus macerans* in recombinant *Escherichia coli*. Prot Express Purif 41: 426-432 (2005).

Kling J: Fresh from the biotech pipeline-2013. Nat Biotech 32: 121-124 (2014).

Kreilgaard L, Jones LS, Randolph TW, Frokjaer S, Flink JM, Manning MC and Carpenter JF: Effect of Tween 20 on freeze-thawing- and agitation-induced aggregation of recombinant human factor XIII. J Pharm Sci 87: 1597-1603 (1998).

Kueltzo LA, Wang W, Randolph TW and Carpenter JF: Effects of solution conditions, processing parameters, and container materials on aggregation of a monoclonal antibody during freeze-thawing. J Pharm Sci 97: 1801-1812 (2008).

Lam XM, Yang JY and Cleland JL: Antioxidants for prevention of methionine oxidation in recombinant monoclonal antibody HER2. J Pharm Sci 86: 1250-1255 (1997).

Laurence JS and Middaugh CR: Fundamental structures and behaviors of proteins. In: Aggregation of therapeutic proteins, p. 1-61, 1st edition. Edited by Wang W and Roberts CJ, John Wiley & Sons Ltd, New Jersey 2010.

Leader B, Baca QJ and Golan DE: Protein therapeutics: a summary and pharmacological classification. Nature Rev Drug Discovery 7: 21-39 (2008).

Levy Y and Onuchic JN: Water mediation in protein folding and molecular recognition. Annu Rev Biophys Biomol Struct 35: 389-415 (2006).

Li Y, Lubchenko V and Vekilov PG: The use of dynamic light scattering and Brownian microscopy to characterize protein aggregation. Rev Sci Instrum 82: 053106-1 – 053106-8 (2011).

Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ and Tan S: Comparison of affinity tags for protein purification. Prot Purif Express 41: 98-105 (2005).

Liu J, Andya JD and Shire SJ: A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. The AAPS Journal 8: E580-E589 (2006).

Liu L, Randolph TW and Carpenter JF: Particles shed from syringe filters and their effects on agitation-induced protein aggregation. J Pharm Sci 101: 2952-2959 (2012).

Mahler H-C, Friess W, Grauschopf U and Kiese S: Protein aggregation: pathways, induction factors and analysis. J Pharm Sci 98: 2909-2934 (2009).

Mahler H-C, Fischer S, Randolph TW and Carpenter JF. Protein aggregation and particle formation: Effects of formulation, interfaces, and drug product manufacturing operations. In: Aggregation of therapeutic proteins, p. 301-331, 1st edition. Edited by Wang W and Roberts CJ, John Wiley & Sons Ltd, New Jersey 2010.

Mahler H-C, Huber F, Kishore RSK, Reindl J, Rückert P and Müller R: Adsorption behavior of a surfactant and a monoclonal antibody to sterilizing-grade filters. J Pharm Sci 99: 2620-2627 (2010).

Majumdar S, Ford BM, Mar KD, Sullivan VJ, Ulrich RG and D'souza AJM: Evaluation of the effect of syringe surfaces on protein formulations. J Pharm Sci 100: 2563-2573 (2011).

Malencik DA and Anderson SR: Dityrosine as a product of oxidative stress and fluorescent probe. Amino Acids 25: 233-247 (2003).

Manning MC, Chou DK, Murphy BM, Payne RW and Katayama DS: Stability of protein pharmaceuticals: an update. Pharm Res 27: 544-575 (2010).

de Marco A and de Marco V: Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. J Biotechnol 109: 45-52 (2004).

de Marco A, Deuerling E, Mogk A, Tomoyasu T and Bukau B: Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. BMC Biotechnol 7: 32-40 (2007).

Maury M, Murphy K, Kumar S, Mauerer A and Lee G: Spray-drying of proteins: effects of sorbitol and trehalose on aggregation and FT-IR amide I spectrum of an immunoglobulin G. Eur J Pharm Biopharm 59: 251-261 (2005).

Miroux B and Walker JE: Over-production of proteins in *Escherichia coli*: mutants hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol 260: 289-298 (1996).

Narhi LO, Schmit J, Bechtold-Peters K and Sharma D: Classification of protein aggregates. J Pharm Sci 101: 493-498 (2012).

Nath S, Goodwin J, Engelbohrs Y and Pountney DL: Raised calcium promotes α-synuclein aggregate formation. Mol Cell Neurosci 46: 516-526 (2011).

Pawar AP, DuBay KF, Zurdo J, Chiti F, Vendruscolo M and Dobson CM: Prediction of "Aggregation-prone" and "Aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol 350: 379-392 (2005).

Pekar A and Sukumar M: Quantitation of aggregates in therapeutic proteins using sedimentation velocity analytical ultracentrifugation: Practical considerations that affect precision and accuracy. Anal Biochem 367: 225-237 (2007).

Peti W and Page R: Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. Protein Expres Purif 51: 1-10 (2007).

Phillips J, Drumm A, Harrison P, Bird P, Bhamra K, Berrie E and Hale G: Manufacture and quality control of CAMPATH-1 antibodies for clinical trials. Cytotherapy 3: 233-242 (2001).

Philo JS and Arakawa T: Mechanisms of protein aggregation. Curr Pharm Biotechn 10: 248-251 (2009).

Philo JS: Characterizing the aggregation and conformation of protein therapeutics. Am Biotechnol Lab 22-26 (2003).

Philo JS: Is any measurement method optional for all aggregates sizes and types? The AAPS Journal 8: E564-E571 (2006).

Philo JS: A critical review of methods for size characterization of non-particulate protein aggregates. Curr Pharm Biotechn 10: 359-372 (2009).

Piedmonte DM, Summers C, McAuley A, Karamujic L and Ratnaswamy G: Sorbitol crystallization can lead to protein aggregation in frozen protein formulations. Pharm Res 24: 136-146 (2007).

Rader RA: FDA Biopharmaceutical product approvals and trends in 2012. Up from 2011, but innovation and impact are limited. Bioprocess Int 11: 18-27 (2013).

Richardson JS: The anatomy and taxonomy of protein structure. Adv Prot Chem 34: 167-339 (1981).

Rosenberg AS: Effects of protein aggregates: An immunologic perspective. The AAPS Journal 8: E501-E507 (2006).

Roughton BC, Iyer LK, Bertelsen E, Topp EM and Camarda KV: Protein aggregation and lyophilization: Protein structural descriptors as predictors of aggregation propensity. Comput Chem Eng 58: 369-377 (2013).

Rudolph R and Lilie H: *In vitro* folding of inclusion body proteins. FASEB J 10: 49-56 (1996).

Sahdev S, Khattar SK and Saini KS: Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem 307: 249-264 (2008).

Salnikova MS, Middaugh CR and Rytting JH: Stability of lyophilized human growth hormone. Int J Pharm 358: 108-113 (2008).

Saluja A and Kalonia DS: Nature and consequences of protein-protein interactions in high protein concentration solutions. Int J Pharm 358: 1-15 (2008).

San-Miguel T, Pérez-Bermúdez P and Gavidia I: Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. SpringerPlus 2: 89-92 (2013).

Sarciaux J-M, Mansour S, Hageman MJ and Nail SL: Effects of buffer composition and processing conditions on aggregation of bovine IgG during freeze-drying. J Pharm Sci 88: 1354-1361 (1999).

Sauerborn M, Brinks V, Jiskoot W and Schellekens H: Immunological mechanism underlying the immune response to recombinant human protein therapeutics. Trends Pharmacol Sci 31: 53-59 (2010).

Schellekens H: Factors influencing the immunogenicity of therapeutic proteins. Nephrol Dial Transplant 20: vi3-vi9 (2005).

Sharma B: Immunogenicity of therapeutic proteins. Part 2: Impact of container closures. Biotechnol Adv 25: 318-324 (2007).

Sheridan C: Fresh from the biologic pipeline-2009. Nat Biotech 28: 307-310 (2010).

Shiraki K, Kudou M, Fujiwara S, Imanaka T and Takagi M: Biophysical effect of amino acids on the prevention of protein aggregation. J Biochem 132: 591-595 (2002).

Shire SJ, Shahrokh Z and Liu J: Challenges in the development of high protein concentration formulations. J Pharm Sci 93: 1390-1402 (2004).

Shukla AA, Gupta P and Han X: Protein aggregation kinetics during Protein A chromatography. Case study for an Fc fusion protein. J Chromatogr A 1171: 22-28 (2007).

Singh SK, Afonina N, Awwad M, Bechtold-Peters K, Blue JT, Chou D, Cromwell M, Krause H-J, Mahler H-C, Meyer BK, Narhi L, Nesta DP and Spitznagel T: An industry perspective on the monitoring of subvisible particles as a quality attribute for protein therapeutics. J Pharm Sci 99: 3302-3321 (2010).

Singh SK, Kolhe P, Mehta AP, Chico SC, Lary AL and Huang M: Frozen state storage instability of a monoclonal antibody: Aggregation as a consequence of trehalose crystallization and protein unfolding. Pharm Res 28: 873-885 (2011).

Singh SM and Panda AK: Solubilization and refolding of bacterial inclusion body proteins. J Biosci Bioeng 99: 303-310 (2005).

Smith MI, Sharp JS and Roberts CJ: Nucleation and growth of insulin fibrils in bulk solution and at hydrophobic polystyrene surfaces. Biophys J 93: 2143-2151 (2007).

Speed MA, King J and Wang DIC: Polymerization mechanism of polypeptide chain aggregation. Biotechnol Bioeng 54: 333-343 (1997).

Stadtman ER: Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. Annu Rev Biochem 62: 797-821 (1993).

Stathopulos PB, Rumfeldt JAO, Scholz GA, Irani RA, Frey HE, Hallewell RA, Lepock JR and Melering EM: Cu/Zn superoxide dismutase mutants associate with amyotrophic lateral sclerosis show enhanced formation of aggregates *in vitro*. PNAS 100: 7021-7026 (2003).

Stefani M and Dobson CM: Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. J Mol Med 81: 678-699 (2003).

Strickley RG and Anderson BD: Solid-state stability of human insulin II. Effect of water on reactive intermediate partitioning in lyophiles from pH 2–5 solutions: Stabilization against covalent dimer formation. J Pharm Sci 86: 645-653 (1997).

Studier FW: Protein production by auto-induction in high-density shaking cultures. Prot Express Purif 41: 207-234 (2005).

Swami R and Shahiwala A: Impact of physiochemical properties on pharmacokinetics of protein therapeutics. Eur J Drug Metab Pharmacokinet 38: 231-239 (2013).

Sørensen HP and Mortensen KK: Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115: 113-128 (2005).

Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F and Vendruscolo M: Prediction of aggregation-prone regions in structured proteins. J Mol Biol 380: 425-436 (2008).

Treuheit MJ, Kosky AA and Brems DN: Inverse relationship of protein concentration and aggregation. Pharm Res 19: 511-516 (2002).

Trivedi MV, Laurence JS and Siahaan TJ: The role of thiols and disulfides on protein stability. Curr Protein Pept Sci 10: 614-625 (2009).

Tu W, Cai K, Gao X, Xiao L, Chen R, Shi J, Liu H, Hou X, Wang Q and Wang H: Improved production of holotoxin Stx2 with biological activities using a single-promoter vector and an auto-induction expression system. Prot Express Purif 67: 169-174 (2009).

Tyagi AK, Randolph TW, Dong A, Maloney KM, Hitscheric C Jr and Carpenter JF: IgG particle formation during filling pump operation: A case study of heterogenous nucleation on stainless steel nanoparticles. J Pharm Sci 98: 94-104 (2009).

Tyedmers J, Mogk A and Bukau B: Cellular strategies for controlling protein aggregation. Nature 11: 777-788 (2010).

Vasina JA and Baneyx F: Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli* cspA and tac promoter systems. Protein Express Purif 9: 211-218 (1997).

Vera A, González-Montalbán N, Arís A and Villaverde A: The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotech Bioeng 96: 1101-1106 (2007).

Villaverde A and Carrió MM: Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol Lett 25: 1385-1395 (2003).

Villegas V, Zurdo J, Filimonov VV, Avilés FC, Dobson CM and Serrano L: Protein engineering as a strategy to avoid formation of amyloid fibrils. Prot Sci 9: 1700-1708 (2000).

Walsh G: Biopharmaceutical benchmarks 2010. Nat Biotechnol 28: 917-924 (2010).

Wang G, Johnson AJ and Kaltashov IA: Evaluation of electrospray ionization mass spectrometry as a tool for characterization of small soluble protein aggregates. Anal Chem 84: 1718-1724 (2012).

Wang W: Instability, stabilization, and formulation of liquid protein pharmaceuticals. Int J Pharm 185: 129-188 (1999).

Wang W: Protein aggregation and its inhibition in biopharmaceutics. Int J Pharm 289: 1-30 (2005).

Waugh DS: Making the most of affinity tags. Trends Biotechnol 23: 316-320 (2005).

Webb SD, Cleland JL, Carpenter JF and Randolph TW: Effects of annealing lyophilized and spray-lyophilized formulations of recombinant human interferon-γ. J Pharm Sci 92: 715-729 (2003).

Weiss IV WF, Young TM and Roberts CJ: Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. J Pharm Sci 98: 1246-1277 (2009).

Xu Y, Yan Y, Seeman D, Sun L and Dubin PL: Multimerization and aggregation of native-state insulin: Effect of Zinc. Langmuir 28: 579-586 (2012).

Yoshioka S: Aggregates formed during storage of β-galactosidase in solution and in the freeze-dried state. Pharm Res 10: 687-691 (1993).

Zölls S, Tantipolphan R, Wiggenhorn M, Winter G, Jiskoot W, Friess W and Hawe A: Particles in therapeutic protein formulations, part 1: Overview of analytical methods. J Pharm Sci 101: 914-935 (2012).

The 15 best-selling drugs of 2012: FiercePharma, 2012. Retrieved from the internet 08.07.2014: http://www.fiercepharma.com/special-reports/15-best-selling-drugs-2012

SOLUBLE EXPRESSION OF HUMAN CILIARY NEUROTROPHIC FACTOR IN *Escherichia coli*

Pro Gradu

Jaakko Itkonen
University of Helsinki
Faculty of Pharmacy
Division of Pharmaceutical Biosciences
January 2013 – August 2014

1 INTRODUCTION

Neurotrophic factors are proteins that have a multitude of functions within the nervous system. They regulate survival, proliferation, differentiation of neurons and other cell types in the nervous system, such as dendrocytes and glial cells (Chao 2003, Wen *et al*. 2012). Ciliary neurotrophic factor (CNTF) was initially identified in chick embryo extract (Adler *et al*. 1979, Varon *et al*. 1979). Human CNTF (hCNTF), a 23 kDa polypeptide, has 200 amino acid residues (Richardson 1994, Wen *et al*. 2012). It belongs to the interleukin-6 (IL-6) family of neuropoietic cytokines (McDonald *et al.* 1995, Wen *et al.* 2012). It is a cytosolic protein under normal physiological conditions and is released upon damage to the cells expressing CNTF, such as glial cells and astrocytes (Sleeman *et al.* 2000, Cognet *et al*. 2004, Wen *et al*. 2012).

CNTF exerts its biological function on different types of cells in the nervous system, e.g., photoreceptors and skeletal muscle (Wen *et al*. 2006). CNTF promotes the survival of neurons (Dutta *et al*. 2007), photoreceptors and is known to have a protective influence on the outer segments of the cone receptor cells (Li *et al*. 2010). CNTF is also known to influence the energy balance (Ott *et al*. 2002, Wen *et al*. 2012). The therapeutic potential of CNTF therefore ranges from neurodegenerative diseases (amyotrophic lateral sclerosis [ALS], Huntington's disease [HD], age-related macular degeneration [AMD]) to metabolic disorders such as, obesity and type II diabetes (Sleeman *et al*. 2000, Ettinger *et al.* 2003, Cognet *et al*. 2004). Successful clinical studies with hCNTF-secreting implants for the treatment of retinitis pigmentosa and dry AMD have been conducted (Sieving *et al*. 2006, Wen *et al*. 2012).

The tertiary structure of CNTF consists of a bundle of four anti-parallel α-helices (A [Arg13-His41], B [Glu69-Val96], C[Phe105-Leu129] and D [Phe152-Ser180]) with cross-over loops between helices A-B and C-D and a single short loop between helices B-C (Figure 1a). (McDonald *et al*. 1995, Kallen *et al*. 1999). hCNTF lacks a signal peptide and a consensus sequence for glycosylation.

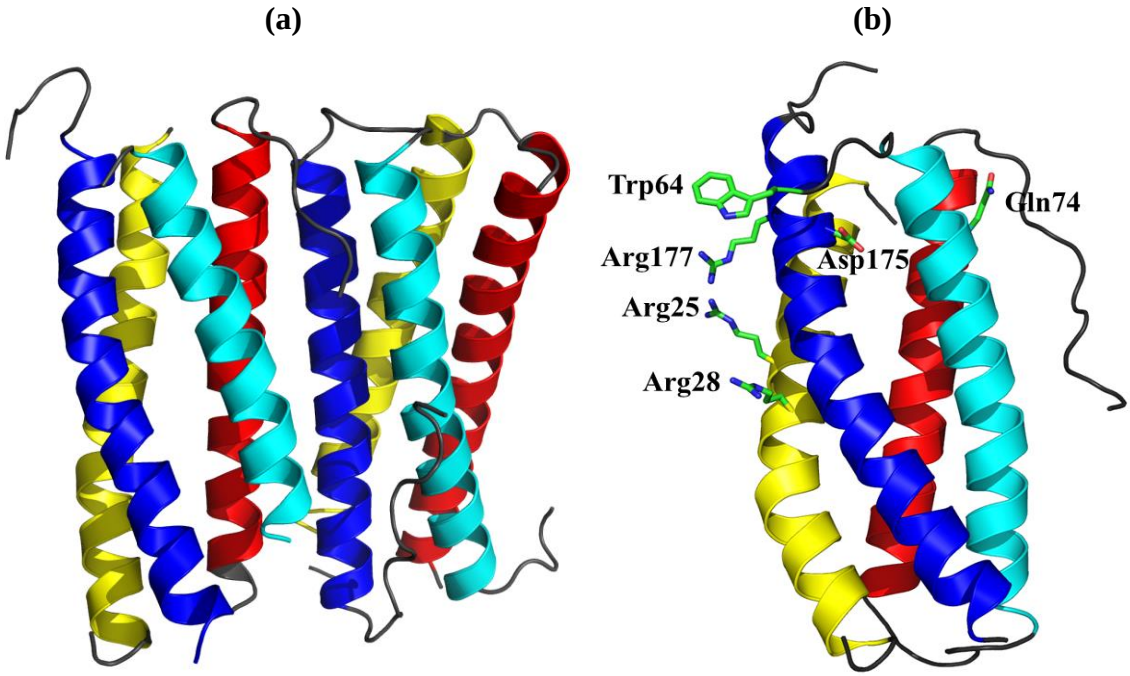


Figure 1: (a) Crystal structure of hCNTF (2-187) dimer. Helices A, B, C and D are depicted with yellow, cyan, red and blue, respectively; **(b)** Residues attributed to binding to CNTFRα.

CNTF mediates its activity by binding to a tripartite receptor complex consisting of CNTF receptor subunit alpha (CNTFRα) (Davis *et al*. 1991) and the trans-membrane signal transducing subunits gp130 and Leukemia inhibitory factor (LIF) receptor β (LIFRβ) (Davis *et al.* 1993b). The CNTFRα-binding surface epitope of CNTF was studied with random mutagenesis and identified to consist of residues Arg25, Arg28, Gln63, Trp64, Gln74, Asp175 and Arg177 (Panayotatos *et al.* 1995). These residues are located in helix A, the loop between helices A-B, helix B and helix D, are spatially clustered and surface accessible (Figure 1b). CNTFRα is a peripheral membrane protein linked to the cell membrane with a glycosyl-phosphatidylinositol-linkage (Davis *et al*. 1991) and is also known to exist as a soluble receptor (sCNTFRα) after phospholipase C mediated cleavage (Davis *et al*. 1993a). In humans, binding of CNTF to CNTFRα results in the recruitment of gp130 and LIFRβ that eventually activates signal transduction.

hCNTF is also known to bind with lower affinity to both, membrane-bound and soluble human IL-6 receptor α chain (IL-6Rα), a structural relative of CNTFRα (Davis *et al*. 1991, Davis *et al*. 1993b. This is possibly an alternate arrangement to form receptor complex to activate downstream signal transduction processes (Schuster *et al.* 2003). These observations offer an explanation as to why cells expressing gp130 and LIFRβ but not CNTFRα (e.g. human liver cells) are still responsive to CNTF (Sleeman *et al*. 2000, Schuster *et al*. 2003, Cognet *et al*. 2004). The side effects caused by higher doses of CNTF have also been partially attributed to the formation of IL-6Rα/gp130/LIFRβ receptor complex (Schuster *et al*. 2003, Wagener *et al*. 2014).

Clinical and investigative biological studies are often limited by the amount of pure protein available (Peti & Page 2007). Isolation and purification of protein from natural sources such as tissues, is often cumbersome, limited in amounts and poses challenges in purification. This necessitates heterologous recombinant expression of the protein of interest. The recombinant technology to express proteins has been successfully utilized to produce many therapeutic proteins. Some examples from the recent-past include human insulin (Goeddel *et al.* 1979), human growth hormone (Olson *et al.* 1981), erythropoietin (Winearls 1998) and antibodies (Better *et al.* 1988, Skerra and Plückthun 1988). Currently, therapeutic proteins are routinely expressed in mammalian cells, in yeasts, and in bacterial cells, such as *Escherichia coli* (Gengross 2004, Nayak 2010, Walsh 2010, Bandranayake and Almo 2014). Recent advances in the production of protein therapeutics include producing therapeutic proteins in plants and transgenic animals (Nayak 2010, Walsh 2010). Notable accomplishments include approvals for the first proteins expressed in baculovirus-infected insect cells and in the yeast *Pichia pastoris* (Walsh 2010); taliglucerase alpha was the first recombinant protein produced in plant cells (procuded in cultured carrot cells) to receive US Food and Drug Administration (FDA) approval (Aggarwal 2012), while Ecallantide and the proteins in the human papilloma virus (HPV) vaccine Cervarix were the first approved therapeutic proteins produced in baculovirus-infected insect cells and in the yeast *Pichia pastoris*, respectively (Walsh 2010).

Different recombinant expression systems exist for protein production and the choice for a particular system is governed by the characteristics of the protein of interest. Eukaryotic expression systems (Jarvis 2003, Gräslund *et al.* 2008, Yin *et al.* 2007, Tworak *et al.* 2011) include yeast (*Pichia pastoris, Saccharomyces cerevisiae*), baculovirus-infected insect cells (Sf9, Tn-5B1-4) and mammalian cells (CHO, HEK). The yeast system offers operational simplicity and is inexpensive. It is capable of forming disulphide bonds and carrying out post-translational modifications (PTMs) such as glycosylation. The extent of glycosylation is usually different from that achieved in mammalian systems and this might limits its applicability. Baculovirus-infected insect cells are able to carry out PTMs similar to higher eukaryotic systems; however, their transient transfection renders them unsuitable for continuous fermentation. Mammalian cells produce proteins with correct PTMs and structural fold. The system is however complex, time consuming and costly. Among the prokaryotes, the enterobacterium *Escherichia coli* is the most versatile and economical expression system. They are commonly referred to as the 'laboratory workhorses' for their utility in protein production (Baneyx and Mujacic 2004). However, recombinant expression of eukaryotic proteins (especially of human origin) in *E. coli* is a challenging issue (Sahdev *et al*. 2008). Inefficient or improper folding of recombinant proteins due to lack of eukaryotic transcriptional machinery in *E. coli* (e.g., disulphide bond formation and glycosylation) might result in the expression of non-native, non-functional protein. Improper folding often results in the formation of insoluble aggregates known as inclusion bodies (Baneyx and Mujacic 2004, Sørensen & Mortensen 2005). *E. coli*'s inability to effect PTMs limits its utility as a robust expression system for eukaryotic proteins (Baneyx and Mujacic 2004, Sahdev *et al*. 2008).

Earlier efforts to produce recombinant hCNTF in *E. coli* have been limited by low yields and the need to refold the protein from inclusion bodies (McDonald *et al.* 1991, Masiakowski *et al.* 1991, Negro *et al.* 1991, McDonald *et al.* 1995, Wagener *et al.* 2014). For therapeutic protein production, refolding from inclusion bodies is not desirable; it not only complicates the manufacturing process but also the possibility of protein aggregation raises the risk of immunogenic reactions. An effective, economical expression system for soluble production of hCNTF is highly desirable towards its development as a therapeutic agent.

2 AIM

The objective of our work was to express soluble and functional hCNTF in *Escherichia coli*. Towards the realization of our goal, the following experiments were carried out:

- Codon optimization of the hCNTF sequence for expression in *E. coli*
- Expression plasmids with nine different fusion partners and codon optimized hCNTF were constructed for enhanced solubility/facilitated folding of the expressed '*fusion tag* – hCNTF' construct
- Small-scale expression screening for soluble expression of hCNTF.
- Large-scale expression, purification of hCNTF.
- *In vitro* binding of hCNTF to hCNTFRα

3 MATERIALS AND METHODS

3.1 Codon optimization

The codons of the hCNTF gene were optimized based on codon preference usage in *E. coli* using a commercial proprietary algorithm, OptimumGene^{TM} (GenScript, NJ, USA). Variety of parameters were optimized, including codon usage bias, GC content, CpG dinucleotides content, mRNA secondary structure, cryptic splicing sites, premature PolyA sites, internal chi sites and ribosomal binding sites, negative CpG islands, RNA instability motif, repeat sequences. The synthetic gene was purchased from GenScript, US.

3.2 Construction of expression vectors

Preparation of *E. coli* competent cells, bacterial transformation, plasmid DNA purification and plasmid DNA digestion are prescribed in APPENDICES 1, 2, 3 and 4, respectively.

3.2.1 Polymerase chain reaction (PCR) amplification of hCNTF construct

Primers for amplifying the hCNTF sequence were designed using the NetPrimer software (PREMIER Biosoft). The forward and reverse primers were designed to contain overhang sequences AAGTTCTGTTTCAGGGCCCG and ATGGTCTAGAAAGCTTTA, respectively (Bird 2011). The primers were procured from Eurofins MWG Operon; Table 1 lists the full sequence of the primers.

Table 1: Primers used for amplification of hCNTF.

Primer	Sequence (5' → 3')	T_m (°C)
CNTF_1_fp	AAGTTCTGTTTCAGGGCCCGATGGCGTTTACCGAACATTCC	60
CNTF_200_rp	ATGGTCTAGAAAGCTTTACATCTTCTTGTTGTTTGCGATGTAG	60

Sequences in blue and red denote overhangs, sequences in green denote complementary sequences of hCNTF; fp and rp denotes forward and reverse primers, respectively.

PCR amplification was carried out using KOD Hot Start DNA Polymerase (Novagen). The reactions were set-up on ice as in Table 2**.**

Table 2: PCR reaction set-up.

Component	Final concentration in 50 µl
KOD Hot Start Buffer	1 x
dNTP mix	0.2 mM
$MgSO_4$	1.5 mM
Forward primer	0.3 µM
Reverse primer	0.3 µM
Template DNA	0.2 ng/µl
KOD Hot Start DNA Polymerase	0.02 U/µl

The thermal cycling carried out is depicted in Figure 2. The reaction mixture was analyzed by gel electrophoresis (0.7 % agarose, Bionordika; TAE buffer). The gel contained 0.5 mg/ml of ethidium bromide and the run was carried out at 110 V for 50 minutes. Samples were prepared with 5 µl of each PCR reaction mixture and 6X sample buffer. 5µl of each sample was loaded to the gel.

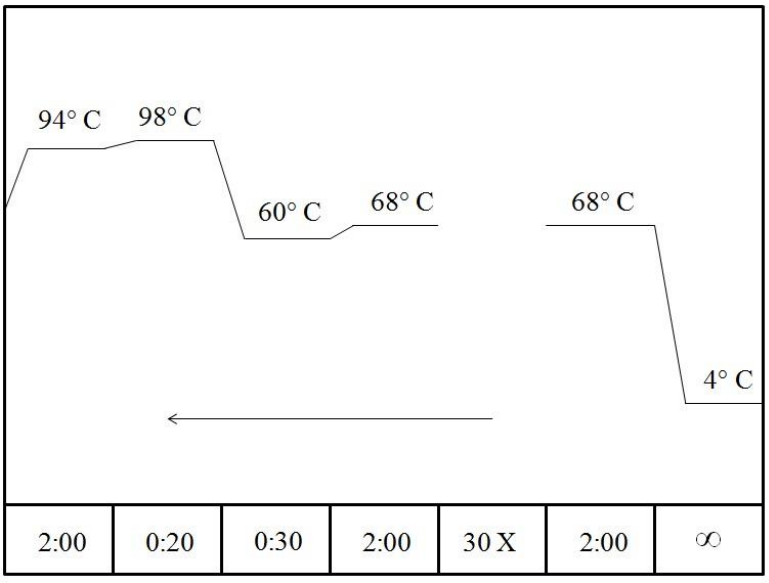


Figure 2: PCR thermal cycle

3.2.2 Cloning of hCNTF into pOPIN vectors

The suite of nine pOPIN vectors (Oxford Protein Production Facility, UK [OPPF-UK]) used for cloning hCNTF gene is listed in Table 3. All nine fusion partners contained an N-terminal His_6-affinity tag to facilitate the subsequent purification of the fusion protein, and a Rhinovirus 3C protease cleavage site between the N-terminal fusion partner and the target protein to facilitate the removal of said fusion partner. All the vectors contained common restriction sites, HindIII and KpnI; this facilitated parallel cloning.

Linearized plasmid vectors and PCR-amplified hCNTF insert was purified by gel extraction. Details of gel extraction protocol are provided in APPENDIX 5. hCNTF gene was inserted into linearized pOPIN vectors using the Gibson Assembly™ Cloning Kit (NEB).

20 µl of the cloning reaction mixture consisted of 75 - 100 ng of the linearized vector, inserts in a 2-fold molar excess of the vector, 10 µl of 2 X Gibson Assembly Master Mix,. The reactions were incubated at 50°C for 25 minutes. After incubation, the reaction mixtures were placed on ice (to arrest the reaction) and subsequently 2 µl was used to transform NEB 5-alpha (New England BioLabs) *E. coli* competent cells. Details of the transformation protocol are provided in APPENDIX 2.

The transformed cells were plated on LB-agar plates supplemented with ampicillin[1], 80 µg/ml X-gal and 0.3 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). The plates were incubated overnight at 37°C. Separate colonies were used to inoculate 5 ml liquid cultures for DNA plasmid preparation (detailed in APPENDIX 3).

[1] Throughout the course of this work, antibiotic concentrations used in cultures were 100 µg/ml and 34 µg/ml for ampicillin (ICN) and chloramphenicol (ICN), respectively.

3.2.3 Validation of positive clones

Validation of the cloned constructs were carried out by PCR amplifying the full-length fusion construct (fusion tag – hCNTF) using *Pfu* DNA Polymerase (Thermo Scientific). The forward primers used were specific to each fusion tag in the vector and the reverse primer was specific to the C-terminus of hCNTF. The primers used for validation are listed in Table 3. PCR reaction set up and thermal cycle is depicted in Table 4 and Figure 3, respectively.

Table 3: Primers for PCR amplification of the constructs. fp denotes forward primer, while rp denotes reverse primer.

Primer	Sequence (5' → 3')	T_m (°C)
pOPINF_2357_fp	AGCAGCGGTCTGGAAGTTCTGTTT	65
pOPINS3C_2357_fp	GGGAGCGATAGCGAAGTGAACCA	67
pOPINTRX_2357_fp	AGCGATAAAATTATTCACCTGACTGACGAC	68
pOPINMSYB_2357_fp	ACCATGTACGCAACGCTTGAAGAAG	66
pOPINJ_2357_fp	ATGTCCCCTATACTAGGTTATTGGAAAATTAAG	65
pOPINHALO7_2333_fp	ATGGCACACCATCACCACCATCAC	67
pOPINM_2357_fp	AGCAGCGGCATGAAAATCGAAG	65
pOPINTF_2357_fp	CAAGTTTCAGTTGAAACCACTCAAGGC	66
pOPINNusA_2333_fp	ATGGCACACCATCACCACCATCAC	67
CNTF_200_rp	CATCTTCTTGTTGTTTGCGATGTAGTGC	67

Table 4: PCR reaction set-up

Component	Final concentration in 50 µl
Pfu Buffer with $MgSO_4$	1 x
dNTP mix	0.2 mM
Forward primer	0.3 µM
Reverse primer	0.3 µM
Template DNA	0.2 ng/µl
Pfu DNA Polymerase	0.02 U/µl

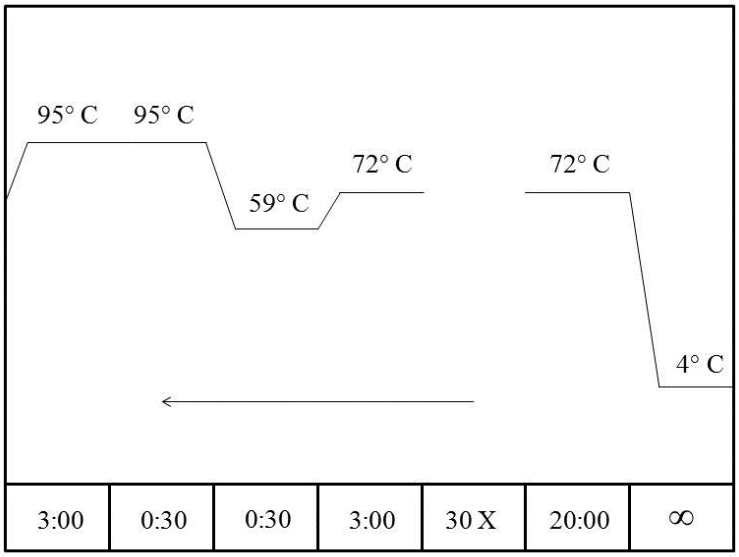


Figure 3. PCR thermal cycle for the PCR-check

After PCR-amplification, 5 µl of each reaction mixture was loaded to an agarose gel (0.7 % agarose, TAE buffer). The electrophoresis run was carried out at 110 V for 70 minutes. The size of the constructs as obtained from PCR validation is listed in Table 5.

Table 5: Sizes of tag sequences coupled with hCNTF sequence

Vector	Size of 'fusion tag – hCNTF' constructs (bp)
pOPINF	668
pOPINS3C	956
pOPINTRX	992
pOPINMSYB	1037
pOPINJ	1331
pOPINHALO7	1550
pOPINM	1778
pOPINTF	1961
pOPINNusA	2150

3.3 Protein expression screening

3.3.1 Small scale expression

Expression strains BL21(DE3)pLyS (Novagen) and Rosetta 2(DE3)pLysS (Novagen) – a derivative of BL21 engineered to supply tRNAs for 7 rare codons; AGA, AGG, AUA, CUA, GGA, CCC and CGG – were transformed with the plasmids. Single colonies were used to inoculate 0.7 ml of (Power *Prime* Broth [PPB] [AthenaEs]) (supplemented with ampicillin and chloramphenicol) on a multiwall plate and grown overnight at 37°C with shaking at 250 rpm.

150 µl and 250 µl of seed cultures from BL21(DE3)pLysS, and Rosetta 2(DE3)pLysS, respectively were used to inoculate 3 ml of expression culture (PPB / Overnight Express™ Instant TB Media [TBONEX] [Novagen] supplemented with ampicillin and chloramphenicol) on a multiwall plate. The cultures were grown at 37°C with shaking at 250 rpm to a OD_{600} value of 0.5.

For protein expression at 37°C in PPB, induction was carried out with 1mM IPTG and grown for 4 h with shaking at 250 rpm. For expression at 20°C with PPB, the cultures were cooled on the bench-top for 15 minutes before adding 1mM IPTG. The culture was grown at 20°C for 18 h with shaking at 250 rpm. Expression in TBONEX media was carried out at 25°C for 24 h with shaking at 225 rpm.

The cells were harvested by centrifugation (3650 rpm) for 35 minutes at 4°C. The supernatants were discarded and pellets washed with 210 µl of NPI-10 (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 10 mM Imidazole) and centrifuged (3650 rpm) for 35 minutes at 4°C. The wash liquid was discarded and the cell pellets were stored at -80°C.

3.3.2 Estimation of soluble fraction

Cell pellets were thawed at room temperature. The pellets were resuspended in 250 µl of buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 0.2 % Tween 20) and the suspended cells were transferred to microcentrifuge tubes. The samples were freeze-thawed three times (liquid nitrogen and 37°C water bath). 50 µl of cell mixtures were transferred to microcentrifuge tubes followed by 1 µl of Lysonase™ Bioprocessing Reagent (EMD Millipore). The cell mixtures were incubated at room temperature with shaking at 270 rpm for 45 minutes.

After incubation, the homogenized samples were divided into two equal fractions of 25 µl each. For the fraction ‘total protein’; 25 µl of water and 10 µl of 6x SDS-PAGE sample buffer was added and the mixture heated at 95°C for 15 minutes. The mixture was then centrifuged at 21,100 g for 15 minutes. The supernatants were transferred to microcentrifuge tubes and stored at -20°C. For the fraction, ‘soluble protein’; 25 µl cell suspensions were centrifuged at 21,100 g for 15 minutes and the supernatants transferred to microcentrifuge tubes. 25 µl of water and 10 µl of 6X SDS-PAGE sample buffer were added to the supernatants and the samples were heated at 95°C for 15 minutes. The samples were stored at -20°C and further analyzed by SDS-PAGE gel electrophoresis.

3.3.3 Affinity purification of soluble fraction

Frozen cell pellets were thawed on 42°C water bath for 15 minutes followed by 210 µl of resuspension buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 10 mM Imidazole, 1 mM DL-dithiothreitol [DTT, Sigma], Protease Inhibitor Cocktail [PIC, Sigma][2]) was added to the cell suspension. The freeze thaw cycle was repeated three times with mixing in between cycles. After freeze-thawing, 1 µl of Lysonase was added to the pellet suspensions and incubated with shaking (270 rpm) at room temperature for 45 minutes. After incubation, the lysates were transferred to microcentrifuge tubes and centrifugation (21,100 g) was carried out at 4°C for 40 minutes.

20 µl of fully suspended nickel nitriloacetic (Ni-NTA) magnetic beads (Genscript, NJ, USA) was dispensed to the wells of a flat bottom plate. 200 µl of NPI-10-Tween buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 10 mM Imidazole, 1% Tween 20) was added to each well and the plate was gently agitated (270 rpm) at room temperature for 7 minutes. The plate was then placed on a 96-well Magnetic Separation Rack (New England Biolabs) for 5 minutes following which the buffer was removed by aspiration. The washing of the beads was repeated 3 times.

Supernatants from cell lysates were transferred to the wells containing the magnetic beads and incubated with shaking (270 rpm) at room temperature for 55 minutes. The plate was then placed on the magnetic rack for 5 minutes and the supernatant discarded. The beads were washed 4X with NPI-20-Tween buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 20 mM Imidazole, 0.05% Tween 20, 1 mM DTT) to remove unbound proteins.

[2] Throughout the course of this work, whenever buffers were supplemented with PIC, this was done in a 1:100 dilution

The bound proteins were eluted by adding 50 µl of NPI-250-Tween buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 250 mM Imidazole, 0.05% Tween 20, 1 mM DTT). The plate was shaken (270 rpm) at room temperature for 10 minutes followed by placing it on the magnetic rack for 7 minutes. 45 µl of supernatants were transferred to microcentrifuge tubes. 9 µl of 6X SDS-PAGE sample buffer was added to each tube and heated at 95°C for 7 minutes and stored at -20°C for further analysis with SDS-PAGE gel electrophoresis.

3.3.4 Analysis with SDS-PAGE

Percentage of soluble protein fraction was estimated by using GelQuant.NET software (V 1.7.8) from biochemlabsolutions.com. Band intensities of soluble and total protein fractions in the gel were measured three times and averaged. Procedural details of SDS-PAGE gel electrophoresis are outlined in APPENDIX 6.

3.4 Large scale protein expression and purification

6His-hCNTF construct was used for large-scale protein expression and purification. Transformed single separate Rosetta 2 (DE3) pLysS colonies were used to inoculate 20 ml of LB media supplemented with ampicillin and chloramphenicol. The cultures were grown at 25°C for 15 hours with shaking at 225 rpm. 450 ml of TBONEX was inoculated with overnight culture (in a ratio of 30:1) and the culture grown at 25°C with shaking (225 rpm) for 24 hours. The cells were harvested by centrifugation (10,000 g) at 4°C for 30 minutes and the supernatant removed. The pellets were washed and re-suspended in 20 ml of NPI-10 (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 10 mM Imidazole) followed by centrifugation (10,000 g) at 4°C for 30 minutes. The supernatant was removed, the pellets weighed and stored at -80°C.

The frozen cell pellet was thawed and re-suspended in lysis buffer[3]. The cell suspension was freeze-thawed 3X (alternating between -80°C and 37°C). After the freeze thaw cycle, Lysonase was added to the mixture and incubated at room temperature for 90 minutes with gentle agitation. The lysate was then centrifuged (14,500 g) at 4°C for 50 minutes and the supernatant reserved for purification.

7.0 g of Ni-IDA was suspended in 8 bed volumes of NP-buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 5 % glycerol, 0.2 % Tween 20 supplemented with 1mM DTT and PIC). The suspension was rocked gently at room temperature for 15 minutes followed by centrifugation at 500 g for 3 minutes. The supernatant was removed and this process was repeated one more time. The resin was finally re-suspended in 8 bed volumes of NP-buffer.

Supernatant from cell lysate was added to Ni-IDA resin and incubated at 4°C for 120 minutes with gentle agitation. Centrifugation was then carried out at 500 g for 3 minutes and the supernatant removed.
Unbound proteins were removed by washing (gently agitated for 10 minutes followed by centrifugation at 500 g for 3 minutes) the resin with 4 bed volumes of NPI-10-Tween (50 mM NaH2PO4, 300 mM NaCl, pH 8.0; 10 mM Imidazole, 5 % glycerol, 0.2% Tween 20 supplemented with 1mM DTT and PIC). Further washings were carried out 1X with NPI-20-Tween (50 mM NaH2PO4, 300 mM NaCl, pH 8.0; 20 mM Imidazole, 5 % glycerol, 0.2% Tween 20 supplemented with 1 mM DTT and PIC) and NPI-40-Tween (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 40 mM Imidazole 5 % glycerol, 0.2 % Tween 20, supplemented with 1 mM DTT and PIC).

[3] Estimation of resin amount for optimal IMAC purification is described in APPENDIX 10

Bound protein(s) was eluted in a stepwise manner (1X with 1 bed volume, 2X with 2 bed volumes and 3X with 1.33 bed volumes; incubation at room temperature for 15 minutes with gentle agitation followed by centrifugation at 500 g for 3 minutes) with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 250mM Imidazole, 5% glycerol, 0.2% Tween 20 supplemented with 1 mM DTT and PIC). 7 µl of each elution fraction was used for SDS-PAGE gel electrophoresis analysis. Relevant elution fractions were pooled, buffer exchanged (100 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) and concentrated to 6 ml using Amicon® Ultra-4 Centrifugal filter units (Merck Millipore) with a 3 kDa molecular weight cut off.

Concentrated protein sample after Ni-IDA purification was applied to HiLoad Superdex 200 prep grade gel filtration column (2.6 x 60 cm) equilibrated with buffer containing 100 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 1 mM DTT. Relevant protein fractions were pooled, concentrated and buffer exchanged in 100 mM NaH_2PO_4, 50 mM NaCl pH 8.0; 2 mM DTT using Amicon® Ultra-4 centrifugal filter units.

3.5 *In vitro* activity assay of hCNTF

3.5.1 Fluorescent labeling of hCNTF

hCNTF was labeled with biotin using EZ-Link™ NHS-PEG_4-Biotinylation Kit (Thermo Scientific). The biotinylation reaction was carried out in a carbonate buffered solution (100 mM $NaHCO_3$, 100 mM Na_2CO_3) in dimethylformamide (DMF) in 20:1 molar ratio of NHS-PEG_4 Biotin to hCNTF at room temperature for 2 h with gentle agitation. Buffer exchange was carried out in 100 mM NaH_2PO_4, 50 mM NaCl, pH 8.0 to remove excess NHS-PEG_4 Biotin.

3.5.2 Binding of biotinylated hCNTF to hCNTFR

Nunc MaxiSorp® flat-bottom 96-well plate was washed 2X with BupH™ Phosphate Buffered Saline (PBS, ThermoScientific). 1 µg of human CNTFRα (hCNTFRα; Sino Biological Inc.) in PBS was added to the wells of the plate and incubated overnight at +4°C with gentle shaking. The wells were washed 3X with 50 mM Tris-HCl, 150 mM NaCl, 0.05% Tween 20 (TBS-T); pH 7.0. To reduce unspecific binding to the plate surface, blocking was carried out by adding 300 µl of 2% bovine serum albumin (BSA) in TBS-T to each well followed by incubation at room temperature for 90 minutes with gentle agitation. The wells were then washed 4X with TBS-T.

0.2, 0.5, 1.0, 1.5, 2.0 and 3.0 µg of biotinylated hCNTF (bCNTF) (contained in 100 µl of 0.1 % BSA in TBS-T) were added to the wells containing hCNTFRα. To account for unspecific binding, 3 µg of unlabeled hCNTF (contained in 100 µl of 0.1 % BSA in TBS-T) was added to the wells. The plate was incubated for 2 h at room temperature with gentle shaking followed by washing 4X with TBS-T. 50 ng of HRP (horseradish peroxidase) – conjugated streptavidin (Thermo Scientific) in 100 µl of 0.5% BSA in TBS-T was added to the wells and incubated at room temperature for 70 minutes with gentle agitation. This was followed by 4X washings with TBS-T.

200 µl of TMB (3,3’,5,5’-tetramethylbenzidine) substrate solution (Thermo Scientific) was added to the wells and incubated at room temperature with shaking for 20 minutes. The reaction was stopped with 2 M H_2SO_4. After gentle mixing, optical density of the well solutions was measured at 450 nm using a spectrophotometer (Varioskan Flash, Thermo Scientific, Waltham, MA, USA).

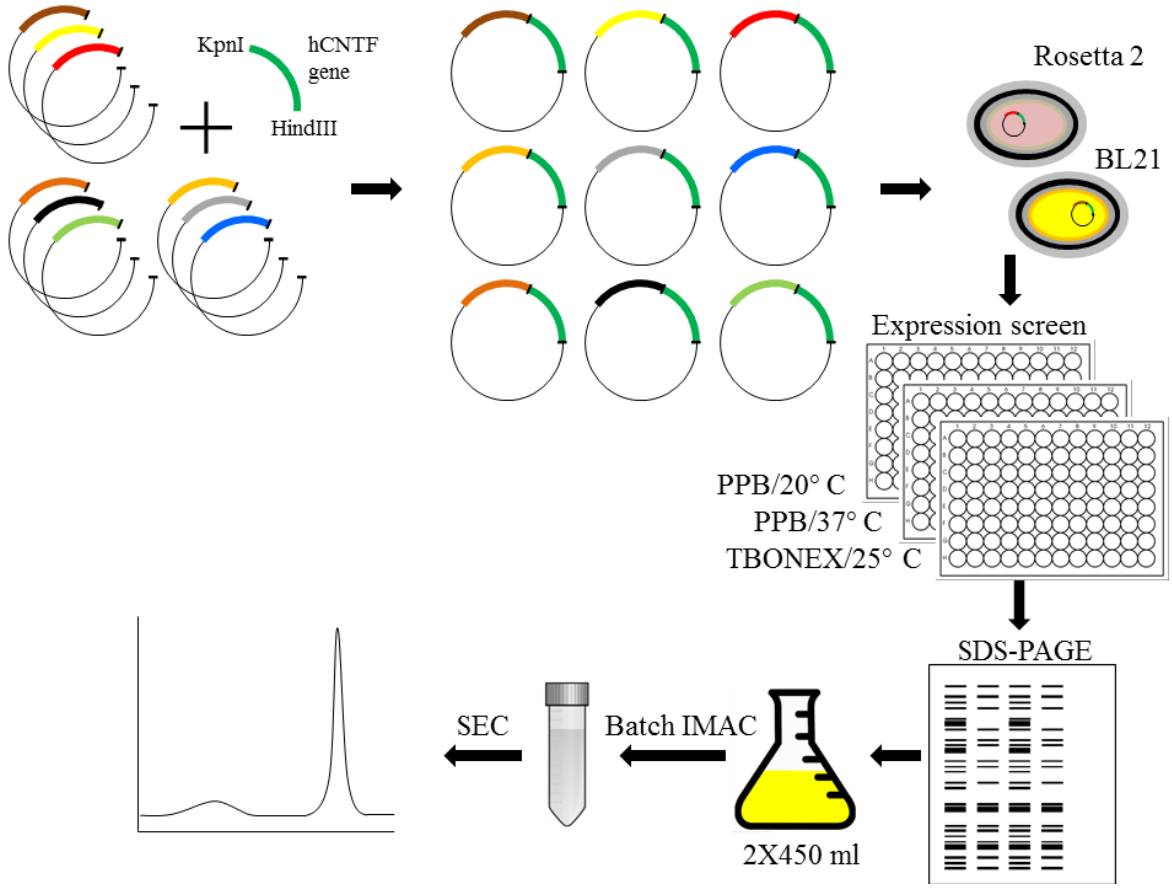
4 RESULTS AND DISCUSSION

Previous studies of hCNTF expression in *E. coli* (BL21) have reported low soluble amounts and the need to purify from inclusion bodies (McDonald *et al.* 1991, Masiakowski *et al.* 1991, Negro *et al.* 1991, McDonald *et al.* 1995, Wagener *et al.* 2014). In one study, only 13% of soluble hCNTF was obtained from cell extracts. Other studies have reported the presence of hCNTF in amounts as high as 80 % in the insoluble fraction (Negro *et al*. 1991) although in one report, translation of hCNTF was 20 – 40% of the total protein (Masiakowski *et al.* 1991). Challenges of recombinant expression of eukaryotic proteins in bacterial hosts such as *E. coli*, stem from the fact that they are often limited in amounts of tRNA for codons that are used less frequently. This might result in non-optimal translation, termination, frame shifting and low levels of protein expression (Kane 1995, Calderone *et al.* 1996). Non-optimal structural features of mRNA are also believed to adversely affect protein expression (Kudla *et al.* 2009, Welch *et al.* 2009). To mitigate issues arising from codon preference usage, cell strains engineered to supply tRNAs for codons rarely found in *E. coli* have been used or the gene of interest has been codon optimized for preferential use in *E. coli* (Peti & Page 2007, Burgess-Brown *et al*. 2008, Sahdev *et al*. 2008). Codon optimization enhanced expression of pigment epithelium-derived factor (Gvritishvili *et al.* 2010) and thirty human short chain dehydrogenase/reductase genes showed improved expression with optimized codon and use of *E. coli* strains containing rare codon tRNAs (Burgess-Brown *et al.* 2008).

The other limiting factor with heterologous expression in *E. coli* arises from improper folding of the polypeptide chain and subsequent formation of insoluble aggregates, commonly referred to as inclusion bodies. Refolding of proteins from inclusion bodies often requires extensive optimization and might often result in non-native conformational states with propensity to aggregate; this is not desirable for therapeutic protein production. Co-expression of molecular chaperones (proteins that assist folding) along with the target protein of interest and using cell strains that have been engineered to aid protein folding *via* the formation of disulfide bonds have proved beneficial towards soluble expression of natively folded proteins (Baneyx and Mujacic 2004, Sørensen & Mortensen 2005, Sahdev *et al*. 2008).

Soluble protein expression in *E. coli* has also been facilitated by the fusion of soluble ‘tags’ (protein / peptide fragments) at the N-terminus of the protein of interest. The fusion tags can be cleaved off with a site-specific protease after successful expression of soluble fusion construct. Fusion ‘tags’ are known to improve expression yield, reduce susceptibility to proteolysis and might also serve as affinity ‘tags’ to facilitate purification of the expressed proteins (Waugh 2005, Arnau *et al*. 2006, Esposito & Chatterjee 2006, Peti & Page 2007, Sahdev *et al*. 2008, Huang *et al*. 2012). However, no general rules apply to the choice of a particular fusion ‘tag’ for a protein of interest and the process of trial-and-error is usually followed. Comparisons of the utility of various fusion tags with different proteins have been reported (Hammarström *et al*. 2002, Bird 2011, Huang *et al*. 2012). Other factors such as, decreased temperature of growth is also known to be helpful for soluble expression of proteins (Vasina and Baneyx 1997).

Successful protein expression in *E. coli* remains largely a trial-and-error process. Small-scale expression screens have emerged as an effective approach to test and identify various conditions (culture media, temperature, expression strains, fusion tags) in parallel for optimal protein expression (Berrow *et al*. 2006, Gräslund *et al*. 2008). In our efforts to express soluble hCNTF, we combined the twin strategies of codon optimization and using a set of nine different soluble tags to set up a factorial screen covering conditions of temperature, growth media and bacterial strains. The overall strategy is depicted in Scheme 1.



Scheme 1: Depiction of the strategy of parallel cloning, small scale expression screening and large scale expression of hCNTF

4.1 Construction of expression vectors

Full-length hCNTF gene was successfully amplified by PCR (Figure 4) and purified by gel extraction. Details of gel extraction and purification have been provided in APPENDIX 5.

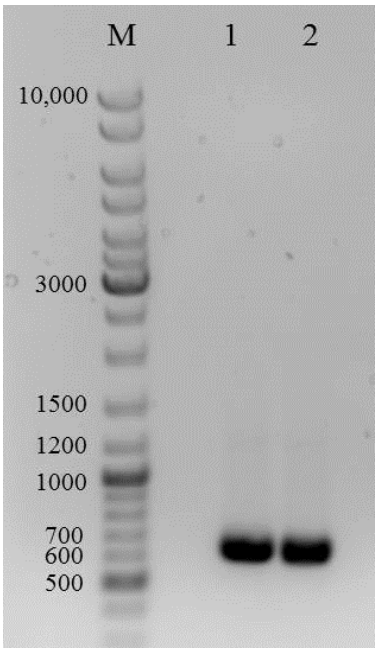
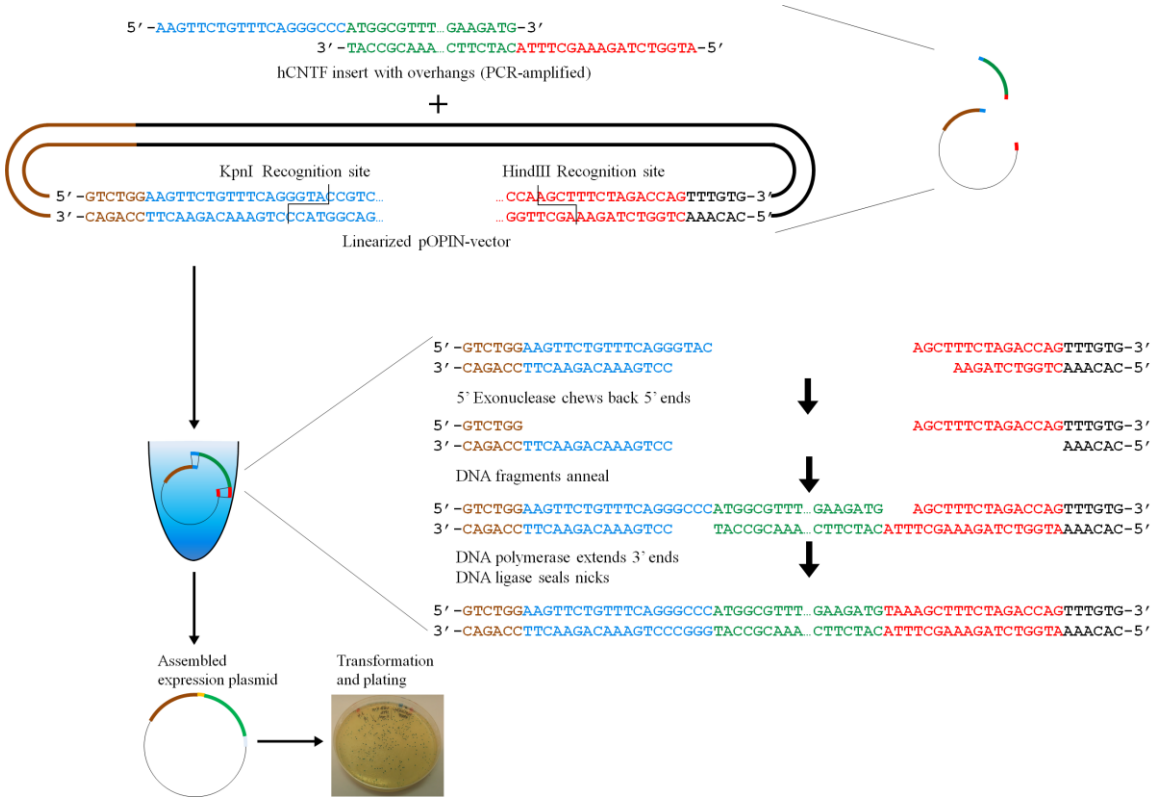


Figure 4. PCR-amplified hCNTF. M denotes DNA marker (*bp*); lanes 1 and 2 denote amplified hCNTF from reaction conditions without and with 3 % DMSO, respectively.

4.2 Cloning of hCNTF into pOPIN vectors

Purified hCNTF gene was cloned into nine different pOPIN vectors (Scheme 2 and Table 6) and three positive clones were picked based on blue / white screening (Matthews 2005). Plasmids were isolated from each of the three clones for all the nine fusion constructs and validated for the presence of the insert (fusion construct) by PCR. A cloning efficiency of 100% was achieved (Figure 5).



Scheme 2: Scheme depicting the cloning strategy[4]

Table 6: Sizes of expression vectors

Vector	Fusion partner	Tag-hCNTF insert size (bp)
pOPINF	MAHis$_6$ -SSGLEVLQF↓GP tag	5531
pOPINS3C	Small ubiquitin-modifier	5824
pOPINTRX	Thioredoxin	5860
pOPINMSYB	MsyB tag	5905
pOPINJ	Glutathione-S-Transferase	6218
pOPINHALO7	HaloTag	6413
pOPINM	Maltose binding protein	6665
pOPINTF	Trigger Factor	6824
pOPINNusA	N-utilisation substance A	7013

[4] The discrepancy caused by a single nucleotide mismatch is a remnant from when the expression vectors were engineered. As it doesn't interfere with the cloning process, it has been left in the sequence.

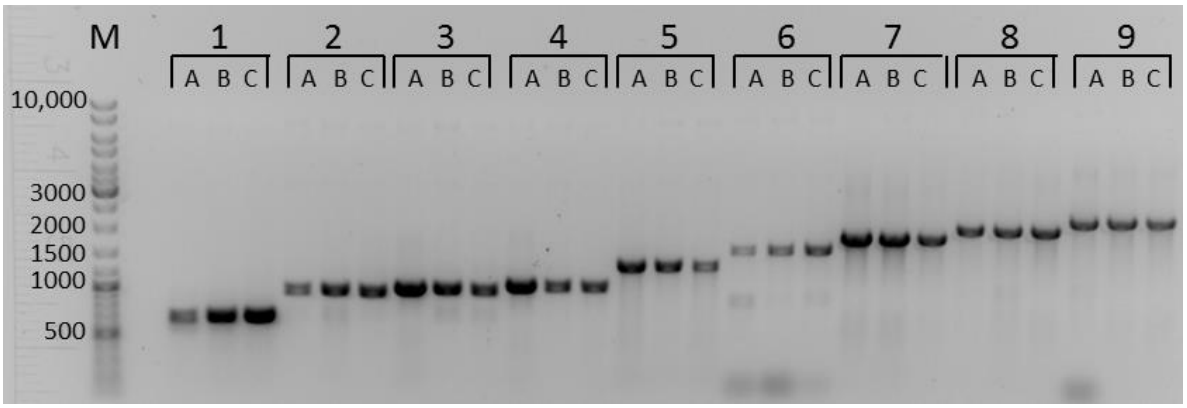


Figure 5. PCR validation of Tag-hCNTF constructs. M denotes marker; lane groupings 1 - 9 denote hCNTF fusion constructs cloned into vectors pOPINF, pOPINS3C, pOPINTRX, pOPINMSYB, pOPINJ, pOPINHALO7, pOPINM, pOPINTF and pOPINNusA, respectively. A, B and C denote the three colonies screened for each construct.

4.3 Small-scale protein expression screen

4.3.1 Estimation of percentage soluble fraction

All-in-all 54 different combinations of expression conditions were assessed. Most conditions (constructs/temperature/media/strains) screened showed the presence of soluble hCNTF constructs. The percentage of hCNTF in the soluble fraction (relative to the total protein) was over 80 – 90 % in most of the expression conditions tested (Figure 6). Our results highlight the huge improvement achieved in soluble expression of hCNTF in *E. coli* as compared to previous reports of less than 20 % (Masiakowski *et al.* 1991, McDonald *et al.* 1991, Negro *et al.* 1991, Wagener *et al.* 2014).

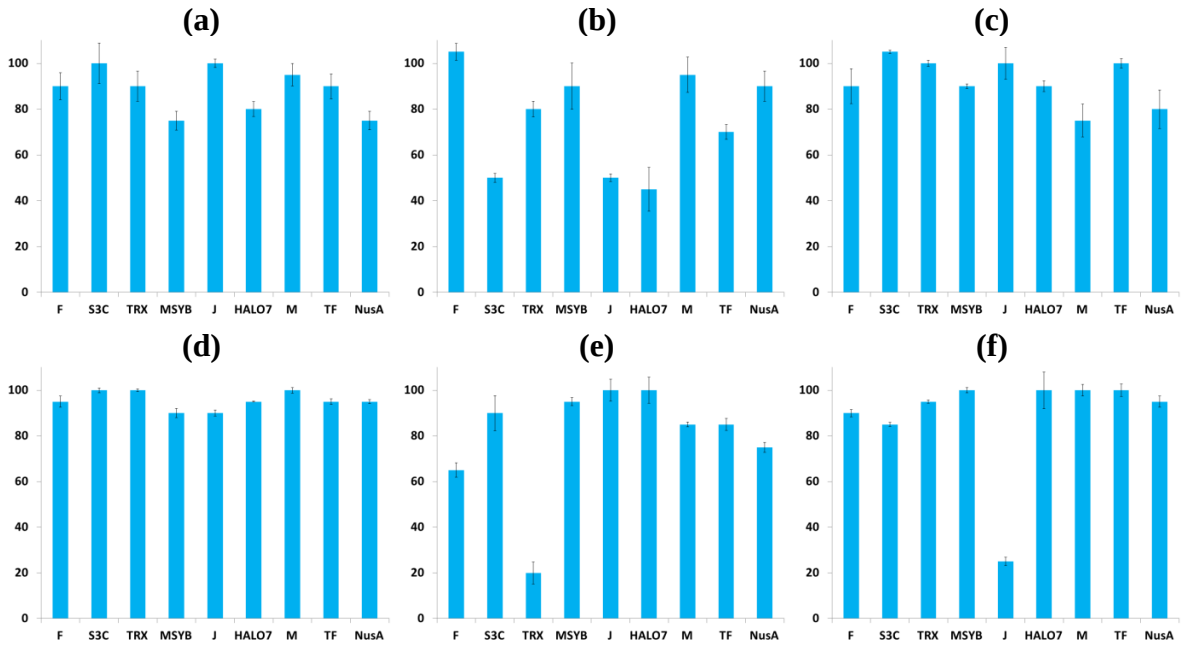


Figure 6. Histograms showing percentage soluble fraction of hCNTF in expression screen[5]**.** (a) BL21/20°C/ PPB; (b) BL21/25°C/ TBONEX; (c) BL21/37°C/ PPB; (d) Rosetta 2/20°C/ PPB; (e) Rosetta 2/25°C/ TBONEX; (f) Rosetta 2/37°C/ PPB.

4.3.2 Affinity purification of soluble protein

Table 7: Expression vectors and their fusion tags.

Lane	Vector	Fusion tag	BL21			Rosetta 2		
			20° C PPB	25° C TBONEX	37° C PPB	20° C PPB	25° C TBONEX	37° C PPB
1	pOPIN**F**	6-His-	1	1	1	1	1	1
2	pOPIN**S3C**	6-His-SUMO-	0.7	1.5	2.1	1.3	2.7	1.1
3	pOPIN**TRX**	6-His-Thioredoxin	1.2	1.7	4.9	1.1	0.7	0.1
4	pOPIN**MSYB**	6-His-MsyB-	0.5	0.8	2.9	1.8	3.6	2.0
5	pOPIN**J**	6-His-GST-	0.3	1.3	0.4	0.1	0.5	0.1
6	pOPINH**ALO7**	6-His-Halo-	1.5	1.6	1.7	1.6	3.0	2.0
7	pOPIN**M**	6-His-MBP-	0.8	1.4	2.5	0.9	1.1	0.7
8	pOPIN**TF**	6-His-Trigger Factor-	0.5	1.1	1.7	0.3	0.6	0.5
9	pOPIN**NusA**	6-His-NusA-	0.6	1.2	1.1	0.7	0.8	0.1

Empirical scores of soluble fraction yield in small-scale expression relative to 6-His-hCNTF construct[6].

[5] Images of the SDS-PAGE gels used for the estimation of the ratio of soluble to total protein are presented in APPENDIX 9.

Empirical scores of soluble fraction yield in small-scale affinity purification relative to 6-His-hCNTF.Affinity purified soluble fractions were analyzed by SDS-PAGE (Figure 7). The intensity of the bands was compared to estimate the relative amounts of soluble hCNTF with respect to the 6-His-construct in each condition (Table 7). Band intensity depends on concentration and also on the size and residue content of the protein. The empirical scores should therefore be used as estimates for easy summarization of results from small-scale screening experiments and not for direct comparisons of protein yield. Efficacy of the fusion tags (scores averaged over six expression conditions) was ranked; the trend followed MsyB / Halo (1.9- fold); SUMO / Trx (1.6-fold); MBP / TF / NusA (approx. 1-fold) and GST (0.5 fold). For some of the larger fusion constructs (Figure 7; Halo, MBP, TF), presence of lower molecular weight fragments hinted at possible proteolysis. In a recent publication, similar observations was made and attributed to improper folding of the larger fusion constructs (Bird 2011). A noteworthy observation of the small-scale screening experiment was that under all conditions of expression tested, 6-His-hCNTF constructs showed appreciable expression. It is highly desirable to be able to express soluble recombinant proteins without solubility enhancing 'tags' as downstream purification processes can avoid additional steps of tag cleavage and protease removal. Also, artifacts of low protein solubility after 'tag' removal can be avoided. The 6-His-tag is an affinity tag to aid purification and does not have solubility enhancing property. In fact, a study on 20 human proteins reported that 6-His-tags are known to have a negative impact on protein solubility when present at either the N- or C-terminus (Woestenek *et al.* 2004). This fact reinforces the notion that soluble expression of hCNTF in *E. coli* can be attributed to codon optimization. In this context, it is worthwhile to mention that in studies reported earlier (Masiakowski *et al.* 1991, McDonald *et al.* 1991, Negro *et al.* 1991, Wagener *et al.* 2014), conditions of expression as varied (soluble tags, lower temperatures, expression strain rich in tRNAs for rare codons) as in our experimental set up were not explored. It seems both translation (low yields 5 – 10 mg/L) and folding (inclusion body formation) were limiting factors in the earlier studies. Expression studies of the wild type sequence in

[6] Images of the SDS-PAGE gels used for the comparison of soluble fraction yield in small-scale expression relative to 6-His-hCNTF construct are presented in APPENDIX 8

a strain rich in tRNAs for rare codons or with soluble fusion tags would be of interest to see if soluble expression of hCNTF could be achieved. Even if a positive outcome from such experiments were to be expected, our results with the codon-optimized sequence still presents a significant improvement when comparing the results of expression in BL21 at 37° C. In our experimental set up, we achieved close to 90 % soluble fraction of hCNTF with the codon optimized sequence but without the aid of soluble fusion tags as compared to less than 20 % with the wild type sequence under similar conditions (BL21/ 37° C) reported in previous studies (McDonald *et al.* 1991, Negro *et al.* 1991, Masiakowski *et al.* 1991, Wagener *et al.* 2014). Synthetic genes are easy options nowadays and our results present an ideal option of recombinant hCNTF production in BL21 (common laboratory expression strain) without a soluble fusion tag.

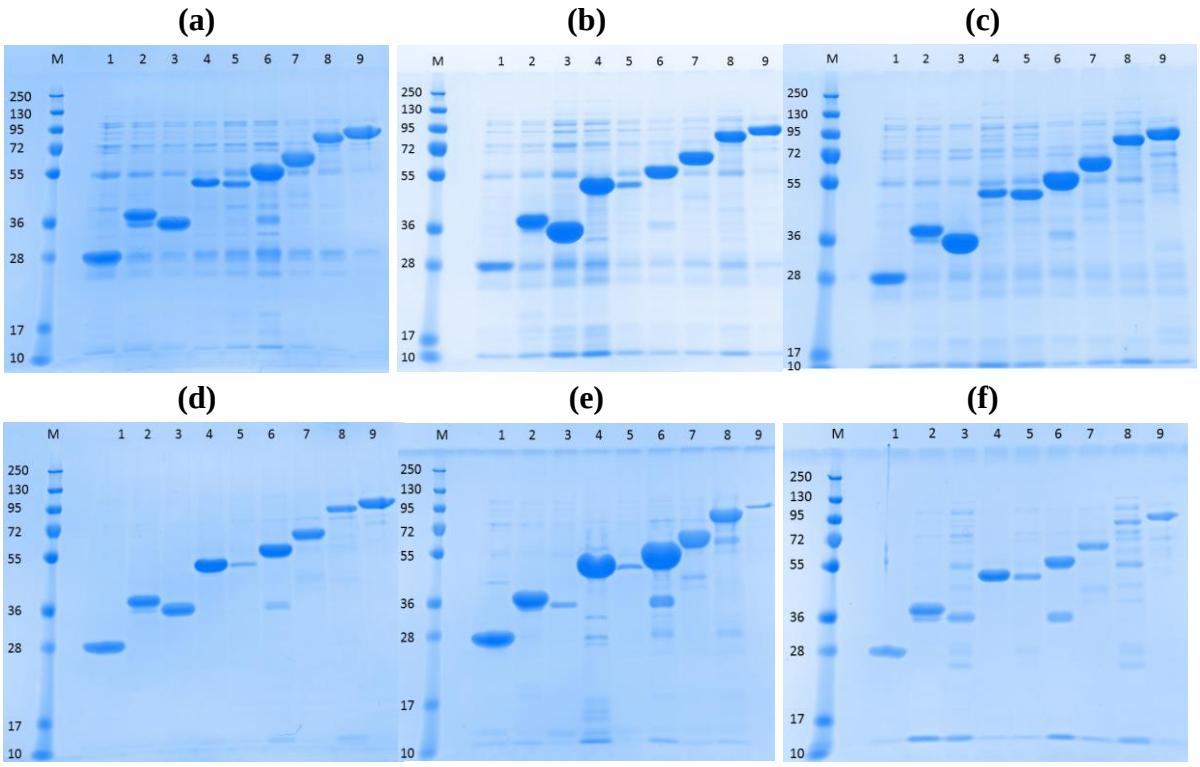


Figure 7. SDS-PAGE analysis of affinity purification from small-scale expression screen. (a) BL21/20°C/PPB; (b) BL21/25°C/TBONEX; (c) BL21/37°C/PPB; (d) Rosetta 2/20°C/PPB; (e) Rosetta 2/25°C/TBONEX; (f) Rosetta 2/37°C/PPB. Lanes M and 1 - 9 denote protein marker and soluble Ni-NTA purified hCNTF from expression constructs as listed in Table 7, respectively.

4.4 Large scale protein expression and purification

As a representative example of large scale protein production, 6-His-hCNTF was chosen for expression in Rosetta 2(pLysS) using auto-inducing media (TBONEX) at 25° C. Our choice was based on results from small scale expression screen with considerations of the following factors: (a) preferable construct without a soluble fusion tag to avoid additional tag cleavage and protease removal steps (b) high soluble yield and purity (c) easy experimental set up; such as, using auto-inducing media that does not require adding external inducing agent (Studier and Moffatt 1986).

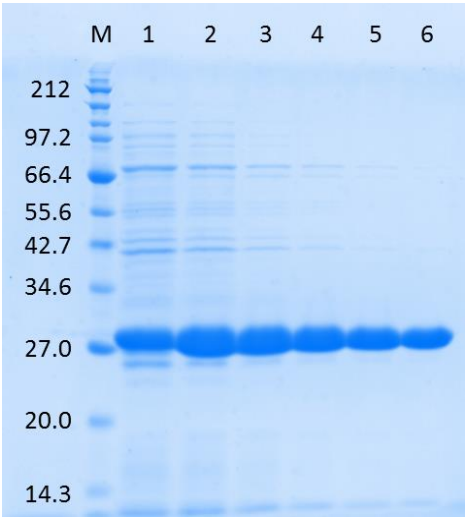


Figure 8. Batch IMAC purification of hCNTF. Lanes M and 1 - 6 denote protein marker and representative elution fractions from the resin Ni-IDA.

hCNTF was extracted from cytosol by cell lysis and subjected to batch purification using Ni-IDA resin followed by size exclusion chromatography (SEC). Results of purification are depicted in Figures 8 and 9.

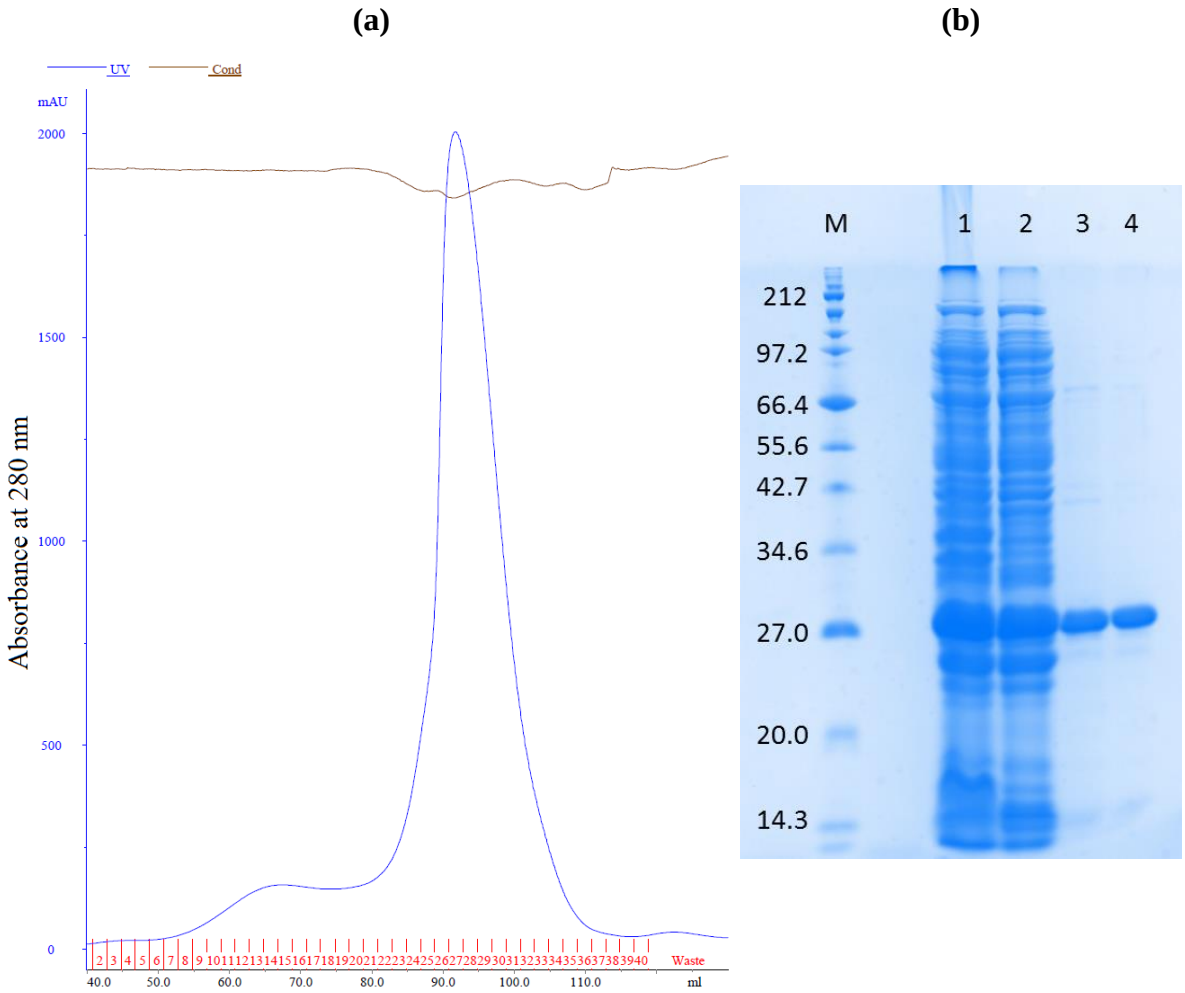


Figure 9. (a) Chromatogram of SEC-purification of hCNTF. The red line portrays the conductivity curve, while the blue line depicts the UV absorbance monitored at 280 nm; **(b) SDS-PAGE analysis of hCNTF purification.** M denotes protein marker and lanes 1 - 4 represent total cell lysate, soluble protein fraction, Ni-IDA purified and SEC purified hCNTF, respectively.

After final purification, 112 mg/L of hCNTF was obtained. This marks a significant improvement in overall yield as compared to previous reports of 5 – 10 mg/L (Masiakowski *et al.* 1991, McDonald *et al.* 1991, Negro *et al.* 1991, Wagener *et al.* 2014).

4.5 Functional activity assay of hCNTF

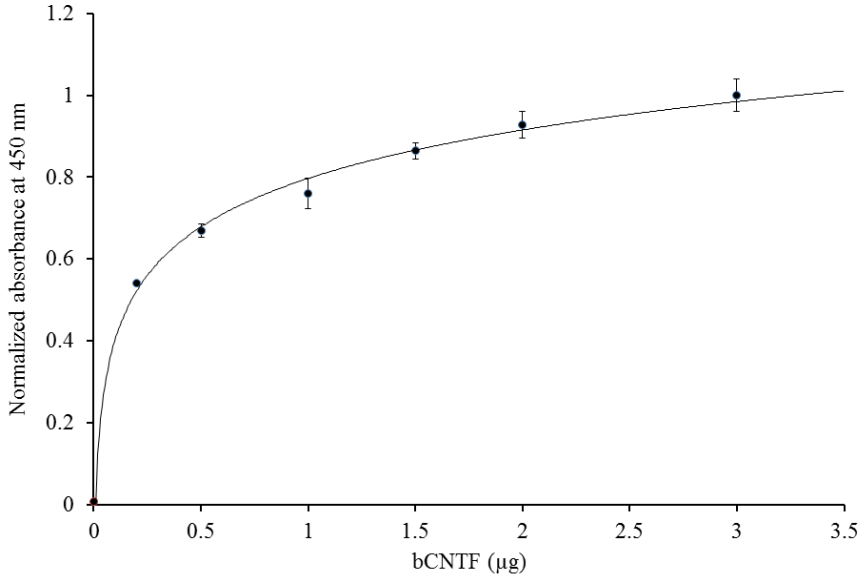


Figure 10. Binding of biotinylated hCNTF to hCNTFRα

In vitro binding assay of bCNTF to hCNTFRα was carried out to assess functionality of the soluble recombinant hCNTF produced in *E. coli*. Biotinylated hCNTF bound hCNTFRα with an EC_{50} value of 36 nM (Figure 10), in agreement with a previous report (Saggio *et al.* 1994).

5 CONCLUSIONS

We report here successful expression of high yielding soluble hCNTF in *E. coli*. Our strategy of codon optimization and factorial screening of expression conditions with nine soluble tags identified several conditions for soluble hCNTF expression. Soluble expression and yield achieved was significantly higher (8 – 9 fold) and (10 – 20 fold), respectively as compared to earlier studies. Soluble expression of 6-His-hCNTF in all the conditions tested hint to the fact that codon optimization was a sufficient criteria for soluble expression in *E. coli*. Our strategy to combine codon optimization with factorial screening might be useful to produce proteins of pharmaceutical relevance in *E. coli*.

6 REFERENCES

Adler R, Landa KB, Manthorpe M and Varon S: Cholinergic neuronotrophic factors: Intaocular distribution of trophic activity for ciliary neurons. Science 204: 1434-1436 (1979).

Aggarwal S: What's fueling the biotech engine – 2011 to 2012. Nat Biotechnol 30: 1191-1197 (2012).

Arnau J, Lauritzen C, Petersen GE and Pedersen J: Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expres Purif 48: 1-13 (2006).

Bandranayake AD and Almo SC: Recent advances in mammalian protein production. FEBS Lett 588: 253-260 (2014).

Baneyx F and Mujacic M: Recombinant protein folding and misfolding in *Escherichia coli*. Nat Biotechnol 22: 1399-1408 (2004).

Berrow NS, Büssow K, Coutard B, Diprose J, Ekberg M, Folkers GE, Levy N, Lieu V, Owens RJ, Peleg Y, Pinaglia C, Quevillon-Cheruel S, Salim L, Scheich C, Vincentelli R and Busso D: Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. Acta Crystallogr D 62: 1218-1226 (2006).

Better M, Chang CP, Robinson RR and Horwitz AH: *Escherichia coli* secretion of an active chimeric antibody fragment. Sci 240: 1041-1043 (1988).

Bird LE: High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. Methods 55: 29-37 (2011).

Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U and Gileadi O: Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. Protein Expres Purif 59: 94-102 (2008).

Calderone TL, Stevens RD and Oas TG: High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli.* J Mol Bio 262: 407-412 (1996).

Chao MV: Neurotrophins and their receptors: A convergence point for many signaling pathways. Nat Rev Neurosci 4: 299-309 (2003).

Cognet I, Guilhot F, Chevalier S, Guay-Giroux A, Bert A, Elson GCA, Gascan H and Gauchat J-F: Expression of biologically active mouse ciliary neurotrophic factor (CNTF) and soluble CNTFRα in *Escherichia coli* and characterization of their functional specificities. Eur Cytokine Netw 15: 255-262 (2004).

Davis S, Aldrich TH, Valenzuela DM, Wong V, Furth ME, Squinto SP and Yancopoulos GD: The receptor for Ciliary neurotrophic factor. Science 253: 59-63 (1991).

Davis S, Aldrich TH, Ip NY, Stahl N, Scherer S, Farruggella T, DiStefano PS, Curtis R, Panayotatos N, Gascan H, Chevalier S and Yancopoulos GD: Released form of CNTFRα component as a soluble mediator of CNTF response. Science 259: 1736-1739 (1993).

Davis S, Aldrich TH, Stahl N, Pan L, Taga T, Kishimoto T, Ip NY and Yancopoulos GD: LIFRβ and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor. Science 260: 1805-1808 (1993).

Dutta R, McDonough J, Chang A, Swamy L, Kidd GJ, Rudick R, Mirnics K and Trapp BD: Activation of the Ciliary neurotrophic factr (CNTF) signaling pathway in cortical neurons of multiple sclerosis patients. Brain 130: 2566-2576 (2007).

Esposito D and Chatterjee DK: Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotech 17: 353-358 (2006).

Ettinger MP, Littlejohn TW, Schwartz SL, Weiss SR, McIlwain HH, Heymsfield SB, Bray GA, Roberts WG, Heyman ER, Stambler N, Heshka S, Vicary C and Guler H-P: Recombinant variant of Ciliary neurotrophic factor for weight loss in obese adults. A randomized, dose-ranging study. JAMA 289: 1826-1832 (2003).

Gengross TU: Advances in the production of human therapeutic proteins in yeast s and filamentous fungi. Nat Biotechnol 22: 1409-1414 (2004).

Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K and Riggs AD: Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Biochem 76: 106-110 (1979).

Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, won Park H, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim S-H, Rao Z, Shi Y, Terwilliger TC, Kim C-Y, Hung L-W, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi1 T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schütz A, Heinemann U, Yokoyama S, Büssow K, Gunsalus KC: Protein production and purification. Nat Methods 5: 135-146 (2008).

Gvritishvili AG, Leung KH and Tombran-Tink J: Codon preference optimization increases heterologous PEDF expression. PLoS One 5: 1-13 (2010).

Hammarström M, Hellgren N, van den Berg S, Berglund H and Härd T: Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. Protein Sci 11: 313-321 (2002).

Huang J, Cao L, Guo W, Yuan R, Jia Z and Huang K: Enhanced soluble expression of recombinant *Flavobacterium heparinum* heparinase I in *Escherichia coli* by fusing it with various soluble partners. Protein Expres Purif 83: 169-176 (2012).

Jarvis DL: Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production. Virology 310: 1-7 (2003).

Kallen K-J, Grötzinger J, Lelièvre E, Vollmer P, Aasland D, Renné C, Müllberg J, Meyer zum Büschenfelde K-H, Gascan H and Rose-John S: Receptor recognition sites of cytokines are organized as exchangeable modules: Transfer of the leukemia inhibitory factor receptor-binding site from Ciliary neurotrophic factor to interleukin-6. J Biol Chem 274: 11859-11867 (1999).

Kane JF: Effects of rare codon clusters on high-level expression of heterologous protein in *Escherichia coli*. Curr Opin Biotechnol 6: 494-500 (1995).

Kudla G, Murray AW, Tollervey D and Plotkin JB: Codin-sequence determinants of gene expression in *Escherichia coli.* 324: 255-258 (2009).

Laemmli UK: Cleavage of structural proteins during the assembly of the head of Bacteriophage T4. Nature 227: 680-685 (1970).

Li Y, Tao W, Luo L, Huang D, Kauper K, Stabila P, LaVail MM, Laties AM and Wen R: CNTF induces regeneration of cone outer segments in rat model of retinal degeneration. PLos One 5: 1-7 (2010).

Masiakowski P, Liu H, Radziejewski C, Lottspeich F, Oberthuer W, Wong V, Lindsay RM, Furth ME and Panayotatos N: Recombinant human and rat Ciliary neurotrophic factor from *E. coli*. J Neurochem 57: 1003-1012 (1991).

Matthews BW: The structure of *E. coli* β-galactosidase. C.R. Biologies 328: 549-556 (2005).

McDonald JR, Ko C, Mismer D, Smith DJ and Collins F: Expression and characterization of recombinant human ciliary neurotrophic factor from *Escherichia coli*. Biochim Biophys Acta (BBA) – Gene Structure and Expression 1090: 70-80 (1991).

McDonald NQ, Panayotatos N and Hendrickson WA: Crystal structure of dimeric human ciliary neurotrophic factor determined by MAD phasing. EMBO J 14: 2689-2699 (1995).

Nayak AK: Advances in therapeutic protein production and delivery. Int J Pharm Pharm Sci 2: 1-5 (2010).

Negro A, Corona G, Bigon E, Martini I, Grandi C, Skaper SD and Callegaro L: Synthesis, purification, and characterization of human ciliary neuronotrophic factor from *E. coli*. J Neurosci Res 29: 251-260 (1991).

Olson KC, Fenno J, Lin N, Harkins RN, Snider C, Kohr WH, Ross MJ, Fodge D, Prender G and Stebbing N: Purified human growth hormone from *E. coli* is biologically active. Nature 293: 408-411 (1981).

Ott V, Fasshauer M, Dalski A, Klein HH and Klein J: Direct effects of ciliary neurotrophic factor on brown adipocytes: evidence for a role in peripheral regulation of energy homeostasis. J Endocrinol 173: R1-R8 (2002).

Panayotatos N, Radziejewska E, Acheson A, Somogyi R, Thadanit A, Hendrickson WA and McDonald NQ: Localization of functional receptor epitopes on the surface of Ciliary neurotrophic factor indicates a conserved, function-related epitope topography among helical cytokines. J Biol Chem 270: 14007-14014 (1995).

Peti W and Page R: Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. Protein Expres Purif 51: 1-10 (2007).

Richardson PM: Ciliary neurotrophic factor: A Review. Pharmac Ther 63: 187-198 (1994).

Saggio I, Paonessa G, Gloaguen I, Graziani R, De Serio A, Laufer R: Nonradioactive receptor binding assay for ciliary neurotrophic factor. Anal Biochem 221: 387-391 (1994).

Sahdev S, Khattar SK and Saini KS: Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem 307: 249-264 (2008).

Schuster B, Kovaleva M, Sun Y, Regenhard P, Matthews V, Grötzinger J, Rose-John S and Kallen K-L: Signaling of human Ciliary neurotrophic factor (CNTF) revisited: The interleukin-6 receptor can serve as an α-receptor for CNTF. J Biol Chem 278: 9528-9535 (2003).

Sieving PA, Caruso RC, Tao W, Coleman HR, Thompson DJS, Fullmer KR and Bush RA: Ciliary neurotrophic factor (CNTF) for human retinal degeneration: Phase I trial of CNTF delivered by encapsulated cell intraocular implants. PNAS 103: 3896-3901 (2006).

Skerra A and Plückthun A: Assembly of a functional Immunoglobulin F_v Fragment in *Escherichia coli*. Sci 240: 1038-1041 (1988).

Sleeman MW, Anderson KD, Lambert PD, Yancopoulos GD and Wiegand SJ: The ciliary neurotrophic factor and its receptor, CNTFRα. Pharm Acta Helv 74: 265-272 (2000).

Studier FW and Moffatt BA: Use of bacteriophage T7 RNA Polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189: 113-130 (1986),

Sørensen HP and Mortensen KK: Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115: 113-128 (2005).

Tworak A, Podkowiński J and Figlerowicz M: High throughput protein production. BTA – JBCBB 92: 119-131 (2011).

Varon S, Manthorpe M and Adler R: Cholinergic neuronotrophic factors: I. Survival, neurite outgrowth and choline acetyltransferase activity in monolayer cultures from chick embryo ciliary ganglia. Brain Res 173: 29-45 (1979).

Vasina JA and Baneyx F: Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli* cspA and tac promoter systems. Protein Expr Purif 2: 211-218 (1997).

Wagener E-M, Aurich M, Aparicio-Siegmund S, Floss DM, Garbers C, Bresing K, Rabe B, Schwanbeck R, Grötzinger J, Rose-John S and Scheller J: The amino acid exchange R28E in Ciliary neurotrophic factor (CNTF) abrogates IL-6R- but retains CNTFR-dependent signaling via gp130/LIFR. J Biol Chem 289: (2014).

Walsh G: Biopharmaceutical benchmarks 2010. Nat Biotechnol 28: 917-924 (2010).

Waugh DS: Making the most of affinity tags. Trends Biotechnol 23: 316-320 (2005).

Welch M, Villalobos A, Gustafsson C and Minshull Y: You're in a googol: optimizing genes for protein expression. J R Soc Interface 6 suppl 4: S467-476 (2009).

Wen R, Song Y, Kjellstrom S, Tanikawa A, Liu Y, Zhao L, Bush RA, Laties AM and Sieving PA: Regulation of rod phototransduction machinery by ciliary neurotrophic factor. J Neurosci 26: 13523-13530 (2006).

Wen R, Tao W, Li Y and Sieving PA: CNTF and retina. Prog Retin Eye Res 31: 136-151 (2012).

Winearls CG: Recombinant human erythropoietin: 10 years of clinical experience. Nephrol Dial Transplant 13 (Suppl 2): 3-8 (1998).

Woestenek EA, Hammarström M, van den Berg S, Härd T and Berglund H: His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors. J Sctructu Funct Genomics 5: 217-229 (2004).

Yin J, Li G, Ren X and Herrler G: Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. J Biotechnol 127: 335-347 (2007).

APPENDIX 1 – Preparation of *E. coli* competent cells

Competent *E. coli* bacterial cells were prepared according to a protocol adopted from an article by Bird (Bird 2011). NEB Turbo competent *E. coli* and NEB 5-alpha colonies were grown on a lysogeny broth (LB) agar plate for approximately 9 hours at 37°C. A single colony was subsequently used to inoculate 5 ml of 2X LB culture medium grown overnight at 37°C with shaking at 220 rpm. BL21(DE3)pLysS and Rosetta 2 cells were grown on LB agar plates supplemented with chloramphenicol overnight at 37°C. Single colonies were used to inoculate 5 ml of 2x LB culture medium supplemented with chloramphenicol and subsequently incubated for 14 hours at 37°C with shaking at 220 rpm.

1 ml of the overnight culture was used to inoculate 100 ml of 2X LB media (1:100). Culture media for growing BL21(DE3)pLysS and Rosetta 2 strains were supplemented with chloramphenicol. Cultures were grown at 37°C with shaking at 220 rpm until the OD_{600} had reached a value 0.25 - 0.3. The cultures were then cooled on ice bath for 5 min, divided into two 50 ml fractions and centrifuged at 4000 rpm at 4°C for 7 min. The supernatants were discarded. The pellets were gently re-suspended in 5 ml of Transformation buffer I (Tfb I, Table 8) and centrifuged at 4000 rpm at 4°C for 7 min. After discarding the supernatants, the pellets were re-suspended in 10 ml of Tfb I and incubated on ice for 5 min. After incubation, cell suspensions were centrifuged at 4000 rpm at 4°C for 7 min. The supernatants were discarded and the pellets gently re-suspended in 1 ml of Transformation buffer II (Tfb II, Table 9) and incubated on ice for 15 min. The suspensions were flash frozen in liquid nitrogen and stored as 50 µl aliquots at -80°C.

Table 8: Composition of Tfb I

Compound	Concentration
Potassium acetate (KOAc)	30 mM
Rubidium chloride (RbCl)	100 mM
Calcium chloride ($CaCl_2$)	10 mM
Manganese chloride ($MnCl_2$)	50 mM
Hexammine cobalt chloride	3 mM
Glycerol	15 %
pH adjusted to 5.8 with 0.2 M acetic acid and sterilized by filtration (0.22 µm)	

Table 9: Composition of Tfb II

Compound	Concentration
MOPS	10 mM
Calcium chloride ($CaCl_2$)	75 mM
Rubidium chloride ($RbCl_2$)	10 mM
Glycerol	15 %
pH adjusted to 6.5 with KOH and sterilized by filtration (0.22 µm)	

APPENDIX 2 – Bacterial transformation

Frozen *E. coli* competent cells were thawed on ice and as soon as the last trace of ice disappeared, 1 - 10 ng of DNA was added and gently mixed with the cells. The cells were incubated on ice for 40 min. After incubation, a heat-shock was given to the cells at 42°C for 30 seconds and immediately transferred to an ice bath to incubate for 5 min. After cooling, 800 µl of SOC (Super Optimal broth with Catabolite repression) medium was added to the cells and incubated at 37°C with shaking at 220 rpm for 60 min. After incubation, 200 µl of cell culture was pipetted and spread evenly onto a LB plate (supplemented with appropriate antibiotics) and incubated overnight (12 - 14 h) at 37°C. To increase the efficiency of colony formation, some cell cultures were centrifuged at 4°C at 4000 rpm for 10 minutes and 600 µl of the supernatant was discarded. The cells were re-suspended in 200 µl of media and plated to grow overnight (12 - 14 h) at 37°C.

APPENDIX 3 – Plasmid DNA purification

Single separate transformed cell colonies were used to inoculate 5 ml of LB medium (supplemented with ampicillin) and grown overnight at 37°C with shaking at 220 rpm. Plasmid DNA purifications were performed using Wizard® *Plus* SV Minipreps DNA Purification System (Promega).

After the overnight cultures had reached an OD_{600} value between 2 and 4, they were centrifuged at 4000 rpm (4°C) for 10 min. After centrifugation, the supernatants were discarded and the pellets re-suspended in 250 µl of Cell Resuspension Buffer and transferred to sterile 1.5 ml eppendorf tubes. 250 µl of Cell Lysis Solution was added, gently mixed and incubated at RT for 5 min. After incubation, 10 µl of Alkaline Protease Solution was added, gently mixed and incubated at RT for 5 min. To neutralize the protease activity, 350 µl of Neutralization buffer was added, gently mixed and centrifuged at 14,000 g (RT) for 10 minutes.

The supernatants were carefully transferred (without disturbing the debris from cell lysis) to spin columns and centrifuged at max speed (RT) for 1 min. The flow-through was discarded. 750 µl of Column Wash Solution was added to each spin column and centrifuged at max speed (RT) for 1 min. After discarding the flow-through, the washing step was repeated with 250 µl of Column Wash Solution. After centrifugation, the spin columns were transferred to sterile 1.5 ml eppendorf tubes and centrifuged at max speed for 2 min. The spin columns were then transferred to sterile 1.5 ml eppendorf tubes and 70 µl of pre-warmed Nuclease Free Water (NFW) was added to each spin column and allowed to stand for 2 min. The spin columns were then centrifuged at max speed for 2 min. To increase the extraction efficiency, the filtrate was applied to the spin column followed by centrifugation. The filtrate was recovered and stored at -20°C.

Plasmids were purified with a yield ranging roughly from 50 ng/µl to 80 ng/µl. To assess the purity and size of the isolated plasmids, gel electrophoresis was performed on a 1x TAE, 0.7 % agarose (Bionordika) gel containing 0.5 mg/ml of ethidium bromide. DNA samples were prepared to contain approx. 25 ng of purified plasmid DNA in the loading volumes of 5 µl.

APPENDIX 4 – Plasmid DNA digestion

The purified pOPIN plasmids were digested with HindIII-HF™ (New England BioLabs) and KpnI-HF™ (New England BioLabs) restriction enzymes. Depending on the vector, 3 µg to 3.5 µg of each plasmid DNA was used in a total digestion reaction volume of 100 µl. The amount of restriction enzymes was adjusted to 10 units per 1 µg of vector DNA. The reaction mixture was incubated at 37°C for 3.5 h. The reaction products were analyzed by gel electrophoresis (1x TAE, 0.7 % agarose, Bionordika) containing 0.5 mg/ml of ethidium bromide. DNA samples were prepared with 1 µl of reaction mixture diluted 6-fold. 5µl of each sample were loaded to the gel.

APPENDIX 5 – Gel extraction and purification

Linearized vectors and PCR-amplified inserts were purified using a NucleoSpin ® Gel and PCR Clean-up (Macherey-Nagel) kit. After the electrophoretic runs, agarose slices containing bands of DNA were excised and weighed. Buffer NTI was added to each sample and subsequently allowed to incubate at 50°C for 10 min with intermittent vortex to dissolve the gel slices. After dissolution, 700 µl of sample was loaded to each clean-up column and centrifuged at 11,000 g for 30 s after which the flow-throughs were discarded. Remaining samples were loaded onto the clean-up columns and the centrifugation step repeated.

700 µl of Buffer NT3 was added to each clean-up column as a wash. The samples were subsequently centrifuged at 11,000 g for 30 s and the flow-through discarded. The washing step was repeated. The samples were incubated at 70°C for 5 min and subsequently centrifuged at 11,000 g for 1 min to remove the Buffer NT3 and to dry the silica membrane. The flow-throughs were subsequently discarded.

The clean-up columns were placed into new eppendorf tubes. 25 µl of preheated 0.5 X Buffer NE was added to each column and incubated at 70°C for 5 min and subsequently centrifuged at 11,000 g for 1 min. This was repeated twice with 25 µl of fresh 0.5 X Buffer NE to increase the recovery yield. The concentration of the isolated DNA sample was measured by UV spectrophotometer.

APPENDIX 6 – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

1 Casting the gels

For sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), 12% gels were cast. The resolving gels were casted first, consisting of 30% acrylamide/Bis (37.5:1, Bio-Rad), 1.5 M Tris-HCl pH 8.8, 10% SDS (Bio-Rad), MilliQ water, 10% APS (Bio-Rad) and N,N,N',N'-tetramethylethylenediamine (TEMED, Sigma-Aldrich). After the resolving gels had polymerized (30 minutes), the 4 % stacking gels were casted on top of it, consisting of 30% acrylamide/Bis solution, 0.5 M Tris-HCl pH 6.8, 10% SDS, MilliQ water, APS and TEMED. After 60 min, the gels were stored at +4º C.

2 Sample buffers

SDS-PAGE sample buffer was prepared as a 6 X concentrate containing 62.5 mM Tris-HCl pH 6.8, 2 % SDS, 3 % glycerol, 5 % β-mercaptoethanol, 0.0125 % bromophenol blue and stored at -20°C.

3 Running the gels

For SDS-PAGE, PageRuler™ Plus Prestained Protein Ladder (Thermo Scientific) and/or Protein Marker, Broad Range, (2 - 212 kDa, New England Biolabs) were used. The electrophoretic runs were carried out in discontinuous buffer system of Laemmli (Laemmli *et al*. 1970) at 180 V and 160 V, for the stacking and resolving layers, respectively. The samples were run till the dye front reached the bottom of the gel and the markers were visibly well resolved.

4 6X His tag gel staining

Staining of the gels of the small-scale expression screen was carried out using Pierce™ 6X His Protein Tag Stain Reagent Kit (Thermo Scientific) at RT. After electrophoresis, the gels were washed by gently agitating with Milli-Q water for 25 minutes while changing the wash every 5 minutes. After washing, the gels were covered with 6xHis Tag Stain and gently agitated for 5 minutes. After removing the staining solution, the gels were subjected to another washing step with Milli-Q water followed by addition of the 6xHis Protein Tag Developer. The gel was gently agitated for 15 minutes. The gels were finally washed with Milli-Q water for 25 minutes; the wash was changed every 5 minutes.

5 Documentation

SDS-PAGE gels were scanned and imaged (302 nm) with a ChemiDoc™ XRS+ Molecular Imager® system (Bio-Rad)/ Image Lab™ software (Bio-Rad).

All gels were stained with Coomassie brilliant blue R-250 solution (Bio-Rad) at RT by gently shaking for 60 min followed by destaining with Milli-Q for 120 min while gently shaking and subsequently documented.

APPENDIX 7 – Plasmid DNA purification and digestion

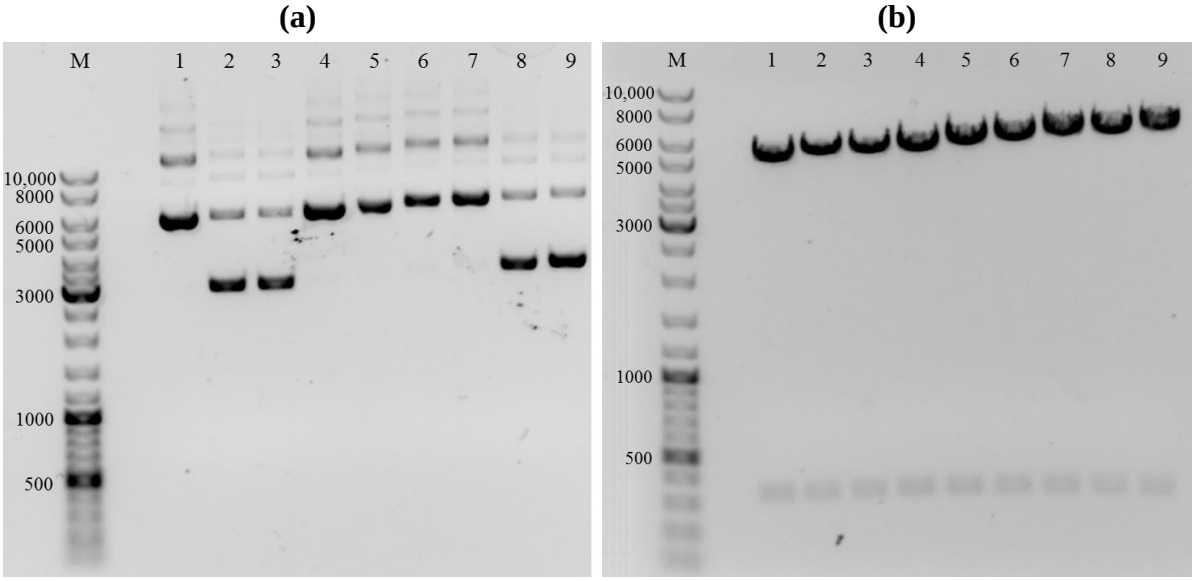


Figure 11. (a) Amplified and purified pOPIN vectors; (b) HindIII-HF and KpnI-HF linearized pOPIN vectors. M denotes DNA marker (in bp); lanes 1-9 denote the amplified and Miniprep-purified plasmids pOPINF, pOPINS3C, pOPINTRX, pOPINMSYB, pOPINJ, pOPINHALO7, pOPINM, pOPINTF and pOPINNusA, respectively.

All nine plasmid vectors were amplified and purified with a high recovery. The multiple bands visible for each plasmid (Figure 11a) represent different DNA conformations such as, circular and supercoiled DNA.

Linearized plasmids (Figure 11b) were purified with gel extraction.

APPENDIX 8 – 6X His-stained SDS-PAGE for soluble affinity purified hCNTF constructs

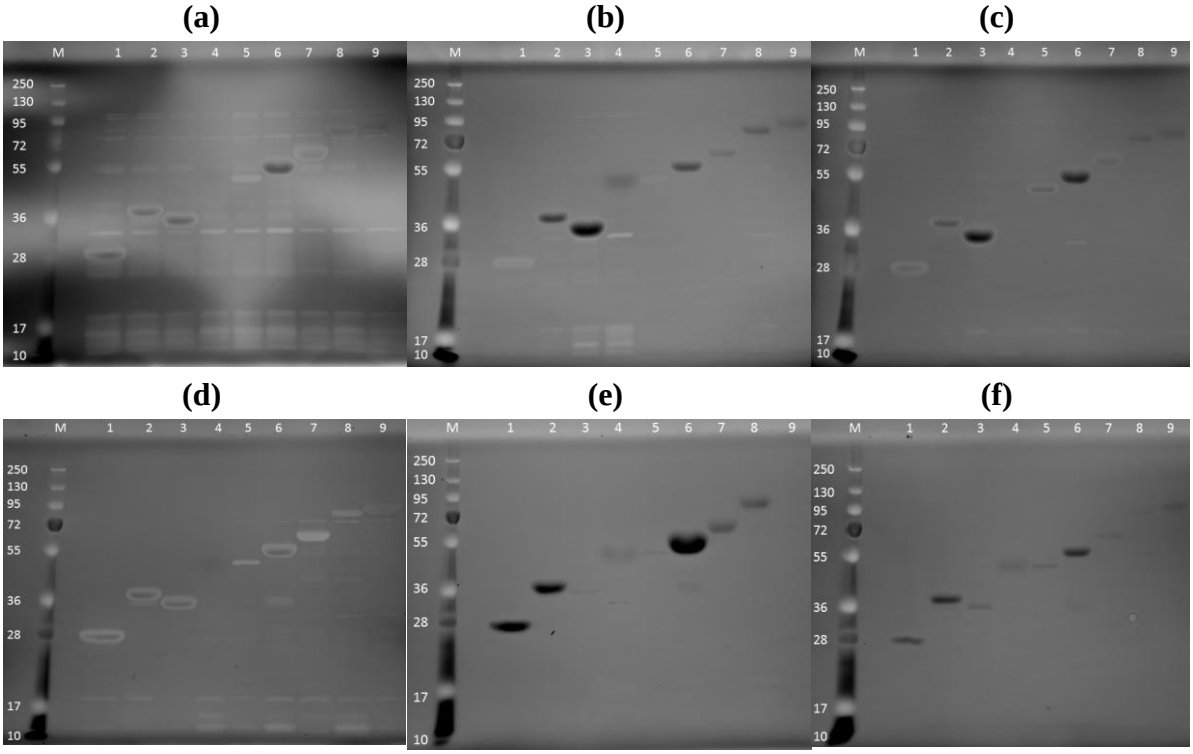


Figure 12. (a) BL21/20°C/PPB; (b) BL21/25°C/TBONEX; (c) BL21/37°C/PPB; (d) Rosetta 2/20°C/PPB; (e) Rosetta 2/25°C/TBONEX; (f) Rosetta 2/37°C/PPB. Lanes M and 1 - 9 denote the protein marker and soluble Ni-NTA purified hCNTF using expression constructs listed in Table 7, respectively.

APPENDIX 9 – SDS-PAGE analysis for hCNTF constructs estimation of ratio of soluble to total protein

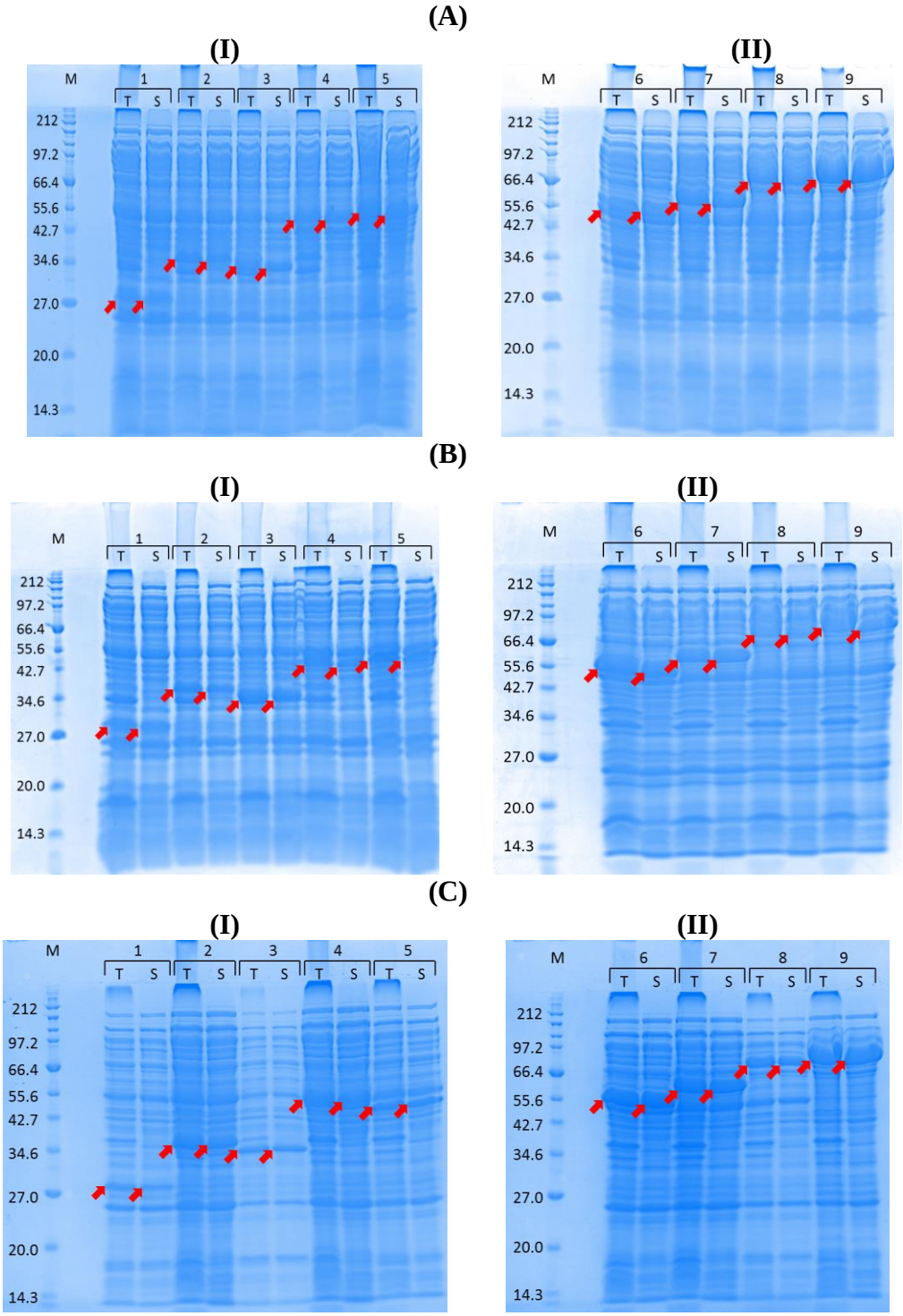


Figure 13. SDS-PAGE analysis of fusion hCNTF constructs in total (T) cell lysate and soluble (S) fraction. (A) BL21/20°C/PPB; (B) BL21/25°C/TBONEX; (C) BL21/37°C/PPB. Lane M and lane groupings 1 - 9 denote the protein marker and hCNTF constructs.

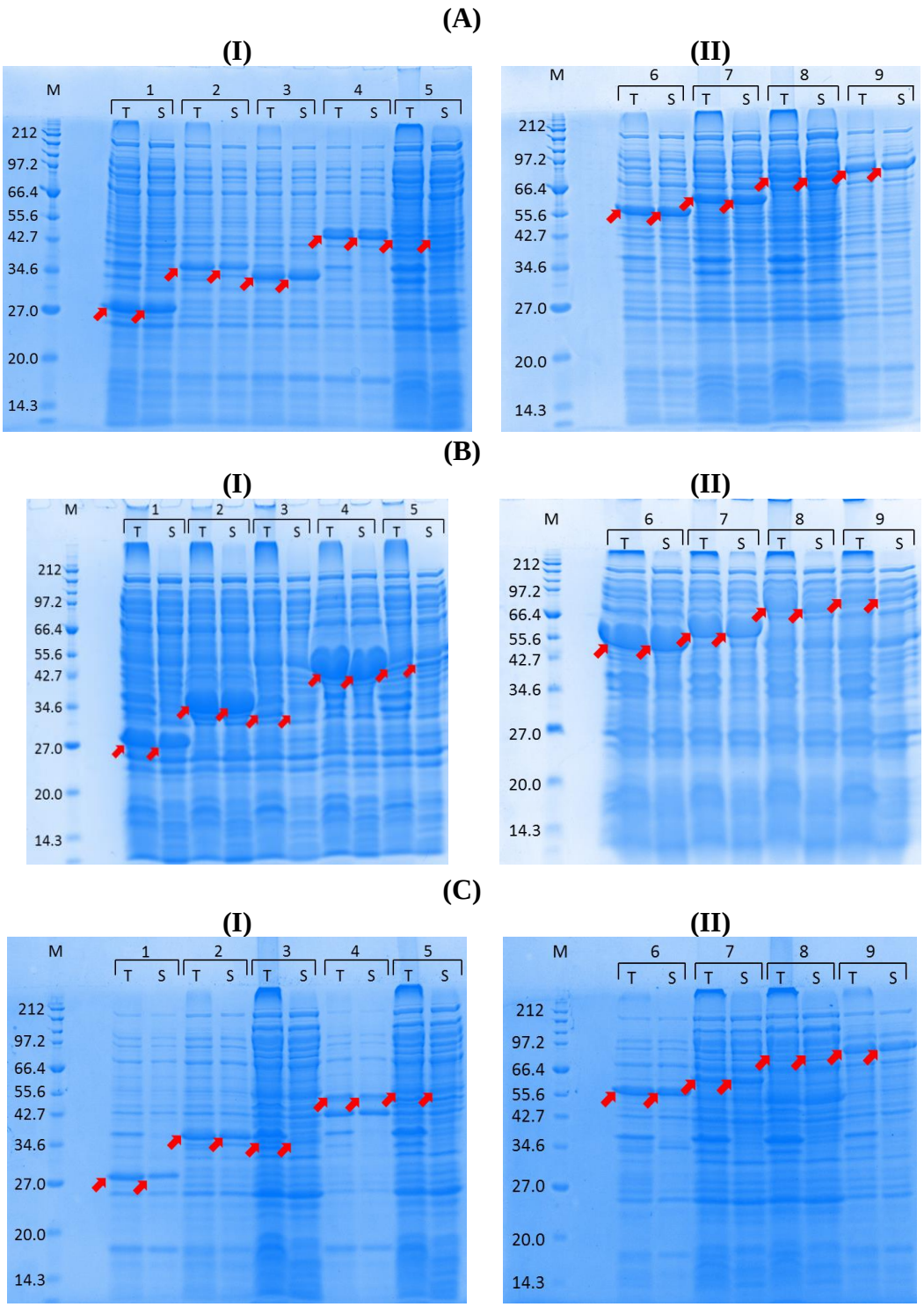


Figure 14. SDS-PAGE analysis of fusion hCNTF constructs in total (T) cell lysate and soluble (S) fraction. (A) Rosetta 2/20°C/PPB; (B) Rosetta 2/25°C/TBONEX; (C) Rosetta 2/37°C/PPB. Lane M and lane groupings 1-9 denote the protein marker and hCNTF constructs.

APPENDIX 10 – Estimation of resin amount for optimal IMAC purification

To estimate the amount of Protino® nickel iminodiacetic acid (Ni-IDA) resin (Macherey-Nagel) necessary for optimal affinity purification, the following experiment was carried out.

The thawed cell pellet (0.41 g) was re-suspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 10 mM Imidazole, 1 mM DTT and PIC). The suspension was frozen at -80°C and subsequently thawed at 37°C. This freeze-thaw cycle was repeated twice. Lysonase was then added and the suspension was incubated at RT for 60 min with gentle agitation. The lysate was subsequently centrifuged at 14,500 g at 4°C for 50 minutes.

Approximately 42 mg of Ni-IDA was weighed into a microcentrifuge tube and suspended in 8 bed volumes[7] of NP-buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0; 10 mM Imidazole, 1 mM DTT, PIC). The suspension was mixed gently at RT for 10 minutes. The resin was subsequently sedimented by centrifugation at 500 g for 3 min. The supernatant was removed and discarded. The previous step was repeated once (washing) and the resin was subsequently re-suspended in 8 bed volumes of NP-buffer. The total volume of suspension was estimated and divided into microcentrifuge tubes in fractions 0.0303, 0.0606, 0.1515, 0.3030 and 0.4545 of the total suspension. The fractions were centrifuged at 500 g and the supernatants discarded.

Supernatant from the cell lysate was divided into five equal fractions; each added to the different resin amounts and allowed to incubate at 4°C for 90 min with gentle agitation. After incubation, the suspensions were centrifuged at 500 g for 3 min and the supernatants removed.

[7] 1 g of resin corresponds to 2 ml bed volume (Macherey-Nagel)

Washings were carried out with 4 bed volumes of NPI-10-Tween (supplemented with 1mM DTT and PIC) by gently mixing for 10 min and subsequently centrifuged at 500 g for 3 minutes. Supernatants were discarded. This step was repeated using appropriate volumes of NPI-25-Tween (supplemented with 1 mM DTT and PIC).

Elution of bound proteins was carried out by adding 3 bed volumes of elution buffer NPI-250-Tween (supplemented with 1 mM DTT and PIC). The suspensions were incubated at RT for 20 min with gentle shaking followed by centrifugation at 500 g for 3 min. The elution step was repeated twice and supernatants from each elution were reserved for gel analysis.

Results

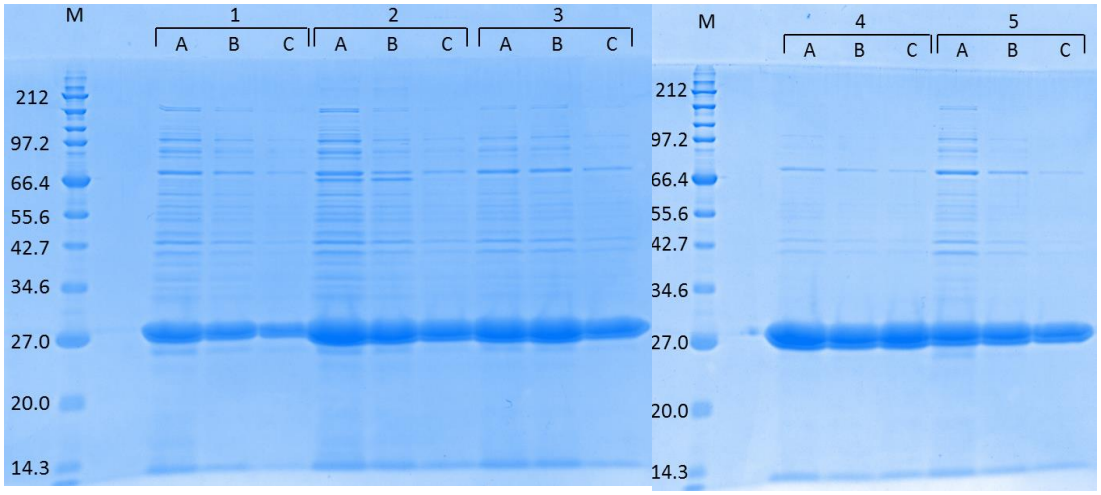


Figure 15. SDS-PAGE analysis of small-scale IMAC purification with Ni-IDA resin. Lane M and lane groupings 1 - 5 denote the protein marker and Ni-IDA resin amounts of 1.25 mg, 2.5 mg, 6.25 mg, 12.5 mg and 18.75 mg, respectively. A, B and C denote elution fractions.

The amount of Ni-IDA resin to be used in large scale IMAC purification was decided based on visual inspection of SDS-PAGE analysis (Figure 15) for yield and purity.