Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Stable mammalian producer cell lines for structural biology

Konrad Büssow

The mammalian cell lines HEK293 and CHO have become important expression hosts in structural biology. Generating stable mammalian cell lines remains essential for studying the function and structure of recombinant proteins, despite the emergence of highly efficient transient transfection protocols. Production with stable cell lines can be scaled up easily and high volumetric product yield can be achieved. Protein structure reports of the past two years that used stable cell lines were surveyed for this review. Well-established techniques and novel approaches for generating stable cell lines and stable cell pools are presented, including cell sorting, site-specific recombination, transposons, the Lentivirus system and phage integrases. Host cell line optimization by endoglycosidase overexpression and sequence-specific genome engineering is highlighted.

Address
Helmholtz Centre for Infection Research, Structure and Function of Proteins, Inhoffenstr. 7, 38124 Braunschweig, Germany

Corresponding author: Büssow, Konrad (konrad.buessow@helmholtz-hzi.de)

Current Opinion in Structural Biology 2015, **32**:81–90

This review comes from a themed issue on **New constructs and expressions of proteins**

Edited by **Imre Berger** and **Roslyn M Bill**

For a complete overview see the Issue and the Editorial

Available online 23rd March 2015

http://dx.doi.org/10.1016/j.sbi.2015.03.002

Introduction

Generating pure, soluble and homogeneous protein is a major step in the overall process of protein structure determination. The choice of the expression system has a great influence on the quality and quantity of the produced recombinant protein. The Human Embryonic Kidney cell line HEK293 and the Chinese Hamster Ovary cell line CHO are excellent host cells for robust secretion of mammalian proteins with appropriate post-translational modifications [1]. These cell lines are used for production of secreted mammalian and viral proteins and soluble ectodomains of transmembrane proteins, but also for complete membrane proteins (Table 1). Cytosolic proteins and complexes can be produced with stable mammalian cell lines, but the yield is usually low compared with other expression systems. The long-awaited crystal structure of the cytosolic mTORC complex was obtained upon overexpression of its subunits in a stable HEK293 cell line [2••].

For recombinant protein overexpression, an expression vector for the protein of interest is transferred into the cell's nucleus by transfection. In a transient transfection experiment, the protein of interest is harvested a few days later. Alternatively, a stable cell line is generated from transfected cells that have integrated the vector into their genome. Stable cell lines overexpress the target protein uniformly and indefinitely. Protein production with a stable cell line is therefore reproducible and can be scaled up easily. Recent protein structure reports using HEK293 or CHO cell lines were surveyed for this review and it was found that transient transfection and stable cell lines were used with around the same frequency. New and improved technologies for generating stable cell lines are expected to increase their use in the future.

Establishing stable cell lines requires substantial time and effort in comparison to transient transfection processes. An expression vector with the gene of interest has to be inserted into the host cell genome. Using standard methods, the efficiency of genome integration is low. Moreover, only very few cells will integrate the vector into a highly transcribed region and will produce sufficient amounts of recombinant protein. Even then, transgene expression is often silenced upon long term cell culture. Isolating and characterizing a large number of clones is therefore required, which can take several months of laboratory work. Fortunately, stable cell line technology is improving rapidly on the levels of host cell line, integration process and selection of high-producer cells.

Crystal structures of proteins produced by stable cell lines reported during the past two years are listed in Table 1. The table provides an overview of the host cell lines and experimental techniques currently used in structural biology. The most common approach involves transfection with vectors carrying a selectable marker for random chromosomal integration, followed by isolation and screening of single cell clones. This process was reviewed in three recent publications [3,4,5] and protocols for establishing stable cell lines for structural biology with antibiotic selection markers have been published [6,7]. The performance of the antibiotics hygromycin B, neomycin, puromycin and Zeocin as selection markers for stable cell line development was recently compared [8].

Table 1

The Protein Data Bank (PDB) was searched for mammalian protein structures, excluding antibodies, that were released later than July 2012. Selected proteins that had been produced by stable cell lines are listed

	Crystallized protein	Host cell line[a]	Vector	Stable cell line generation[b]	PDB entry and Ref.
Cytosolic proteins	mTOR:LST8 complex	HEK293-F	pcDNA 3.1 (+)	Co-expression of mTORC subunits using different antibiotics	PDB: 4JSN [2••]
Membrane proteins	Rhodopsin	HEK293 GnTI$^-$			PDB: 4BEY [54]
	GABAA receptor	HEK293S GnTI$^-$			PDB: 4COF [55]
	5-HT3A receptor	T-REx-293	pcDNA5/TO		PDB: 4PIR [56]
Secreted proteins	Prothrombin	BHK	pNUT	Lipofection, MTX	PDB: 4HZH [57]
	Aminopeptidase N	CHO Lec3.2.8.1	pBJ5-GS	MSX	PDB: 4F5C [22]
	Integrin α_4	CHO Lec3.2.8.1	pcDNA3.1/Hygro		PDB: 4IRZ [58]
	MHC class II HLA-DQ1: antigen:TCR complex	CHO Lec3.2.8.1	pEE13.1	MSX	PDB: 4GRL [21]
	CD200 receptor	CHO Lec3.2.8.1	pEE14		PDB: 4BFE [59]
	DC-LAMP	CHO Lec3.2.8.1	pFS-sigHis	Electroporation, RMCE, G418	PDB: 4AKM [39•]
	Neuroligin-1	CHO Lec3.2.8.1	pSGHV0	Electroporation, puromycin	PDB: 3VKF [60]
	HLA-DM	CHO Lec3.2.8.1			PDB: 4FQX [61]
	ICAM-5	CHO Lec3.2.8.1		MSX	PDB: 4OI9 [20]
	Tyrosine kinase receptor RET	CHO Lec8	pcDNA3	Lipofection (Effectene), G418	PDB: 4UX8 [62]
	Insulin receptor	CHO Lec8	pEE14	Lipofection (Lipofectamine 2000), MSX	PDB: 3W11 [23]
	Cholesteryl ester transfer protein	CHO DG44			PDB: 4F2A [63]
	Receptor tyrosine-kinase erbB-4	CHO Lec1	pSGHV0	Lipofection (Fugene), pcDNA3.1 co-transfection, G418	PDB: 3U7U [64]
	Folate receptor α	CHO duk$^-$ (kifu.)	pSGHV0	Electroporation, MTX	PDB: 4KM6 [14]
	Acetylcholinesterase	CHO-K1	pGS	jetPEI, MSX	PDB: 4BDT [24]
	LIMP-2	HEK293S GnTI$^-$	pNeoSec	ΦC31 integrase (stable cell pool)	PDB: 4Q4B [42]
	Glutamate receptor 4	HEK293 GnTI$^-$	pHLsec		PDB: 4GPA [65]
	Tumor antigen 5T4/WAIF1	HEK293 GnTI$^-$	pURD	ΦC31 integrase, puromycin	PDB: 4CNC [48]
	Transcobalamin-1	HEK293 GnTI$^-$		Lipofection (lipofectamine)	PDB: 4KKI [66]
	Integrin $\alpha_X\beta_2$ heterodimer	HEK293S GnTI$^-$	ET1 (pIRES2-EGFP), pEF1-puro	Ca-phosphate, FACS, puromycin, G418	PDB: 4NEH [29]
	von Willebrand factor	HEK293S GnTI$^-$	ET8	PEI, G418	PDB: 4NT5 [67]
	Aminopeptidase N	HEK293S GnTI$^-$			PDB: 4FYQ [68]
	NPP1 pyrophosphatase	HEK293S GnTI$^-$			PDB: 4B56 [69]
	Angiopoietin-2	HEK293	pcDNA3.1 hygro	Lipofection	PDB: 4JZC [70]
	Netrin-1:neogenin complex	HEK293	pcDNA3.1+		PDB: 4PLN [71]
	RNaseT2	HEK293 (kifu.)	pcDNA3.1/myc-His	Lipofection (Effectene)	PDB: 3T0O [15]
	TGF-β-receptor-3	HEK293 EBNA		Lipofection (Lipofectamin 2000), puromycin	PDB: 4AJV [72]
	Folate receptor α	HEK293	pcDNA6	blasticidin	PDB: 4LRH [73]
	IgG receptor FcRn subunit p51	HEK293-6E	pLVX	Lentivirus	PDB: 4K71 [46]
	Nectin1, 2, 3, Necl-5.	HEK293-F	pCEP4	Lipofection (Lipofectamin 2000), hygromycin B	PDB: 4FMF [74]
	Monocyte differentiation antigen CD14	HEK293-F	pDisplay	G418	PDB: 4GLP [75]
	Acetylcholinesterase	HEK293-H	pJT1 Fast Dest	Lipofection (Lipofectamin 2000), ΦC31, hygromycin B	PDB: 4EY4 [49]
	Platelet glycoprotein Ib	HEK293T	ET-6		PDB: 4C2A [76]
	Galactocerebrosidase	HEK293T	pSecTag2B		PDB: 4CCC [77]
	O-fucosyltransferase 2	HEK293T (kifu.)	pSecTagB	jetPEI, zeocin	PDB: 4AP5 [13]

[a] kifu., the GnTI inhibitor kifunensine was used during protein production.

[b] Cells were transfected by lipofection, electroporation or with polyethylenimine (PEI, jetPEI). Recombinant cells were selected with antibiotics (blasticidin, G418, hygromycin B, puromycin, Zeocin), with the glutamine synthase inhibitor methionine sulfoximine (MSX) or with the DHFR inhibitor methotrexate (MTX). FACS, GFP-positive cells were isolated by cell sorting.

Engineering of host cell lines for stable cell line generation

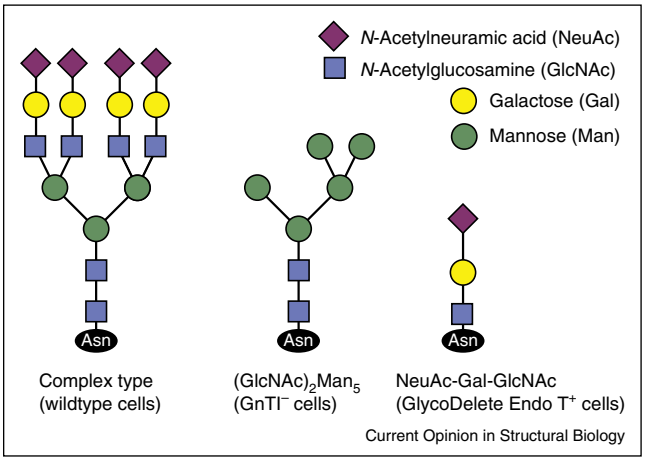
Glycosylation-deficient cell lines

Most secreted mammalian proteins are glycosylated, which can interfere with crystallization [9••]. Glycosylation sites that are not required for folding or secretion are therefore removed by mutagenesis [9••]. The processing of N-linked glycans from the high-mannose type to the larger, complex type requires the enzyme N-acetylglucosaminyl-transferase I (GnTI, MGAT1) (Figure 1). The GnTI-deficient host cell lines HEK293S GnTI$^-$ and CHO Lec3.2.8.1 produce glycoproteins with high-mannose type glycans. These glycans can be readily trimmed further to a single GlcNAc sugar unit by endoglycosidase treatment [10,11]. The HEK293S GnTI$^-$ cell line [10] (ATCC CRL-3022) is currently the most popular cell line in structural biology (Table 1). It is used for transient and stable expression. High-mannose type glycosylation is also obtained by cultivation with the GnTI inhibitor kifunensine [12]. Kifunensine was applied for crystal structure determination of the glycoproteins O-fucosyltransferase [13], folate receptor α [14] and RNaseT2 [15] (Table 1).

Intracellular endoglycosidase overexpression

Recently the 'GlycoDelete' HEK293 cell line was developed that carries a heterologous endoglycosidase in its Golgi apparatus [16••]. This resulted in robust secretion of deglycosylated glycoproteins (Figure 1). The GlycoDelete study demonstrates that protein deglycosylation is possible in the Golgi apparatus, where glycoproteins traveling the secretory pathway have already passed the quality control in the endoplasmic reticulum. Cells that secrete glycoproteins with minimal glycosylation would enable crystallization without further *in vitro* deglycosylation.

Figure 1



N-linked sugar classes produced by mammalian cell lines. Complex-type glycans are synthesized as a heterogeneous mixture of bi-antennary, tri-antennary and tetraantennary forms. Lack of N-acetylglucosaminyl-transferase I (GnTI) prevents the formation of complex sugars and results in high-mannose type protein glycosylation. High-mannose sugars consist of a variable number of mannose residues attached to a chitobiose ($GlcNAc_2$) core. Overexpression of endoglycosidase T (Endo T) in the Golgi apparatus of the GnTI$^-$ 'GlycoDelete' HEK293 cell line leads to intracellular cleavage of N-linked sugars after the first GlcNAc residue. The cleavage product is partly modified with galactose and N-acetylneuramic acid residues.

Sequence-specific genome engineering

The cell lines CHO Lec3.2.8.1 and HEK293S GnTI$^-$ were created by chemical mutagenesis, followed by selection for glycosylation deficiency. Chemical mutagenesis lacks specificity, leading to random mutations throughout the genome. Sequence-specific genome engineering represents an elegant alternative that greatly reduces unwanted mutations. Meganucleases with long recognition sequences cleave genomic DNA at rare sites and can be used to introduce gene-inactivating mutations more specifically [17]. Nucleases linked to programmable, sequence-specific DNA-binding modules, such as zinc finger nucleases (ZFN), TALE nucleases and the CRISPR/Cas9 nucleases allow for modification of arbitrary genetic loci with excellent specificity [18]. Both alleles of a gene can be inactivated with high frequency with these nucleases. Especially with CRISPR/Cas9, mammalian genome engineering has become simple, reliable and cheap [19].

Glutamine synthetase knockout cells

The glutamine synthetase (GS) gene is used widely as a selection marker for stable CHO cells. Cells overexpressing GS can be selected by inhibiting the endogenous glutamine synthetase with the inhibitor methionine sulfoximine (MSX). The GS marker and MSX-selection of stable CHO cell lines were used to produce protein for structure determination of the ICAM-5 ectodomain [20], a TCR:MHC:antigen heterotetramer [21], aminopeptidase N [22], the insulin receptor ectodomain [23] and acetylcholinesterase [24]. A novel CHO host cell was created by knocking out the GS gene and the GnTI gene with specific zinc finger nucleases [25•]. With GS-deficient CHO host cells, MSX selection of cells stably transfected with a GS vector is much more efficient than with normal cells. The selection of low-producing cell clones was largely prevented by using a GS-deficient host cell line [26]. Moreover, among MSX-selected, stably transfected cell clones, a six-fold higher proportion of top producing clones (>2 g/L) was found when a GS-deficient host cell line was used, in comparison to the original host cell line.

New cell lines developed by sequence-specific genome engineering such as GnTI and GS double knockouts are potentially useful for structural biology. Genome engineering could also be used for removing genes of

unwanted host cell proteins that are co-purified with the target protein.

Methods of stable cell line generation

GFP and cell sorting

Transgene insertion into a host cell chromosome upon transfection is a rare event and, to make things worse, most of the integrated transgenes will be inactivated by epigenetic mechanisms. Cells that received an active genome-integrated transgene therefore have to be selected by a marker. Antibiotic resistance is commonly used, but can be an unreliable reporter for high level, uniform transgene expression [27]. Green fluorescent protein (GFP) is a useful alternative [28]. Cells that stably integrate a GFP expression vector are directly identified by intracellular fluorescence and can be isolated by preparative FACS cell sorting. Cell sorting can isolate a small number of high-producer clones from millions of cells. By repeating the process at different time points, cells that express GFP stably over time can be isolated. By this method, clonal cell lines with constantly high GFP expression over several months can be obtained without applying selective pressure [27].

GFP expression can be coupled to expression of the gene of interest by constructing internal ribosome entry site (IRES)-based bicistronic vectors comprising the gene of interest and a GFP gene. IRES-based GFP co-expression in combination with cell sorting and antibiotic selection was used to generate a stable HEK293S GnTI$^-$ cell line for production of a soluble integrin $\alpha_X\beta_2$ heterodimer for structure determination [29]. In an alternative approach, the GFP gene can be excised by site-specific recombination upon clone isolation, thereby bringing a gene of interest, located downstream, under the control of the promoter driving transgene expression (Figure 2) [27,30].

Cell sorting may require optimization to maintain cell viability. In our own experience, the viability of CHO Lec3.2.8.1 cells grown in suspension can be low upon cell sorting. Shear stress was reported to cause low cell viability upon sorting of insect cell lines. Addition of Pluronic acid F-68 improved the survival of sorted insect cells [31•] and may also have a positive effect on mammalian cells.

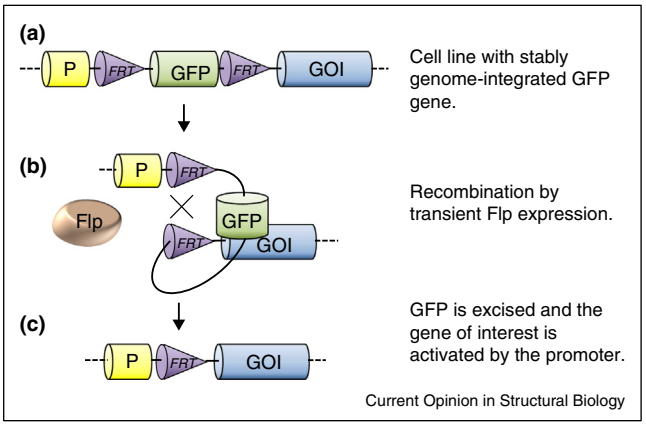
Secreted GFP fusion proteins

A secreted GFP marker was used for studying the structure of a sialyltransferase [32•]. The sialyltransferase was fused to a codon-optimized, folding-enhanced GFP version called 'superfolder' [33]. The fusion protein was secreted by a stable HEK293S GnTI$^-$ cell line with high yield (75 mg/L). The GFP tag allowed for direct protein quantification by fluorescence spectroscopy during cell line development and protein purification. It was removed by proteolysis from the sialyltransferase before crystallization.

Recombinase-mediated cassette exchange

The productivity of a stable cell line depends on the genetic locus of transgene integration. Instead of random integration, it would be desirable to target the transgene to a specific locus that allows for strong and stable transgene transcription. This can be achieved by recombinase-mediated cassette exchange (RMCE) using site-specific recombinases [34,35]. RMCE requires a 'master' cell line carrying a single copy of a reporter gene at a suitable genetic locus. By RMCE, the reporter gene is exchanged against the gene of interest (GOI). For RMCE with the

Figure 2



Replacing a GFP marker with a gene of interest by Flp-mediated excision for generating stable producer cell lines [27]. **(a)** First, a stable GFP-positive cell line is established by cell sorting. **(b)** Then, the GFP gene is replaced by the gene of interest. This is achieved by transient overexpression of Flp recombinase. Flp mediates recombination of the *FRT*-sites flanking the GFP-gene, which leads to the gene's excision. **(c)** The gene of interest now becomes activated by the transgene's promoter. P: promoter, GOI: gene of interest, *FRT*: Flp recognition target site.
Figure adapted from Wilke *et al.* [30].

site-specific Flp recombinase, the reporter gene is flanked by two distinct Flp recognition target (*FRT*) sites that have been engineered so that they cannot recombine with each other (Figures 3 and 4). RMCE is initiated by introducing the Flp recombinase and a vector with the gene of interest (GOI), flanked by *FRT* sites, into the master cell line. Recombination of *FRT* sites leads to exchange of the reporter gene and the GOI, placing the GOI at a highly active, stable genetic locus (Figure 4). The same master cell line can be used for generating producer cell lines for many different target proteins [34,35].

The stable CHO Lec3.2.8.1 cell line SWI3a-26 is a master cell line for RMCE that was generated by random integration of a *FRT*-flanked GFP reporter gene [36]. It contains a single copy of the GFP reporter transgene at a genetic locus that is protected from silencing. The integrated GFP cassette contains a 'selection trap' that allows for selection of recombinant cells upon RMCE [37]. The selection trap is an inactive selection marker, lacking a promoter and a start codon (Figure 4). It is complemented upon RMCE and allows for antibiotic selection of recombinant cells. Using SWI3a-26, production cell lines were established by RMCE for different mammalian glycoproteins [36,38], including the ectodomain of the lysosomal membrane protein DC-LAMP. In consequence, the DC-LAMP domain structure was solved by X-ray crystallography [39•].

In our experience, RMCE takes about 7 weeks from the day of transfection to cryopreservation of clonal production cell lines [36]. In comparison to random integration, the process is faster and the effort of screening of large numbers of clones is avoided. The multi-host expression vector pFlpBtM allows for protein production in *E. coli*, transiently transfected mammalian cells and Baculovirus-infected insect cells and for construction of stable cell lines by RMCE with a single vector [38].

RMCE-derived production cell lines contain only a single copy of the transgene. Nevertheless, high-yield antibody production of up to 2 g/L shake flask culture was achieved with an RMCE system that uses Cre-*lox* recombination [40•]. With this system, about 5 weeks were required from transfection to completion of stable pool production cultures.

Two independent genetic loci can be targeted with two different transgenes by RMCE with a single transfection. This has been achieved by designing new synthetic *FRT* site variants [41].

Transfection efficiency and stable pools

Transfection of HEK293 and CHO cells leads to integration of transgene DNA at random chromosomal loci, but the frequency of these integration events is very low with commonly used vectors. Highly efficient systems for chromosomal integration of transgenes accelerate cell line generation and allow for protein production with stable pools, thereby eliminating time-consuming cloning steps. For protein production with a stable pool, the bulk of stably transfected cells are selected and used directly for protein production. The structure of lysosomal integral membrane protein (LIMP2) was solved using a stable pool obtained by transfection with the ΦC31 integrase system (described below) and antibiotic selection [42]. Efficient chromosomal integration is also achieved with Lentivirus particles and transposons (Figure 3).

Lentivirus

Lentiviral transduction of mammalian cells is very efficient and highly productive cells are generated at a high frequency. The usefulness of the Lentivirus system was demonstrated by establishing stable cell lines and stable pools for production of antibodies and blood coagulation factor VIII [43,44,45]. The Lentivirus efficiently transports the transgene cDNA into the nucleus, where it is integrated into the host cell genome by the viral integrase (Figure 3). A stable cell line for production of an IgG receptor subunit, which resulted in crystal structure determination, was established with a recombinant Lentivirus [46]. Stable cell line generation by Lentivirus transduction and by non-viral plasmid transfection was compared [45]. Gene delivery into nearly 100% of CHO cells grown in serum-free suspension culture was obtained by Lentivirus. GFP overexpression was up to five times higher in comparison to plasmid transfection. Potential drawbacks of the Lentivirus system are safety concerns, the error-prone replication of the viral RNA genome by reverse transcription and the extra step of virus particle preparation.

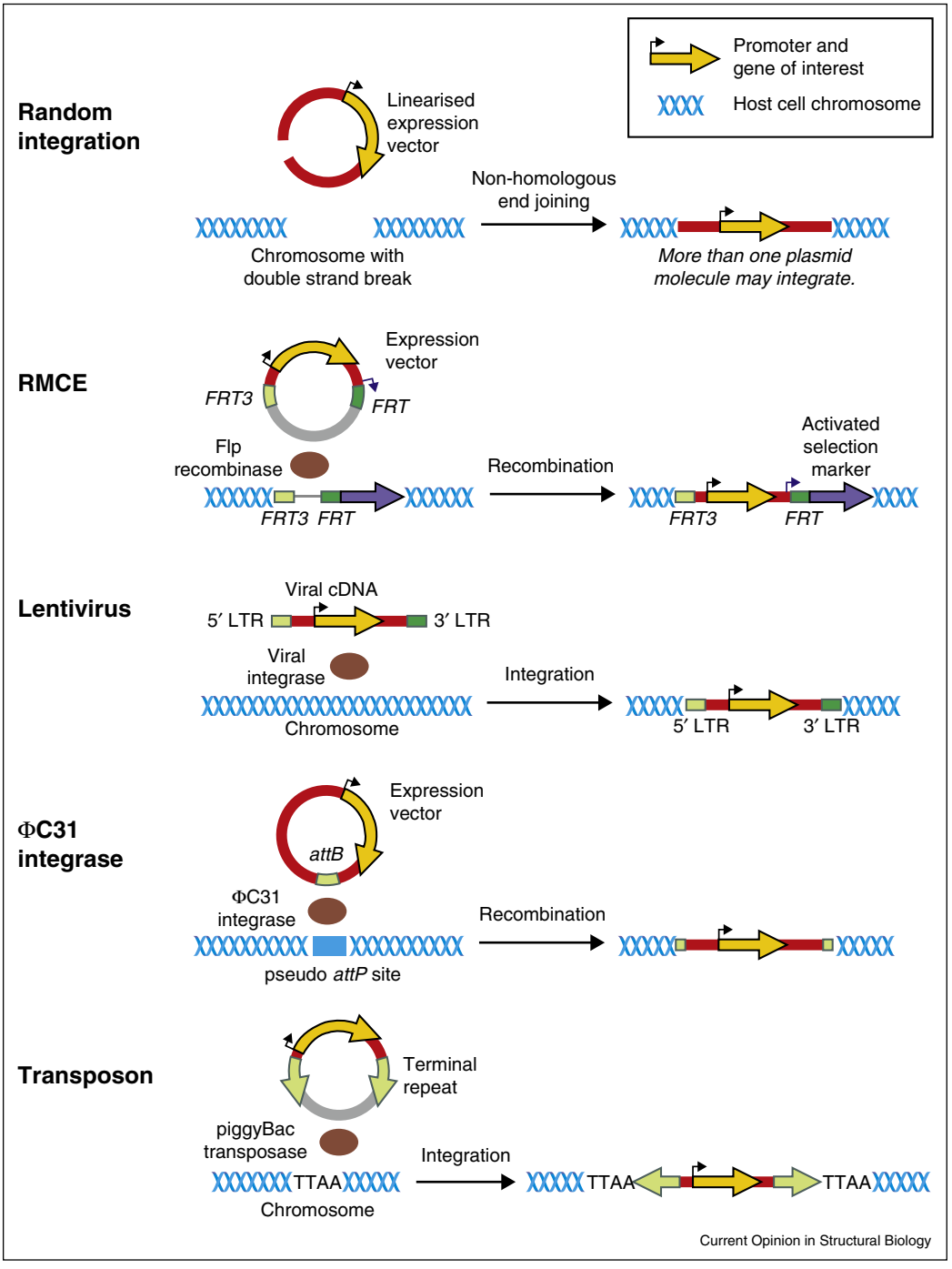
Phage ΦC31 integrase

The integration system of the *Streptomyces* phage ΦC31 represents a non-viral alternative for active transgene integration. The ΦC31 integrase performs recombination between the *attP* site of the phage genome and the *attB* site in the host bacterial chromosome. In mammalian cells, it mediates integration of plasmids bearing an *attB* site into chromosomal sequences that have sequence similarity with *attP*, termed pseudo *attP* sites [47]. Stable mammalian cell lines are generated by co-transfection with a ΦC31 integrase expression vector and an expression vector for the gene of interest that has an *attB* site (Figure 3). This system was used for protein production and structure determination of LIMP-2 [42], tumor antigen 5T4 [48] and acetylcholinesterase [49].

Transposons

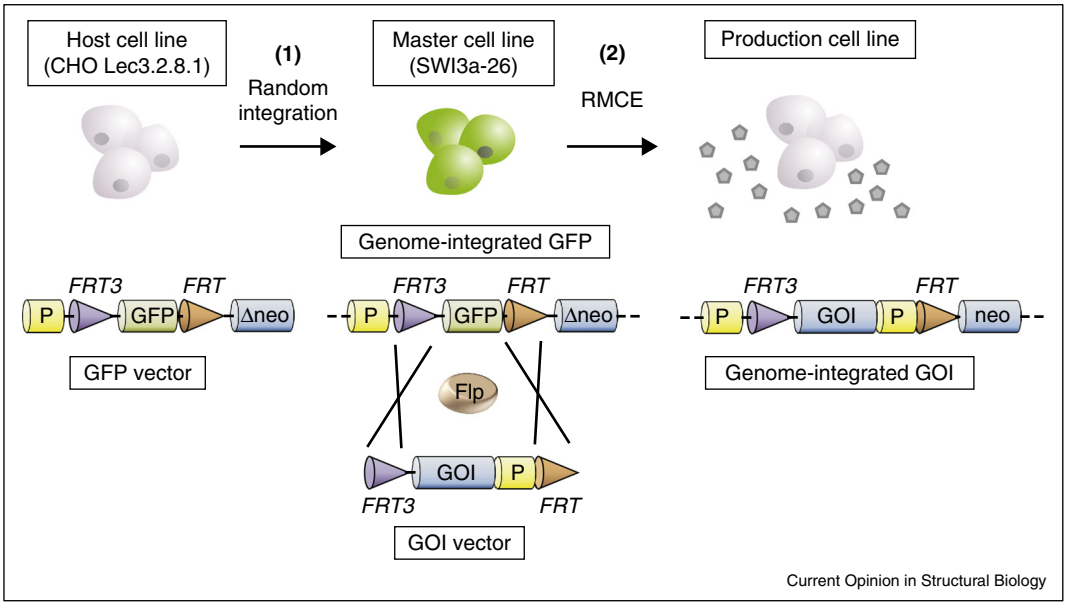
High rates of chromosomal integration have also been achieved with transposon vectors. The terminal inverted repeats of the 'piggyBac' transposon are recognized by the transposon's integrase, which leads to integration of the flanked sequence into a chromosomal TTAA site

Figure 3



Techniques of transgene genome integration for generating stable cell lines. *Random integration*: The linearized expression vector with the gene of interest (yellow) is integrated by non-homologous end joining into the host cell genome at the position of a random chromosome break. *RMCE* (recombinase-mediated cassette exchange): A master cell line, engineered to carry *FRT* sites at a suitable locus of its genome, is co-transfected with a vector with an *FRT*-flanked gene of interest (GOI) and a second vector for overexpression of Flp recombinase. The Flp recombinase catalyses site-specific genome integration of the GOI (see also Figure 4). *Lentivirus*: The host cell line is transduced by a Lentivirus that carries the GOI in its genome. The viral RNA genome is reverse-transcribed into cDNA, which is integrated into the host cell genome by the viral integrase at random locations. *ΦC31 integrase*: The host cell line is co-transfected with an expression vector for phage ΦC31 integrase and an expression vector for the GOI which contains an *attB* recombination site. The ΦC31 integrase is expressed and catalyses recombination of the *attB* site with genomic sequences that resemble the *attP* recombination site (pseudo *attP* sites). *Transposase*: The host cell line is co-transfected with an expression vector for the transposase of the piggyBac transposon and a vector containing the GOI, flanked by the transposon's inverted terminal repeats. The integrase recognizes the repeats and integrates the flanked sequence into a chromosomal TTAA site.

Figure 4



Recombinase-mediated cassette exchange (RMCE) with the CHO Lec3.2.8.1 cell line SWI3a-26 [36]. **(1)** The SWI3a-26 master cell line was generated from CHO Lec3.2.8.1 cells by random integration of a vector with a GFP reporter gene, flanked by a wild-type *FRT* site and a synthetic variant (*FRT3*). The cell line was isolated by cell sorting and contains a single copy of the GFP vector integrated at a genomic locus that allows for high-level gene expression. The GFP vector contains a selection trap consisting of an inactive neomycin resistance gene, Δneo, that lacks a promoter and a start codon. **(2)** Cassette exchange is initiated by co-transfecting the master cell line with vectors for the Flp recombinase and the gene of interest (GOI), flanked by compatible *FRT* sites. The transiently expressed Flp recombinase exchanges the *FRT*-flanked cassettes. The selection trap is triggered by activating the Δneo gene with a promoter and a start codon from the GOI vector, thereby enabling selection of recombinant clones.
Figure adapted from Wilke *et al.* [36].

(Figure 3). In stable CHO cell development, the piggyBac system strongly increased the frequency of stable integration and lead to up to fourfold higher protein yield from pools of transfected cells [50]. Similarly positive results were obtained with the 'Sleeping Beauty' transposon system and HEK293 cells [51]. A stable cell line overexpressing the four subunits of a γ-secretase complex was established in one step with the piggyBac system [52••]. A newly designed vector set for doxycycline-inducible overexpression utilizes multiple-copy integration by the piggyBac integrase and was used with HEK293S GnTI$^-$ cells for high-level secretion of 14 proteins with stable pools [53•].

Conclusions

Protein production with stable cell lines for structural biology relies strongly on glycosylation-deficient host cells. A novel HEK293 cell line, called 'GlycoDelete', was equipped with an intracellular endoglycosidase for secretion of deglycosylated glycoproteins. Host cells can be improved by sequence-specific genome engineering, allowing for efficient and highly specific knock out of multiple genes. A CHO cell was reported that lacks both GnTI and glutamine synthetase (GS) activity, which allows for highly efficient selection of stable cell lines with GS markers.

A survey of recent reports of protein crystal structures indicated that transient transfection of mammalian cells and stable mammalian cell lines were used with around the same frequency. In comparison to transient transfection protocols, the generation of stable cell lines has several bottlenecks. The frequency of stable genome integration upon transfection with plasmid vectors is low and, moreover, most integrated transgenes will be silenced. Top producers are usually rare in the pool of stably transfected cells and their identification requires isolating and characterizing a large number of clones.

Techniques for improved genome integration of transgenes and for improved selection of high-producing stable cell lines have been developed. Using fluorescent proteins as selection markers allows for isolating high-producing stable cells among millions of transfected cells by cell sorting. Novel glutamine synthetase knock out cells generated by sequence-specific genome engineering allow for a more efficient selection of high-producer cells with the glutamine synthetase selection marker. The problem of transgene silencing has been addressed by the recombinase-mediated cassette exchange (RMCE) technique. Here, site-specific recombination is used for targeting the gene of interest to specific genetic loci that

are protected from silencing and that allow for high-level gene expression.

Plasmid DNA introduced into the nucleus of host cells is integrated into the genome by host cell factors at random sites with a low frequency. Much more efficient genome integration is achieved when additional, heterologous integration factors are introduced. This is accomplished with Lentiviruses, piggyBag and Sleeping Beauty transposons and the phage ΦC31 integrase. Transfection of host cells with these highly efficient systems results in a pool of cells with a high proportion of stable high-producers. Such a stable pool can be used directly for protein production, thereby avoiding time-consuming clone isolation and characterization steps.

Acknowledgements

T. Böldicke, M. Gossen, J. Krausze and J. van den Heuvel are acknowledged for critical reading of the manuscript. This work was funded by the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement number 279039 (ComplexInc) and the Helmholtz Association funded Protein Sample Production Facility (PSPF). Support by Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI), is also acknowledged.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Almo SC, Love JD: **Better and faster: improvements and optimization for mammalian recombinant protein production**. *Curr Opin Struct Biol* 2014, **26**:39-43.

2. •• Yang H, Rudge DG, Koos JD, Vaidialingam B, Yang HJ, Pavletich NP: **mTOR kinase structure, mechanism and regulation**. *Nature* 2013, **497**:217-223.
The crystal structure of the cytosolic mTORC complex was obtained upon co-expression of the mTOR and LST8 subunits in a stable HEK293 cell line.

3. Jostock T, Knopf HP: **Mammalian stable expression of biotherapeutics**. *Methods Mol Biol* 2012, **899**:227-238.

4. Kim JY, Kim YG, Lee GM: **CHO cells in biotechnology for production of recombinant proteins: current state and further potential**. *Appl Microbiol Biotechnol* 2012, **93**:917-930.

5. Lai T, Yang Y, Ng SK: **Advances in Mammalian cell line development technologies for recombinant protein production**. *Pharmaceuticals (Basel)* 2013, **6**:579-603.

6. Chaudhary S, Pak JE, Gruswitz F, Sharma V, Stroud RM: **Overexpressing human membrane proteins in stably transfected and clonal human embryonic kidney 293S cells**. *Nat Protoc* 2012, **7**:453-466.

7. Vinothkumar KR, Edwards PC, Standfuss J: **Practical aspects in expression and purification of membrane proteins for structural analysis**. In *Electron Crystallography of Soluble and Membrane Proteins*. Edited by Schmidt-Krey , Cheng .. **Methods in Molecular Biology**. Humana Press; 2013:17-30.

8. Lanza AM, Kim do S, Alper HS: **Evaluating the influence of selection markers on obtaining selected pools and stable cell lines in human cells**. *Biotechnol J* 2013, **8**:811-821.

9. •• Nettleship EJ: **Structural biology of glycoproteins**. In *Glycosylation*. Edited by Petrescu . Intech; 2012 http://dx.doi.org/10.5772/48154.
This book chapter provides a comprehensive introduction on the structural analysis of glycoproteins. Expression systems for glycoproteins and strategies for seleno-methionine labelling and manipulating protein glycosylation are reviewed.

10. Reeves PJ, Callewaert N, Contreras R, Khorana HG: **Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line**. *Proc Natl Acad Sci U S A* 2002, **99**:13419-13424.

11. Stanley P: **Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity**. *Mol Cell Biol* 1989, **9**:377-383.

12. Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, Yu C, Boles KS, Evans EJ, Stuart DI *et al.*: **Glycoprotein structural genomics: solving the glycosylation problem**. *Structure* 2007, **15**:267-273.

13. Chen CI, Keusch JJ, Klein D, Hess D, Hofsteenge J, Gut H: **Structure of human POFUT2: insights into thrombospondin type 1 repeat fold and O-fucosylation**. *EMBO J* 2012, **31**:3183-3197.

14. Wibowo AS, Singh M, Reeder KM, Carter JJ, Kovach AR, Meng W, Ratnam M, Zhang F, Dann CE 3rd: **Structures of human folate receptors reveal biological trafficking states and diversity in folate and antifolate recognition**. *Proc Natl Acad Sci U S A* 2013, **110**:15180-15188.

15. Thorn A, Steinfeld R, Ziegenbein M, Grapp M, Hsiao HH, Urlaub H, Sheldrick GM, Gärtner J, Krätzner R: **Structure and activity of the only human RNase T2**. *Nucleic Acids Res* 2012, **40**:8733-8742.

16. •• Meuris L, Santens F, Elson G, Festjens N, Boone M, Dos Santos A, Devos S, Rousseau F, Plets E, Houthuys E *et al.*: **GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins**. *Nat Biotechnol* 2014, **32**:485-489.
The 'GlycoDelete' CHO cell line was engineered to carry a heterologous endoglycosidase in its Golgi apparatus, resulting in robust secretion of glycoproteins that had been deglycosylated in the cell.

17. Silva G, Poirot L, Galetto R, Smith J, Montoya G, Duchateau P, Paques F: **Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy**. *Curr Gene Ther* 2011, **11**:11-27.

18. Gaj T, Gersbach CA, Barbas CF 3rd: **ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering**. *Trends Biotechnol* 2013, **31**:397-405.

19. Hsu PD, Lander ES, Zhang F: **Development and applications of CRISPR-Cas9 for genome engineering**. *Cell* 2014, **157**:1262-1278.

20. Recacha R, Jiménez D, Tian L, Barredo R, Gahmberg CG, Casasnovas JM: **Crystal structures of an ICAM-5 ectodomain fragment show electrostatic-based homophilic adhesions**. *Acta Crystallogr D Biol Crystallogr* 2014, **70**:1934-1943.

21. Sethi DK, Gordo S, Schubert DA, Wucherpfennig KW: **Crossreactivity of a human autoimmune TCR is dominated by a single TCR loop**. *Nat Commun* 2013, **4**:2623.

22. Reguera J, Santiago C, Mudgal G, Ordoño D, Enjuanes L, Casasnovas JM: **Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies**. *PLoS Pathog* 2012, **8**:e1002859.

23. Menting JG, Whittaker J, Margetts MB, Whittaker LJ, Kong GK, Smith BJ, Watson CJ, Žáková L, Kletvíková E, Jiráček J *et al.*: **How insulin engages its primary binding site on the insulin receptor**. *Nature* 2013, **493**:241-245.

24. Nachon F, Carletti E, Ronco C, Trovaslet M, Nicolet Y, Jean L, Renard PY: **Crystal structures of human cholinesterases in complex with huprine W and tacrine: elements of specificity for anti-Alzheimer's drugs targeting acetyl- and butyryl-cholinesterase**. *Biochem J* 2013, **453**:393-399.

25. • Sealover NR, Davis AM, Brooks JK, George HJ, Kayser KJ, Lin N: **Engineering Chinese hamster ovary (CHO) cells for producing recombinant proteins with simple glycoforms by zinc-finger nuclease (ZFN)-mediated gene knockout of mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (Mgat1)**. *J Biotechnol* 2013, **167**:24-32.

Site-specific knock-out of the genes for glutamine synthetase and GnTI (MGAT1) in a CHO cell line demonstrates the potential of host cell line improvement by genome engineering with programmable nucleases.

26. Fan L, Kadura I, Krebs LE, Hatfield CC, Shaw MM, Frye CC: **Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells**. *Biotechnol Bioeng* 2012, **109**:1007-1015.

27. Kaufman WL, Kocman I, Agrawal V, Rahn HP, Besser D, Gossen M: **Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation**. *Nucleic Acids Res* 2008, **36**:e111.

28. Mancia F, Patel SD, Rajala MW, Scherer PE, Nemes A, Schieren I, Hendrickson WA, Shapiro L: **Optimization of protein production in mammalian cells with a coexpressed fluorescent marker**. *Structure* 2004, **12**:1355-1360.

29. Sen M, Yuki K, Springer TA: **An internal ligand-bound, metastable state of a leukocyte integrin, $\alpha_X\beta_2$**. *J Cell Biol* 2013, **203**:629-642.

30. Wilke S, Krausze J, Gossen M, Groebe L, Jäger V, Gherardi E, van den Heuvel J, Büssow K: **Glycoprotein production for structure analysis with stable, glycosylation mutant CHO cell lines established by fluorescence-activated cell sorting**. *Protein Sci* 2010, **19**:1264-1271.

31. • Vidigal J, Dias MM, Fernandes F, Patrone M, Bispo C, Andrade C, Gardner R, Carrondo MJT, Alves PM, Teixeira AP: **A cell sorting protocol for selecting high-producing sub-populations of Sf9 and High Five™ cells**. *J Biotechnol* 2013, **168**:436-439.
Preparative FACS cell sorting was optimized to improve cell survival of sensitive cell lines.

32. • Meng L, Forouhar F, Thieker D, Gao Z, Ramiah A, Moniz H, Xiang Y, Seetharaman J, Milaninia S, Su M *et al.*: **Enzymatic basis for N-glycan sialylation: structure of rat α2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation**. *J Biol Chem* 2013, **288**:34680-34698.
A novel GFP fusion tag for secreted proteins facilitates quantification of the target protein during cell line development and protein purification.

33. Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS: **Engineering and characterization of a superfolder green fluorescent protein**. *Nat Biotechnol* 2006, **24**:79-88.

34. Turan S, Zehe C, Kuehle J, Qiao J, Bode J: **Recombinase-mediated cassette exchange (RMCE) – a rapidly-expanding toolbox for targeted genomic modifications**. *Gene* 2013, **515**:1-27.

35. Nehlsen K, Schucht R, da Gama-Norton L, Krömer W, Baer A, Cayli A, Hauser H, Wirth D: **Recombinant protein expression by targeting pre-selected chromosomal loci**. *BMC Biotechnol* 2009, **9**:100.

36. Wilke S, Groebe L, Maffenbeier V, Jäger V, Gossen M, Josewski J, Duda A, Polle L, Owens RJ, Wirth D *et al.*: **Streamlining homogeneous glycoprotein production for biophysical and structural applications by targeted cell line development**. *PLoS ONE* 2011, **6**:e27829.

37. Verhoeyen E, Hauser H, Wirth D: **Efficient targeting of retrovirally FRT-tagged chromosomal loci**. *Technical Tips Online* 1998, **3**:122-125.

38. Meyer S, Lorenz C, Baser B, Wördehoff M, Jäger V, van den Heuvel J: **Multi-host expression system for recombinant production of challenging proteins**. *PLOS ONE* 2013, **8**:e68674.

39. • Wilke S, Krausze J, Büssow K: **Crystal structure of the conserved domain of the DC lysosomal associated membrane protein: implications for the lysosomal glycocalyx**. *BMC Biol* 2012, **10**:62.
The crystal structure of the human lysosomal membrane glycoprotein DC-LAMP was solved upon protein production by a stable CHO Lec3.2.8.1 cell line that was established by recombinase-mediated cassette exchange (RMCE).

40. • Crawford Y, Zhou M, Hu Z, Joly J, Snedecor B, Shen A, Gao A: **Fast identification of reliable hosts for targeted cell line development from a limited-genome screening using combined φC31 integrase and CRE-Lox technologies**. *Biotechnol Prog* 2013, **29**:1307-1315.
Fast generation of antibody-producing stable cell lines and stable pools was achieved by site-specific recombination with the RMCE approach. Up to 2 g/L volumetric antibody yield was obtained.

41. Turan S, Kuehle J, Schambach A, Baum C, Bode J: **Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology**. *J Mol Biol* 2010, **402**:52-69.

42. Zhao Y, Ren J, Padilla-Parra S, Fry EE, Stuart DI: **Lysosome sorting of beta-glucocerebrosidase by LIMP-2 is targeted by the mannose 6-phosphate receptor**. *Nat Commun* 2014, **5**:4321.

43. Baranyi L, Doering CB, Denning G, Gautney RE, Harris KT, Spencer HT, Roy A, Zayed H, Dropulic B: **Rapid generation of stable cell lines expressing high levels of erythropoietin, factor VIII, and an antihuman CD20 antibody using lentiviral vectors**. *Hum Gene Ther Methods* 2013, **24**:214-227.

44. Mufarrege EF, Antuña S, Etcheverrigaray M, Kratje R, Prieto C: **Development of lentiviral vectors for transient and stable protein overexpression in mammalian cells. A new strategy for recombinant human FVIII (rhFVIII) production**. *Protein Expr Purif* 2014, **95**:50-56.

45. Oberbek A, Matasci M, Hacker DL, Wurm FM: **Generation of stable, high-producing CHO cell lines by lentiviral vector-mediated gene transfer in serum-free suspension culture**. *Biotechnol Bioeng* 2011, **108**:600-610.

46. Schmidt MM, Townson SA, Andreucci AJ, King BM, Schirmer EB, Murillo AJ, Dombrowski C, Tisdale AW, Lowden PA, Masci AL *et al.*: **Crystal structure of an HSA/FcRn complex reveals recycling by competitive mimicry of HSA ligands at a pH-dependent hydrophobic interface**. *Structure* 2013, **21**:1966-1978.

47. Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT, Hoelters J, Calos MP: **Integration specificity of phage ΦC31 integrase in the human genome**. *J Mol Biol* 2006, **357**:28-48.

48. Zhao Y, Malinauskas T, Harlos K, Jones EY: **Structural insights into the inhibition of Wnt signaling by cancer antigen 5T4/Wnt-activated inhibitory factor 1**. *Structure* 2014, **22**:612-620.

49. Cheung J, Rudolph MJ, Burshteyn F, Cassidy MS, Gary EN, Love J, Franklin MC, Height JJ: **Structures of human acetylcholinesterase in complex with pharmacologically important ligands**. *J Med Chem* 2012, **55**:10282-10286.

50. Matasci M, Baldi L, Hacker DL, Wurm FM: **The PiggyBac transposon enhances the frequency of CHO stable cell line generation and yields recombinant lines with superior productivity and stability**. *Biotechnol Bioeng* 2011, **108**:2141-2150.

51. Petrakis S, Rasko T, Mátés L, Ivics Z, Izsvak Z, Kouzi-Koliakou K, Koliakos G: **Gateway-compatible transposon vector to genetically modify human embryonic kidney and adipose-derived stromal cells**. *Biotechnol J* 2012, **7**:891-897.

52. •• Alattia JR, Matasci M, Dimitrov M, Aeschbach L, Balasubramanian S, Hacker DL, Wurm FM, Fraering PC: **Highly efficient production of the Alzheimer's gamma-secretase integral membrane protease complex by a multi-gene stable integration approach**. *Biotechnol Bioeng* 2013, **110**:1995-2005.
A stable CHO cell line producing four subunits of the γ-secretase complex was generated with a single transfection using the piggyBac transposon system.

53. • Li Z, Michael IP, Zhou D, Nagy A, Rini JM: **Simple piggyBac transposon-based mammalian cell expression system for inducible protein production**. *Proc Natl Acad Sci U S A* 2013, **110**:5004-5009.
One-step generating of stable HEK293 GnTI$^-$ cell lines for inducible overexpression was enabled with a piggyBac vector system.

54. Singhal A, Ostermaier MK, Vishnivetskiy SA, Panneels V, Homan KT, Tesmer JJ, Veprintsev D, Deupi X, Gurevich VV, Schertler GF *et al.*: **Insights into congenital stationary night blindness based on the structure of G90D rhodopsin**. *EMBO Rep* 2013, **14**:520-526.

55. Miller PS, Aricescu AR: **Crystal structure of a human GABAA receptor**. *Nature* 2014, **512**:270-275.

56. Hassaine G, Deluz C, Grasso L, Wyss R, Tol MB, Hovius R, Graff A, Stahlberg H, Tomizaki T, Desmyter A *et al.*: **X-ray structure of the mouse serotonin 5-HT3 receptor**. *Nature* 2014, **512**:276-281.

57. Pozzi N, Chen Z, Gohara DW, Niu W, Heyduk T, Di Cera E: **Crystal structure of prothrombin reveals conformational flexibility and mechanism of activation**. *J Biol Chem* 2013, **288**:22734-22744.

58. Yu Y, Schürpf T, Springer TA: **How natalizumab binds and antagonizes alpha4 integrins**. *J Biol Chem* 2013, **288**:32314-32325.

59. Hatherley D, Lea SM, Johnson S, Barclay AN: **Structures of CD200/CD200 receptor family and implications for topology, regulation, and evolution**. *Structure* 2013, **21**:820-832.

60. Tanaka H, Miyazaki N, Matoba K, Nogi T, Iwasaki K, Takagi J: **Higher-order architecture of cell adhesion mediated by polymorphic synaptic adhesion molecules neurexin and neuroligin**. *Cell Rep* 2012, **2**:101-110.

61. Pos W, Sethi DK, Call MJ, Schulze MS, Anders AK, Pyrdol J, Wucherpfennig KW: **Crystal structure of the HLA-DM-HLA-DR1 complex defines mechanisms for rapid peptide selection**. *Cell* 2012, **151**:1557-1568.

62. Goodman KM, Kjær S, Beuron F, Knowles PP, Nawrotek A, Burns EM, Purkiss AG, George R, Santoro M, Morris EP *et al.*: **RET recognition of GDNF-GFRalpha1 ligand by a composite binding site promotes membrane-proximal self-association**. *Cell Rep* 2014, **8**:1894-1904.

63. Liu S, Mistry A, Reynolds JM, Lloyd DB, Griffor MC, Perry DA, Ruggeri RB, Clark RW, Qiu X: **Crystal structures of cholesteryl ester transfer protein in complex with inhibitors**. *J Biol Chem* 2012, **287**:37321-37329.

64. Liu P, Cleveland TE 4th, Bouyain S, Byrne PO, Longo PA, Leahy DJ: **A single ligand is sufficient to activate EGFR dimers**. *Proc Natl Acad Sci U S A* 2012, **109**:10861-10866.

65. Dutta A, Shrivastava IH, Sukumaran M, Greger IH, Bahar I: **Comparative dynamics of NMDA- and AMPA-glutamate receptor N-terminal domains**. *Structure* 2012, **20**:1838-1849.

66. Furger E, Frei DC, Schibli R, Fischer E, Prota AE: **Structural basis for universal corrinoid recognition by the cobalamin transport protein haptocorrin**. *J Biol Chem* 2013, **288**:25466-25476.

67. Zhou YF, Springer TA: **Highly reinforced structure of a C-terminal dimerization domain in von Willebrand factor**. *Blood* 2014, **123**:1785-1793.

68. Wong AH, Zhou D, Rini JM: **The X-ray crystal structure of human aminopeptidase N reveals a novel dimer and the basis for peptide processing**. *J Biol Chem* 2012, **287**:36804-36813.

69. Jansen S, Perrakis A, Ulens C, Winkler C, Andries M, Joosten RP, Van Acker M, Luyten FP, Moolenaar WH, Bollen M: **Structure of NPP1, an ectonucleotide pyrophosphatase/ phosphodiesterase involved in tissue calcification**. *Structure* 2012, **20**:1948-1959.

70. Yu X, Seegar TC, Dalton AC, Tzvetkova-Robev D, Goldgur Y, Rajashankar KR, Nikolov DB, Barton WA: **Structural basis for angiopoietin-1-mediated signaling initiation**. *Proc Natl Acad Sci U S A* 2013, **110**:7205-7210.

71. Xu K, Wu Z, Renier N, Antipenko A, Tzvetkova-Robev D, Xu Y, Minchenko M, Nardi-Dei V, Rajashankar KR, Himanen J *et al.*: **Neural migration. Structures of netrin-1 bound to two receptors provide insight into its axon guidance mechanism**. *Science* 2014, **344**:1275-1279.

72. Diestel U, Resch M, Meinhardt K, Weiler S, Hellmann TV, Mueller TD, Nickel J, Eichler J, Muller YA: **Identification of a novel TGF-beta-binding site in the zona pellucida C-terminal (ZP-C) domain of TGF-beta-receptor-3 (TGFR-3)**. *PLOS ONE* 2013, **8**:e67214.

73. Chen C, Ke J, Zhou XE, Yi W, Brunzelle JS, Li J, Yong EL, Xu HE, Melcher K: **Structural basis for molecular recognition of folic acid by folate receptors**. *Nature* 2013, **500**:486-489.

74. Harrison OJ, Vendome J, Brasch J, Jin X, Hong S, Katsamba PS, Ahlsen G, Troyanovsky RB, Troyanovsky SM, Honig B *et al.*: **Nectin ectodomain structures reveal a canonical adhesive interface**. *Nat Struct Mol Biol* 2012, **19**:906-915.

75. Kelley SL, Lukk T, Nair SK, Tapping RI: **The crystal structure of human soluble CD14 reveals a bent solenoid with a hydrophobic amino-terminal pocket**. *J Immunol* 2013, **190**:1304-1311.

76. Blenner MA, Dong X, Springer TA: **Structural basis of regulation of von Willebrand factor binding to glycoprotein Ib**. *J Biol Chem* 2014, **289**:5565-5579.

77. Hill CH, Graham SC, Read RJ, Deane JE: **Structural snapshots illustrate the catalytic cycle of beta-galactocerebrosidase, the defective enzyme in Krabbe disease**. *Proc Natl Acad Sci U S A* 2013, **110**:20479-20484.