doi:10.1016/j.jmb.2012.01.027 J. Mol. Biol. (2012) 417, 309–335

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Engineering Aggregation Resistance in IgG by Two Independent Mechanisms: Lessons from Comparison of *Pichia pastoris* and Mammalian Cell Expression

Jonas V. Schaefer and Andreas Plückthun*

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Received 7 November 2011; received in revised form 16 January 2012; accepted 20 January 2012
Available online 27 January 2012

Edited by F. Schmid

Keywords:
protein aggregation;
antibody engineering;
Pichia pastoris;
signal sequence;
α-factor pre-pro sequence

Aggregation is an important concern for therapeutic antibodies, since it can lead to reduced bioactivity and increase the risk of immunogenicity. In our analysis of immunoglobulin G (IgG) molecules of identical amino acid sequence but produced either in mammalian cells (HEK293) or in the yeast *Pichia pastoris* (PP), dramatic differences in their aggregation susceptibilities were encountered. The antibodies produced in *Pichia* were much more resistant to aggregation under many conditions, a phenomenon found to be mainly caused by two factors. First, the mannose-rich glycan of the IgG from *Pichia*, while slightly thermally destabilizing the IgG, strongly inhibited its aggregation susceptibility, compared to the complex mammalian glycan. Second, on the *Pichia*-produced IgGs, amino acids belonging to the α-factor pre-pro sequence were left at the N-termini of both chains. These additional residues proved to considerably increase the temperature of the onset of aggregation and reduced the aggregate formation after extended incubation at elevated temperatures. The attachment of these residues to IgGs produced in cell culture confirmed their beneficial effect on the aggregation resistance. Secretion of IgGs with native N-termini in the yeast system became possible after systematic engineering of the precursor proteins and the processing site. Taken together, the present results will be useful for the successful production of full-length IgGs in *Pichia*, give indications on how to engineer aggregation-resistant IgGs and shed new light on potential biophysical effects of tag sequences in general.

*Corresponding author. E-mail address: plueckthun@bioc.uzh.ch.

Abbreviations used: CDR, complementarity-determining region; DMEM, Dulbecco's modified Eagle's medium; DPAPase A, dipeptidyl aminopeptidase A; DSC, differential scanning calorimetry; DSF, differential scanning fluorimetry; FcγR, Fcγ receptor; FBS, fetal bovine serum; GAP, glyceraldehyde-3-phosphate dehydrogenase; GdnHCl, guanidine hydrochloride; HEK, human embryonic kidney; HT, high-tension; IEF, isoelectric focusing; IgG, immunoglobulin G; MALS, multi-angle light scattering; MRE, mean residue ellipticity; MS, mass spectrometry; MST, microscale thermophoresis; AOX1, alcohol oxidase 1; PBS, phosphate-buffered saline; POI, protein of interest; PP, *Pichia pastoris*; SEC, size-exclusion chromatography.

Introduction

Antibodies and their derivatives have found a broad range of applications, from basic research to medical therapy. With the use of many different techniques for their generation,[1] several therapeutic immunoglobulins, usually in the whole immunoglobulin G (IgG) format, have been approved by the Food and Drug Administration and applied to the treatment of a wide range of diseases.[2–4] One major challenge encountered in the development of antibody-based therapeutics, however, is their aggregation susceptibility under both formulation conditions[5,6] and *in vivo*, where aggregation is involved in a significant degradation pathway for

0022-2836/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.

monoclonal antibodies.[7] Irreversible aggregation thereby leads to reduced bioactivity[8] and presents an increased risk of immunogenicity.[9–14] Aggregation of even only a few percent of the administered therapeutic antibody can cause severe problems, as already low amounts of aggregates have long been known to cause anaphylactoid reactions.[15,16] Therefore, engineering IgG to reduce or eliminate its self-association is of great importance and might further extend the potential of therapeutic antibodies.

Since nearly all recombinant therapeutic antibodies are expressed in the IgG format, choices must be made in the molecular composition and expression system. Currently, all approved therapeutic antibodies are produced in mammalian cells,[17] since only this expression system can introduce the complex glycosylation that is necessary for the interaction of the antibodies with some of the activating Fcγ receptors (FcγR).[18] IgGs normally possess a single N-linked glycosylation at Asn297 in the C_H2 domain, maintaining these domains at a critical distance to bind to the FcγR.

In some cases, however, this FcγR interaction is neither needed nor wanted, thus opening the range of possible production hosts also to *Pichia pastoris* (PP). In the past, this methylotrophic yeast has been widely used for the expression of both secreted and intracellular eukaryotic proteins.[19] While it is able to carry out efficient disulfide bond formation and isomerization using the eukaryotic secretion quality control machinery,[19–21] *Pichia* expression generates mannose-rich glycosylation of IgG.[22] The attached glycan moieties differ markedly from the product of mammalian cells, resulting in antibodies that lack FcγR binding capacity. However, recently, *Pichia* has also been engineered to introduce complex, human-like glycosylation.[23–26]

Due to its respiratory growth, *Pichia* can grow to very high cell densities[27] and therefore express high levels of recombinant protein through efficient and tightly regulated promoters, accounting for up to 30% of the total cell protein.[28] Moreover, comparatively few endogenous proteins are secreted by this yeast system, which offers further advantages for subsequent protein purification and downstream processing.[29] Since Ridder *et al.* reported the first successful production of functional antibody fragments in *Pichia*,[30] more than 50 reports describing antibody expression in this system have been published.[31] While most of these molecules were single-chain variable fragments and fusions thereof, several Fab fragments have been produced in *Pichia* as well.[32–34] Interestingly, until today, *Pichia* has only been used as an expression host for a handful of full-length IgGs.[25,26,35]

Almost all heterologous proteins produced in *Pichia* have been expressed under the control of the methanol-inducible alcohol oxidase 1 (AOX1) promoter.[20,36] However, throughout this study, the constitutive glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter[37] was used, since the ability to compare mutant proteins for their expression and aggregation tendencies was crucial. The GAP promoter is usually considered to result in lower protein yields compared to AOX1,[38] even though some reports describe the GAP promoter to be superior to AOX1 for certain proteins,[39,40] including Fab fragments.[41] A potential reason could be the lower rate of expression that might better match the corresponding folding rate and, thus, reduces the stress level for the cells.

Previously, our laboratory analyzed sequence modifications in the variable domain in various antibody formats for aggregation propensity and conformational stability.[42–44] Eukaryotic expression systems, where functional expression yields can be compared and mutant proteins can be isolated and biophysically characterized, are required to carry out such studies for whole IgGs. Therefore, it is vital that isogenic strains can be produced, only differing by the sequence of the protein to be tested. As a first step, we investigated homologous recombination in HEK293 cells[45] and *Pichia*. As *Pichia* does not maintain episomal vectors, the expression cassettes are generally designed for integration into the yeast genome, thus leading to stable cell lines expressing the protein of interest (POI). In initial experiments, we were very surprised to find large differences between the aggregation behaviors of purified IgG protein from both systems, even though their amino acid sequence was identical: the same recombinant IgG expressed in HEK293 cells was much more aggregation prone than its counterpart from *Pichia*.

Here, we describe our investigations of this phenomenon and pinpoint it to differences in the processing of the precursor proteins, as well as to differences between the unequal glycosylations. As an initial model system, an antibody that had originally been selected from the HuCAL library[46] by panning against myoglobin from horse skeletal muscle and subsequently been stability engineered was used.[47] For this study, it was converted to a full-length IgG and investigated for its processing and expression in mammalian HEK293 cells and in the yeast *Pichia pastoris*.

Results

Comparison of various IgG constructs

In this study, we compared IgGs of identical amino acid sequence but produced in different eukaryotic expression systems. Unexpectedly, these molecules revealed dramatic differences in some of their biophysical properties. To monitor changes in their structure as a function of temperature, we analyzed

purified IgGs by circular dichroism (CD). Since the signals stemming from the IgG's β-sheets and random coil essentially canceled out to zero at 208 nm, thermal denaturation of the IgG molecules was determined at this wavelength as a function of temperature. These analyses indicated that the IgGs produced in human embryonic kidney (HEK) cells and in *Pichia* not only had a slightly different temperature of unfolding onset but, unexpectedly, also showed very different aggregation susceptibilities (Fig. 1a). The abrupt jump upward in the respective CD *versus* temperature curve indicates that the antibody expressed in mammalian cells started to aggregate at a temperature higher than ~75 °C, whereas the corresponding construct produced in the *Pichia* system did not exhibit any detectable aggregation at all. This finding could also be visually observed, as the mammalian antibodies had turned into a turbid sample after the experiment, compared to a clear solution for the yeast IgG (Fig. 1b, inset: *Pichia* samples on the left and HEK293 samples on the right). To further confirm these results, we performed an aggregation assay based on light scattering (Fig. 1b) in a fluorimeter: when both excitation and emission wavelengths were set to 500 nm, the different IgG molecules showed clear differences in the intensity of scattered light upon exposure to increasing temperatures. Only the molecules produced in mammalian cells aggregated, forming complexes large enough to measurably scatter the incoming light from about ~75 °C onward. In contrast, the *Pichia*-produced IgG did not show any detectable light scattering at all.

While the protein sequences of the mature IgG H and L chains are identical, different sugar moieties are attached to the C_H2 domain of the heavy chain in yeast and in mammalian cells. Therefore, glycan knockout mutants were designed by replacing the glycosylation motif Asn297-Ser298-Thr299 in the C_H2 domain by an Asn297-Ser298-Ala299 sequence. The resulting unglycosylated T299A mutants were analyzed for their behavior in the aggregation assay

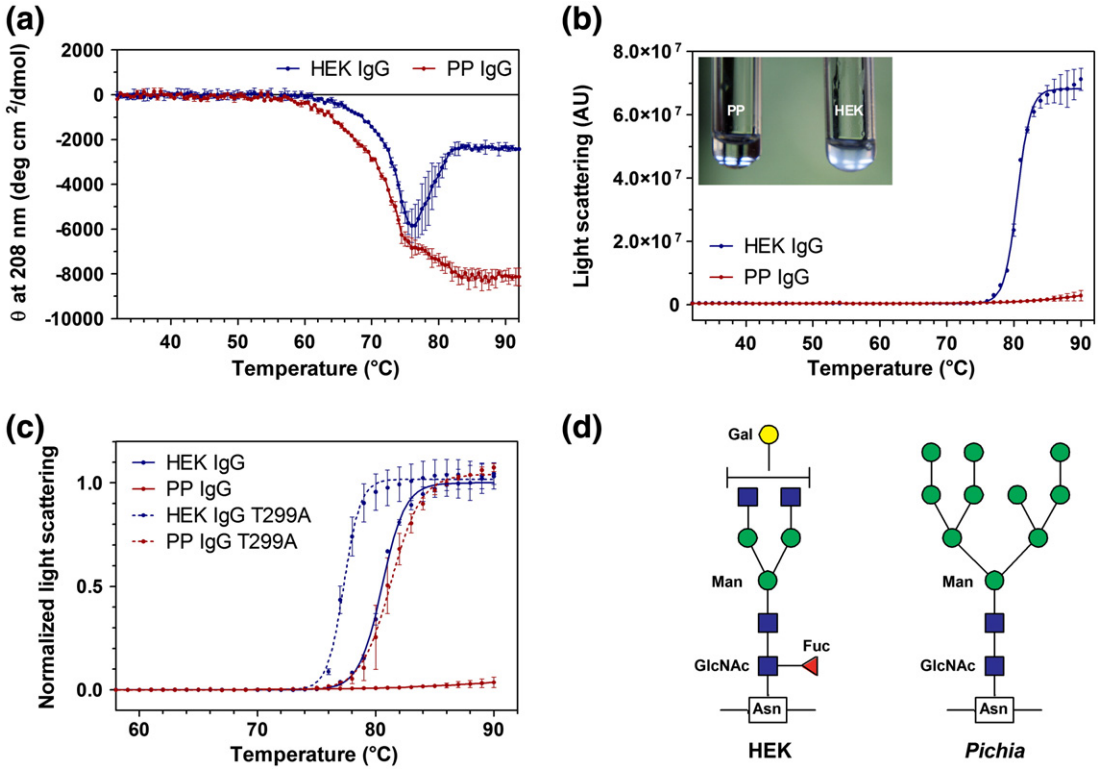


Fig. 1. Aggregation behavior of IgGs produced in eukaryotic expression systems. (a) Thermal denaturation curves of IgGs produced in mammalian cells (HEK293) or in the yeast *Pichia* (PP), respectively. The denaturation was followed by CD, plotting the signals at 208 nm as a function of temperature. The values are reported as mean residue ellipticity (MRE) (θ). The abrupt jump upward at about 79 °C in the HEK-derived signal is caused by the formation of insoluble aggregates. (b) Aggregation of IgG constructs as a function of temperature measured by light scattering at 500 nm. Emission was recorded at the same wavelength as excitation was performed. The inset shows *Pichia* (left) and HEK293 (right) samples after the measurement. (c) Aggregation susceptibility of glycosylated and deglycosylated (glycan knockout T299A) IgGs, analyzed by light scattering at 500 nm. (d) Structures of C_H2-attached glycans of IgGs produced by either the HEK or the *Pichia* system, as determined by ESI (*e*lectro*s*pray *i*onization)–MS. GlcNAc, *N*-acetyl-D-glucosamine; Man, mannose; Fuc, fucose; Gal, galactose.

in comparison to their glycosylated counterparts. As can be seen in Fig. 1c, the glycan moieties did influence the temperature of IgG aggregation onset substantially; however, they seemed to be not the only cause of the different aggregation behavior: the unglycosylated IgG produced in the yeast system did show aggregation, seen as a clear increase in light scattering at 81.3 ± 0.1 °C, while their unglycosylated mammalian counterparts already aggregated at 77.3 ± 0.1 °C. Glycosylated HEK IgG thus aggregates at a temperature of ~3.2 °C higher than the non-glycosylated HEK IgG.

We determined the different glycan structures present in the IgGs from both mammalian and yeast expression systems by mass spectrometry (MS). This was achieved by subtracting the experimentally determined mass of the T299A constructs (49,798.5 Da) from that of the original, glycosylated variant (51,274 Da), correcting it for the Thr-to-Ala mutation (−30 Da). From the resulting mass difference of 1445.5 Da (Fig. 2a), the glycan structure of HEK IgG schematically shown in Fig. 1d can be deduced: the $Gal(GlcNAc)_2(Man)_3(GlcNAc)_2Fuc$ composition typically found for proteins expressed in human cell culture.[48] The yeast glycan, on the other hand, was found to have a mass of 2035 Da (52,399 − 50,334 − 30 Da) and to be rich in mannose units, the total number of which varied between 9 and 15, with the majority of molecules possessing 10 mannose units [molecular formula: $(Man)_{9–15}(GlcNAc)_2$]. These different glycan structures between *Pichia* and mammalian cells were found to be the reason for the slightly decreased temperature of the onset of unfolding recorded in Fig. 1a for the *Pichia*-produced IgG. It could be shown in various experiments that the yeast glycan destabilizes the C_H2–C_H2 domain interface compared to the mammalian counterpart and thereby decreases the thermal stability of the whole IgG (Supplementary Data Fig. S1). Yet, at the same time, it largely prevents irreversible aggregation (Fig. 1c).

Since the different glycan structures were not the only cause of the different aggregation susceptibilities of the HEK- and *Pichia*-produced IgGs, both molecules were further analyzed by MS in their native and PNGase F deglycosylated form. As listed in Fig. 2a, not even in their deglycosylated forms did they correspond to the theoretical expectation from the amino acid sequence. However, considering that the heavy chains of mammalian IgGs generally lack their C-terminal Lys residues (−146 Da),[49] possess pyroglutamate residues at their N-termini [leading to the loss of a water molecule (−18 Da)][50,51] and exhibit oxidation of their methionine residues (+16 Da),[52] the underlined calculated masses, which are in excellent agreement with the measured ones, offer a plausible explanation for the posttranslational modifications of the mammalian IgG heavy chain. The light chain showed the expected mass without any major covalent modifications.

The results for the *Pichia*-produced IgGs, however, indicated the existence of large additional modifications at both the heavy and the light chains. Edman sequencing of their respective amino-termini verified the presence of the tetrapeptide Glu-Ala-Glu-Ala (EAEA) in front of the genuine N-terminus of the proteins, accounting for the additional 400 Da measured by MS (double underlined). Minor fractions with one or both EA pairs deleted were not detected. Since these four amino acids are remains of the α-factor pre-pro sequence (αMFpp) used in the expression setup, the maturation process of secreted proteins within the yeast system was analyzed in more detail. This secretion system consists of a 19-aa signal (pre) sequence followed by a 66-residue (pro) sequence, the latter containing three consensus N-linked glycosylation sites.[53] Directly following a dibasic Kex2 endopeptidase processing site, a spacer peptide of four residues (EAEA) is located upstream of the genuine N-terminus of the mature α-factor peptide. This spacer was included in the present IgG construct as well. As schematically illustrated in Fig. 2b for the heavy chain, proteins fused to the αMFpp sequence undergo various processing steps until being fully matured. Several studies have shown that dipeptidyl aminopeptidase A (DPAPase A) has problems in fulfilling its task in protein overexpression, thereby often leaving residual EAEA residues on the protein to be secreted.[54–56] This also was the case for the *Pichia*-produced IgG of this study, which therefore will be referred to "H-E/L-E" from now on (with the suffix "E" representing the EAEA tetrapeptide), having this extension on both chains.

N-terminal engineering of *Pichia* constructs

To investigate whether antibody chains with genuine N-termini can be produced in *Pichia*, we designed yeast variants lacking the EAEA tetrapeptide on the mature heavy and light chains, respectively. The EAEA coding sequence was deleted from the DNA encoding both chains: either in the individual coding sequences or for both chains simultaneously (depicted in Fig. 2d, exemplified for the light chain).

Surprisingly, upon analysis of purified IgGs by nonreducing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 3a), no distinct band corresponding to the full-length IgG could be detected for the H/L-E and H/L constructs, both lacking the EAEA coding sequence for the heavy chain. Since this is due to a nonhomogeneous composition of the heavy-chain sample as will be described below, these constructs will be referred to as H* from now on. In contrast, deletion of the tetrapeptide from the light chain alone (construct

(a)

construct	measured weight *	PNGaseF treated §	theoretical weight #	theoretical - Lys $	theoretical - Lys - H_2O ß	theoretical - Lys - H_2O + Ox &	theoretical + EAEA %	Δ mass ¶
heavy chain (H)								
HEK	51,274	49,829	49,977	49,831	49,813	49,829	50,377	0
PP	52,399	50,376						1
light chain (L)								
HEK	22,521	22,521	22,525	-	-	-	22,925	4
PP	22,922	22,922						3

* - mass (Da) measured by ESI-MS under reducing conditions

§ - mass (Da) measured by ESI-MS under reducing conditions after deglycosylation by PNGase F

- mass (Da)calculated from the sequence of the unmodified, reduced mature chain

$ - mass (Da) calculated considering truncation of C-terminal Lys residue

ß - mass (Da) calculated considering truncation of C-terminal Lys residue and pyroglutamate formation (loss of a water molecule)

& - mass (Da) calculated considering truncation of C-terminal Lys residue, pyroglutamate formation and Met oxidation

% - mass (Da) calculated considering N-terminal extention by Glu-Ala-Glu-Ala (no pyroglutamate formation)

¶ - mass difference between the best matching theoretical value and the PNGase F treated measured value (underlined or double underlined)

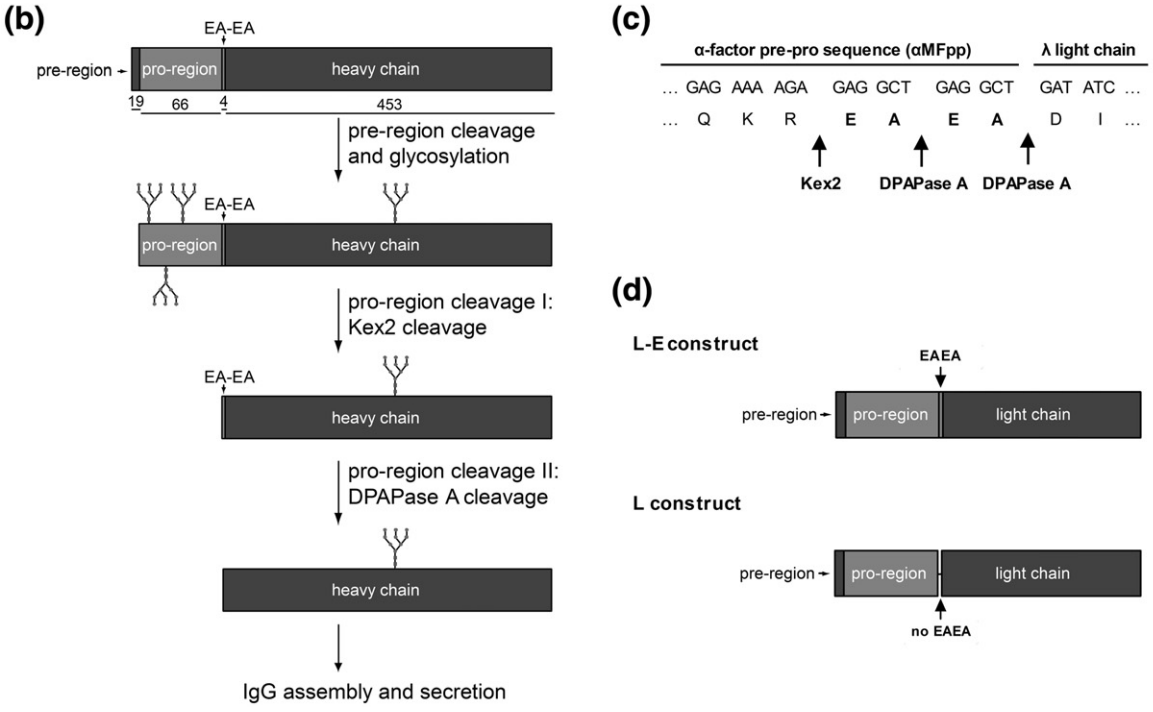
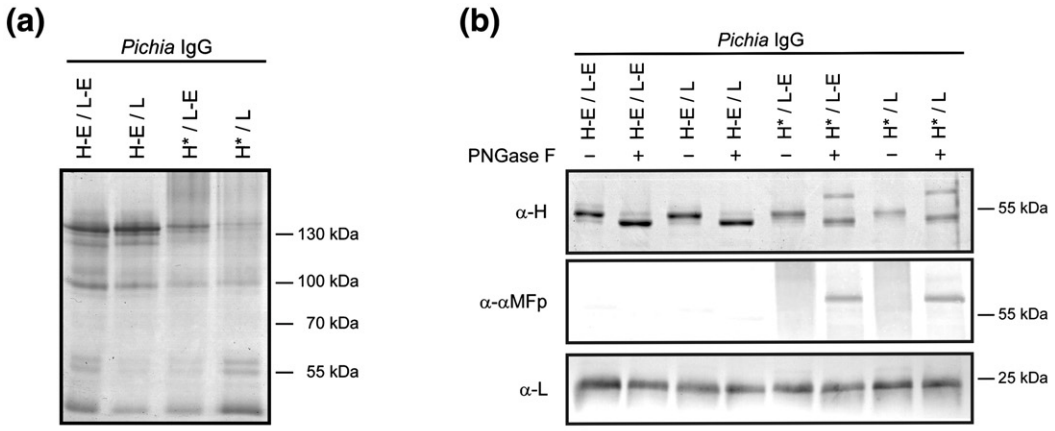


Fig. 2. Overview of detected IgG masses and *Pichia* processing. (a) Summary of molecular masses of HEK- and *Pichia*-produced IgGs, analyzed under reducing conditions by ESI–MS and calculated based on the amino acid sequence of the individual protein chains. All masses are stated in Daltons. Masses matching between calculation and observation are underlined. (b) Schematic overview (not drawn to scale) of N-terminal processing of secreted proteins fused to the α-factor pre-pro sequence (αMFpp) by the endopeptidase Kex2 and the DPAPase A (product of the *STE13* gene). The residue numbers of the different regions of the precursor protein are shown below the top construct. After the pre-region is cleaved off by signal peptidase upon entering the endoplasmic reticulum, the pro-region-containing protein is glycosylated and is further processed in the Golgi apparatus during the translocation process. Once the Kex2 endopeptidase, encoded by the *KEX2* gene, has cleaved off the 66 residues of the pro-region, the DPAPase A removes both EA dipeptides upstream of the mature terminus of the proteins. Afterwards, the mature proteins are released, and the IgG is assembled and finally secreted from the cell into the supernatant. (c) Sequence details of the junction between the αMFpp and the gene coding for the light chain. Arrows indicate the cleavage sites of Kex2 and DPAPase A. (d) Schematic drawing (not to scale) of the L-E and L constructs for *Pichia* expression. The suffix "E" corresponds to the tetrapeptide Glu-Ala-Glu-Ala (EAEA) derived from the pro-region of the αMFpp. In the L construct, the coding sequence for the EAEA tetrapeptide was removed at the DNA level.



(c)

construct	measured weight [§]	theoretical weight [#]	theoretical - H_2O [$]	Δ mass [¶]	N-terminal sequence [%]	name
heavy chain (H)						
H-E	50,376	50,378	-	2	… K R ¦ E A E A Q V Q L V Q …	
H*	49,956	49,977	49,959	3	… K R ¦ Q V Q L V Q S G P G …	"mature"
	51,010	51,005	-	- 5	… A K ¦ E E G V S L E K R Q …	"wrong"
	~ 57,000 [&]	56,916	-	nd	… L A ¦ A P V N T T T E D E …	"pro-region"
light chain (L)						
L-E	22,922	22,925	-	3	… K R ¦ E A E A D I E L T Q …	
L	22,521	22,526	-	5	… K R ¦ D I E L T Q P P S V …	

[§] - mass (Da) measured by ESI-MS under reducing conditions after deglycosylation by PNGase F

[#] - mass (Da) calculated from the sequence of the unmodified, reduced mature chain

[$] - mass (Da) calculated considering pyroglutamate formation (loss of a water molecule)

[¶] - mass (Da) difference between the best matching theoretical value and the PNGase F treated measured value

[%] - the arrow indicates position of cleavage leading to the observed N-terminal sequence

[&] - mass (Da) estimated from the running behavior on SDS-PAGE under reducing conditions

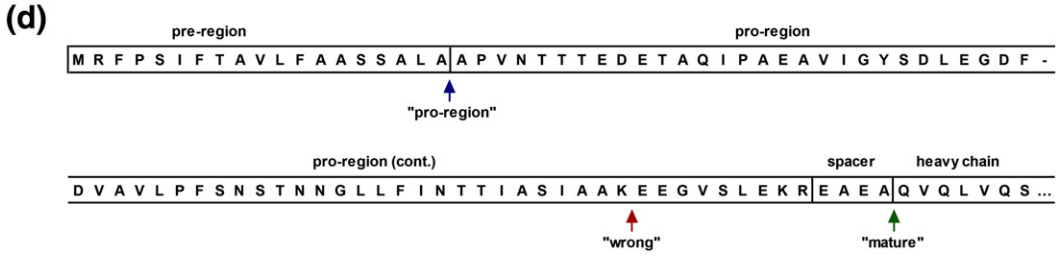


Fig. 3. Analysis of *Pichia*-produced IgG variants. (a) SDS-PAGE analysis of equal amounts of the Protein-A-purified *Pichia* constructs under nonreducing conditions stained with Coomassie Blue. The suffix "E" on the heavy (H)- or the light (L)-chain construct denotes the tetrapeptide Glu-Ala-Glu-Ala (EAEA) derived from the pro-region of the αMFpp. As the constructs without the EAEA turned out to be heterogeneous at their N-terminus (see the main text), they are marked by an asterisk (*). (b) Western blot analysis of IgGs after incubation in the absence or presence of PNGase F. Shown are the blots detected with antibodies specific for the heavy chain (α-H), for the α-factor pro-region (α-αMFp) and for the lambda light chain (α-L). (c) Overview of the measured masses of heavy and light chains. All masses are stated in daltons. The penultimate column shows the corresponding N-terminal protein sequence determined by Edman sequencing. A vertical broken line indicates the position of enzymatic cleavage within the maturation process. The color code highlights the charge of the individual amino acids (blue, positive charge; red, negative charge). The names mentioned in the last column are in agreement with those of Supplementary Data Fig. S2. (d) Schematic representation of the amino acid sequence of the α-factor pre-pro-region and the adjacent heavy-chain protein. Arrows indicate the sites of Kex2 cleavage, resulting in the "pro-region" variants (blue arrow), "wrong" variants (red arrow) and correct "mature" protein chains (green arrow).

H-E/L) still resulted in a clear band for the full-length IgG at the expected size. Detailed analysis of all constructs by reducing Western blots (Fig. 3b) with antibodies specific for the heavy chain, for the α-factor pro sequence (αMFp) or for the lambda light chain indicated a possible explanation. In the H-E-containing constructs (four left lanes in Fig. 3b, middle panel), no αMFp is detected (the light background signal is derived from the secondary antibody used which apparently cross-reacts with the human IgG), whereas in the H*-containing lanes (four right lanes in Fig. 3b, middle panel), a strong signal is found. This appeared as a smear, when left untreated, but distinct bands become visible upon treatment with PNGase F. Thus, part of the heavy chain still possesses the complete α-factor pro signal sequence at an apparent deglycosylated size of ~57 kDa and thus almost 7000 Da too high compared to the correctly processed heavy chain. The presence of the complete pro-region also explains the undetectable band for these constructs in Fig. 3a: due to the heterogeneous glycosylation of the pro-region, the IgG molecules appeared rather as a broad smear than as a single band.

The presence of the pro-region could be further confirmed by sequencing the N-termini of the respective heavy-chain proteins. Edman degradation and MS analyses (Fig. 3c) clearly indicated that deleting the EAEA tetrapeptide from the heavy chain resulted in a heterogeneous sample with three major species. Next to the previously mentioned portion still having the complete pro-region attached to it (named "pro-region", blue arrow in Fig. 3d) and a minor correctly processed protein fraction named "mature" (green arrow in Fig. 3d), also a third "wrong" species was found. This latter fraction was derived from an alternative cleavage site at position 57 of the pro-region (see red arrow in Fig. 3d). These molecules possess additional nine residues of the C-terminus of the pro-region and could also be detected by SDS-PAGE, once the gels were run long enough to provide sufficient separation (lane 2 in Supplementary Data Fig. S2c). Comparing the relative signal intensities of the Edman sequencing results, the fraction containing a correctly matured amino-terminus was determined to be only approximately 20%.

We also investigated whether the processing of the heavy chain can be improved by charge engineering, that is, making it more similar to the beginning of the light chain (Fig. 3c) (see Supplementary Data for details). However, since the charge engineering did not have the desired effect of creating homogeneous processing of the heavy chain (see Supplementary Data Fig. S2), the complete pro-region was deleted from the heavy-chain construct (illustrated in Fig. 4a). The resulting construct, named "H-pro-del", has the signal sequence (pre-region) fused directly to the N-terminus of the mature heavy chain. Examination of the new H-pro-del construct in its native and deglycosylated form by Western blot (Fig. 4b) revealed that its expression resulted in homogeneous chains of the correct size, which was also subsequently confirmed by MS analysis and Edman sequencing (data not shown).

We also wanted to determine how the amounts of the correctly assembled IgG constructs found in the supernatant compared to all other previously described constructs (Fig. 4c). These levels could be analyzed by sandwich enzyme-linked immunosorbent assay (ELISA), capturing the secreted IgGs by antibodies specific for the heavy chain and detecting the recombinant IgG with antibodies specific for the light chain. With the use of this assay, it could be guaranteed that only correctly assembled IgGs were detected and that, for example, no light-chain dimers affected the results. Strikingly, the heterogeneous H* constructs were found to be secreted at approximately only half the yield of the original H-E/L-E construct (49.3% for the H*/L-E and 59.3% for the H*/L construct, respectively). The new H-pro-del variant, however, seemed to overcome this limited expression and led to a total IgG amount even improved by~25%, compared to the original H-E/L-E construct. Interestingly, the additional omission of the pro-region from the light chain ("H-pro-del/L-pro-del" construct) led to a further decrease in the IgG yield found in the supernatant of the yeast cultures.

To confirm these findings, we analyzed the effects of the different pro-regions attached to the heavy chain in the context of the corresponding Fab fragment as well (Fig. 4d). In these Fab fragments, the heavy-chain V_H–C_H1 domains are connected to the complete lambda light chain by a C-terminal disulfide bond. In agreement with the data derived from the full-length IgGs, the expression of the H*/L Fab construct resulted in molecules having the complete pro-region attached at the N-terminus of the heavy chain (data not shown) and expression levels of only half the level of the original H-E/L-E variant (52.6%). The H-E/L-E level could again be increased by 41.3% by using the H-pro-del version.

Expression and analysis of mammalian IgG constructs

To examine the influence of the EAEA tetrapeptides on the aggregation susceptibilities of various IgG constructs, we needed to generate the corresponding molecules independent of the processing preferences of the expression system. Therefore, new constructs for HEK expression were designed. Since the process of protein maturation in the mammalian system is comparatively simple (Fig. 5a), different extensions could easily be attached to the amino-termini of both protein chains.

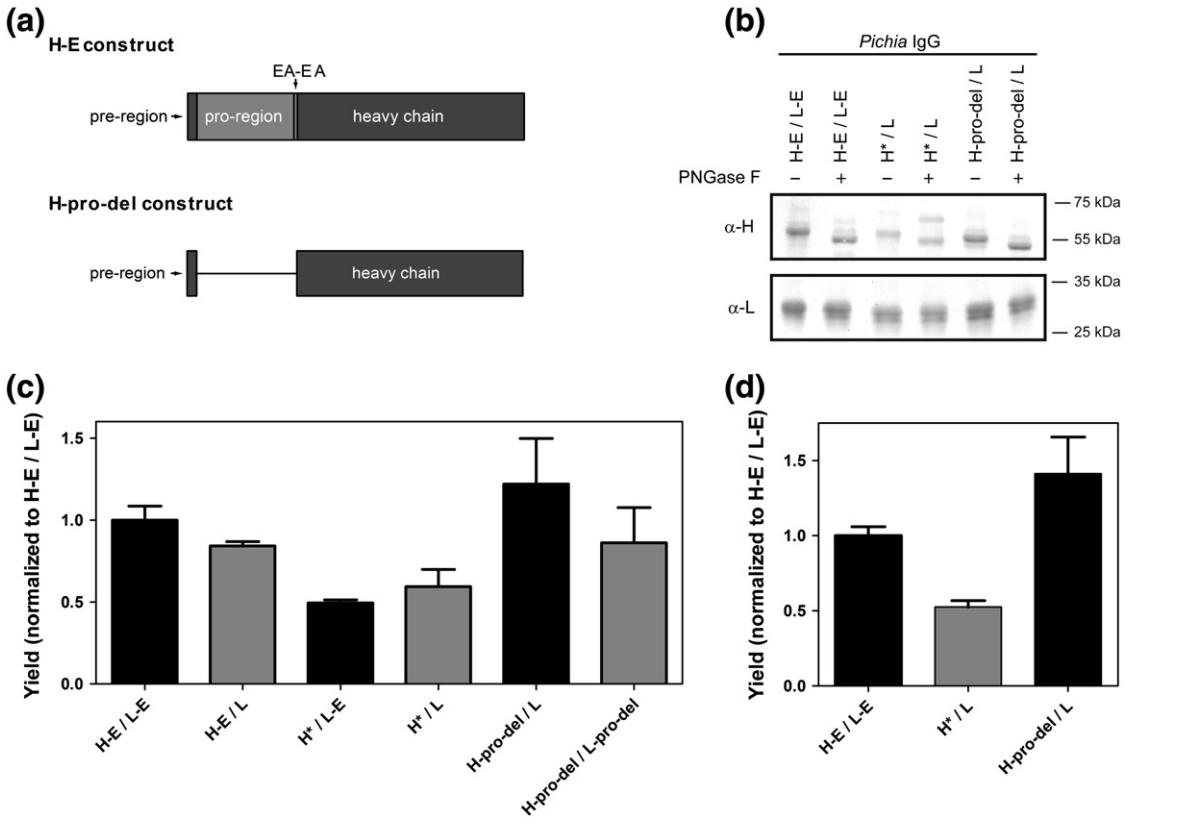


Fig. 4. Design of correctly processed IgG heavy chain in *Pichia*. (a) Schematic representation (not drawn to scale) of the H-E and H-pro-del constructs for *Pichia* expression. The latter construct was designed by deleting the complete pro-region at the DNA level. (b) Western blot analysis of IgG constructs after incubation in the absence or presence of PNGase F. Shown are the blots detected with antibodies specific to the heavy chain (α-H) and to the lambda light chain (α-L). (c) Secreted IgG levels of various *Pichia* constructs, detected by sandwich ELISA in the supernatant of stable *Pichia* clones. H* indicates the incorrectly processed, heterogeneous heavy chain; H-pro-del indicates the deletion of the complete pro-region from the heavy-chain construct; L-pro-del implies that this region is also deleted from the light chain. (d) Secretion level of corresponding Fab fragments expressed in stable *Pichia* clones analyzed by ELISA.

Figure 5b compares the variations made apart from the EAEA tetrapeptide. To examine whether just any random tetrapeptide might influence the biophysical characteristics of aggregation or whether the *Pichia*-produced EAEA is of special relevance, we also attached the completely unrelated and uncharged control peptide Ala-Gly-Ile-Gln (AGIQ; represented by the suffix "A") to both heavy and light chains in different combinations. Furthermore, the light-chain variant L-EA, carrying both tetrapeptides combined, EAEAAGIQ, in the orientation depicted, was included in the study to analyze the potential importance of steric proximity of the charged residues to the original N-terminus. Various combinations of these heavy- and light-chain variants were created, and stable HEK cell lines were established for all of them. All constructs could be expressed and resulted in proteins of expected size and banding pattern (reducing SDS-PAGE analysis in Fig. 5c; MS data not shown). To investigate the influence of the different amino-terminal extensions on the level of secreted IgGs, we determined the amount of IgG found in the supernatant of the stable cell lines by the same sandwich ELISA used for the yeast-produced IgGs. The extended mature N-terminus of the analyzed IgG chains only minimally influenced the levels of secreted IgGs (Fig. 5d). After 24 h of expression, the level of H-E/L-E found in the media reached almost 90% of the amount derived from the original H/L construct. This is in clear contrast to the yeast αMFpp system, as detailed above.

The main biochemical and biophysical characteristics of the N-terminal extension mutants from mammalian and yeast origins were analyzed to determine potential impacts of the various tetrapeptides on the IgG structures and stabilities. The addition of the EAEA tetrapeptides was found not to influence the binding of the resulting IgGs to their antigen myoglobin (see Supplementary Data Fig. S3). The different K_d values derived from microscale thermophoresis (MST) measurements are all within the experimental error of each other.

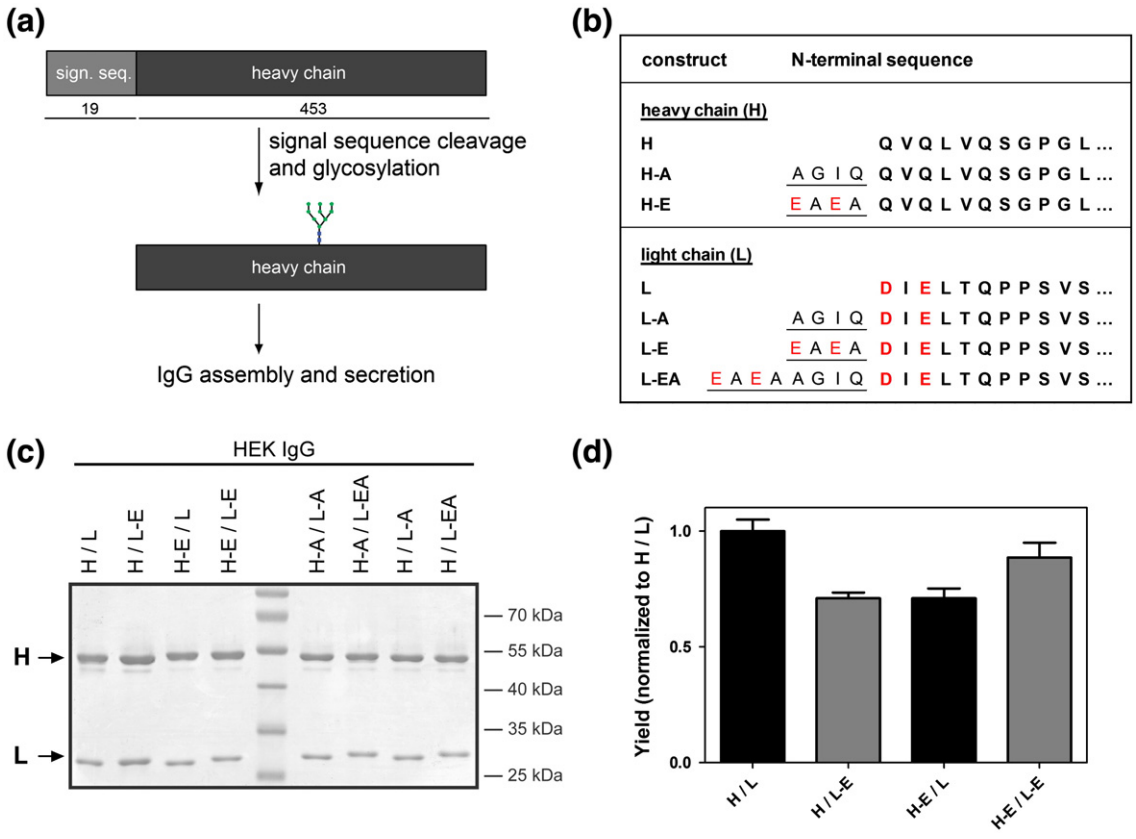


Fig. 5. Overview of different IgG constructs expressed in mammalian cells. (a) Schematic overview (not drawn to scale) of N-terminal processing of proteins in mammalian cells. Sign. seq. indicates the signal sequence. (b) Comparison of the mature N-terminal amino acid sequences of different mammalian heavy (H)- and light (L)-chain variants analyzed in this study. The suffix "A" corresponds to the control tetrapeptide Ala-Gly-Ile-Gln (AGIQ); the tetrapeptide Glu-Ala-Glu-Ala (EAEA) derived from the pro-region of the αMFpp is abbreviated as "E". The combination of both letters indicates the presence of all eight additional residues in the order presented. Red-colored amino acids are negatively charged. (c) SDS-PAGE analysis of HEK-produced IgG variants under reducing conditions, stained by Coomassie Blue. H and L indicate the respective chains. (d) Secretion level of IgG variants expressed by stably transformed HEK293 cells detected by ELISA.

Figure 6 compares structural characteristics of some selected HEK-produced IgGs under native conditions and their behavior under thermal and chemically induced unfolding. CD spectra recorded at room temperature for different mammalian variants are practically superimposable, indicating that the N-terminal extensions had no effect on the overall structure of the IgG, as expected (Fig. 6a). On the other hand, slight differences for the various E variants were revealed in thermal denaturation experiments measured by intrinsic tryptophan fluorescence (Fig. 6b). The unfolding curves were derived from the intensity ratio of the emission spectra at 330 nm (F_{330}) and 350 nm (F_{350}) upon excitation at 295 nm, plotted as a function of temperature.[57] One of the biggest advantages of this technique is the fact that this ratio is fairly robust against the loss of intensity due to light scattering from soluble aggregates. As can be seen in Fig. 6b, the addition of the EAEA tetrapeptide to the light chain seemed to slightly decrease the overall stability of the IgGs.

To further investigate this finding, we examined the H/L and the H-E/L-E constructs by differential scanning calorimetry (DSC) (Fig. 6c). This method is especially suitable for distinguishing different transitions in multidomain proteins (such as IgGs) that are often "silent" or overlapping in spectroscopic methods such as intrinsic tryptophan fluorescence. The previous results of a slightly decreased IgG stability could be validated, indicating that the H-E/L-E Fab fragment has its transition maximum at 74.5 °C, compared to 75.8 °C for H/L. These results are comparable to those obtained by guanidine hydrochloride (GdnHCl)-induced denaturation (Fig. 6d). In these experiments, the denaturation was followed by plotting the F_{330}/F_{350} ratio as a function of increasing GdnHCl concentration. These curves again suggested a slightly reduced stability for

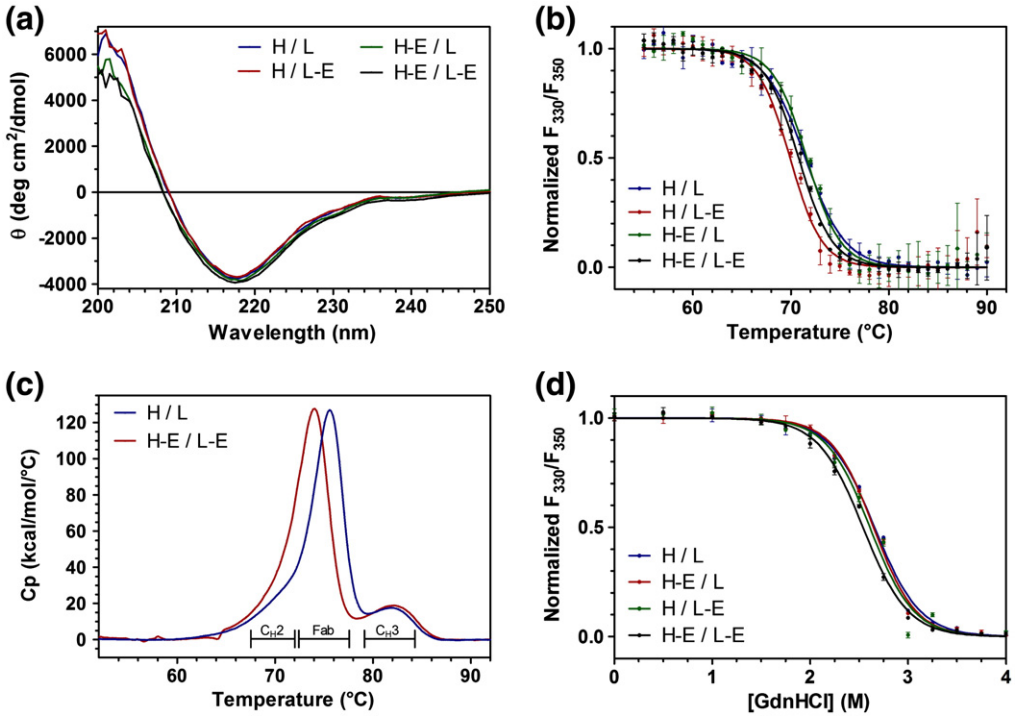


Fig. 6. Biophysical characterization of IgG constructs produced in mammalian cells. (a) CD spectra of different IgGs produced in mammalian cells. The values are reported as MRE (θ). (b) Thermal denaturation curves measured by intrinsic tryptophan fluorescence. The curves were obtained from the intensity ratio of the emission spectrum at 330 nm (F_{330}) and 350 nm (F_{350}) upon excitation at 295 nm, plotted as a function of temperature. A comparison between the different mammalian IgG variants is shown. (c) DSC of the H/L and H-E/L-E constructs. Indicated by the lines under the curves are the regions in which the individual domains (C_H2 domain, Fab fragment and C_H3 domain) undergo their transitions. (d) GdnHCl-induced denaturation of IgGs expressed in mammalian cells. The denaturation was followed by plotting the F_{330}/F_{350} ratio as a function of increasing GdnHCl concentration.

those constructs containing the EAEA tetrapeptide located at the N-terminus of the light or heavy chain. This overall tendency of a slightly reduced stability of the H-E/L-E constructs could further be confirmed by the analysis of thermal and chemically induced unfolding of *Pichia*-produced IgGs (Supplementary Data Fig. S4).

Determination of the aggregation susceptibilities of various IgG constructs

To test whether the EAEA tetrapeptides are indeed the reason for the very different aggregation susceptibilities of HEK- and *Pichia*-produced IgGs (Fig. 1), we analyzed the mammalian variants (equipped with these N-terminal extensions) by thermal unfolding recorded by CD at 208 nm. For clarity, Fig. 7 only compares the signals derived from the most important variants containing the EAEA tetrapeptides. Further data from all tested variants including the AGIQ controls can be found in Supplementary Data Fig. S5. The thermal scans provided in Fig. 7a clearly indicate a dramatic increase in the temperature of aggregation onset for those variants carrying the EAEA appendix at the N-terminus of their light chain. Interestingly, the H-E/L construct, having the identical appendix but at its heavy chain, only showed a minor increase of its aggregation onset temperature (+0.5 °C). The addition of the control tetrapeptide (Supplementary Data Fig. S5) even slightly increased the aggregation susceptibility of mammalian IgGs (−0.5 °C). This indicates that EAEA specifically—and not any random peptide—has this beneficial effect detailed above. Supplementary Data Fig. S5a demonstrate that all constructs represented by reddish symbols have a higher temperature of aggregation onset and carry the EAEA tetrapeptide at their light chain.

The differences in aggregation susceptibilities could also be verified by the analysis of the signals recorded by the photomultiplier connected to the CD spectrometer denoted as high-tension (HT) voltage. The HT signals shown in Fig. 7b belong to the curves shown in Fig. 7a (further curves can be found in Supplementary Data Fig. S5a and b). The changes in the HT signals confirmed the formation of aggregates that induced scattering of the UV light. As this led to the reduction in the number of photons

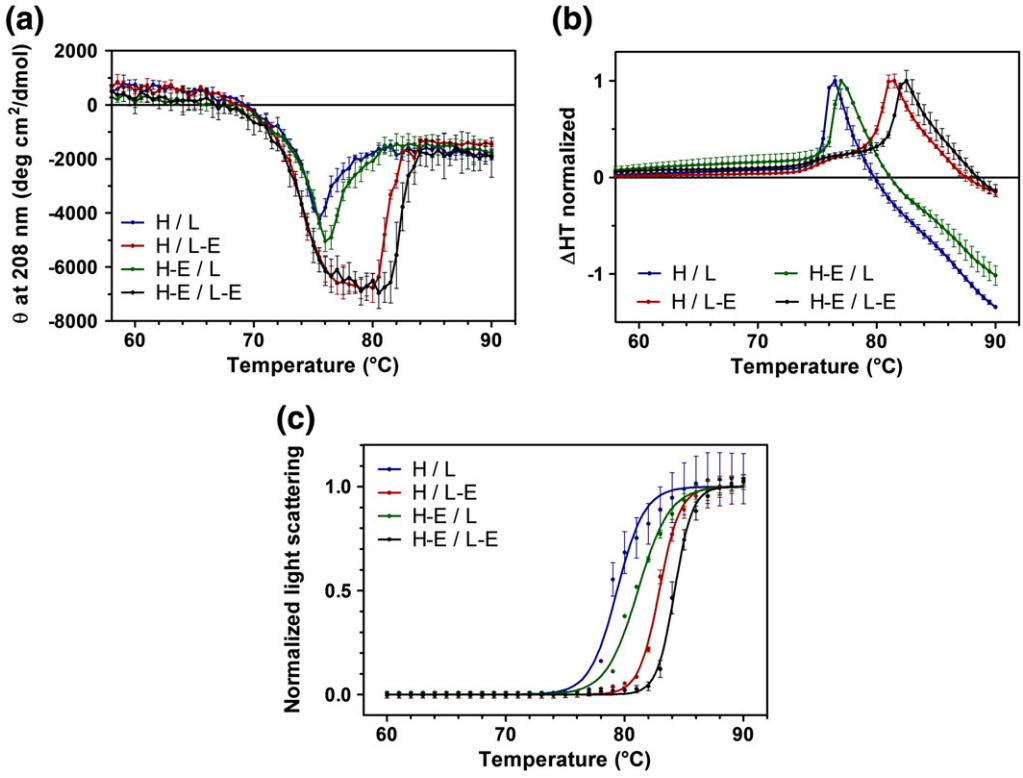


Fig. 7. Aggregation behavior of different IgGs obtained from mammalian expression. Shown are the signals derived from the most important variants containing the EAEA tetrapeptides. Further data from additional variants can be found in Supplementary Data Fig. S5. (a) Thermal denaturation curves. The denaturation was followed by CD measurements, plotting the signals at 208 nm as a function of temperature. A comparison of different mammalian IgG variants is shown, and the values are reported as MRE. (b) The HT voltage signals corresponding to the curves shown in (a) recorded by the photomultiplier connected to the CD spectrometer are a measure of sample transparency and thus aggregation and subsequent clearance by aggregate precipitation. The values were normalized by setting the initial and the highest values of the curves as 0 and 1, respectively. (c) Aggregation of IgG constructs measured by light scattering at 500 nm, with excitation and emission at this wavelength.

reaching the detector, the HT voltage was increased. The drop of signal intensity following the maximum was finally caused by precipitation of the aggregates. The temperatures for the maximal HT signal were determined to be 76.5 °C for the H/L, 77 °C for the H-E/L, 81.5 °C for the H/L-E and 82.5 °C for the H-E/L-E construct, respectively. These results were comparable to those derived from the aggregation assay observed in Fig. 7c. By this method, it could once more be shown that the addition of EAEA to the light chain—but not to the heavy chain—resulted in an apparent increase in the temperature of aggregation onset for the H/L-E construct by ~5 °C, compared to the H/L variant.

The results obtained from the mammalian-cell-produced IgGs could be confirmed by the analysis of their counterparts produced in *Pichia*. Supplementary Data Fig. S6 shows the analysis of the aggregation behavior of various *Pichia*-produced IgGs, analyzed by CD spectroscopy and the aggregation assay. In agreement with the data derived from HEK IgGs, the presence of the EAEA tetrapeptide in the yeast sequences resulted in IgG variants with decreased aggregation tendencies—emphasizing that the effect of this peptide on aggregation susceptibility required its location specifically at the N-terminus of the light chain.

To investigate the effect of the four extraneous amino acids at the light chain under more physiological conditions, we measured the formation of aggregates after extended incubation of freshly prepared IgGs by static multi-angle light scattering (MALS), coupled to size-exclusion chromatography (SEC). Figure 8a and b display the normalized light scattering measured for the HEK-expressed H/L and H/L-E variants, respectively, after a 5-day incubation at either 4 °C, 37 °C or 50 °C. Incubation at elevated temperatures dramatically increased the light scattering of the control IgG (H/L) for an aggregate peak eluting at ~8 ml (Fig. 8a). This larger signal was due to both an increased amount and the larger size of these aggregates. Consequently, the monomeric peak at ~13 ml decreased in its intensity. This

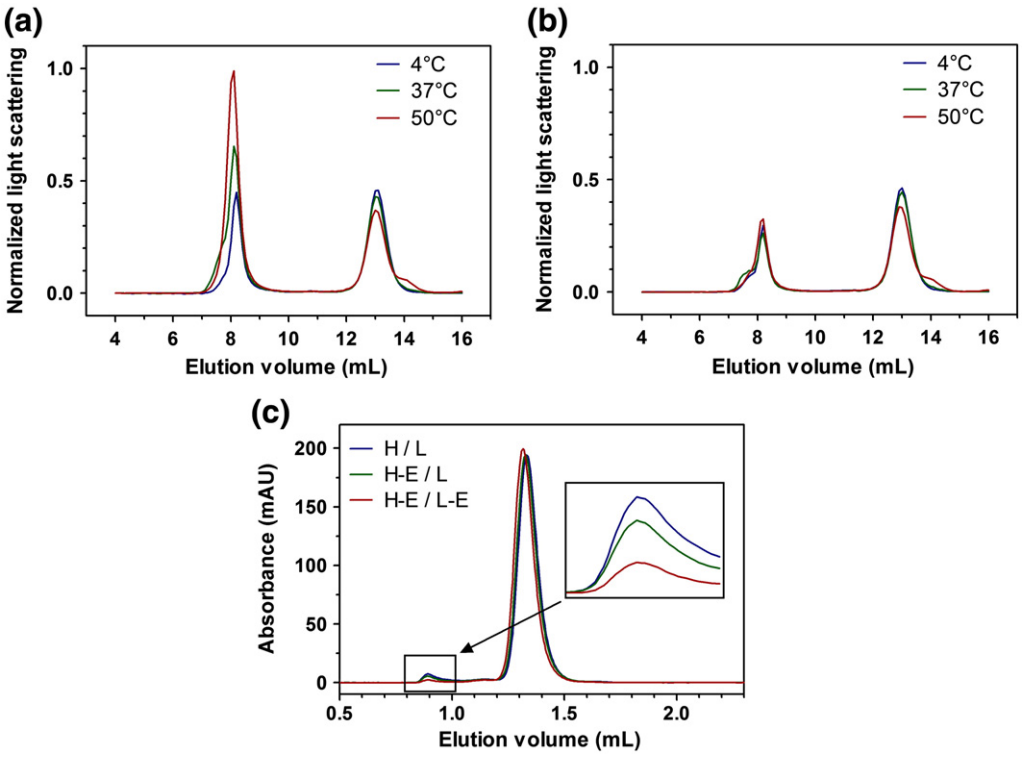


Fig. 8. Aggregation of different IgGs upon incubation at 4 °C, 37 °C or 50 °C analyzed by MALS and SEC. (a) Light scattering of the IgG H/L construct, analyzed by MALS, coupled to SEC. Samples were previously incubated for 5 days at 4 °C, 37 °C or 50 °C, respectively. (b) Light scattering of the IgG H/L-E construct. Plotted are the signals obtained from samples incubated for 5 days at 4 °C, 37 °C or 50 °C, respectively. (c) SEC of various HEK-produced IgG constructs after storage at 4 °C for 6 months. Note that (a) and (b) were performed with a 24-ml Superdex 200 10/300 GL column, while a 2.4-ml Superdex 200 PC 3.2/30 column was used in (c).

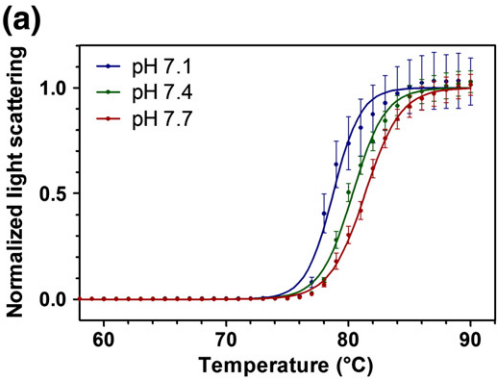
was, however, also partly due to the disassembly of both light chains from the antibody, detectable as a little shoulder toward higher elution volumes and in SDS-PAGE (data not shown). Quite different results were obtained once the EAEA tetrapeptide was attached to the light chain. As depicted in Fig. 8b, the amount (i.e., intensity) of the aggregation peak did hardly change even under more stressful conditions (50 °C), and only a minimal increase in the size of aggregates (i.e., a shift toward smaller elution volumes) was detected. As for H/L, the monomeric IgG peak, however, still was slightly diminished due to IgG disassembly.

To get further insights into the effect of the EAEA peptide on long-term stability, we also incubated IgG samples for 6 months at 4 °C and afterwards run them on an analytical SEC as illustrated in Fig. 8c. Although only a small fraction of the sample turned into aggregates (and eluted at ~0.9 ml), a clear difference in intensity could be seen for the various constructs. While the H/L variant exhibited an aggregated fraction of almost 3%, this peak was noticeably decreased for the H-E/L-E construct, adding up to only 0.5% of the total protein.

Influence of pH values on the temperature of aggregation onset

Since the difference between the EAEA peptide and the ineffective control peptide AGIQ is two negative charges, the influence of pH values on the aggregation susceptibilities was examined more closely. As depicted in Fig. 9a, even minor changes in the pH value dramatically influenced the temperatures of aggregation onset. While the H/L-A construct aggregated at 78.8 ± 0.1 °C at the same pH as in all previous analyses (pH 7.1), changing it to pH 7.4 increased this temperature to 80.3 ± 0.1 °C, and even further to 81.4 ± 0.1 °C in phosphate-buffered saline (PBS) adjusted to pH 7.7. Using a Web-based protein calculator†, we computed the net charge of the different IgG constructs at various pH values. In the second column of Fig. 9b, the net charges of full-length IgG consisting of both heavy and light chains are shown for various pH values. Adding the EAEA extension to the light chain (third column) had the same effect on the protein's overall

† www.scripps.edu/~cdputnam/protcalc.html



(b)

pH	charge of constructs		
	H/L	H/L-E	λ-L
7.0	0.4	-3.6	-6.7
7.1	-0.9	-4.9	-6.9
7.2	-2.0	-6.0	-7.0
7.3	-3.2	-7.2	-7.2
7.4	-4.3	-8.3	-7.4
7.5	-5.5	-9.5	-7.6
7.6	-6.7	-10.7	-7.8
7.7	-8.1	-12.1	-8.0

Fig. 9. Influence of the pH on the aggregation of IgG. (a) Aggregation of IgG H/L-A construct measured by light scattering at 500 nm. Samples were previously equilibrated in PBS buffer at the indicated pH. (b) Overview of the overall charge of the different IgG constructs at various pH values. Analyses were performed using the Web-based proteincalculator†. In the first three columns, the results for the full-length IgG consisting of two heavy and light chains (including indicated N-terminal appendices) are shown. Two positive charges less have been considered due to the formation of N-terminal pyroglutamate residues at the heavy chain, thus lacking the positively charged N-terminal amino group. The last column specifies the charges derived from analyzing only one light chain.

charge as increasing the pH by 0.35 units, namely, adding four negative charges to the molecule (two for each light chain). This addition of negative charge had such a dramatic effect on aggregation susceptibility since the charge of the H/L construct is minimally positive at the pH values generally used throughout this study (pH 7.1). Therefore, lowering the overall charge by 4 units has a rather profound effect on the net charge of the whole molecule.

Transfer of findings onto a second IgG construct

Finally, we analyzed to which extent our findings could be transferred to other IgG constructs. Since the effect seemed to be related to the overall charge of the molecule, we chose a completely unrelated IgG as an alternative model system. Having very different complementarity-determining regions CDRs regions in its kappa light chain and thereby a comparatively high isoelectric point (p*I*) of 6.9 (the previously used lambda light chain has a p*I* of 4.9 due to its acidic CDRs), this construct had a completely different range of overall IgG charges. When tested in the aggregation assay (Fig. 10a), the yeast-produced version with its different glycan moiety indeed aggregated at higher temperatures compared to its mammalian counterparts (88.4 ± 0.1 °C *versus* 82.2 ± 0.1 °C), indicating that this glycan effect seems to be general. However, once the glycans were removed by PNGase F treatment, only minor differences in the temperature of aggregation onset were observed (Fig. 10b; 86.4 ± 0.1 °C *versus* 85.5 ± 0.3 °C). The reason for this might be that the overall charge of this new IgG construct is more positive than for the first analyzed IgG, such that the addition of the EAEA tetrapeptide has a less profound influence on the overall charge in this case (Fig. 10c).

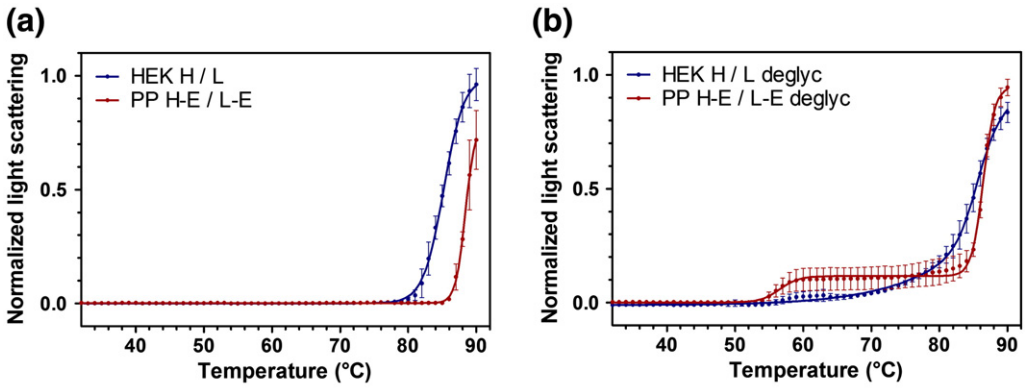
Discussion

When IgG molecules of identical amino acid sequence were produced in different eukaryotic systems, dramatic differences in their aggregation susceptibilities were encountered. This observation turned out to be due to two major factors: (i) differences in the glycan structures and (ii) residues remaining from the *Pichia* secretion system used, the α-factor pre-pro sequence (αMFpp).

Glycan structures

Antibodies of the IgG format are N-glycosylated at the Asn297 position in their C_H2 domains. These domains, having the lowest stability within the IgG molecule,[58] are the only ones not having any direct protein–protein interaction and solely contact each other by their sugar moieties. The attached glycans have proven to fulfill multiple roles in the maturation and function of IgGs. In general, glycosylation is a means of "quality control" in the eukaryotic endoplasmic reticulum.[59,60] Indeed, the glycosylation of the antibody's C_H2 domain has recently been shown to be a critical step in antibody folding.[61,62]

In the mammalian IgG molecule, the sugars have further taken on the functional role of enabling the antibody to bind to the activating FcγR.[18] This has been interpreted as the sugars mediating both the distance and the relative mobility of the domains toward each other, which is necessary since FcγR bind the C_H2 domains near the hinge region.[63–66] Even one single fucose residue can decrease the



(c)

pH	charge of constructs		
	H / L	H / L-E	κ-L
7.0	20.8	16.8	0.3
7.1	19.5	15.5	0.1
7.2	18.3	14.3	0.0
7.3	17.1	13.1	- 0.2
7.4	15.9	11.9	- 0.4
7.5	14.6	10.6	- 0.6
7.6	13.4	9.4	- 0.8
7.7	12.0	8.0	- 1.0

Fig. 10. Aggregation behavior of a second, kappa light chain-containing IgG. Comparison of a different IgG expressed in either *Pichia* or mammalian cells. The mammalian construct is of the H/L design, whereas the IgGs from the *Pichia* system are still of the H-E/L-E format. Shown are the light-scattering data of (a) the original and (b) the PNGase-F-deglycosylated IgGs. The additionally detectable light-scattering signal starting at ~56 °C is due to the unfolding and aggregation of the PNGase F enzyme. (c) Overview of the overall charge of the second IgG constructs at various pH values. Analyses were performed using the Web-based protein calculator†. (as in Fig. 9b). Two additional negative charges have been considered due to the formation of N-terminal pyroglutamate residues at the heavy chain, lacking the positively charged N-terminal amino group.

affinity of IgG for FcγRIII as it sterically interferes with the receptor's own oligosaccharide,[67] while fucosylated IgGs still bind to the inhibitory FcγRII equally well, having important consequences for glycoengineering.

Nonetheless, an increasing number of therapeutic programs are not aiming for FcγR activation but, indeed, actively avoiding it.[68] Thus, it is worthwhile to consider alternative expression systems for IgG production, including the yeast *Pichia*. For this production system, however, the role of glycosylation for the stability of the native IgG under storage and *in vivo* conditions has to be considered. When comparing IgGs with complex glycosylation from HEK293 cells to those containing *Pichia*-produced mannose-rich glycosylation, we found two decisive differences. First, the melting temperature of the corresponding C_H2 domain and thereby of the whole IgG is lowered when produced in *Pichia*, as can be seen by three independent techniques: CD (Fig. 1a), intrinsic tryptophan fluorescence (Supplementary Data Fig. S1f) and differential scanning fluorimetry (DSF; Supplementary Data Fig. S1d). In the latter technique, which offers increased resolution of the thermal unfolding of different domains like DSC, the shift in the stability of the C_H2 domain to lower temperatures due to the *Pichia* glycosylation is clearly visible. The second difference, however, is more than counterbalancing this negative effect: as detailed above, IgGs with *Pichia* glycosylation are much less prone to aggregation.

In agreement with our data, *Pichia* glycans carry more sugar units and are known to be rich in mannose units,[22] which may sterically block the initial docking of IgGs to each other at the onset of aggregation. It is also known that these oligosaccharides may contain several mannosylphosphates, transferred to both the core and the outer sugar chains.[21,22,69,70] However, our analysis showed that our IgGs produced in *Pichia* carried no negative charge—and thus no mannosylphosphate—in their oligosaccharides. We arrived at this conclusion by isoelectric focusing (IEF) analysis of the IgGs. It revealed that the isoelectric points (p*I*) of both *Pichia*- and HEK-produced antibodies before and after deglycosylation by PNGase F were almost identical (see Supplementary Data Fig. S7). If a negative charge had been removed from the IgGs

produced in *Pichia*, this would have been seen as a shift upward to less negative p*I* upon deglycosylation. In addition, the running behavior of *Pichia*-produced IgGs on ion-exchange chromatography did not change upon deglycosylation (data not shown): an effect that would have been expected if a certain fraction of the resulting *Pichia* glycans had possessed negative charges on their mannose units. Therefore, the beneficial effect of the *Pichia* glycans conferring resistance to aggregation does not seem to be based on any glycan contribution to the protein's overall charge.

To highlight the importance of the glycan moiety, we have included unglycosylated T299A IgG constructs in our studies, confirming the role of the sugars for stability and in protection against aggregation (Supplementary Data Fig. S1 and Fig. 1c). Unglycosylated IgGs melt at lower temperatures and are more aggregation prone than IgGs glycosylated with either *Pichia* or mammalian oligosaccharides—a finding consistent with previous observations.[71] Figure 11 compares the aggregate formation for IgGs with identical primary sequences and N-terminal extensions (H-E/L-E constructs) made in either expression system. Upon deglycosylation by PNGase F, the molecules should be of identical composition, whether expressed in either mammalian or *Pichia* cells, except for the removal of the C-terminal lysine from the heavy chains in mammalian cells, which was not observed to occur for the *Pichia* constructs. Indeed, the curves indicating the aggregation behavior are almost identical.

Taken together, the *Pichia* glycosylation has the surprising effect of protecting the IgG against aggregation better than the mammalian complex sugar moiety. For future applications, this must of course be weighed against the therapeutic necessity of FcγR activation. Recently, glycoengineered *Pichia* strains have been developed and meanwhile successfully applied to the production of monoclonal antibodies, such that options for both types of glycosylation exist.[23–26]

Processing in *Pichia*

Most secreted recombinant proteins produced in *Pichia* are constructed as fusions to the α-factor prepro sequence (αMFpp) derived from *Saccharomyces cerevisiae*.[72] The α-factor peptide itself is separated by an EAEA spacer from the pro-sequence, necessitating a two-step cleavage by Kex2 and DPAPase A.[73] Next to making the Kex2 processing site accessible, the spacer could also help to keep the resulting product more aggregation resistant by providing charged extensions. While the natural α-factor peptide is soluble, the precursor contains several copies of the mature peptide, and thus, it is conceivable that some intermediate processing products will benefit from increased solubility.

Once a recombinant protein is to be secreted with this system, the question arises whether or not this spacer should be included. It has previously been shown that its inclusion can increase the amount of correctly processed material by preventing steric hindrance of the Kex2 cleavage site.[74] Whether this is necessary for the specific POI may depend on its primary sequence around the cleavage site,[75] its structure and thus the accessibility to Kex2 and, potentially, its aggregation tendency, as postulated in this study. While the inclusion of the spacer generally enhances the cleavage efficiency of this endopeptidase specific for dibasic sites, several studies have demonstrated problems with the subsequent proteolytic processing of the EA repeats.[54–56,76] Apparently, the quantity and/or activity of the *STE13*-encoded DPAPase A in the secretory pathway is not sufficient to process these

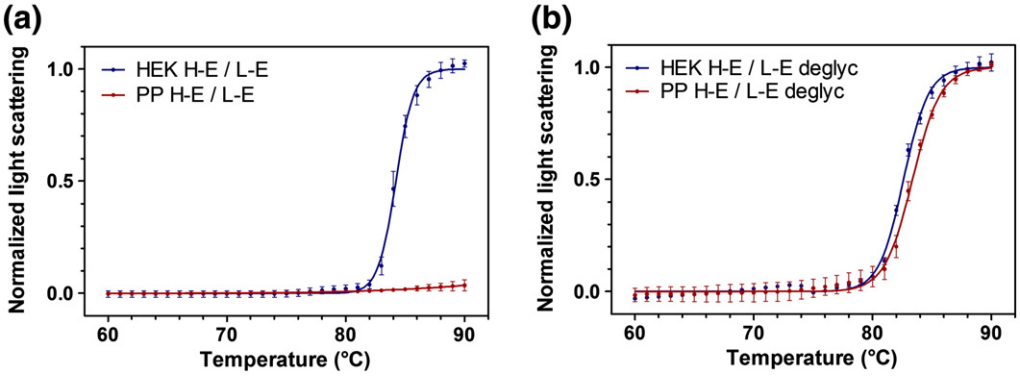


Fig. 11. Aggregation behavior of the same IgG constructs produced in different mammalian expression systems. (a) Aggregation of the IgG H-E/L-E construct produced in either *Pichia* or mammalian cells (HEK293). Having identical amino acid sequences, these samples only differ with respect to their posttranslational modifications such as glycosylation. Aggregation was measured by light scattering at 500 nm and plotted as a function of temperature. (b) Aggregation of the same IgGs as in (a) in their deglycosylated state. Samples were treated by PNGase F prior to analysis by light scattering.

repeats for large amounts of recombinant protein expressed under strong promoters. The resulting secretion of molecules with amino-terminal EAEA extensions was also observed for both the heavy and the light chains of the initial IgG constructs expressed in the *Pichia* system in this study.

Since the presence of these amino-terminal extensions can be undesirable for many applications and even might create a new immunogenic epitope,[77] the omission of the EAEA tetrapeptides was evaluated, aiming to achieve authentic N-termini at both heavy and light chains. This simple strategy worked well for the expression of the light-chain components, leading to proteins of the correct sequence, while causing only a slight reduction in the total amount of secreted IgGs (Fig. 4c, H-E/L-E *versus* H-E/L).

However, deleting the EA dipeptides from the heavy-chain constructs caused severe problems in its maturation process. Three species of resulting heavy-chain proteins were observed: (i) a fraction containing the complete and unprocessed pro-region, (ii) molecules containing the last nine amino acids of the pro-region on its N-terminus (resulting from incorrect processing) and (iii) the correctly processed heavy chain. Similar results were achieved when analogous changes were carried out with Fab fragments (data not shown). The underlying decreased Kex2 cleavage efficiency in the absence of a spacer is in agreement with previous studies reporting that expression of proteins without the EAEA spacer peptide led to the secretion of incompletely or incorrectly processed proteins.[78–80] Remaining precursor was observed when onconase was fused to the αMFpp without the EAEA spacer.[81] The problem of uncleaved pro-regions seems not to be specific for the *Pichia* system, as similar findings had already been observed earlier for the expression of an insulin precursor in *S. cerevisiae* with the αMFpp system in the absence of any EAEA spacer.[82]

The presence of an additional cleavage site within the pro-region was also previously described for a different single-chain antibody in V_H–linker–V_L orientation.[55] However, in that case, 11 residues from the carboxy-terminus of the pro-region were found to be attached to the protein, resulting from a cleavage at position 57 of the pro-region between ..-I-A‖A-K… (where ‖ indicates the Kex2 cleavage site). In our case, this cleavage was shown to have occurred between ..-A-K‖E-E-.. and thereby two residues further downstream in the pro-region, resulting in an extension of only nine residues. This site actually seems to be more reasonable for Kex2 cleavage, taking into account that the "original" Kex2 cleavage site (..-K-R‖E-A-..) has a similar charge distribution, possessing a basic amino acid in the P1 position just upstream of the cleavage site and an acidic residue in the adjacent P1′ position. While general DPAPase A processing can probably be increased by just overexpressing the enzyme, it would be expected that this incorrect cleavage would remain as a side reaction, and thus, DPAPase A overexpression might not lead to homogeneous product either.

Secretion without the pre-pro-α-factor pro sequence

In principle, the signal sequence alone is sufficient for antibody secretion to the medium in eukaryotic cells, and the pro-sequence of the α-factor is not needed. The role of the pro-region for efficient secretion has been investigated before for other proteins such as the lipase Lip1p.[75] This protein could be expressed and secreted under the direction of either the α-factor pre or the pre-pro sequence; however, the highest level of secretion was achieved by the construct with the full pre-pro sequence. In contrast, here, the deletion of the complete pro-region from the heavy chain did not only lead to a product with the correct N-terminus but also increased the amount of correctly assembled antibody found in the cell supernatant by 25% or 40% for the IgG or Fab fragment, respectively (Fig. 4c and d). While this study revealed that deleting the pro-region is the only option for the heavy chain in obtaining homogeneous protein with the genuine N-terminus, correctly processed light chain can be produced either with the pro-region (in the absence of EAEA) or even in the absence of the pro-region. Yet, for the light chain, the deletion of the pro-region resulted in a reduction in total protein yield by ~30% ("H-pro-del/L" *versus* "H-pro-del/L-pro-del" constructs; see Fig. 4c). Thus, the highest yield of genuine IgG without N-terminal extensions can be obtained after deletion of the complete pro-region for the heavy chain and the presence of the pro-region but the absence of the EAEA tetrapeptide for the light chain (Fig. 4c).

The tetrapeptide extensions influence protein aggregation but not stability

In the analysis of the impact of various amino-terminal extensions on the biophysical characteristics of the IgGs, the ability to attach any extension to the N-terminus of any chain was crucial. Therefore, these constructs were expressed in HEK293 cells. Analysis of the sequences of the new constructs (Fig. 5b) by the signal sequence prediction software SignalP 3.0‡[83] predicted that all variants should be processed correctly. Indeed, all constructs could be expressed and were confirmed to be homogeneous and in agreement with their design by MS, SDS-

‡ www.cbs.dtu.dk/services/SignalP

PAGE and sandwich ELISA. Nonetheless, compared to the H/L construct, slightly lower levels of IgG were found in the supernatant (~70% for the H/L-E and H-E/L constructs and 88% for the H-E/L-E variant, respectively) (Fig. 5d). Further investigations will have to clarify which step in biosynthesis is responsible for these differences.

Once stable cell lines were established for all mammalian constructs, their biophysical properties could be analyzed in detail. As expected, the amino-terminal extensions of either the heavy or the light chain—individually or combined in the same construct—only very slightly influenced these characteristics (Fig. 6). Only a slight decrease of ~1 °C was encountered for H-E/L-E compared to the original H/L construct in the Fab transition within the IgG in DSC (Fig. 6c and Supplementary Data Fig. S4). A dramatic difference, however, became apparent for the aggregation susceptibilities of the different constructs. As depicted in Fig. 7, the presence of the amino-terminal EAEA tetrapeptide at either N-terminus generally enhanced the temperature of aggregation onset. Strikingly, this increase was more pronounced for the EAEA extension on the light chain. From a comparison with the extension with the AGIQ control tetrapeptide (Supplementary Fig. S4), these findings indicate that the two negative charges of the EAEA extension could be partially responsible for the reduced susceptibility of the resulting IgG molecules toward aggregation.

The influence of charge and location on the effect of the EAEA extension

Our data indicate that the EAEA tetrapeptide extension has more influence on preventing aggregation once it is added to the light chain rather than to the heavy chain of the antibody studied (Fig. 7c). One obvious difference between the two chains is the presence of pyroglutamate as the N-terminal residue of the heavy chain produced in mammalian cells. MS data and the results of Edman sequencing indicated that the encoded glutamine is indeed converted into pyroglutamate, as commonly seen for IgGs in general.[51] Since pyroglutamate formation from glutamine results in the loss of the positively charged N-terminal amino group, it slightly lowers the isoelectric point of the antibody.

If any tetrapeptide (either AGIQ or EAEA) is N-terminally attached to the heavy chain, the cyclization reaction forming pyroglutamate cannot take place anymore, and the new construct will carry this original positive charge at the amino terminus of its glutamine residue again plus any charges introduced by the peptide side chains. Therefore, the addition of AGIQ adds an additional positive charge to the IgG, while the addition of the EAEA tetrapeptide to the heavy chain introduces only one new net negative charge (the second one is canceled out by the positive charge of the free N-terminal amino group). Interestingly, the newly N-terminally located glutamic acid of the EAEA tetrapeptide was not found to have undergone the cyclization reaction into pyroglutamate (data not shown), even though the formation of pyroglutamate from N-terminal glutamate should in principle be possible as well.[84]

When the EAEA extension is present on the light chain, which carries an N-terminal amino group, both its negative charges are added to the protein's net charge. Taking all findings together, it seems that the EAEA tetrapeptide with its two negative charges is beneficial for the IgG's resistance to aggregation. In general, its exact location within the molecule might *per se* not be of great importance, as the H/L-EA construct (where the EAEA extension is spaced by the AGIQ tetrapeptide from the N-terminus of the light chain) also exhibited an increased aggregation resistance. However, the loss of the pyroglutamate residue might reduce the effect of this extension at the heavy chain.

Our hypothesis that the positive impact of the EAEA tetrapeptide is very likely based on its effect on the overall charge of the antibody molecule was further corroborated by the observation that the exact temperature of aggregation onset depends on the pH of the sample (Fig. 9). As the charge of the H/L construct is almost neutral at the pH value generally used throughout this study (pH 7.1), this sequence extension has a rather dramatic effect, causing the whole molecule to possess a clearly negative net charge. The only marginally negative charge of the original constructs at pH 7.1 originates from the nature of their CDR regions. While the isoelectric points (p*I*) of the backbones are 8.2 and 5.5 for the heavy and light chains, respectively (determined by the analysis tool on the ExPASy server§), the basic p*I* of the CDRs within the heavy chain are "compensated" by a very acidic p*I* of the light chain's CDRs (9.2 and 4.1, respectively). The reason for the pH affecting mostly the net charge of the heavy chain lies in its larger number of histidine groups.

Application aspects of these findings

The beneficial effect of the EAEA tetrapeptide on the aggregation susceptibilities of the analyzed antibodies was not only seen when the IgGs were analyzed by gradual heating. When freshly produced IgG variants were incubated for 5 days at various temperatures, clear differences in the formation of aggregates could be detected by MALS analyses (Fig. 8). Aggregate accumulation was seen for the H/L construct, while for the EAEA-

§ http://web.expasy.org/compute_pi

containing H/L-E construct, neither the amount nor the size of aggregated IgGs noticeably increased. Also, when the antibodies were stored at 4 °C for several months and subsequently analyzed by SEC, the H-E/L-E construct possessed the smallest aggregated fraction. The percentages, comparing this fraction's aggregates with that of H/L aggregates (0.5% *versus* 3%), might sound insignificant and negligible; however, it should be kept in mind that the samples were stored at a relatively low protein concentration of 1 mg/ml. Since antibodies administered in the clinic are generally stored and used at much higher concentrations, this small difference might then increase dramatically.

In the transfer of the results of this study to other antibodies, it is important to understand to what degree the findings can be generally applied. Using another unrelated IgG, whose p*I* is substantially higher due to its CDR regions (light chain alone, 6.9 *versus* 4.8; complete IgG, 8.9 *versus* 7.2) and which is thereby more positively charged at pH 7.1, we found that there was only a small influence of the EAEA tetrapeptide (Fig. 10b). To test these effects, we produced IgGs in HEK293 cells (which leads to correct processing to the genuine N-terminus) or in *Pichia*, where the EAEA tetrapeptide was still attached on both chains, and subsequently analyzed them in glycosylated and deglycosylated states. Compared to the previously analyzed antibody, this IgG has a substantial positive net charge at neutral pH (Fig. 10c), and even though its charge is diminished by the EAEA extension, this has only a very small effect. It thus implies that the EAEA extension will be most promising for antibodies whose net charge can be boosted significantly by this extension.

The results obtained from the second IgG, however, also indicated the transferability of the other finding of this study. The glycosylated *Pichia*-produced IgG generally showed a higher resistance to aggregation compared to its counterpart from mammalian cells. This result suggests that the improved aggregation protection by the mannose-rich oligosaccharide is seen in this antibody too, despite its high isoelectric point. It is thus likely to be a general phenomenon that might be transferable to other molecules as well.

Taken together, the present finding indicate that the convenient production in the *Pichia* host has an unexpected benefit, as it leads to antibodies being more resistant to aggregation due to both the *Pichia* glycosylation and the residual tetrapeptides left by the α-factor pre-pro secretion system. Importantly, through these modifications, aggregation of the IgG was also prevented under physiological conditions, such as after an extended time at 37 °C. Nonetheless, the magnitude of these effects will be antibody dependent, with the net charge seemingly being a good predictor.

Comparison with other proteins

The present findings are in agreement with previous analyses showing that the solubility of proteins can be enhanced by the introduction of charged residues resulting in an altered overall protein charge.[85,86] In addition, our results match previous studies showing that single V_H domains that are by themselves resistant to aggregation generally possess a greater negative net charge than their aggregation-prone counterparts.[86–88] This correlation has, however, never been reported for full-lengths IgGs so far. A similar connection between p*I* values and aggregation susceptibility has been presumed for single-domain camelid V_{HH} antibodies. Although possessing partially lower conformational stabilities compared to their human V_H counterparts, these molecules mainly unfold reversibly and resist aggregation.[42,89] Following a similar reasoning, a human V_H domain prone to aggregation could be converted into an aggregation-resistant variant by introducing negatively charged residues within or near the CDR1 loop.[88] However, in the same study, it was shown that other negatively charged mutations in the most solvent exposed residues outside of the CDRs were incapable of suppressing the parent V_H's aggregation propensity. Therefore, the authors concluded that the simple difference in the net charge of this domain antibody might not be responsible for the improvements of the aggregation propensities. Instead, the specific locations of the mutations seemed to be of great importance. Our results, however, indicate that an N-terminal placement of negatively charged residues might be sufficient to increase aggregation resistance. This strategy circumvents the concern that the introduction of charged residues into the CDR regions may abolish the antibody–antigen recognition directly or indirectly by altering the loop conformation.[88]

Generally, it appears that the N-terminal location is particularly suitable for the addition of charges, as it has diminished the aggregation of a molecule as large as an IgG. Compared to the previously mentioned study by Perchiacca *et al.*,[88] the strategy resulting from our results should be more easily applicable to a large variety of molecules, as the charged EAEA extension can readily be attached to any molecule by simple cloning. In addition, there is no risk that these new residues might disturb the structure of the IgG as they are not inserted within the domains. Importantly, the present study could confirm that the various amino-terminal extensions did not affect the binding of the IgG variants to their antigen. As depicted in Supplementary Data Fig. S3, these interactions were analyzed for several mammalian constructs, and the determined affinities were unaltered within experimental error.

Conclusions

Taken together, our results indicate that the production of full-length IgG in the yeast *Pichia*—being attractive for large-scale preparations due to its convenient handling—is possible and results in molecules with the desired antigen binding and thermal stability. However, obtaining correctly processed IgGs requires the direct fusion of a signal sequence to the heavy chain. When using the commonly utilized α-factor pre-pro secretion system, the secreted antibodies will contain the EAEA extension at both N-termini. Removing this extension from the coding sequence can easily be performed for the light chain but will generate heterogeneous processing for the heavy chain and correct processing requires removal of the whole α-factor pro region. Most surprisingly, antibodies produced in *Pichia* proved to be more resistant to aggregation than their counterparts with identical primary sequences produced in mammalian cells. This was shown to be caused by the mannose-rich glycosylation and the residual EAEA extensions. While the former effect appears to be general, the latter is most pronounced when this tetrapeptide addition can significantly increase the net negative charge of the antibody. Furthermore, these results show that the addition of a mere four amino acids to a protein of several hundred residues can have a dramatic impact on its biophysical characteristics. In a broader view, these findings could have important implications for the common approach of adding tags to a POI, as it is generally assumed that few additional amino acids will not have major effects on the molecule and its properties.

Materials and Methods

Cell culture

Materials and cultivation of mammalian and Pichia cells

All media and supplements for mammalian expression were purchased either from Sigma-Aldrich (Missouri, USA), Invitrogen (California, USA) or Amimed (BioConcept, Switzerland). The antibiotics that were used, Zeocin™ and Hygromycin B, were bought from Invitrogen and PAA (Austria), respectively. All solutions used were sterilized by filtration through 0. 22-μm filters (Millipore, Massachusetts, USA). Stably transformed HEK293 cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Sigma-Aldrich; high glucose, 4.5 g/l) supplemented with 10% (*v*/*v*) heat-inactivated fetal bovine serum (FBS; Amimed) in a humidified incubator under 5% carbon dioxide at 37 °C. Expression of IgGs was carried out in DMEM supplemented with 5% (*v*/*v*) FBS (instead of the more commonly used 10%). For secretion of IgGs, derivatives of the vector pcDNA5 (Invitrogen) containing constitutively active cytomegalovirus promoters upstream of the endogenous IgG signal sequences were used.

For all work with *Pichia*, the strain SMD1163 (*his4 pep4 prb1*; Invitrogen) was used. All media and supplements for this work were purchased from either Sigma-Aldrich or Invitrogen. All work was performed in a sterile laminar flow bench, and yeast growth was generally performed at 30 °C. Selection of clones stably expressing the IgGs was based on Zeocin™ resistance. For secretion of IgGs, derivatives of the pGAPZαB vector (Invitrogen) containing the constitutively active yeast GAP promoter followed by the α-factor pre-pro (αMFpp) region were used. YPD (*y*east extract–*p*eptone–*d*extrose) medium containing 20 g/l peptone, 10 g/l yeast extract and 20 g/l D-glucose (plus 20 g/l agar for YPD-agar) was used for routine growth and subculturing of *Pichia* cells. Zeocin™ was added to a final concentration of 100 μg/ml. IgG expression was performed in buffered medium with glycerol for yeast (BMGY) [20 g/l peptone, 10 g/l yeast extract, 100 mM potassium phosphate (pH 6.0), 1.34% yeast nitrogen base without amino acids, 1% (*w*/*v*) glycerol and 400 μg/l biotin] in the absence of antibiotic.

Construction of expression plasmids

Unless stated otherwise, all molecular biology methods were performed according to standard protocols.[90] All enzymes used for cloning were purchased from New England Biolabs (Massachusetts, USA) or Fermentas (Germany). In general, cloning and propagation of all plasmids was carried out in *Escherichia coli* DH5α (Life Technologies, California, USA), grown at 37 °C in low-salt LB broth [10 g/l bacto-tryptone, 5 g/l yeast extract and 5 g/l NaCl (pH 7.5)] containing 25 μg/ml Zeocin™ for the *Pichia* plasmids or in 2YT broth [16 g/l bacto-tryptone, 10 g/l yeast extract and 5 g/l NaCl (pH 7.5)] containing 100 μg/ml ampicillin (AppliChem, Germany) for mammalian vectors.

Construction of initial expression plasmids will be described in detail elsewhere (Schaefer *et al.*, unpublished). In summary, the vectors used for the creation of stable cell lines contained the genes for the light and heavy chains as complete individual expression cassettes; that is, both genes had their own constitutively active promoter and their own downstream polyadenylation sequence. The attachment of additional appendices to the N-termini of mammalian heavy and light chains was achieved by overhanging primers and PCR. The upstream regions of the heavy- or light-chain coding sequences were amplified with forward primers annealing upstream of the cytomegalovirus promoter region and reverse primers introducing the new N-terminal appendices. The following forward primers were used for the heavy (H) or the light (L) chain: HEK_H_for (5′ GCG TTT CTG GGT GAG CAA AAA CAG GAA GGC 3′) and HEK_L_for (5′ GGC ACT GTC CTC TCA TGC GTT GGG TCC 3′). The reverse primers were HEK_H-E_rev (5′ GCC AGC *CAA TTG* CAC CTG *AGC TTC AGC CTC* GGA CAG GAC CCA TCT GGG AGC TGC CAC CAG CAG G 3′), HEK_H-A_rev (5′ GCC AGC *CAA TTG* CAC CTG *TTG AAT TCC TGC* GGA CAG GAC CCA TCT GGG AGC TGC CAC CAG CAG G 3′), HEK_L-E_rev (5′ GCAT *GATATCAGC TTC AGC CTC* AGC CCA GGA TCC TGT GCC CTG AGT GAG G 3′), HEK_L-A_rev (5′ GCAT *GATATCTTG AAT TCC TGC* AGC CCA GGA TCC TGT GCC CTG AGT GAG G 3′) and

HEK_L-EA_rev (5′ GCAT *GATATCTTG AAT TCC TGCAGC TTC AGC CTC* AGC CCA GGA TCC TGT GCC CTG AGT GAG G 3′). Regions coding for the new appendices are underlined, while restriction sites used (MfeI or EcoRV for the heavy or the light chain, respectively) are printed in italics.

For each heavy- or light-chain fragment, a standard PCR using Phusion high-fidelity DNA polymerase (Finnzymes, acquired by Thermo Scientific) was carried out with a final concentration of 0.5 μM for each forward and reverse primer in the presence of 3% (*v*/*v*) dimethyl sulfoxide. The PCR parameters were as follows: pre-denaturation at 98 °C for 30 s, followed by 27 cycles of denaturation at 98 °C for 10 s, and annealing and extension at 72 °C for 50 s, followed by a final incubation for 5 min at 72 °C. The resulting PCR products were purified using a PCR clean-up kit (Macherey-Nagel, Germany) and digested with BglII and MfeI for the heavy chain or AflII and EcoRV for the light chain, respectively. Finally, these inserts were ligated into the identically treated and dephosphorylated initial expression vectors, yielding the final plasmids subsequently sequenced using standard techniques.

The removal of the EAEA tetrapeptide from the initial yeast heavy-chain construct was performed by assembly PCR in three steps: (i) amplification of both the region upstream of the intersection of the αMFpp with the gene encoding either the heavy or the light chain and the region downstream of it, both lacking the sequence encoding the EAEA tetrapeptide; (ii) assembly of fragments by an overlapping PCR method; and (iii) amplification of full-length products. The upstream region was amplified using the oligonucleotides PP_out_H_for (5′ GGA AGG AGT TAG ACA ACC TGA AGT CTA GGT CCC 3′) and the reverse PP_H_rev (5′ *C CAA TTG CAC CTGTCT TTT CTC GAG AGA TAC CCC TTC TTC TTT AGC* 3′), while the downstream fragment was amplified by using PP_out_H_rev (5′ CCG GAG ACA GGG AGA GGC TCT TCT GC 3′) and PP_H_for (5′ *CT CTC GAG AAA AGACAG GTG CAA TTG GTA CAG TCT GGT CCG GG* 3′). For the deletion of the EAEA tetrapeptide from the light-chain construct, the primers were the following: PP_out_L_for (5′ CGT GGA GGT GCA TAA TGC CAA GAC AAA GCC GC 3′), PP_L_rev (5′ *GGT CAG TTC GAT ATCTCT TTT CTC GAG AGA TAC CCC TTC TTC TTT AGC* 3′), PP_out_L_rev (5′ CTA GGA CGG TTA ACT TCG TGC CGC CGC C 3′) and PP_L_for (5′ *CT CTC GAG AAA AGA GAT ATC GAA CTG ACC CAG CCG CCT TCA GTG* 3′). Regions complementary to the mature N-terminus of the heavy or light chain are underlined, while those annealing to the remaining pro-region of the αMFpp are printed in italics.

The different fragments were generated by PCR using the respective forward and reverse oligonucleotides on the initial expression plasmid. The PCR was performed as follows: 30 s at 98 °C, followed by 25 cycles of 10 s at 98 °C and 50 s at 72 °C, followed by a 5-min final extension cycle at 72 °C. The PCR products were purified using the same PCR purification kit as above and served as templates in the subsequent overlapping PCR. In this second PCR reaction, the outer primers (e.g., PP_out_H_for and PP_out_H_rev for the heavy chain; 0.5 μM each) were added to 15 ng of each purified PCR product. Twenty cycles of amplification (denaturation at 98 °C for 10 s and annealing and extension at 72 °C for 55 s) were performed, resulting in full-length fragments that were MluI/BstEII or HpaI/SacII digested (for the heavy and light chains, respectively) and finally inserted in the appropriately cut initial vector.

The deletion of the complete pro-region in the H-pro-del construct was performed in almost the same manner, albeit using different primers in the first PCR. PP_H-pro-del_for (5′ *CC TCC GCA TTA GCTCAG GTG CAA TTG GTA CAG TCT GGT CCG G* 3′) and PP_H-pro-del_rev (5′ *C CAA TTG CAC CTGAGC TAA TGC GGA GGA TGC TGC GAA TAA AAC AGC* 3′) were used. The oligonucleotides for the pro-region deletion of the light-chain construct were PP_L-pro-del_for (5′ *CC TCC GCA TTA GCT GAT ATC GAA CTG ACC CAG CCG CCT TCA GTG* 3′) and PP_L-pro-del_rev (5′ *GGT CAG TTC GAT ATC AGC TAA TGC GGA GGA TGC TGC GAA TAA AAC AGC* 3′). As in the primers above, regions complementary to the mature N-terminus of the heavy chain are underlined, while this time, those parts annealing to the remaining pre-region of the αMFpp are printed in italics.

Design of glycan knockout mutants T299A

To analyze the influence of the glycan moiety attached to Asn297 in the C_H2 domain of the heavy chains, we designed the glycan knockout construct T299A. The ACG, coding for the threonine in the glycosylation motif Asn297-Ser298-Thr299, was mutated to the alanine encoding GCG using the QuikChange® site-directed mutagenesis kit from Stratagene (acquired by Agilent, USA) according to the manufacturer's instructions. Mutagenesis was performed with the following primers encoding the mutated triplet: IgG_deglyc_for (5′ G GAG CAG TAC AAC AGC **G**CG TAC CGG GTG GTC 3′) and the complementary reverse primer IgG_deglyc_rev (5′ GAC CAC CCG GTA **C**GC GCT GTT GTA CTG CTC C 3′). After selection and sequencing, a correct clone was digested by EcoNI, cleaving twice in the human γ1 constant region within the C_H1 and C_H3 domains, and the resulting fragment was inserted in the original vector cut with the same enzyme. After the selection of a construct with the correct orientation of the insert, the final vectors were re-sequenced and prepared for transformation.

Creation of stable cell lines

Flp-In HEK293 cell lines (Invitrogen), stably expressing the IgG of interest and secreting it into the culture medium, were generated according to the instructions of the Flp-In system manual. For this purpose, the expression cassettes for both heavy- and light-chain genes were cloned into the pcDNA5/FRT vector and finally inserted at a specific location in the genome by the Flp recombinase within the cell. After selection, 10 single clones of each construct were tested for expression of the desired IgG by Western blot analysis of the supernatant. More than 90% of the analyzed clones were positive for the desired IgG.

For the *Pichia* constructs, the corresponding plasmids containing both genes for the light and heavy chains under the control of constitutively active GAP promoters within one plasmid were linearized by BglII and integrated into the yeast genome upon transformation of competent *Pichia*. Finally, several colonies per construct were

replated on fresh YPD-agar plates and subsequently analyzed for IgG expression by Western blot.

Large-scale expression of IgG

For large-scale expression of mammalian IgGs, 3×10^7 cells stably expressing the respective IgG were seeded in 500-cm^2 plates (Nunc), containing 65 ml of regular medium (DMEM and 10% FBS). At confluency (approximately 3 day s after seeding), this medium was removed and replaced by 50 ml expression media, containing only 5% FBS. The supernatant containing the secreted IgGs was harvested 3–4 days later and renewed three times for additional rounds of expression. Detached cells were removed from the harvested supernatant by centrifugation (10 min; 9000*g*; 4 °C) and filtration through a 0. 22-μm filter (Millipore). The filtered supernatant was adjusted to binding conditions by the addition of half the volume of 200 mM sodium phosphate (pH 8.0) and 150 mM NaCl, refiltered and stored at 4 °C until IgG purification by Protein A affinity chromatography.

Large-scale expression in the yeast system was performed using a three-step procedure using two different, consecutive pre-cultures. First, 5 ml YPD medium was inoculated with a single *Pichia* colony and grown overnight. The next day, the second pre-culture of 100 ml YPD media was inoculated with 2 ml of the small-scale overnight culture in a 300-ml Erlenmeyer flask. After 24 h at 30 °C and 250 rpm, the final expression culture (1 l BMGY medium in 5-l Erlenmeyer flasks) was inoculated with the second pre-culture at an optical density (OD_{600}) of 1. For this inoculation, the appropriate amount of cells from the pre-culture was pelleted by centrifugation (10 min; 5000*g*; 4 °C) and resuspended in the expression media. After incubation for 48–60 h at 30 °C and 250 rpm, the expression culture was centrifuged (15 min; 5000*g*; 4 °C), and the supernatant was adjusted to the conditions required for IgG purification. Since the BMGY medium already contained potassium phosphate buffer at a 100-mM concentration, no sodium phosphate but only NaCl was added to the supernatant. The pH was adjusted to 8.0 by the addition of NaOH. Afterwards, the supernatant containing the IgGs was filtered through a 0. 22-μm filter and loaded onto the Protein A column.

Biophysical and biochemical analysis

Purification of IgGs

Antibodies were purified from culture supernatants by Protein A affinity chromatography. For this purpose, the supernatants—adjusted to pH 8.0 and the appropriate [Na^+] (see above)—were loaded onto HiTrap Protein A columns (GE Healthcare, USA) at 4 °C with a flow rate of 1 ml/min. Chromatography was performed using an ÄKTA PrimePlus chromatography system (GE Healthcare) at 4 °C. After loading, the column was washed with 100 mM sodium phosphate buffer (pH 8.0) containing 150 mM NaCl. Elution of IgG was accomplished by using 0.1 M glycine (pH 2.7), followed by immediate neutralization of each fraction to pH 7.5 using 1 M Tris (pH 8.0). The concentrations of the sample fractions were determined by UV spectroscopy at 280 nm with a NanoDrop ND-1000 spectrophotometer (Thermo Scientific), assuming a mass extinction coefficient of 13.7 for a 10-mg/ml solution of IgG. The samples with the highest protein concentration were pooled and dialyzed twice against PBS [Sigma-Aldrich; 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl and 137 mM NaCl (pH 7.1)] at 4 °C. After dialysis, the samples were filtered through 0. 22-μm filters (Millipore) and stored at 4 °C.

SDS-PAGE and Western blotting

SDS-PAGE was performed using Tris–glycine, 7.5% or 12% gels for nonreduced or reduced samples, respectively. Gels and samples were prepared according to standard protocols,[91] and 1.5 μg of IgG antibody was loaded per lane. Nonreduced samples were only mixed with loading buffer, while, for reduced samples, DTT was added to a final concentration of 16 mM followed by incubation for 5 min at 95 °C. As molecular mass marker, the PageRuler prestained protein ladder (Fermentas) was used.

For the specific detection of the IgGs, the proteins were transferred to a polyvinylidene fluoride membrane (Millipore) after their separation by SDS-PAGE. The blotting was performed following the semidry method within a Trans-Blot SD cell (Bio-Rad, California, USA) for 60 min at 11 V, using a Tris–glycine buffer (20 mM Tris, 150 mM glycine and 10% methanol). After blocking the membrane in 5% MPBST [5% (*w*/*v*) skimmed *m*ilk dissolved in *PBST* (*PBS* with 0.05% *T*ween 20)] for at least 1 h, the membrane was incubated for another hour in 2.5% MPBST containing detection antibodies conjugated to alkaline phosphatase. For detection of the heavy chain, a goat anti-human Fc-specific antibody (Sigma-Aldrich; #A9544; 1:4000 dilution) was used; for the light chain, a goat anti-human lambda light-chain-specific antibody (Sigma-Aldrich; #A2904; 1:1500 dilution) was used. The polyclonal rabbit αMFpp antiserum was a kind gift of Dr. Randy Schekman (University of California, Berkeley) and used in a 1:1000 dilution, followed by a secondary anti-rabbit IgG specific antibody conjugated to alkaline phosphatase (Sigma-Aldrich; #A3687; 1:20,000 dilution). Antibodies were detected by the addition of nitro-blue tetrazolium chloride with 5-bromo-4-chloro-3′-indolyphosphate *p*-toluidine salt (Sigma-Aldrich) solution in 100 mM Tris (pH 9.5), 5 mM $MgCl_2$ and 100 mM NaCl; the development was terminated when the first bands became clearly visible by washing the membrane in ultra high pure water.

Isoelectric focusing

Determination of the isoelectric point of the constructs was carried out in IEF Ready Gels (pH 3–10; Bio-Rad) according to the manufacturer's instructions on a Mini-PROTEAN II vertical cell system (Bio-Rad), using increasing voltages (100 V for 1 h, 250 V for 1 h and 500 V for 30 min). Next to the samples, IEF standards, p*I* range 4.45–9.6 (Bio-Rad), were co-electrophoresed. The cathode and anode buffers were 20 mM lysine/20 mM arginine or 7 mM phosphoric acid, respectively. Following electrophoresis, the proteins were visualized by Coomassie Blue—Crocein Scarlet staining [0.04% (*w*/*v*) Coomassie R-250 and 0.05% (*w*/*v*) Crocein Scarlet 3B in 27% isopropanol/10% acetic acid] for 45 min and destaining (40% methanol/10% acetic acid).

N-terminal sequence analysis

Sequence analysis was performed after incubating 50 μg purified IgGs with 0.3 mU of *Pfu* pyroglutamate aminopeptidase (TaKaRa Biomedicals, Japan) as described previously[92] to remove blocking pyroglutamate residues. Reduced antibodies were separated using SDS-PAGE and transferred to polyvinylidene fluoride membranes using Tris–glycine buffer as described above. After transfer, the membrane was stained with freshly prepared Coomassie stain (0.1% Coomassie Blue R-250, 40% methanol and 1% acetic acid) for 30 s, followed by destaining with 50% methanol until protein bands were clearly visible. Protein bands corresponding to the expected molecular weight of the heavy and light chains, respectively, were excised and submitted for N-terminal sequencing analysis. The first 5–8 aa of purified IgG from *Pichia* was sequenced at the Functional Genomic Center Zurich using an Applied Biosystems Procise 492 cLC protein sequencer.

Analysis of IgG expression levels by ELISA

To analyze the influence of N-terminal variations on the IgG expression and secretion levels, we seeded the same number of HEK293 cells stably expressing the corresponding constructs in 12-well plates. Upon reaching confluency, the medium was removed and replaced by 1 ml expression media, containing 5% FBS. Twenty-four hours later, the supernatant was analyzed for its IgG content by ELISA. For the *Pichia*-produced IgGs, optical-density-normalized aliquots were taken after 24 h of expression for ELISA analysis. For ELISA, a capture antibody recognizing the human IgG heavy chain (Jackson ImmunoResearch, Pennsylvania, USA; #209-005-098) was immobilized on MaxiSorb plates (Nunc) overnight at 4 °C. After 1 h of blocking in 5% skimmed milk in PBST, 100 μl of *Pichia*- or HEK-derived supernatant (diluted in a range of 1:25 to 1:100 in fresh BMGY or DMEM media) was incubated for 1 h at room temperature. The expressed IgG molecules were detected by incubating with an anti-human lambda light-chain-specific antibody conjugated to alkaline phosphatase (Sigma-Aldrich; #A2904; 1:2000 dilution) for 1 h and subsequent addition of *p*-nitrophenyl phosphate (Sigma-Aldrich). Absorbance at 405 nm was measured using a Perkin Elmer HTS 7000 Plus plate reader for up to 1 h.

CD spectroscopy

CD measurements were performed on a Jasco J-810 Spectropolarimeter (Jasco, Japan) equipped with a computer-controlled water bath (refrigerated circulator FS18; Julabo, Germany), using a 0.5-mm cylindrical thermocuvette. CD spectra were recorded from 200 to 250 nm with a data pitch of 0.5 nm, a scan speed of 20 nm/min, a response time of 4 s and a bandwidth of 2 nm. Measurements were performed at 25 °C, and each spectrum was recorded three times and averaged. The CD signal was corrected by buffer subtraction and converted to MRE (*m*ean *r*esidue *e*llipticity; θ) using the concentration of the sample determined spectrophotometrically at 280 nm. Heat denaturation curves were obtained by measuring the CD signal at 208 nm at temperatures increasing from 25 °C to 90 °C (heating rate, 1 °C/min; response time, 4 s; bandwidth, 2 nm). Data were collected and processed as described above. CD spectra and denaturation curves of the purified IgGs were measured in PBS (Sigma-Aldrich; pH 7.1) at a protein concentration of 5 μM.

Fluorescence spectroscopy

This method essentially was used to record the red shift of the emission maximum λ_{max} upon heat-induced unfolding of the large IgG proteins with numerous Trp residues. As water penetrates the unfolding IgG structure, the polarity in the vicinity of the Trp residues changes toward a more hydrophilic environment, thus causing an increase in λ_{max}, recorded as a loss in the F_{330}/F_{350} ratio upon unfolding. Fluorescence spectra were measured with a Jobin-Yvon Fluoromax-4 spectrofluorimeter (Horiba Scientific, New Jersey, USA) equipped with a Peltier-controlled cuvette holder. The temperature was controlled by an LFI3751 5A digital temperature control instrument (Wavelength Electronics Inc., Montana, USA). Upon excitation at 295 nm, Trp emission spectra were recorded from 300 to 400 nm ($\Delta\lambda = 1$ nm; scan rate, 1 nm/s) in 0.5 °C steps from 25 °C to 90 °C. The sample cuvette was equilibrated for 2 min at each temperature to ensure that the desired temperature was reached within the cell. Protein concentrations were 1 μM in every case, and all measurements were performed in PBS (pH 7.1). The intensity of the emission spectrum at 330 nm (F_{330}) and 350 nm (F_{350}) was determined at each temperature, and the ratio F_{330}/F_{350} was calculated and subsequently plotted as a function of temperature.

GdnHCl-induced equilibrium unfolding

GdnHCl-induced denaturation measurements were carried out with protein/GdnHCl mixtures containing a final protein concentration of 1 μM and denaturant concentrations ranging from 0 to 4 M (99.5% purity; Fluka, Missouri, USA). These mixtures were prepared from a GdnHCl stock solution (6 M) (in PBS, pH adjusted to 7.1) and equilibrated overnight at 25 °C. Each final concentration of GdnHCl was determined by measuring the refractive index. The intrinsic fluorescence emission spectra were then recorded from 300 to 400 nm with an excitation wavelength of 295 nm. Slit widths of 2 nm were used for both excitation and emission. Individual GdnHCl blanks were recorded and automatically subtracted from the data. The emission ratio F_{330}/F_{350} was calculated and plotted as a function of GdnHCl concentration.

Aggregation assay (light scattering at 500 nm)

The temperature of aggregation was determined by light scattering using a Jobin-Yvon Fluoromax-4 spectrofluorimeter (Horiba Scientific). Excitation and emission wavelengths were set to 500 nm (slit width, 2 nm each; integration time, 1 s). Protein concentrations were 2 μM for HEK-produced IgGs and 4 μM for antibodies produced in *Pichia*. A heating rate of 1 °C/min was applied starting from 25 °C up to 90 °C; the intensities were measured every 1 °C. Each data point depicted in the plots is an average of five measured intensity values. The time period necessary to collect this set of five data points was about ~10 s; therefore, only a negligible intensity change

occurred due to the concomitant rise in temperature (about 0.1 °C). All measurements were performed in triplicates in PBS (pH 7.1), and averaged values are given.

Deglycosylation by PNGase F treatment

The aggregation behavior was analyzed after enzymatic removal of the glycan moiety from Thr299. For this purpose, 80 μl of 2 μM purified IgGs was incubated with 50 U PNGase F (New England Biolabs) at 37 °C for 3 h in PBS (in the absence of any detergents, in contrast to the instructions of the manufacturer). Afterwards, PNGase-F-treated proteins were compared to their glycosylated counterparts by reducing SDS-PAGE for confirmation of complete deglycosylation. Then, these samples were analyzed by the aggregation assay as described before.

Size-exclusion chromatography

SEC experiments were performed using a Superdex 200 PC 3.2/30column (GE Healthcare). The runs were performed in PBS buffer (Sigma-Aldrich; pH 7.1) at a flow rate of 0.06 ml/min at 25 °C on an ÄKTA Micro system (GE Healthcare). Samples of 50 μl containing 6.7 μM IgG were injected, and protein elution was monitored at 280 nm. Cytochrome *c* (12.4 kDa), albumin (66 kDa) and β-amylase (200 kDa) were used as standards to calibrate the column.

Static MALS

The temperature-induced increase of aggregation was analyzed by SEC MALS. Purified IgGs were incubated in PBS (pH 7.1) at a protein concentration of 3 μM at 4 °C, 37 °C or 50 °C for 5 days. Afterwards, SEC was carried out on an LC 1100 Series HPLC System (Agilent) using a Superdex 200 10/300 GL column (GE Healthcare) with a flow rate of 0.5 ml/min. MALS measurements were performed using a miniDAWN MALS instrument (Wyatt Technology, California, USA) to determine the Rayleigh ratio. Three discrete photodetectors are spaced around the flow cell and enabled simultaneous measurements at 45°, 90°and 135°. An Optilab REX Refractometer (Wyatt Technology) was in-line with the MALS detector. All measurements were carried out at 23 °C. Chromatographic data were collected and processed using the ASTRA software (Version 5.3.4.14; Wyatt Technology). Bovine serum albumin (Sigma-Aldrich; 1 mg/ml) was used for the alignment and normalization of various detectors' signals relative to the 90° detector signal.

Electrospray ionization–MS

MS analyses were undertaken by the Functional Genomic Center Zurich. IgG samples were reduced with DTT (50 mM final concentration), desalted using a C4 ZipTip (Millipore) and measured in 50% acetonitrile/0.2% formic acid (pH 2). The m/z data were deconvoluted into MS data using the MaxEnt1 software (Waters/Micromass, Massachusetts, USA).

Differential scanning calorimetry

DSC measurements were performed using a VP-Capillary DSC system (Microcal Inc., acquired by GE Healthcare). The antibody concentrations were adjusted to 0.5 mg/ml prior to the measurement. The corresponding buffer was used as a reference. The samples were heated from 8 °C to 90 °C at a rate of 1 °C/min after an initial 15 min of equilibration at 8 °C. A filtering period of 16 s was used, and data were analyzed using the Origin 7.0 software (OriginLab1 Corporation, Massachusetts, USA). Thermograms were corrected by subtraction of buffer-only scans, and the corrected thermograms were normalized to the molar concentration of the protein. The final excess heat capacity thermogram was obtained by interpolating a cubic baseline in the transition region. The midpoint of a thermal transition temperature (T_m) was obtained by analyzing the data using the Origin 7 software provided with the instrument. As all measured transitions are irreversible, all the experimental values reported in this study for melting temperatures have to be regarded as "apparent" values.

Differential scanning fluorimetry

DSF was performed using the Rotor-Gene Q real-time PCR cycler (QIAGEN, Germany), and fluorescence data were collected using the instrument's HRM channel settings (λ_{ex}, 460 nm; λ_{em}, 510 nm). The SYPRO Orange dye (Molecular Probes) was supplied in dimethyl sulfoxide and diluted 500-fold from the supplied stock solution into the appropriate buffers just prior to being added to the protein solutions. The samples with a final protein concentration of 3.5 μM in a 20-μl reaction mixture were subjected to a temperature ramp from 30 °C to 90 °C at a heating rate of 1 °C/min and at 0.5 °C increments with an equilibration time of 30 s at each temperature prior to measurement. The T_m was determined as the temperature corresponding to the maximum value of the first derivative of the fluorescence changes, calculated by the software. When multiple unfolding transitions are observed, only the T_m value of the first transition can be accurately determined, as the transitions at higher temperatures overlap. Prior to the DSF analysis, several IgGs were analyzed by fluorescence spectroscopy in the presence and absence of SYPRO Orange. In agreement with the literature, the dye (at 1:200- to 1:500-fold dilutions of the original reagent) did not induce any changes in the thermal stability determined by the intrinsic fluorescence acquired during heating.

MST measurements

Binding affinities of purified IgGs to their antigen myoglobin (Sigma-Aldrich) were measured using MST (NanoTemper, Germany) as described previously.[93] Myoglobin was fluorescently labeled according to the manufacturer's instructions with a reactive NT-647 dye using *N*-hydroxysuccinimide ester chemistry that reacts with primary amines to form dye–protein conjugates. For each analyzed construct, a titration series with constant antigen concentration (20 μM) and varying IgG concentrations between 10^{-11} and 10^{-6} M was prepared in PBS. The mixed samples were equilibrated for 1 h at room temperature, and approximately 4 μl of each sample was loaded in the capillary. An infrared laser diode within the Monolith NT.115 instrument (NanoTemper) was used to increase the temperature by 4 K in the beam center. Throughout the measurement, the

fluorescence inside the capillary was recorded by a charge-coupled device camera, and the normalized fluorescence was afterwards plotted against the IgG concentrations. The K_d values were subsequently obtained from fitting the binding curves using Prism 5 (GraphPad, California, USA).

Acknowledgements

We want to thank Dr. I. Jelezarov, Dr. P. Gimeson, Dr. B. Dreier for their help with the DSC and MALS analyses. We are grateful to the personnel of the Functional Genomics Center at University of Zurich for their help with MS analyses and protein sequencing. Our thanks are also due to Dr. S. Duhr from NanoTemper, Germany, for his help and the chance to determine affinities by MST. The polyclonal rabbit αMFpp antiserum was a kind gift of Dr. R. Schekman (University of California, Berkeley). We further thank Dr. Y. L. Boersma and Dr. P. Lindner for critically reading the manuscript and for valuable suggestions and the other members of the Plückthun laboratory for fruitful discussions. J.S. was a recipient of a Kekulé predoctoral fellowship of the German Chemical Industry Association and a member of the Molecular Life Science PhD program. This work was supported by the Schweizerische Nationalfonds grant 3100A0-128671/1 (to A.P.).

Supplementary Data

Supplementary data to this article can be found online at doi:10.1016/j.jmb.2012.01.027

References

1. Plückthun, A. & Moroney, S. E. (2005). Modern antibody technology: the impact on drug development. In *Modern Biopharmaceuticals* (Knäblein, J., ed.), Vol. 3, pp. 1147–1186. Wiley-VCH, Weinheim, Germany.
2. Carter, P. J. (2006). Potent antibody therapeutics by design. *Nat. Rev., Immunol.* **6**, 343–357.
3. Elbakri, A., Nelson, P. N. & Abu Odeh, R. O. (2010). The state of antibody therapy. *Hum. Immunol.* **71**, 1243–1250.
4. Fontoura, P. (2010). Monoclonal antibody therapy in multiple sclerosis: paradigm shifts and emerging challenges. *mAbs*, **2**, 670–681.
5. Demarest, S. J. & Glaser, S. M. (2008). Antibody therapeutics, antibody engineering, and the merits of protein stability. *Curr. Opin. Drug Discov. Dev*, **11**, 675–687.
6. Shire, S. J. (2009). Formulation and manufacturability of biologics. *Curr. Opin. Biotechnol.* **20**, 708–714.
7. Baynes, B. M. & Trout, B. L. (2004). Rational design of solution additives for the prevention of protein aggregation. *Biophys. J.* **87**, 1631–1639.
8. Di Paolo, C., Willuda, J., Kubetzko, S., Lauffer, I., Tschudi, D., Waibel, R. *et al.* (2003). A recombinant immunotoxin derived from a humanized epithelial cell adhesion molecule-specific single-chain antibody fragment has potent and selective antitumor activity. *Clin. Cancer Res.* **9**, 2837–2848.
9. Braun, A., Kwee, L., Labow, M. A. & Alsenz, J. (1997). Protein aggregates seem to play a key role among the parameters influencing the antigenicity of interferon alpha (IFN-α) in normal and transgenic mice. *Pharm. Res.* **14**, 1472–1478.
10. Hermeling, S., Crommelin, D. J., Schellekens, H. & Jiskoot, W. (2004). Structure–immunogenicity relationships of therapeutic proteins. *Pharm. Res.* **21**, 897–903.
11. Schellekens, H. (2005). Factors influencing the immunogenicity of therapeutic proteins. *Nephrol., Dial., Transplant.* **20**, vi3–vi9.
12. Rosenberg, A. S. (2006). Effects of protein aggregates: an immunologic perspective. *AAPS J.* **8**, E501–E507.
13. Kessler, M., Goldsmith, D. & Schellekens, H. (2006). Immunogenicity of biopharmaceuticals. *Nephrol., Dial., Transplant.* **21**(Suppl. 5), v9–v12.
14. Maas, C., Hermeling, S., Bouma, B., Jiskoot, W. & Gebbink, M. F. (2007). A role for protein misfolding in immunogenicity of biopharmaceuticals. *J. Biol. Chem.* **282**, 2229–2236.
15. Ring, J., Seifert, J., Jesch, F. & Brendel, W. (1977). Anaphylactoid reactions due to non-immune complex serum protein aggregates. *Monogr. Allergy*, **12**, 27–35.
16. Ring, J., Stephan, W. & Brendel, W. (1979). Anaphylactoid reactions to infusions of plasma protein and human serum albumin. Role of aggregated proteins and of stabilizers added during production. *Clin. Allergy*, **9**, 89–97.
17. Farid, S. S. (2006). Established bioprocesses for producing antibodies as a basis for future planning. *Adv. Biochem. Eng./Biotechnol.* **101**, 1–42.
18. Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M. & Dwek, R. A. (2007). The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu. Rev. Immunol.* **25**, 21–50.
19. Cregg, J. M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K. & Chappell, T. (2009). Expression in the yeast *Pichia pastoris*. *Methods Enzymol.* **463**, 169–189.
20. Cereghino, G. P., Cereghino, J. L., Ilgen, C. & Cregg, J. M. (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr. Opin. Biotechnol.* **13**, 329–332.
21. Bretthauer, R. K. & Castellino, F. J. (1999). Glycosylation of *Pichia pastoris*-derived proteins. *Biotechnol. Appl. Biochem.* **30**, 193–200.
22. Montesino, R., Garcia, R., Quintero, O. & Cremata, J. A. (1998). Variation in N-linked oligosaccharide structures on heterologous proteins secreted by the methylotrophic yeast *Pichia pastoris*. *Protein Expression Purif.* **14**, 197–207.
23. Jacobs, P. P., Geysens, S., Vervecken, W., Contreras, R. & Callewaert, N. (2009). Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. *Nat. Protoc.* **4**, 58–70.

24. Vervecken, W., Callewaert, N., Kaigorodov, V., Geysens, S. & Contreras, R. (2007). Modification of the N-glycosylation pathway to produce homogeneous, human-like glycans using GlycoSwitch plasmids. *Methods Mol. Biol.* **389**, 119–138.
25. Potgieter, T. I., Cukan, M., Drummond, J. E., Houston-Cummings, N. R., Jiang, Y., Li, F. *et al.* (2009). Production of monoclonal antibodies by glycoengineered *Pichia pastoris*. *J. Biotechnol.* **139**, 318–325.
26. Li, H., Sethuraman, N., Stadheim, T. A., Zha, D., Prinz, B., Ballew, N. *et al.* (2006). Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. *Nat. Biotechnol.* **24**, 210–215.
27. Cregg, J. M., Cereghino, J. L., Shi, J. & Higgins, D. R. (2000). Recombinant protein expression in *Pichia pastoris*. *Mol. Biotechnol.* **16**, 23–52.
28. Couderc, R. & Baratti, J. (1980). Oxidation of methanol by the yeast, *Pichia pastoris*. Purification and properties of the alcohol oxidase. *Agric. Biol. Chem.* **44**, 2279–2289.
29. Digan, M. E., Tschopp, J., Grinna, L., Lair, S. V., Craig, W. S., Velicelebi, G. *et al.* (1988). Secretion of heterologous proteins from the methylotrophic yeast, *Pichia pastoris*. In *Development in Industrial Microbiology* (Pierce, G., ed.), Vol. 29, pp. 59–65. Elsevier Science, Amsterdam, The Netherlands.
30. Ridder, R., Schmitz, R., Legay, F. & Gram, H. (1995). Generation of rabbit monoclonal antibody fragments from a combinatorial phage display library and their production in the yeast *Pichia pastoris*. *Biotechnology (N. Y.)*, **13**, 255–260.
31. Nett, J. H. (2009). Production of antibodies in *Pichia pastoris*. In *Therapeutic Monoclonal Antibodies: From Bench to Clinic* (An, Z., ed.), pp. 569–584, John Wiley & Sons, Inc., Hoboken, NJ.
32. Takahashi, K., Yuuki, T., Takai, T., Ra, C., Okumura, K., Yokota, T. & Okumura, Y. (2000). Production of humanized Fab fragment against human high affinity IgE receptor in *Pichia pastoris*. *Biosci., Biotechnol., Biochem.* **64**, 2138–2144.
33. Lange, S., Schmitt, J. & Schmid, R. D. (2001). High-yield expression of the recombinant, atrazine-specific Fab fragment K411B by the methylotrophic yeast *Pichia pastoris*. *J. Immunol. Methods*, **255**, 103–114.
34. Gach, J. S., Maurer, M., Hahn, R., Gasser, B., Mattanovich, D., Katinger, H. & Kunert, R. (2007). High level expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in *Pichia pastoris*. *J. Biotechnol.* **128**, 735–746.
35. Ogunjimi, A. A., Chandler, J. M., Gooding, C. M., Recinos, A. & Choudary, P. V. (1999). High-level secretory expression of immunologically active intact antibody from the yeast *Pichia pastoris*. *Biotechnol. Lett.* **21**, 561–567.
36. Ellis, S. B., Brust, P. F., Koutz, P. J., Waters, A. F., Harpold, M. M. & Gingeras, T. R. (1985). Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. *Mol. Cell. Biol.* **5**, 1111–1121.
37. Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V. & Cregg, J. M. (1997). Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene*, **186**, 37–44.
38. Zhang, A. L., Luo, J. X., Zhang, T. Y., Pan, Y. W., Tan, Y. H., Fu, C. Y. & Tu, F. Z. (2009). Recent advances on the GAP promoter derived expression system of *Pichia pastoris*. *Mol. Biol. Rep.* **36**, 1611–1619.
39. Boer, H., Teeri, T. T. & Koivula, A. (2000). Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters. *Biotechnol. Bioeng.* **69**, 486–494.
40. Hohenblum, H., Gasser, B., Maurer, M., Borth, N. & Mattanovich, D. (2004). Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. *Biotechnol. Bioeng.* **85**, 367–375.
41. Gasser, B., Maurer, M., Gach, J., Kunert, R. & Mattanovich, D. (2006). Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol. Bioeng.* **94**, 353–361.
42. Ewert, S., Cambillau, C., Conrath, K. & Plückthun, A. (2002). Biophysical properties of camelid V_{HH} domains compared to those of human V_H3 domains. *Biochemistry*, **41**, 3628–3636.
43. Knappik, A. & Plückthun, A. (1995). Engineered turns of a recombinant antibody improve its *in vivo* folding. *Protein Eng.* **8**, 81–89.
44. Nieba, L., Honegger, A., Krebber, C. & Plückthun, A. (1997). Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved *in vivo* folding and physical characterization of an engineered scFv fragment. *Protein Eng.* **10**, 435–444.
45. Bode, J., Schlake, T., Iber, M., Schubeler, D., Seibler, J., Snezhkov, E. & Nikolaev, L. (2000). The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. *Biol. Chem.* **381**, 801–813.
46. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G. *et al.* (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* **296**, 57–86.
47. Ewert, S., Honegger, A. & Plückthun, A. (2003). Structure-based improvement of the biophysical properties of immunoglobulin VH domains with a generalizable approach. *Biochemistry*, **42**, 1517–1528.
48. Jefferis, R. (2005). Glycosylation of recombinant antibody therapeutics. *Biotechnol. Prog.* **21**, 11–16.
49. Harris, R. J. (1995). Processing of C-terminal lysine and arginine residues of proteins isolated from mammalian cell culture. *J. Chromatogr., A*, **705**, 129–134.
50. Kaplan, A. P., Hood, L. E., Terry, W. D. & Metzger, H. (1971). Amino terminal sequences of human immunoglobulin heavy chains. *Immunochemistry*, **8**, 801–811.
51. Chelius, D., Jing, K., Lueras, A., Rehder, D. S., Dillon, T. M., Vizel, A. *et al.* (2006). Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. *Anal. Chem.* **78**, 2370–2376.
52. Roberts, G. D., Johnson, W. P., Burman, S., Anumula, K. R. & Carr, S. A. (1995). An integrated strategy for structural characterization of the protein and carbohydrate components of monoclonal antibodies: application to anti-respiratory syncytial virus MAb. *Anal. Chem.* **67**, 3613–3625.

53. Kurjan, J. & Herskowitz, I. (1982). Structure of a yeast pheromone gene (*MFα*): a putative α-factor precursor contains four tandem copies of mature α-factor. *Cell*, **30**, 933–943.
54. Kozlov, D. G. & Yagudin, T. A. (2008). Antibody fragments may be incorrectly processed in the yeast *Pichia pastoris*. *Biotechnol. Lett.* **30**, 1661–1663.
55. Emberson, L. M., Trivett, A. J., Blower, P. J. & Nicholls, P. J. (2005). Expression of an anti-CD33 single-chain antibody by *Pichia pastoris*. *J. Immunol. Methods*, **305**, 135–151.
56. Daly, R. & Hearn, M. T. (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J. Mol. Recognit.* **18**, 119–138.
57. Garidel, P., Hegyi, M., Bassarab, S. & Weichel, M. (2008). A rapid, sensitive and economical assessment of monoclonal antibody conformational stability by intrinsic tryptophan fluorescence spectroscopy. *Biotechnol. J.* **3**, 1201–1211.
58. Garber, E. & Demarest, S. J. (2007). A broad range of Fab stabilities within a host of therapeutic IgGs. *Biochem. Biophys. Res. Commun.* **355**, 751–757.
59. Helenius, A. & Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049.
60. Roth, J., Zuber, C., Park, S., Jang, I., Lee, Y., Kysela, K. G. *et al.* (2010). Protein N-glycosylation, protein folding, and protein quality control. *Mol. Cell*, **30**, 497–506.
61. Feige, M. J., Walter, S. & Buchner, J. (2004). Folding mechanism of the C_H2 antibody domain. *J. Mol. Biol.* **344**, 107–118.
62. Feige, M. J., Hendershot, L. M. & Buchner, J. (2010). How antibodies fold. *Trends Biochem. Sci.* **35**, 189–198.
63. Krapp, S., Mimura, Y., Jefferis, R., Huber, R. & Sondermann, P. (2003). Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J. Mol. Biol.* **325**, 979–989.
64. Hulett, M. D., Witort, E., Brinkworth, R. I., McKenzie, I. F. & Hogarth, P. M. (1994). Identification of the IgG binding site of the human low affinity receptor for IgG FcγRII. Enhancement and ablation of binding by site-directed mutagenesis. *J. Biol. Chem.* **269**, 15287–15293.
65. Huber, R., Deisenhofer, J., Colman, P. M., Matsushima, M. & Palm, W. (1976). Crystallographic structure studies of an IgG molecule and an Fc fragment. *Nature*, **264**, 415–420.
66. Sondermann, P., Huber, R., Oosthuizen, V. & Jacob, U. (2000). The 3.2-Å crystal structure of the human IgG1 Fc fragment–FcγRIII complex. *Nature*, **406**, 267–273.
67. Ferrara, C., Stuart, F., Sondermann, P., Brunker, P. & Umana, P. (2006). The carbohydrate at FcγRIIIa Asn-162. An element required for high affinity binding to non-fucosylated IgG glycoforms. *J. Biol. Chem.* **281**, 5032–5036.
68. Ravetch, J. V. & Bolland, S. (2001). IgG Fc receptors. *Annu. Rev. Immunol.* **19**, 275–290.
69. Grinna, L. S. & Tschopp, J. F. (1989). Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, *Pichia pastoris*. *Yeast*, **5**, 107–115.
70. Martinet, W., Saelens, X., Deroo, T., Neirynck, S., Contreras, R., Min Jou, W. & Fiers, W. (1997). Protection of mice against a lethal influenza challenge by immunization with yeast-derived recombinant influenza neuraminidase. *Eur. J. Biochem.* **247**, 332–338.
71. Kayser, V., Chennamsetty, N., Voynov, V., Forrer, K., Helk, B. & Trout, B. L. (2011). Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. *Biotechnol. J.* **6**, 38–44.
72. Li, P., Anumanthan, A., Gao, X. G., Ilangovan, K., Suzara, V. V., Duzgunes, N. & Renugopalakrishnan, V. (2007). Expression of recombinant proteins in *Pichia pastoris*. *Appl. Biochem. Biotechnol.* **142**, 105–124.
73. Cereghino, J. L. & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* **24**, 45–66.
74. Brake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F. R., Mullenbach, G. T. *et al.* (1984). Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **81**, 4642–4646.
75. Brocca, S., Schmidt-Dannert, C., Lotti, M., Alberghina, L. & Schmid, R. D. (1998). Design, total synthesis, and functional overexpression of the *Candida rugosa lip1* gene coding for a major industrial lipase. *Protein Sci.* **7**, 1415–1422.
76. Lombardi, A., Bursomanno, S., Lopardo, T., Traini, R., Colombatti, M., Ippoliti, R. *et al.* (2010). *Pichia pastoris* as a host for secretion of toxic saporin chimeras. *FASEB J.* **24**, 253–265.
77. Monsalve, R. I., Lu, G. & King, T. P. (1999). Expressions of recombinant venom allergen, antigen 5 of yellow-jacket (*Vespula vulgaris*) and paper wasp (*Polistes annularis*), in bacteria or yeast. *Protein Expression Purif.* **16**, 410–416.
78. Brake, A. J. (1989). Secretion of heterologous proteins directed by the yeast alpha-factor leader. *Biotechnology*, **13**, 269–280.
79. Piggott, J. R., Watson, M. E., Doel, S. M., Goodey, A. R. & Carter, B. L. (1987). The secretion and post translational modification of interferons from *Saccharomyces cerevisiae*. *Curr. Genet.* **12**, 561–567.
80. Zsebo, K. M., Lu, H. S., Fieschko, J. C., Goldstein, L., Davis, J., Duker, K. *et al.* (1986). Protein secretion from *Saccharomyces cerevisiae* directed by the prepro-α-factor leader region. *J. Biol. Chem.* **261**, 5858–5865.
81. Zhao, H. L., He, Q., Xue, C., Sun, B., Yao, X. Q. & Liu, Z. M. (2009). Secretory expression of glycosylated and aglycosylated mutein of onconase from *Pichia pastoris* using different secretion signals and their purification and characterization. *FEMS Yeast Res.* **9**, 591–599.
82. Kjeldsen, T., Brandt, J., Andersen, A. S., Egel-Mitani, M., Hach, M., Pettersson, A. F. & Vad, K. (1996). A removable spacer peptide in an α-factor-leader/insulin precursor fusion protein improves processing and concomitant yield of the insulin precursor in *Saccharomyces cerevisiae*. *Gene*, **170**, 107–112.
83. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795.
84. Liu, Y. D., Goetze, A. M., Bass, R. B. & Flynn, G. C. (2011). N-terminal glutamate to pyroglutamate conversion *in vivo* for human IgG2 antibodies. *J. Biol. Chem.* **286**, 11211–11217.

85. Lawrence, M. S., Phillips, K. J. & Liu, D. R. (2007). Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129**, 10110–10112.
86. Arbabi-Ghahroudi, M., To, R., Gaudette, N., Hirama, T., Ding, W., MacKenzie, R. & Tanha, J. (2009). Aggregation-resistant VHs selected by *in vitro* evolution tend to have disulfide-bonded loops and acidic isoelectric points. *Protein Eng. Des. Sel.* **22**, 59–66.
87. Jespers, L., Schon, O., Famm, K. & Winter, G. (2004). Aggregation-resistant domain antibodies selected on phage by heat denaturation. *Nat. Biotechnol.* **22**, 1161–1165.
88. Perchiacca, J. M., Bhattacharya, M. & Tessier, P. M. (2011). Mutational analysis of domain antibodies reveals aggregation hotspots within and near the complementarity determining regions. *Proteins*, **79**, 2637–2647.
89. Vincke, C., Loris, R., Saerens, D., Martinez-Rodriguez, S., Muyldermans, S. & Conrath, K. (2009). General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. *J. Biol. Chem.* **284**, 3273–3284.
90. Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd edit. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
91. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
92. Werner, W. E., Wu, S. & Mulkerrin, M. (2005). The removal of pyroglutamic acid from monoclonal antibodies without denaturation of the protein chains. *Anal. Biochem.* **342**, 120–125.
93. Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* **1**, 100.