

Harmonization of the Genetic Code Effectively Enhances the Recombinant Production of the Major Birch Pollen Allergen Bet v 1

Claudia Asam^a Anargyros Roulias^a Maria A. Parigiani^a Astrid Haab^a
Michael Wallner^a Martin Wolf^a Peter Briza^a Barbara Bohle^b Fatima Ferreira^a
Michael Hauser^a

^aDepartment of Biosciences, University of Salzburg, Salzburg, Austria; ^bDepartment of Pathophysiology and Allergy Research, Medical University of Vienna, Vienna, Austria

Keywords

Bet v 1 · Codon harmonization · Heterologous protein expression · Expression yield · Recombinant protein

Abstract

Background: Enhancing the quality and yield of protein production in heterologous expression systems is an important issue for developing new biopharmaceuticals. It has been shown that the dynamics of protein folding is influenced by codon frequencies. As codon usage frequencies are species specific, this can affect heterologous protein expression. In this respect, “codon harmonization,” that is, the usage of synonymous codons with usage frequencies in the host resembling the usage frequencies in the native organism, is a promising strategy. As recombinant proteins are important tools in the area of allergy research, we investigated in this study the influence of codon harmonization on the production of the major birch pollen allergen Bet v 1.0101. **Methods:** To accomplish this task, parallel production of several batches of rBet v 1, BWT, together with a harmonized vari-

ant, BH, was applied. The expression yield of soluble and insoluble protein was assayed via densitometric analysis of SDS-PAGEs for every batch. The quality of purified proteins was assessed with a variety of physicochemical methods including mass spectrometry, circular dichroism, dynamic light scattering, Fourier transform infrared spectroscopy, in vitro degradation, and 1-anilino-8-naphthalene sulfonate-binding assays. Patients’ IgE reactivity was tested in enzyme-linked immunosorbent assays and rat basophil mediator release experiments. **Results:** No significant differences in the ligand-binding capacity and secondary structure elements, as well as, in immunological assays could be found; however, the production yield was drastically increased for BH. **Conclusion:** We could show that codon harmonization is a powerful method to enhance protein yields in heterologous expression systems and should be considered especially for difficult-to-express proteins.

© 2018 S. Karger AG, Basel

Edited by: O. Palomares, Madrid.
C.A. and A.R. contributed equally to this work.

Introduction

For decades, large-scale production of recombinant proteins in heterologous expression systems is routinely applied in a variety of biological research areas, whereas *E. coli* has established itself as the most popular expression platform in the field [1]. Besides tremendous advances in developing *E. coli* as a versatile host for recombinant protein production, depending on the protein of interest the system still harbors problems such as low/no expression, protein truncation, aggregation, misfolding, insolubility, or non-functionality of the target protein [2]. Thus, many approaches have been followed to tune the expression process at all levels to improve quality and yields of the recombinant protein. Several of these approaches directly address the genetic code by altering the nucleotide sequence of the gene of interest [1, 2].

In most organisms, 61 of the 64 codons encoding for one of the 20 amino acids are used in a degenerate manner. Therefore, it has been generally believed that silent mutations or synonymous substitutions of codons have no effect on the translation process of the respective protein [3]. However, the codon usage among different species is extremely variable. Consequently, Ikemura et al. [4] described a bias in codon usage by demonstrating that specific codons are used more frequently in highly expressed genes. However, codon usage not only varies between genes within the genome of a certain species, but also distinct positions within a single gene show codon usage variations. Moreover, different species use different codons [5–7]. Regions in a nucleotide sequence that separate elements of higher ordered structure often contain less frequently used codons, leading to a slowing down of the ribosomal translation process. This alteration in translation speed was shown to play a critical role in the process of protein folding [8, 9]. Nowadays, it is well accepted that the genetic code itself carries the capacity to regulate the rate of protein expression as well as protein folding and that “silent” mutations influence this process after all [10]. Consequently, this knowledge can be used to harmonize the “genetic melody” between the donor and the host organism to improve recombinant protein production, solubility, and functionality [9, 11]. Thus, codon harmonization of the host tRNA pool to the tRNA frequency of the donor organism will synchronize the speed of ribosomal movement along the mRNA sequence [11].

Therefore, the aim of the study was to adapt the method of codon harmonization to investigate the in-

fluence on the allergen production. By taking the major birch pollen allergen Bet v 1 as model, we produced 2 batches of codon harmonized and 2 batches of conventional Bet v 1. All proteins were expressed in the same *E. coli* strain, purified to homogeneity, and analyzed in detail.

Material and Methods

Patients and Sera

Birch pollen-allergic patients ($n = 20$) were selected based on case history, positive in vivo skin prick test, and in vitro IgE detection (CAP system, Thermo Fisher Scientific, Phadia AB). Experiments with patients' sera were approved by the Ethics Committee of the University of Vienna (EK028/2006) and informed consent was obtained from all participants in the study.

Construction of the Codon Harmonized Bet v 1 Sequence

Codon harmonization was performed by analyzing codon usage frequencies with an online database available at <http://www.kazusa.or.jp/codon>. First, the codon usage frequency of Bet v 1 in *Betula verrucosa* was determined. Thereafter, synonymous codons were replaced with codons of *E. coli* K12, which showed the same codon usage frequency. As a template, the DNA sequence of Bet v 1.0101, referred to as Bet v 1, was used. The gene was ordered from ATG:biosynthetics GmbH, Merzhausen, Germany.

Expression and Purification of BWT and BH

The wild type gene sequence of Bet v 1, termed BWT thereafter, as well as the *E. coli* codon harmonized sequence of Bet v 1, termed BH in the following, were cloned into the vector pET28b (EMD Biosciences Inc, San Diego, CA, USA), transformed in *E. coli* BL21 Star™ (DE3) cells (Invitrogen, Carlsbad, CA, USA), and grown at 37 °C in LB supp. medium (1% [w/v] peptone, 0.5% [w/v] yeast extract, 0.5% [w/v] NaCl, 2 mM MgSO₄, 1% [v/v] glycerol, 0.2% [w/v] ammonium sulfate, 10 mM sodium phosphate pH 7.4) supplemented with 25 mg/L kanamycin. Protein expression was induced by the addition of 0.5 mM isopropyl- β -D-thiogalactopyranoside at OD_{600 nm} of 0.8, and growth was continued at 37 °C for 5 h. Two expression batches were produced for each construct. After harvesting, cells were dissolved in 1/50 culture volume of 20 mM sodium phosphate buffer pH 7.4 containing 0.8 M NaCl and 15 mM NaH₂PO₄, and protein purification was performed as previously described [12]. The purity of recombinant proteins was analyzed by SDS-PAGE and protein identity was verified by mass spectrometry using a Q-Exactive LC-MS/MS system (Thermo Fisher Scientific). Recombinant proteins were dialyzed against 10 mM sodium phosphate buffer and stored in lyophilized form at –20 °C.

Physicochemical Characterization

Secondary structure analysis of recombinant proteins was performed by circular dichroism (CD) spectroscopy using a JASCO J-815 spectropolarimeter fitted with a PTC-423S Peltier-type single position cell holder (Jasco, Tokyo, Japan), as well as Fourier transform infrared spectroscopy (FTIR), and dynamic

light scattering (DLS). Accessibility and functionality of the intrinsic cavity of Bet v 1 were determined by 1-anilino-8-naphthalene sulfonate (ANS) displacement assays, as previously described [13].

Stability against proteolytic activity was assessed by endolysosomal degradation assays as described elsewhere [14].

Enzyme-Linked Immunosorbent Assays Experiments

For enzyme-linked immunosorbent assays (ELISA) experiments, plates were coated with BWT or BH (100 ng/well in 50 µL PBS) overnight at 4 °C. After blocking, plates were incubated with patients' sera (*n* = 20) diluted 1:10 overnight at 4 °C. Bound human IgE (Immunoglobulin E) was detected with alkaline phosphatase-

conjugated monoclonal anti-human IgE antibodies (BD Biosciences, Franklin Lakes, NJ, USA). For human IgE binding, statistics was calculated with repeated measures analysis of variance; *p* values <0.05 were considered significant.

Mediator Release Assay

The allergenic potential of BWT and BH was assessed by rat basophil degranulation assays performed as shown elsewhere [15]. In brief, RBL-2H3 cells transfected with the human high-affinity IgE receptor (FcεRI) were passively sensitized with serum IgE from birch pollen-allergic donors (*n* = 8). Antigen-dependent β-hexosaminidase release into the supernatant was measured by enzymatic cleavage of the fluorogenic substrate 4-methyl

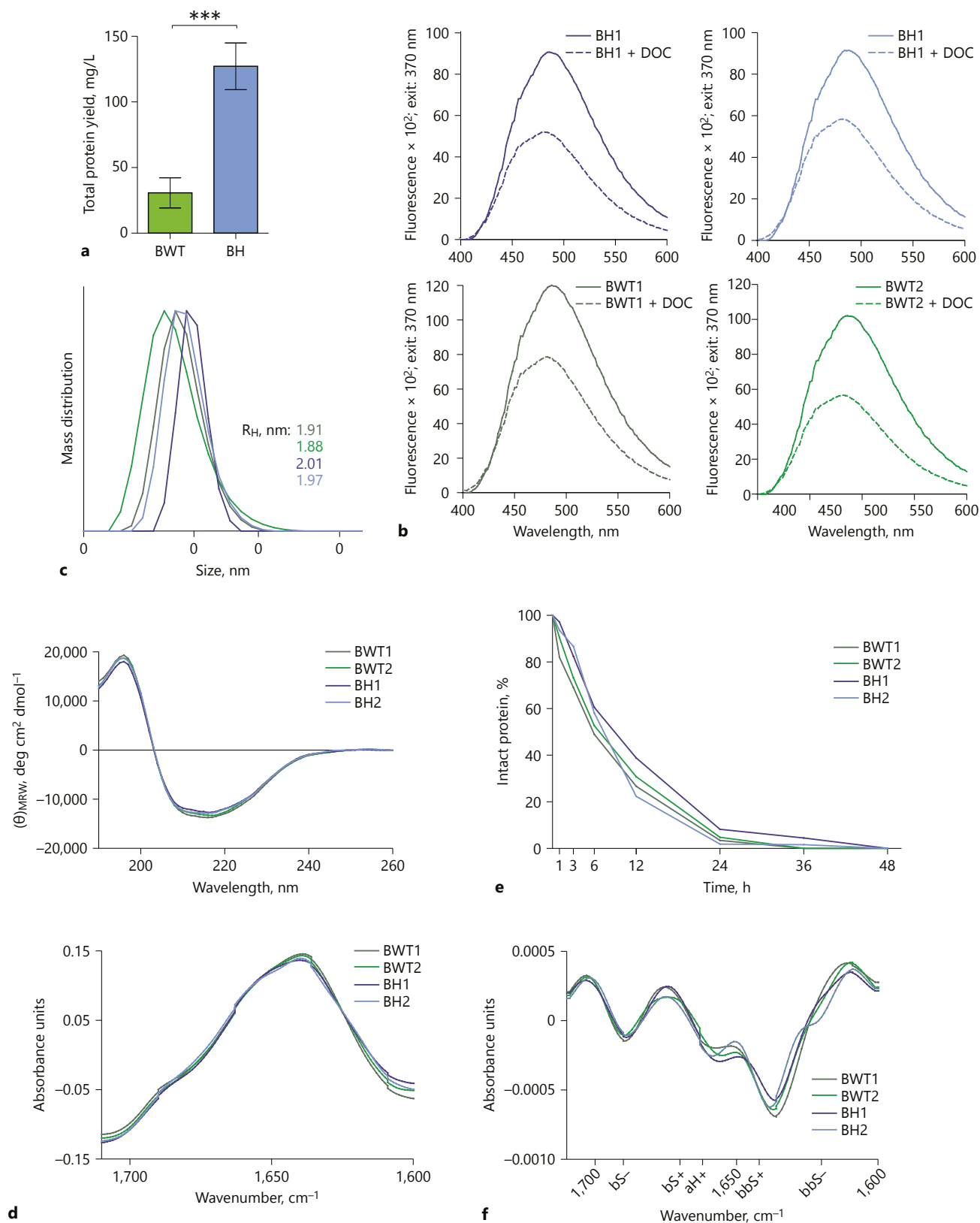
Protein:	G	V	F	N	Y	E	T	E	T	T	S	V	I	P	A	A	R	L	F	K
BWT:	GGT	GTT	TTC	AAT	TAC	GAA	ACT	GAG	ACC	ACC	TCT	GTT	ATC	CCA	GCG	GCT	CGA	CTG	TTC	AAG
BH:	GGT	GTG	TTT	AAT	TAT	GAG	ACT	GAA	ACC	ACC	TCG	GTG	ATT	CCT	GCT	GCA	CGA	CTC	TTT	AAA
Protein:	A	F	I	L	D	G	D	N	L	F	P	K	V	A	P	Q	A	I	S	S
BWT:	GCC	TTT	ATC	CTT	GAT	GGC	GAT	AAT	CTC	TTT	CCA	AAG	GTT	GCA	CCC	CAA	GCC	ATT	AGC	AGT
BH:	GCA	TTT	ATT	TTA	GAC	GGT	GAC	AAT	TTA	TTT	CCT	AAA	GTG	GCA	CCT	CAA	GCA	ATC	AGC	TCG
Protein:	V	E	N	I	E	G	N	G	G	P	G	T	I	K	K	I	S	F	P	E
BWT:	GTT	GAA	AAC	ATT	GAA	GGA	AAT	GGA	GGG	CCT	GGA	ACC	ATT	AAG	AAG	ATC	AGC	TTT	CCC	GAA
BH:	GTG	GAG	AAC	ATC	GAG	GGT	AAT	GGT	GGA	CCT	GGT	ACC	ATC	AAA	AAA	ATT	AGC	TTC	CCT	GAG
Protein:	G	F	P	F	K	Y	V	K	D	R	V	D	E	V	D	H	T	N	F	K
BWT:	GGC	TTC	CCT	TTC	AAG	TAC	GTG	AAG	GAC	AGA	GTT	GAT	GAG	GTG	GAC	CAC	ACA	AAC	TTT	AAA
BH:	GGT	TTT	CCT	TTT	AAA	TAT	GTG	AAA	GAC	CGA	GTG	GAC	GAA	GTG	GAC	CAT	ACG	AAC	TTT	AAG
Protein:	Y	N	Y	S	V	I	E	G	G	P	I	G	D	T	L	E	K	I	S	N
BWT:	TAC	AAT	TAC	AGC	GTG	ATC	GAG	GGC	GGT	CCC	ATA	GGC	GAC	ACA	TTG	GAG	AAG	ATC	TCC	AAC
BH:	TAT	AAT	TAT	AGC	GTG	ATT	GAA	GGT	GGT	CCT	ATC	GGT	GAC	ACG	TTA	GAA	AAA	ATT	AGC	AAC
Protein:	E	I	K	I	V	A	T	P	D	G	G	S	I	L	K	I	S	N	K	Y
BWT:	GAG	ATA	AAG	ATA	GTG	GCA	ACC	CCT	GAT	GGA	GGA	TCC	ATC	TTG	AAG	ATC	AGC	AAC	AAG	TAC
BH:	GAA	ATC	AAA	ATC	GTG	GCA	ACC	CCG	GAC	GGC	GGC	AGC	ATT	TTA	AAA	ATT	AGC	AAC	AAA	TAT
Protein:	H	T	K	G	D	H	E	V	K	A	E	Q	V	K	A	S	K	E	M	G
BWT:	CAC	ACC	AAA	GGT	GAC	CAT	GAG	GTG	AAG	GCA	GAG	CAG	GTT	AAG	GCA	AGT	AAA	GAA	ATG	GGC
BH:	CAT	ACC	AAG	GGT	GAC	CAC	GAA	GTG	AAA	GCA	GAA	CAA	GTG	AAA	GCA	TCG	AAG	GAG	ATG	GGT
Protein:	E	T	L	L	R	A	V	E	S	Y	L	L	A	H	S	D	A	Y	N	
BWT:	GAG	ACA	CTT	TTG	AGG	GCC	GTT	GAG	AGC	TAC	CTC	TTG	GCA	CAC	TCC	GAT	GCC	TAC	AAC	TAA
BH:	GAA	ACG	TTA	TTA	CGT	GCA	GTG	GAA	AGC	TAT	TTA	TTA	GCA	CAT	AGC	GAC	GCA	TAT	AAC	AAA

Fig. 1. DNA sequence alignment of BWT and BH. The respective amino acid sequence is shown (Protein). The recoded nucleotides of the BH sequence are indicated by a dark color.

Fig. 2. Comparison of expression yields of BWT and BH calculated densitometrically (a). Competitive binding of ANS and sodium-deoxycholate (DOC) in a 1/2 molar ratio protein/DOC was analyzed by fluorescence spectroscopy; data are baseline-corrected (b). DLS measurements of BWT (BWT1 and BWT2) and BH (BH1 and BH2) were performed to determine the hydrodynamic radius of the proteins in solution (c). CD spectra of BWT (BWT1 and

BWT2) and BH (BH1 and BH2); data are baseline corrected and presented as mean residue molar ellipticity (d). In vitro degradation assays were performed to determine proteolytic stability of BWT (BWT1 and BWT2) and BH (BH1 and BH2; e). FTIR spectra of amide I peak and corresponding second derivatives recorded at 20 °C (f). Analysis of variance calculated *p* values <0.05 were considered significant.

(For figure see next page.)



umbelliferyl-N-acetyl- β -D-glucosaminide and expressed as percentage of total enzyme content of Triton X-100-treated cells. For the statistics of RBL-2H3 cell mediator release, the half-maximum release for all samples was determined and statistics was calculated with analysis of variance; p values <0.05 were considered significant.

Results

Codon Harmonization of Bet v 1 Sequence

In total, 73.6% (117 out of 159) of the original “birch-codons” were replaced within this process of codon harmonization (Fig. 1).

Expression Yields of BWT and BH

Protein expression was performed as described above. For the determination of expression yields and solubility, *E. coli* lysates of 5 different independent expression attempts of either BWT or BH were densitometrically analyzed. The quantification of total Bet v 1 revealed 35 mg/L for BWT and 120 mg/L for BH, which means a more than threefold higher expression level of soluble Bet v 1 for the harmonized versus the wild type gene ($p < 0.001$; Fig. 2a).

BWT and BH Showed Similar Structural Properties

Two batches of each, BWT and BH, were purified from the soluble fraction of *E. coli* lysates by hydrophobic interaction and anion exchange chromatography and purity was determined by SDS-PAGE. Mass spectrometric analysis confirmed protein identity and indicated that all batches were not modified during the purification process. Bet v 1 harbors an intrinsic hydrophobic cavity; thus, the allergen has the ability to bind the sulfonic acid ANS. As the cavity is solvent-accessible, sodium deoxycholate can replace ANS-binding in a competitive manner. In ANS-displacement assays, we found that both batches of BWT and BH were able to bind this ligand, whereas the addition of deoxycholate in a 1:2 molar ratio reduced ANS-binding in all batches (Fig. 2b). CD and FTIR spectra were recorded to compare secondary structure elements of Bet v 1 batches. Analyses of the proteins resulted in essentially the same spectra (Fig. 2d, f), indicating identical folding behavior. DLS measurements of Bet v 1 batches revealed similar hydrodynamic radii as well as a globular shape of the proteins (Fig. 2c). In vitro degradation assays revealed comparable susceptibility to endolysosomal proteolysis for all Bet v 1 batches (Fig. 2e).

Immunological Comparison of BWT and BH

To gain a deeper insight in the immunologic behavior of BWT versus BH, we analyzed the antibody-binding properties in ELISA using serum IgE of Bet v 1 allergic patients ($n = 20$). We found some variability in IgE binding of the 4 production batches (Fig. 3a). Thereafter, all Bet v 1 batches were tested in humanized basophil mediator release assays using the sera of birch allergic patients ($n = 8$). Here, the results were essentially the same. Overall, we could not detect any significant differences in the allergenic potential of the Bet v 1 batches (BH vs. BWT), whereas we found some significant variation between specific batches (Fig. 3b, c).

Discussion

It is widely accepted that codon usage, silent codon exchanges, and thus the “genetic melody” may directly impact not only protein expression rates but also influence the behavior of newly synthesized gene products [16]. Not frequently used codons at well-placed positions could lead to a stabilization of the ribosomal initial complex and consequently to higher expression yields [17]. Furthermore, during the logarithmic growth phase, extreme expression speeds can induce mRNA loss due to ribosome bottlenecks, also affecting the expression yield. Codon harmonization renders the problem of these deleterious host cell responses [8].

In this study, we generated a codon harmonized DNA sequence of the major birch pollen allergen Bet v 1 by adapting the original plant codon usage to the codon frequency of *E. coli*. The protein was expressed in bacteria and compared with production batches also produced in *E. coli* but without adaptation of the original plant-derived sequence. Using this strategy, we achieved a more than threefold increase in production yields compared to protein expression of the non-harmonized gene.

After purification, detailed physicochemical analyses [18] of 2 batches, BWT and BH, did not reveal any structure-related differences in CD, DLS, FTIR, ANS replacement assays, or proteolytic stability testing. Although, the 2 BH batches showed a slightly reduced IgE-binding capacity in ELISA using human sera, we observed that rather assay variations than changes within the production batches accounted responsible for the divergence in IgE-binding. The findings were further supported by data from mediator release assay, where all Bet v 1 batches showed a very similar ability to induce

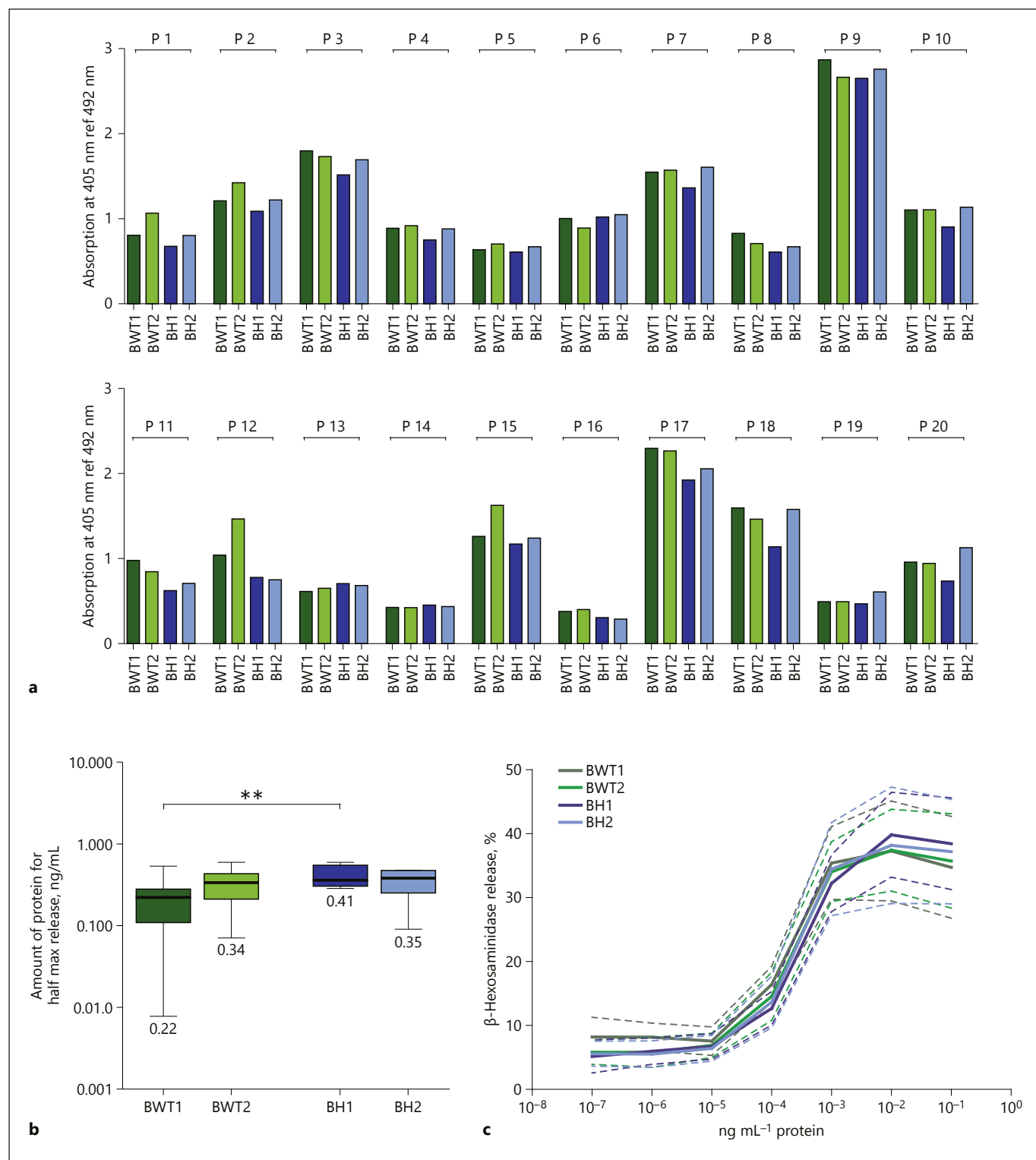


Fig. 3. ELISA was performed to measure the binding of human serum IgE ($n = 20$) to immobilized BWT (BWT1 and BWT2) or BH (BH1 and BH2), numbers represent patient number (**a**). RBL cells were passively sensitized with sera of patients with Bet v 1-mediated birch pollen allergy ($n = 8$) and mediator release was trig-

gered by the addition of the 2 BWT (BWT1 and BWT2) or the 2 BH (BH1 and BH2) batches (**c**) and half maximum release was calculated; numbers below boxes and whiskers represent mean values (**b**). Analysis of variance calculated p values <0.05 were considered significant.

IgE-dependent basophil activation. Thus, we conclude that harmonization of the DNA sequence has a dramatic impact on expression yields, while in our hands the structural properties of Bet v 1 remained unaffected by the codon usage.

Thus, codon harmonization constitutes a promising approach to improve recombinant production of allergens. Moreover, we believe that this approach may have positive effects on the solubility of more complex allergen structures during recombinant production, which would be especially interesting for allergens or proteins with higher structural complexity like the trimeric soybean allergen

Gly m 5 or the disulfide bridged peanut allergen Ara h 2. In fact, we have also harmonized the gene sequence of the major ragweed allergen Amb a 1, a pectate lyase with a complex beta-helix fold, and obtained promising results for the expression of soluble recombinant protein (unpublished results).

Thus, codon harmonization provides an effective approach toward increasing expression levels of soluble and functional proteins and should be considered a strategy for overcoming production issues, especially in industrial-scale production of recombinant allergens.

References

- Samuelson JC: Recent developments in difficult protein expression: a guide to *E. coli* strains, promoters, and relevant host mutations. *Methods Mol Biol* 2011;705:195–209.
- Rosano GÁ L, Ceccarelli EA: Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* 2014; 5:172.
- Spencer PS, Barral JÉ M: Genetic code redundancy and its influence on the encoded polypeptides. *Comput Struct Biotechnol J* 2012; 1:e201204006.
- Ikemura T: Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985;2:13–34.
- Saunders R, Deane CM: Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res* 2010;38:6719–6728.
- Angov E: Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol J* 2011;6:650–659.
- Plotkin JB, Kudla G: Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 2011;12:32–42.
- Angov E, Hillier CJ, Kincaid RL, Lyon JA: Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One* 2008;3:e2189.
- Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, Nussinov R: Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol* 2008;383:281–291.
- Spencer PS, Siller E, Anderson JF, Barral JM: Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol* 2012;422:328–335.
- Angov E, Legler PM, Mease RM: Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods Mol Biol* 2011;705:1–13.
- Hoffmann-Sommergruber K, Susani M, Ferreira F, Jertschin P, Ahorn H, Steiner R, Kraft D, Scheiner O, Breiteneder H: High-level expression and purification of the major birch pollen allergen, Bet v 1. *Protein Expr Purif* 1997;9:33–39.
- Asam C, Batista AL, Moraes AH, de Paula VS, Almeida FC, Aglas L, Kitzmuller C, Bohle B, Ebner C, Ferreira F, Wallner M, Valente AP: Bet v 1—a Trojan horse for small ligands boosting allergic sensitization? *Clin Exp Allergy* 2014;44:1083–1093.
- Egger M, Jurets A, Wallner M, Briza P, Ruzek S, Hainzl S, Pichler U, Kitzmuller C, Bohle B, Huber CG, Ferreira F: Assessing protein immunogenicity with a dendritic cell line-derived endolysosomal degradome. *PLoS One* 2011;6:e17278.
- Vogel L, Luttkopf D, Hatahet L, Haustein D, Vieths S: Development of a functional in vitro assay as a novel tool for the standardization of allergen extracts in the human system. *Allergy* 2005;60:1021–1028.
- Nissley DA, O'Brien EP: Timing is everything: unifying codon translation rates and nascent proteome behavior. *J Am Chem Soc* 2014;136:17892–17898.
- Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Furman I, Pilpel Y: An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 2010, 141:344–354.
- Himly M, Wallner M, Zaborsky N, Hauser M, Gadermaier G, Harrer A, Ferreira F: Physicochemical characterization of recombinant allergens and hypoallergenic variants. Arbeiten aus dem Paul-Ehrlich-Institut (Bundesinstitut für Impfstoffe und biomedizinische Arzneimittel). *Langen/Hessen* 2013;97:101–105.