



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on a thermostable enzymatic one-pot reaction for the production of a high-value compound from L-arabinose

Maria Bawn^a, Fabiana Subrizi^b, Gary J. Lye^a,
Tom D. Sheppard^b, Helen C. Hailes^b, John M. Ward^{a,*}^a The Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Bernard Katz Building, London WC1E 6BT, UK^b Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK

ARTICLE INFO

Article history:

Received 11 May 2018

Accepted 28 May 2018

Available online 31 May 2018

ABSTRACT

The dataset presented in this article is related to the research article entitled “One-pot, two-step transaminase and transketolase synthesis of L-gluco-heptulose from L-arabinose” (Bawn et al., 2018 in press) [1]. This article presents data on initial experiments that were carried out to investigate new thermostable transketolase (TK) activities with L-arabinose. Transaminase (TAM) sequences from an in-house library of thermophilic strains were analyzed to compare homologies to characterized TAMs with desired activity. DNA and amino acid sequences are presented for all the enzymes investigated. Calibration curves for products of the TK and TAM reactions are also presented along with chromatographic analysis of the various one-pot reactions.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

DOI of original article: <https://doi.org/10.1016/j.enzmictec.2018.05.006>

* Corresponding author.

E-mail address: j.ward@ucl.ac.uk (J.M. Ward).<https://doi.org/10.1016/j.dib.2018.05.140>2352-3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Biology
More specific subject area	Biocatalysis
Type of data	Tables, text file, figures
How data was acquired	Experiments/ in-vitro assays and high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD)
Data format	Analyzed and tabulated
Experimental factors	All enzymes and substrates were freshly prepared before use
Experimental features	Experiments were carried out in triplicate
Data source location	United Kingdom, London, University College London (UCL)
Data accessibility	The data are accessible only within this article

Value of the data

- The data presented in this article gives new insight into the activities of thermostable enzymes not published before.
- The data represents a rationale behind why TKs and TAm were selected for the one-pot reaction.
- Product of one-pot reaction, L-glucO-heptulose, is a pharmaceutically-relevant compound.

1. Data

L-Arabinose is a major monosaccharide of sugar beet pulp (SBP), a by-product of sucrose extraction which is currently produced and sold as a low value animal feed [1]. The main focus of this work was to create a value-added product from the monosaccharides that make up SBP via enzymatic routes. Building on previous work [2,3], this present study produces L-glucO-heptulose, a high value, pharmaceutically relevant compound from L-arabinose using a two-step thermostable enzyme cascade. A thermostable TK catalyzed the synthesis of L-glucO-heptulose from L-arabinose and β -hydroxypyruvate (HPA) in which the latter was produced *in situ* from L-serine and α -ketoglutaric acid using a thermostable TAm.

Table 1 identifies thermostable TKs utilized and whether they were active towards L-arabinose via the Seliwanoff assay [4]. Table 2 describes the TAm investigated and compares sequence homologies to TAm previously showing activities required for this reaction. Examples of HPAEC-PAD traces (Figs. 2 and 3) demonstrate how the TK and TAm one-pot reactions were monitored for the presence of L-glucO-heptulose.

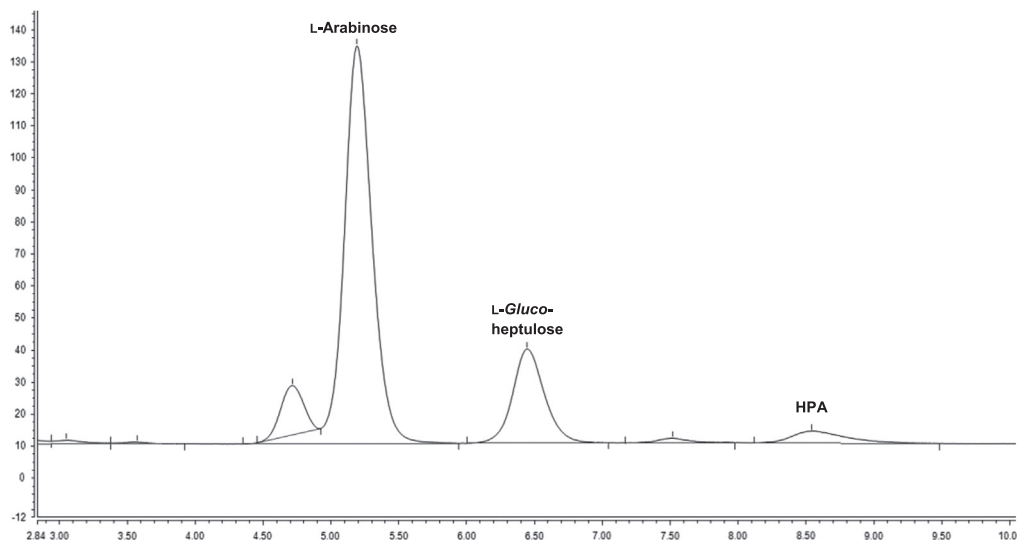
Table 1
TKs showing activity with L-arabinose via Seliwanoff assay.

TK Strain	UniProtKB accession code	Plasmid name and abbreviated name	Active towards L-arabinose from Seliwanoff assay
<i>Deinococcus geothermalis</i> DSM 11300	Q11W07	pQR1758 (TK _{Dgeo})	✓
<i>Deinococcus radiodurans</i> DSM 20539	Q9RS71	pQR1759 (TK _{Drad})	✓
<i>Geobacillus stearothermophilus</i> DSM 22	KFL15812.1	pQR1743	✓
<i>Thermobifida fusca</i> strain YX	Q47ND4	pQR1744	✗
<i>Thermotoga maritima</i> DSM 3109	Q9X283	pQR1745	✗

Table 2

Sequence similarity values between new cloned TAmS and DGEO_0713, SPAT and CV2025.

TAm Strain	UniProtKB accession code	Plasmid name and abbreviated name	Homology to DGEO_0713 <i>Deinococcus geothermalis</i> (%)	Homology to SPAT <i>Sulfolobus solfataricus</i> (%)	Homology to CV2025 <i>Chromobacterium violaceum</i> (%)
<i>Deinococcus radiodurans</i> DSM 20539	Q9RWP3	pQR1746	78	38	31
<i>Geobacillus stearothermophilus</i> DSM 22	Q59228	pQR1756 (TAm _{Gste})	25	38	40
<i>Thermobifida fusca</i> strain YX	Q47LH8	pQR1748	26	37	41
<i>Thermotoga maritima</i> DSM 3109	G4FE93	pQR1749	30	58	31
<i>Deinococcus geothermalis</i> DSM 11300	Q11ZC2	pQR1757 (TAm _{Dgeo})	29	29	44
<i>Xanthomonas campestris</i> pv. <i>Campestris</i> DSM 3586	Q8PDQ2	pQR1751	28	25	30
<i>Thermotoga maritima</i> DSM 3109	Q9X1C0	pQR1752	33	34	36
<i>Pectobacterium carotovorum</i> subsp. <i>Carotovorum</i> DSM 30168	A0A0B3YSH6	pQR1755	24	31	33

**Fig. 1.** HPAEC-PAD trace showing the elution of L-arabinose, L-gluco-heptulose and HPA.

2. Experimental design, materials and methods

2.1. TK activity

Thermostable TKs were cloned and subsequently expressed in *E.coli* BL21 DE3. Cell lysates were used to determine activity towards L-arabinose using the colorimetric assay, Seliwanoff assay. The Seliwanoff assay distinguishes between ketoses and aldoses using 6M HCl and resorcinol

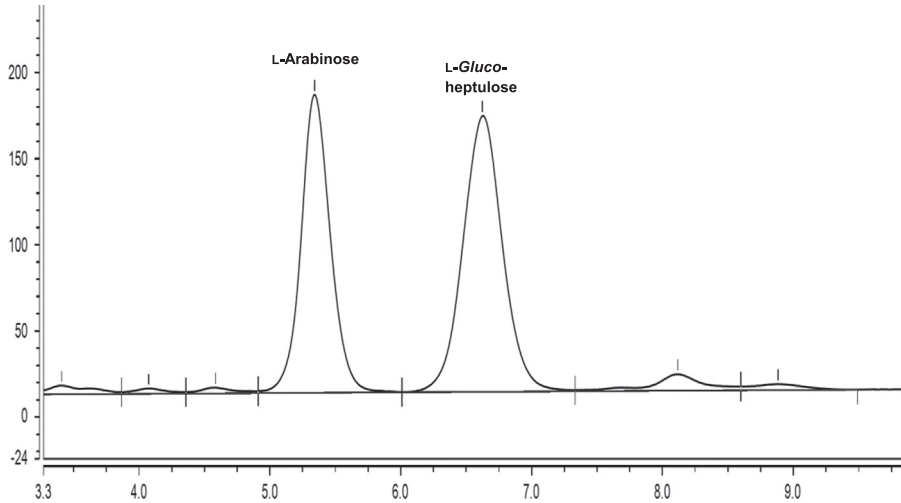


Fig. 2. HPAEC-PAD trace showing *L-gluco*-heptulose production from one-pot reaction with TAM_{Dgeo} and TK_{Dgeo} after 24 h.

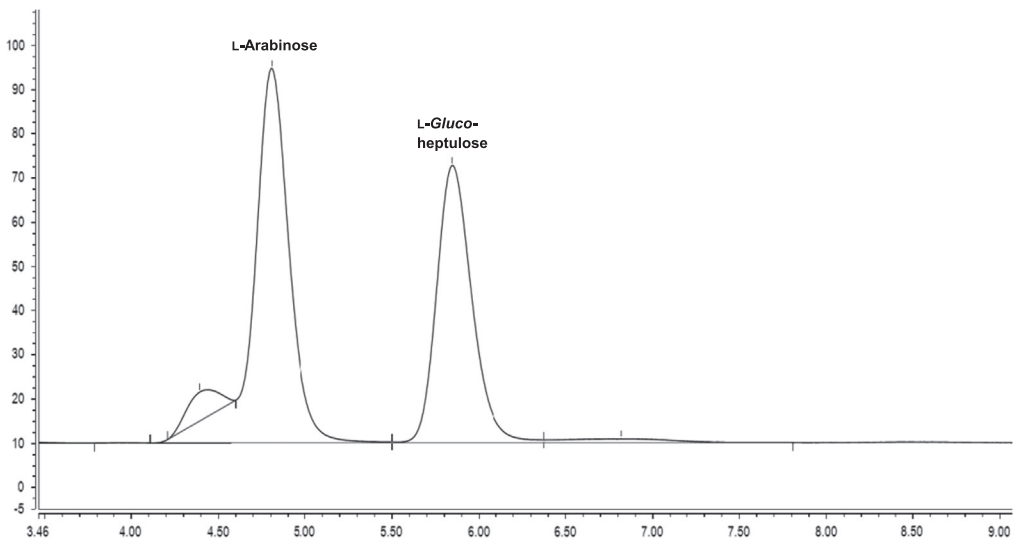


Fig. 3. HPAEC-PAD trace showing *L-gluco*-heptulose production from one-pot reaction with TAM_{Dgeo} and TK_{Drad} after 24 h.

(Seliwanoff's reagent) [4]. After 24 h incubation of enzyme and *L*-arabinose, Seliwanoff reagent was added to the reaction and heated at 100 °C. Colour formation due to the presence of the ketose, *L-gluco*-heptulose, was observed within 15 min (Table 1).

2.2. TAM sequence analysis

TAM sequences were obtained from the NCBI database [5] and the UniProt Knowledgebase (UniProtKB) [6] followed by a sequence alignment using Clustal W [7].

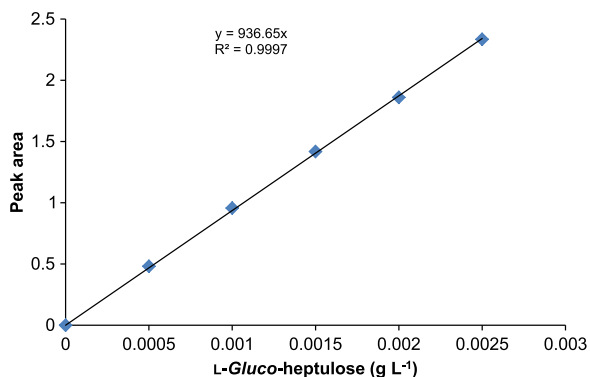


Fig. 4. Calibration curve for determination of *L*-gluco-heptulose yield.

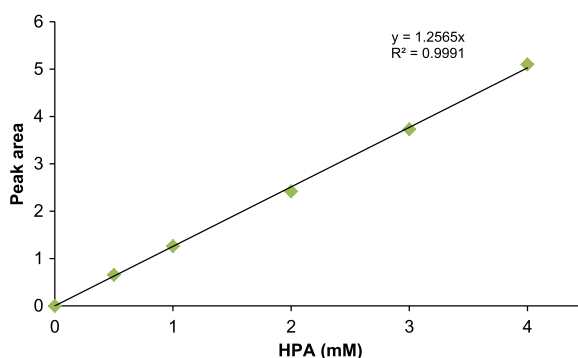


Fig. 5. Calibration curve for determination of HPA yield.

2.3. Product analysis using HPAEC-PAD

Quantitative analysis of *L*-arabinose, *L*-gluco-heptulose and HPA was performed using HPAEC-PAD (ICS 5000+, Dionex) equipped with a Dionex AminopacTM PA1 anion exchange column 4 × 250 mm² fitted with a Dionex AminopacTM PA1 guard column 4 × 50 mm², an electrochemical detector system, and an eluent generator with a KOH 500 cartridge. The elution times of each compound can be observed in Fig. 1. Figs. 3 and 4 are examples of a one-pot reaction analysis with various TKs and TAM_{Dgeo}. Standard calibration curves of *L*-gluco-heptulose and HPA were used for quantification purposes (Figs. 4 and 5).

2.4. TK DNA/ amino acid sequences

DNA sequences were retrieved from the NCBI database [5] and amino acid sequences were obtained through the UniProtKB [6].

> TKDgeo

```

ATGAGTCCCGAACAGCAGGCCGTGCGTCAGGATGTCGATCAGCTGAGCATCAACACCATCAGGACGCTTGCCATCG
ATGCGGTGCAGCGGCCAACAGTGGCCACCCGGCGCGCCGCTCGGCATGGCCCGATGGGCTACGTGCTGTGGCA
GCGCTTCTGCGCCACAATCCGAAACATCCCGAGTGGCCGGGCGCGACCGCTTCGTGCTGTCGGCAGGGCAGCC
AGCATGCTGATCTACTCGCTGCTGCACCTCACCGGTACGACCTGCGCTGGAGGACATCAAGAACTTCCGCCAGT
GGGGCAGCAAGACGCCTGGGCATCCCGAGTTCTTCCACACCCAGGCCTAGACGCCACCACCGGCCGCTCGGTCA
GGGTGCGCGATGACGGTGGGCATGGCGATGGCCGAAGCGCACCTCGCCGCACGCTACAACCGCCCGGCTTCAAG

```

GTCTTTTGACAACACTACACCTACGCGATCTTGGGGGACGGCGACCTGCAAGAAGGCGTCAACCACGAGGCCGCGTTCG
 TGGCAGGGCACCTCAAGCTGGGCAAGCTGATCTGGCTGCACGACGACAACCAGGTGCAGCTGGACACCGCCACGTT
 CAAGGGCCCAACGAGGATACTGCGGAGCGTTACCGCGCCTACGGCTGGGAAAGTTCTGCGTGTGCAAGACGGCAAC
 AATCTCACGAGATCGAGAACGCGATCCGCCAGGCACGGATGAACACCAGCAGCCACCCTGATCCAGGTTTCGCA
 CGGTGATCGGCTTCGGCAGTCCCCGTGCGGGCACAGCAAGGCGCACGGCGAGCCGTGGGCGAGGAAGGCGTGCA
 GGAGACCAAGGCGCCCTGGGCTGGGACTACCCGCCCTTCACGGTGCCCGACGAGGTCAAGGCGCATATGGACGCG
 ACTGAGCGTGGCGCGGAGTGGGAGGCCGACTGGAACCGCTGATGGAGCGCTACCGTGCCGAGTACCCCGATCTCG
 CGGGCGAGGTTGACGCGCTGTGGCGCGCAACTGCCCGCAAATCTCGCCGAAGTGTCCCTCTACGAAGTGGG
 CAGCAAGGCCATCGCCACCCGCAACGCGAGCGGTGAAGTCATCAATGCGCTGGCGCAGGTGGTGGCCGGCCGTGATG
 GGGGGCAGTGCAGTCTCTCCGGCAGCACCAAAACCACCATCAAGGACGCGCGCAGTTCCTGCCAGGAAACTACG
 GGGGCCGCAACGCTACTTTGGCGTCCGCGAGTTTGGGATGGCCGACGCGGGCAATGGCCTTTTCGCTCTACGGAGG
 TGTTTCGGCCCTTGGTGGGGACCTTCTCGTGTGTTCGCGACTACCTCAAGCCCGCCTTCCGCTCTCCGCCCTTACG
 TTCCAGCCGTTTACCTATGTCCTGACCCATGACTCCATTGGCTGGGCGAAGACGGCCCAACCCACCAGCCTATTG
 AGCAGCTCGCCATGCTGCGCGCCGTGCCGGTGCACCGTGTATCGCCCCGCGCAGCCAAACGAGACGGCGCGGC
 CTGGCAGATGGCGCTGGAGTACGACAAGGGACCAACCGCTCTGGCCCTCTCCCGCAGGATCTCCAGTGTGCC
 CGCAACCACGCGGGCGTGAAGAAGGGCGCTACGTGGTTTCGCGACGCCGAAGGGGGCGGCGACAGATCATCTTGA
 TCGCCACCCGCTCGGAGGTGAGCTGGCGCTGGATGCTGCCAAGCGCTGGCGGAGGAAGGCATCCAGGCTCGGGT
 CGTCTCAATGCCCTGCATGGAAGTCTTCCGCCAGCAGGACGCCAGTTATCGGGACAGCGTGTCAACCCCGCGCTG
 AAACGCTGGCCATCGAGGCTGCCAGCCGCTCCCTGGTATGAGTGGTGGGCTTTGACGCGCGGTGATCGGAA
 TGACCACCTTTGGCGCTTCGGCCCGACCAAAGTCTCTTTGAGAAATTCGGCTTCAACGTGCCGAACGTCGTGCA
 GGTGCTCAAGGGCGTTTTGACAGAGTGA

> TKDgeo

MSPEQQAVRQDQDQLSINTIRTLAIDAVQRANSRHPGAPLGMAMPGYVLWQRFRLRHPKHPPEWPRDRFVLSAGHA
 SMLIYSLHLTGYDLPLEDIKNFRQWGSKTPGHPEFFHTPGLDATTGPLQGGAAMTVGMAMAEHLAARYNRPGFK
 VFDNYTYAILGDLQEGVNHEAASLAGHLKGLKIWLHDDNQVQLDTATFKAANEDTAERYRAYGWEVLRVQDGN
 NLTEIENAIRQARMNTEQPTLIQVVRTVIGFSPRAGTSKAHGEPLGEEGVQETKAALGWDPFPTVPDEVKAHMADA
 TERGAEWEDWNALMERYRAEYPLAAEVDALLARELPANLAEVLPSEYVGSKAIATRNASGEVINALAQVVPGLM
 GGSADLSGSKTTIKDGGEFPLPGNYGGRNVYFVREFGMAAAGNLSLYGGVRLVGTFLVFADYLPKPAFRLSALQ
 FQPVTYVYLTHDSIGLGEDGPTHQPIEQQLAMLRAVPGAHVIRPADANETAAAWQMALEYDKGPTALALSRQDLPLVP
 RNHAGVKKAYVVRDAEGGPAQIILIIATGSEVSLALDAAQALAEIEGIQARVVSMPCEVFRQQDASVYRDSVLTGPV
 KRVAIEAASPLPWYEWVDFGAVIGMTTFGASAPAKVLFKFKGFNPNVQVVKVGLQR

> TKDrad

ATGACAGACCAGAGCGTTTCCAAAACGTGGCGCGGCTGAGTGTGAACACCATTTCGCACGCTCGCCATTGACGCGG
 TGCAGGCCGCCAACTCGGGCCACCCCGGTGCGCCGTGGGCATGGCCCGATGGGCTACGTGCTGTGGCACAAGTT
 CCTGCGCCACAACCCCGCGCACCTGAGTGGCCGGCGCGACCGCTTCTGTGCTTCGGCGGGGACGCCTCCATG
 CTGATCTACAGCCTGTGCACCTGACCGGCTACCAGGAAATGACGCTCGACGACCTGCGCCACTTCCGGCAGTGGG
 GCTACCACACACCCCGGCCACCCGAGTTTTTTCCACACCAAGGGTCTGGACGCGACACCCGGCCCGCTTGGGCAAGG
 CGCGCGATGACGCTGGCGATGGCGATGGCCGAAGCACACTCGCCGCCGCTACAACCCGGAAGGCTTTCCGATT
 TTCGACAACCGCACTACGCCATCATGGCGACGGCATCTGCAAGAAGGCATCAACACGAAAGCCGCGCTCG
 CCGGGCACCTGAAACTCGCAAGCTGATCTGGCTGCACGACGACAACCACATCCAGCTCGACACGCCACGAACAA
 GGCCGAGTCCGAGGACACCCCGCCCGCTTCCGCGCTATGGCTGGAACGTGTGAAGGTGGAAGACGGCGACAAT
 CTGGACGAAATTGAAAAGGCGATTGCCGAGGCCCGCAGCCAAAGCGAGCGGCCACGCTGATTCAGGTGCGCACC
 TCATCGCTTTCGCGAGCCCGCGCGCGGACGAGCAAGGCGCACGCGAGCCGCTCGGCGAAGAGGGCGTGGCCGA
 GACGAAGAAAGCGTTGGGCTGGGAGTACCCCGCTTTTACCCTGCCGACGAAGTGGCTGCGCACATGAACGCTCGC
 GCTAAGGGTGTCAACTCGAAGCCGACTGGGAAAACTGATGGCCGACTACCGCACCGCGCACCCCGACTTCGGCA
 AGGAAGTGGACGCGCTGCTCGCCGTGAACCTGCCCGCAACCTCGCCGACTGCTGCCAAGTACGAAGTTCGGCGG
 CAAGGCGCGGCCACCCGCAACGCGAGCGGCGAAGTCAACGCGCTGGCGAAGGTGCTTCCCGTGTGATGGGCG
 GGCAGCGCGACCTCTCGGCTCGACCAAGACCACCATCAAGGACGGCGCGGAGATGGAAGCGGGCACCATGGGCG
 GGCGCAACGTGCTGTTCGGCTGCGCGAGTTCCGCATGAGCGCCCGGGCAATGGCTGAGCCTCTACGGCGGCT
 GCACCCGATGGTAGGCACCTTCTGTTATTCGCCGACTACCTCAAGCCGCTTCCGCTCTCGGCGCTGCAAAATG

CAGCCGGTGACTTACGTGCTGACCCACGACTCCATCGGTCTGGGCGAAGACGGGCCGACCCACCAGCCGGTGGACC
AGCTCGCCATGCTGCGAGCGGTGCCGGGCGCCACGTCATTCGCCCCGCCGACGCCAACGAAACCGCCCGCGGTG
GCTGATGGCGCTGGAATACGACAAGGGCCCCACCGCCTCGCCCTCTCGCGCCAGGATCTGCCGATTCGCCCGCC
AACATCGAAGGCGTGAAGAAGGGCGCGTATGTCTCCGAGATGTGGACGGTGC CGATGGT CAGGGGCTCAAGTCA
TCCTGATC

GCCAGCGGCTCGGAAGTCGCCCTGGCCCTGAGCAGCGCCGAGCGGCTGGCCGAAGAGGGCGTGCAGGCCCGCGT
GGTGTCCATGCCGTGCATGGAGGTC TTTCCGCCAGCAGGAGCAGAGCTACCGCGACAGCGTCTGACCCCCGGCGTG
AAGCGCGTCGCCATCGAGGCCGCCAGCCCGCAGCCCTGGTACGAGTGGACGCTCGCGGCCAGTCAATCGGCATGA
CGACCTTCGGTGCCTCGGCCCGGCCAAGGTGCTGTTTGAGAAGTTCGGCTTCAGCGTGGAAAACGTGGTGAAGGT
GGTCCACTCCGTGCTGTAA

> **TKDrad**

MTDQSVSQNVARLSVNTIIRTLAIDAVQAANS GHPGAPLGMAPMGYVLWHKFLRHNPAHPEWPGDRDFVLSAGHASM
LIYSLHLHTGYQEMTLDDLRFHQWGYHTPGHPEFFHTKGLDATTGPLGQGAAMTVGMAMAEHLAARYNREGFPI
FDNRTYAIMGDGLQEGINHEAAALAGHLKLGKLIWLHDDNHIQLDTPNKAES EDTAARFRAYGWNVLKVEDGDN
LDEIEKAI AEARSQSERPTLIQVRTIIGFSPRAGTSKAHGEPLGEEGVAETKKALGWEYPAFTVPEVAAHMNAR
AKGAQLEADWEKLMADYRTHAPDLGKEVDALLARELPANLADLLPKYEVGGKAAATRNASGEVINALAKVLPGLMG
GSADLSGSTKTTIKDGGEMEAGTMGGRNVLFVREFGMSAAGNLSLYGGLHPMVGTFVLFADYLPAPFRLSALQM
QPVTYVLT HDSIGLGEDGPTHQPV DQLAMLRAVPGAHVIRPADANETA AAWLMALEYDKGPTALALSRQDLPILPA
NIEGVKKGAYVLRDVGADGQGAQVILIASGSEVALALSSAERLAE EGVQARVVSMPCMEVFRQEQSYRDSVLT P
GVKRVAIEAASPQWYEWTLGGPVI GMTTFGASAPAKVLF EKFGFSVENVVKVVHSLV

> **pQR1743**

ATGGCGCATTCGATCGAAGAATTAGCGATTACCACCATT CGAACGCTGTCGATTGACGCGATCGAAAAAGCGAAAT
CCGGGCACCCGGGCATGCCGATGGGCGCGGCCCGATGGCGTATACGCTCTGGACGAAATTTATGAACCATAATCC
AGCCAATCCAACTGGTTTAAACCGCAGCCGGTTTGT TTTGTCCGCTGGGCACGGGTCGATGCTGCTTTACAGCCTG
CTTCATCTAAGCGCTACGATGTCACGATGGACGACTTGAAACAGTTCGGGCAATGGGGAAGCAAAACCGCCGGCC
ATCCGGAATACGCCATACGCTCAGGCGTGGAGGCAACGACCCGCCCGCTCGGCCAAGGGATTCGATGGCGGTCCG
CATGGCGATGGCGGAACGGCATT TGGCGGCTGCATACAATCGCGATGGATTGACATTATCAACCCTACACGTAT
GCGATT TGGCGGACGGCGATTGATGGAAGGAGTGGCGAGCGAAGCGGCTCACTCGCCGGCCACTTGAAGCTCG
GCCGTCTGATCGTCTGTATGACTCGAACGACATTTCTGCTGGACGGCGAGCTCAACTTGTCTGTTTTCGAAAACGT
CGCCCAACGTTTTCCAAGCGTACGGCTGGCAATATTTGCGCGTTGAGGACGGCAACAATATGAAGAAATCGCCAAA
GCGCTCGAGGAGCGCGGACGGACCTCAGCCGGCCGACGCTCATTGAAGTAAAAACGACGATTTGGCTACGGCGCGC
CAAATAAAGCGGGCAGCTCCGGCGTCCACGGCGCTCCGCTCGGCGCCAAAGAGGCGAAGTTGACGAAAAGAAGCGTA
CCGCTGGACGTTTTCCGAAGATTTCTACGTGCCGGATGAAGTGTACGCTCATT TCCGGGAAACGGTGAAGAAGCC
GGAGCGAGAAAAGAAGCGGAGTGAATGAGCGCTTCGTTGCTTACGAGCGGGCGCATCCGGAATGGCCCGCGAGC
TGAAGCAGCGCATTGAAGGGAAGCTTCCGGATGGCTGGGAAACATCGCTGCCGGTATGAAGCGGGCAAAGCTT
GGCGACCCGCTCATCGTCCGGGGAAGTGATCAACGCCATCGCCAAAGCGGTGCCGATTTGTTGGCGGTTCCGGCG
ACTTGGCAAGCTCGAATAAAACGCTTATCAAAGGCGGCGCAACTTCTTGGCGGACAGCTACGAAGGGCGCAACAT
TTGGTTTGGCGTGC GCGAGTTTGCATGGGCGCGGCTTAAACGGCATGGCGCTTACGGCGGGTTGAAAGTGTTC
GGCGGCAGCTTCTCTGTTCTCCGACTACTTGC GCCGGCGATTTCGGCTGGCGGCGCTCATGGGCTTGGCGGTGA
CGTACGTGCTGACGCACGACAGCATCGCCGTGGGGAAGACGGCCCAGCATGAGCCGGTTCGAGCATCTCGCTTC
ACTTCCGGCGATGCCAACTTGTAGTCATCCGGCCGGCTGACGCAAACGAAACGGCGGCCGCTTGGCGGCTGGCG
CTCGAGTCGACGAACAAGCCGACTGCGCTCGTCTTGACGCGTCAAGATGTGCCGACATTGCCGACAACCGCTCAGT
TGGCGTATGAAGGCGTGA AAAAAGGCGCTACGTCGTTTACC GGCGAAAACGGCGCTCCGGAGGCGCTGTTGTT
GGCAGCTGGCTCGAAGTCGCTTGGCCGTC AAAGCGCAAGAAGCGCTCGCCGCTGAGGGCATCCATGTCTCCGTC
ATCAGCATGCCATCGATGGGACCGCTTCGAAGCGCAGCCAAAATCGTACCGCATGAAGGTGCGCCGCGCCGCTGGA
CGAAGCGGCTCGCCATTGAAATGGGCGCTCGCTCGGTTGGGAGCGCTACGTGCGCGCCGAGGGCGACATTTTGGC
CATCGACCGATTCCGTTGCTTCCGCTCCGGGAGAGAAAATCATGGCCGAGTATGGCTTTACGGTTGACAACGTCGTC
CGCCGCACAAAAGCGCTGCTCGGCAAGTAA

> **pQR1743**

MAHSIEELAIITTTIRTLSIDAIEKAKSGHPGMPMGAAPMAYTLWTKFMNHNPNANPNWFRDRFVLSAGHGSMLLYSL
 LHLSGYDVTMDDLKQFRQWGSKTPGHPEYGHTPGVEATGPLGQGIAMAVGMAMAERHLAAAYNRDGFDI INHYTY
 AICGDLMEGVASEAASLAGHLKLRLLVLYDSNDISLDGELNLSFSENAQRQFQAYGWQYLRVEDGNNIEEIAK
 ALBEEARTDLSRPTLIEVKTTIGYGAPNKAGTSGVHGAPLGAQEAKLTKEAYRWTFSEDFYVPEVYAHFRETVEA
 GARKEAEWNERFVAYERAHPLELAELKQAI EGKLPDGWETSLPVYEAGKSLATRSSSSEVINAIKAVPQLFGGSA
 DLASSNKTLIKGGGNFLPDSYEGRNIFWGVREFAMGAALNGMALHGGLKVFVGGTFFVFSYDLRPAIRLAALMGLPV
 TYVLTHDSIAVGEDGPTHEPVEHLASLRAMPNLSVIRPADANETAAAWRLALESTNKPTALVLRQDVPTLPTTAQ
 LAYEGVKKGAYVSPAKNGAPEALLLATGSEVLAVKAQEALAAEGIHVSVISMPSWDRFEAQPKSYRDEVLP
 PAVTKRLAIEMGASLWERYVGAEGDILAI DRFGASAPGEK IMAEYGF TVDNVVRRTKALLGK

> **pQR1744**

ATGAACACCGGCACCCCAAAGACCTGGACTGGTCTGATCTCGATAGACGTACCGTAGACGTGGTTCGTGCCCTGG
 CGATGGACGCGTTCGAAGAAGCGGGATCCGGGCACCTGGAACCGCGATGAGTCTGGCGCCTGTGGCCTACCTGCT
 CTTCAGAAAGGTGATGCGGCACGATCCGACAGATCCGAAGTGGATCGGCCCGCACCGCTTCGTCTGTCTGCGGG
 CACTCCAGCCTCACGCTCTACATCCAGCTCTACCTGGCTGGCTACGGGCTGAGCCTGAACGACATCAAGCGGCTGC
 GCCAGTGGGGCAGCCTACCCCGGGCACCCCGAATACGGGCACACCGCCGGGGTGGAAACCACCCGCGCCCTT
 GGGCAGGGCATCGGCAACCGCGTCCGATGCGCATGCCCCGCCCGCCCGGGAGCGGGCATCTGCGCCGACACC
 CCGATCGGGGAAAGCCGTTTCGACCACTACATCTACGTCTGTGCTCTGACGGCAGCTCCAGGAGGCATCAGCC
 ACGAAGTAAAGTGCCTCGCCGACGCGAGAAGCTCGCAACCTCATCGTCATCTGGGACGACAACCGCATCTCCAT
 CGAAGACGACACCCAGATCGCATTCACCGAAGACGTCTGTCGCCGCTACGCCGCTACGGCTGGCACGTCCAAGAG
 GTCGAGTGGTTCGGCGAGGACGGCTCCTACCACGAAGACGTGGCGGCGTGTACGACGGATCCGGGCCGCCAGG
 CGGAGACGGAACGTCCCTCTTCATCCGGCTGCGCACCATCATCGGCTGGCCGTCCCCGAACAAGCAGAACACGGG
 GGCGATCCACGGCGCCGCGCTGGGGGCTGAAGAGGTGCGCCACCAAGCGGGTGTGGCCTCAACCTGAGGCG
 CAGTTCGACGTGCCAACGAGCTGTGGAGCACGCCCGGGGCGTGGTGGCGCGGGCCGCGCCGCCAGGAAT
 GGGAGGCCTTGTTCGCCAAGTGGCGGGCCAAACCGGGCGAGCGTGCCTGACTGTTCGACCGGCTGATGGCAGGCTC
 GCTCCCGGACGGTGGGGAAGCGCATCCGACCTTCGAGCCACGCGCTAAGGGCATGGCCACCCGGAAGCGTCC
 GGTGAGGTGCTGAGCGCGATCGCCCGGTGCTGCCGAGCTGTGGGGCGGCTCGCGGACCTPGCGCGGATCCAA
 ACACCACGCCTAAGGGCGAGCCGTCGTTTCATCCCGAGGAGCGGTCCACGAAGCGGTTCTCCGGCCACCGCTACGG
 CCGGGTGTGCACTTCGGGATCCGTGAACACGGCATGGGGGCGATCCTCAACGGGATCGCGCTGCACGGCCCCACC
 CGCCCCACGGTGGCACCTTCCTCGTGTTCAGCGACTACATGCGGCCGTCGGTGC GGCTGGCTGCCCTGATGAAGC
 TGCCGGTACGTACGTGTGACCCACGACTCGATCGGTCTGGGCGAAGACGGACCACCCACAGCCGTGGAGCA
 CCTGTGTGTCGTCGCGCCATCCCGGCCCTGGCGGTGTGCGTCCCGCCGACGCCAACGAGACGGCAGTGGCCTGG
 CGCACCATCCTGGAACGCAATGACGGCCCGGTGGCGCTGCGCTGACCCGGCAGTGGTTCGGTTCGACCGCT
 CCGAGCTCGCCTCTGCGGAGCTGGTCTCCCGCGCGGGTACATCTGGCCGAAGCCAGCAACGGCCGCTCCGGAGGC
 GATCATCATCGCCACCGGAAGTGAAGTGCAGATCGCGTGGAGGCGGTTCCCGCCTGGAGGAGTCCGGTACTCCT
 ACCCGTGTGGTGTGATGCCGTCCTGGAGTGGTTCAACGAGCAGGACGACGCTACCGCCAGCAGGTGCTTCCAC
 CGTCCGTCGGGTCCGGTCTCCGTGGAAGCCGGGGTCCCTTGGGCTGGCGCGAGCTGGTGGGCGAGTATGGCGA
 GTCGGTGTGCTGGAACACTTCGGCGCTTCGGCTCCGTACCGGACTCTCTACGAGCAGTTCGGGCTCACCGCCGAC
 CGGGTAGTGGCAGCGGTACTACTCCAGCGCTGCCAAGCTCGCGGTGACCGTGGATCAACGACCGGCAACTGA

> **pQR1744**

MNTGTPKTLWDSDLDRRTVDVVRALAMDAVEEAGSGHPGTAMSLAPVAYLLFQKVMRHDPTDPKWI GRDRFVLSG
 HSSLTLYIQLYLAGYGLSLNDIKRLRQWGS LTPGHPEYGH TAGVETTTGPLGQGI GNAVGMAMAARRERGLFDPDT
 PIGESPFDPHYIYVLCSDGDVQEGISHEVSALAGTQKLGNI LVIWDDNRI SIEDDTQIAFTEDVVARYAAYGWHVQE
 VEUVGEDGSYHEDVAALYDAIRAAQAETERPSFIRLRTI IGWSPNKQNTGAIHGAALGAEVAATKRVLG LNP
 EAQFDVPNELLEHARGVVARGRAARQWEALFAKWRANAGERAE LFDRLMAGSLPDGWEKAIPTFEP
 SAKGMATRKASGEVLSAIAPVLP ELLWGSADLAGSNNTTPKGEPSF IPEERSTKAFSGHRYGRV LHF
 GIREHGMGAILN GIALHGPT RPYGGTFLVFS DYM RPSVRLAALMKLPV TYVWTHDSIGLGEDGPT
 HQPVEHLWSLRAI PGLAVVRPADANETA VAWRTILERN DGPVALALTRQSVPLDRSELASAE LVS
 RGGYILAEASNGRPEAII IATGSEVQI ALEARSRLEESGTP

TRVVSMPCLEWFNEQDDAYRQOVLPPSVRVRVSVEAGVALGWRELVGEYGESVSLEHFASAPYATLYEQFGLTAD
RVVAHVHSSAAKLGDRGSTTGN

> pQR1745

ATGGAAAGGTTTCCCTATGAAAACTTCCAGAAAGCGAACTCAAAGAGTTGAAAGAACTCGGAAGGCTCTGCCGTG
GCGACATACTGAAAAAGACTACATAGCTAACTCAGGCCATCCTGGAGGATCCATGTCTTCGATCGATCTTTATCT
TACCGTCTTCAAGTACGCAAACTCAGACCCGTCGATGATCCTGCAAGAGACAGAATCGTGATCAGCCATGGACAC
ACTTCTCCGGGTGTCTACGCAGCTATGGCTCGTTTGGGGTTTGTTCGATCTCGATGAAGTCTCGCAGGATTCAGAC
ACCCCGCTCCGTTTGTGAAGGACACGTGACCCGAGGTGTTGGGATCATCGACTGGACAACCGGAAACCTCGGTCA
GGGTCTTTCAGCCGACTCGGTTTTGCCCTCGCATCCAGGTTTACAGGAAAAGATTACCACGTCTTTGTCTCATG
AGTGACGCAGAACAGGCAAAAGGACAGGTGGCGGAGGCAAGAAGAGTGGCGAAAAAGTACGGTGTACGAATCTCA
CAGTGATCATCGACTACAACGACGCCAGATCAGTGGCCGTGCCAGAGACGTATGCCCGTGAACATAAAGGAAAA
CTACTTAGCGGACGGCTGGAGGGTATCGAGATCGATGGGCACGACTACGAACAGATCTATCTCGCACTGAAAGAA
CGGGTAGAAGACGAAC

TGAATCCCGTTGCCATAATCGCCAAAACGGTTCATGGGAAAAGGCGTATCTTTTCATGGAAAACGAGGTGAAATAC
CACGGAAAGCCTTTGAAACAGAGAAGAACTCGAAAAAGCCCTCGCGGAACTCGGAATTGAAAACGATGTGTGATGTG
ACATCGAAATAAGAAAACACTTCCAGTGGAAAAACACAAGAAAGTCTACAAAACCTTACCAGTCAAGATCGACAC
GGGAGAGCCCATACCTACACCTCACCCACTGACAACAGAAGCGCATTCGGAAGAGCTATTTCTGGATCTGGTGAAG
AAGAACGTAACAATCCAGAAACCACCCATCGTCGCTGTGGACTGCGACCTGAAGGGATCGGTCAAACCTCGACC
TGCTCGACAAAGAGTTCCCTGAGAGACTCCTGGAAGTGGGCGTTTCAGGAACACAACCGTGCCTGATGGCGGGGGC
ACTCTCCGAGAGGGTGTGATCAGGTTCTTCGCTGATTTTGGTGTTTTTGGAATTTCTGAAACCTACACCAGCAC
AGGCTGAACGCCATCAATGGAACGAACCTCAAAGTCGTTGTACACACTGCGGACTCAACGTGGGAGAGGACGGAA
AAACTCATACGGACTCGACTACGTTTCCGGGCCGATGAACTGGTACGGTTTCAAAGTGATCGTCCCTGGTATCC
CAACCAGACGGATAGAGTTGTCAGATACGCCCGGAAGGAATACGGGAACTTCGTAATCGCCATGGGAAGATCTAAG
CTTCCCATCATCTCGATGAAAACGGGAAACCTTTCTTCGGAGAGGGTTACACCTTCGAATATGGGAAGATCGATG
TCGTTAGAAAAGGTGACGACCGGTGATCATAACTTACGGTTCTACACTCTGTGAAGCCGTAAATGCCGACAGCGA
ACTCAAGAAAAGAGGAGTAAACGTAGCCGTTCTGAATGTCTCTGTCCGGTGGATCTCGATATAGAGACCTTGAAG
ATGGTCGATGGAAAACCCGTTCTCGTTGTGGAGGATCACAACGTTTTTCACAGGACTTGAAGCTTCTCTGGAACCA
CCCTTCTGAAAACGGCATCATCCGGAAGAAATACGTGAGAGTAGGTGTTCAGAAATTCGCCGTGTCCGGCAGTTA
CACGATGCTCTACAACTCTACGGCCTGGATAAAGATGGAATAATTTCCAGACTCAGAGAGATGCTCTAA

> pQR1745

MERFPYEKLPESSELKELKELGRRLCRGDLKMTYIANSHPGSSMSSIDLVLTVFKYAKLRPVDDPARDRIVISHGH
TSPGVYAAARLGFVDLDEVLAGFRHPASVFEHGVTRGVGIIIDWTTGNLQGLSAGLGFALASRFTGKDYHVFVLM
SDAEQAKQVAEARRVAKKYGVNTLTVIIDYNDAQISGRARDVMPVNIKENYLDAGWRVIEIDGHDYEQIYLALKE
AVEDELNPFVAIIAKTVMGKGVSMENEVYHGKPLNREELEKALAEELGIENDVDVYIEKRKQLPVEKHKVKYKTY
IKIDTGEPIITYTSPDNRSAFGKAILDLVKKNVNNPETTPIVAVDCDLKGSVKLDDLDFEFLERLLEVGVEHNAAA
MAGALSAEAGVITFFADFGVFGISETYNQHRLNAINGTNLKVVTTHCGLNVGEDGKTHHGLDYVSGPMNWYGFVIV
PGDPNQTRVRYAAKEYGNFVIAMGRSKLPILDENKPPFFGEGYTFEYKIDVVRKGDDAVITTYGSTLCEAVN
AADELKKEGVNVAVLNVSCFVDLDIETLKMVDGKPVLVVEDHNVFTGLGSFLGTLLLENGIIPKKYVRVGVPEFAV
SGSYTMLYKLYGLDKDGIISRLREML

2.5. TAm DNA/ amino acid sequences

DNA sequences were retrieved from the NCBI database [5] and amino acid sequences were obtained through UniProtKB [6].

> TAm_{Gste}

ATGAAATTGGCAAAAACGGGTGGCGTCGCTGACGCCATCGCGCAGCTTTGGCCATTACGGAGAAAAGCAAAAAGAACTAA
AAGCGGCCGGCATGACGTGATTGGTCTCGGAGCTGGCGAACCGGATTTCAACACGCCACAGCACATTTCTTGATGC
CGCCATCAAGGCAATGAACGAAGGACATACGAAATATACACCATCGGGCGGTTTCCGGCGCTTAAAGGAGGAAATT

ATAAAAAAATTCGCCCGCGACCAAGGCTTGGATTATGAGCCGGCTGAAGTGATTGTATGCGTTCGGAGCGAAGCAGC
 CCCTTTACACGCTGTTCCAAGTATTGCTCGATGAAGCGACGAAGTGATCATTCCGACGCCATACTGGGTGAGCTA
 TCCGGAACAAGTGAAACTGGCGGGCGGTGTTCCGGTTTACGTCGAAGGGCTTGAACAAAATCATTTTAAAAATTACG
 CCGGAGCAGCTGAAACAGGAATCACGCCGCGGACGAAAGCGGTATCATCAACTCGCCGAGCAACCCGACTGGCA
 TGATTTATACAGCCGAAGAGTTGAAGGCGCTTGGTGAGGTGTGCCTAGCGCATGGTGTATTGATCGTGTGAGATGA
 AATTTACGAAAAATTGACTTACGGCGGGGCGAAGCATGTGTCCATCGCTGAGTTGTCCGGAGCTGAAGGCGCAG
 ACAGTCAATCATTACGGCGTGTCAAAGTCGCATTTCGATGACGGGCTGGCGCATTGGTTATGCGGGCGGGCCGAAAG
 ATATTATTAAGGCAATGACAGATTTGGCGAGCCACAGCACGTCGAACCCGACGTCATCGCCCAATACGCGGCCAT
 CGCTGCTTACAGCGGGCCGAGGAGCCGGTCAACAAAATGCGCCAAGCGTTTGAACAACGGCTCAATATCATTTTAC
 GACAAGCTCGTGCAAAATTCAGGATTCACGTGCGTTAAGCCACAAGGGGCGTTTTATTGTTTCCGAACGCCCGCG
 AAGCGGCTGCAATGGCCGGCTGCCGACGGTCGACGAGTTCGTCGCTGCCTTGTGGAGGAAGCGAAAGTCGCGCT
 TGTGCCCGGCTCTGGGTTTGGAGCGCCGATAACGTTCCGCTTGTGCATACGCGACATCGCTCGATGCACTGGAACCC
 CCGTGAACGCATCCACCGGTTTATGGAAGCGCGCGCTTAA

> TAm_{Gste}

MKLAKRVASLTPSATLAI TEKAKELKAAGHDVIGLGAGEPDFNTPQHILDAAIKAMNEGHTKYTPSGGLPALKEEI
 IKKFARDQGLDYEPAEVI VCVGAKHALYTLFQVLLDEGDEVI IPTPYWVSYPEQVKLAGGVPVYVEGLEQNHFKIT
 PEQLKQAITPRTKAVI INSPSNPTGMITAEELKALGEVCLAHGVLIVSDEIYEKLTGGAKHVSIAELSPELKAQT
 VIINGVSKSHSMTGWRIGYAAQPKDI IKAMTDLASHSTSNPTSIAQYAAIAAYSQPEPVEQMRQAFEQRLNI IYD
 KLVQIPGFTCVKPGAFYLFPNAREAAAMAGCRTVDEFVAALLEAKVALVPGSGFGAPDNVRLSYATSLDALETA
 VERIHRFMEARA

> TAm_{DGeo}

ATGTTCCGAGGACACGCCCGCACCTTTCACCCGACATTTCTGCTGACGCCCGGTCCGACACCGATTACCCCCGGG
 CCCAGCGGGCGCTGCTGCGCGAGATGCTGGGGCACATGGACCCCTGAGGTGTTTCGCCCTGAACCCGCGAGATCCAGGC
 GGACTTGCGGGTGATGTACGGGACGGGGCCCCAGACCTTTACGGCGCTGCTGGCGGGCACCCGGAGCCTGGGCATG
 GAGGCGGGCTTCGCCAATTTGTTGGAGAGGGGAGACGACGCTGATCTGCGTCAATGGTGCCTTTGGTACAGCGCA
 TGGCCGAGATGGCGGCGCTACGGTGCGAATGTACGGCGGGTGACCGCGCCGCTGGGCGAGCCGATCGACCCGGC
 CGACGTGGCTGCGCGGTTGAGCGGCGCGCGCTGGTGGCGGTGTCATGGGGAGACGAGCACGGGTGTGCTCAAT
 CCGCTTCCGAGATTCGCGAGCCGTTGCGCGGGAGCGGGCATTTGCTGGCCGTGGACGCCGTGACGACCCCGGGA
 TGGAACCTTCCATATGGCGGACTGGGGCGTGGACTACGCCTATACGGCGCGCAGAAGTGCCTCTCGGCACCCGCC
 CGGCTGGCCCCGGTGGCGATCAGCGACCGTGTCTCGCTCGCCACGCGGCCCGCCGACGCCACGCCGCTGTGG
 TACTGCGATTTTGGAGGCTGCGCGACTACTGGGACCGGCACAGCTACCACCACAGGTCCCGGTGAATCTGCACT
 ACGCCTTCCACGCGCCCTGCGCGCCGACTCGAAGAAGGCCCTCAAGCCCGGACGGCCCGGTGCGCGACCTTGG
 CCAGGCGGTGCTGGCGGCCCTGACGCCGCTGGGCTTACGCGCTATGTGGCCGATCCCGCCGCGCGGCTGCCACC
 GTCTTGGCCCTGCGTCTTCTCCCGGCTTCGACGACGCGGGCGTTCGCCAGGCCCTACGGGAACGCGGGATCAGCG
 TGACCGGCGGCTGGGACCGACGGCAGGGCTGATCTGGCGTCTGGGCTGATGGGGGAAGCGGCTCGCCCCGCGCC
 CTACCGCGCGCTGATCTCGCCCTGGAAGACCTGCTGGCGGAGCGGGGCTTGGTGGCGCGCTTCGAGGAGGCGCTG
 GCGTCCGCGCCTGA

> TAm_{DGeo}

MFEDTPAPFPPIHLLTPGPTPIHPRAQRALLREMLGHMDPEVFNREIQADLRVMYGTGPQTFTALLAGTGSGLM
 EAGFANLVERGDDVLICVNGAFGQRMAEMAARYGANVRRVTA PLGEPIDPADVAARLSGARLVAVVHGETSTGVLN
 PLPEIAEAVRGSALLAVDAVTTAGMEPFHMADWGVYAYTGAQKCLSAPPGLAPVAISDRALARHAARRTPPTPLW
 YCDFEGLRDRYSYHHTVFNLHYAFHAALRAALEEGLQARQARVRLDQAVLAALTPLGFTPYVADPAARLPT
 VLALRLPGFDDAGVRQALRERGISVTGGLGPTAGLIWRLGLMGEAARPA PYRALMLALEDLGERGLVARFEAL
 GVAA

> pQR1746

ATGACCTCTCCTTTCCGCCCTCTCCGCCCGCGCCAGAGCCTCAAGCCGTCTGCGACAGTGCGGGTACAGTCCCGCG
 CCCTGGAACTCCAGCGTCAGGGCCTGGACGTGATTTCCATGAGCGTGGGCGAGCCGGATTTTCGACACGCCGCCACA
 TGTCAAGGCCCGCCGCATCGCCGCCATCGAGGAAGGCAAGACCAAATACACCCCGGTACAGCGCATTTCCGAACTG
 CGCGAGGCCATCAGCGCCAGTTTTCGGCGCGAAAACGGCCTGGACTACGCGCCGAACGCCGTGACGGTAACGAGCG
 GCGGTAACAGCGCTGTTCACGCCTTTTTCGCGTTGCTGAACCCCGCGACGAGGTGCTGATTTCCCGCGCCCA
 CTGGGTTCAGTACCCCGAAATGGTCGCGCTGACCGCGCGGTGCCGGTAACCGTACCCACTACGCCGACGAGGGC
 TTTCAACTCGACCCGGACGCCCTCGCCGCCCATCACGCCGCGCACCCGCATGGTGATTTCTCAACAGCCCGGCA
 ACCCGACGGGCGCGGTGTTTCCGCCGAAACCTTGCGGGCGGTGGCCGACCTCGCCACGCAGCACGGCTTGATGAT
 CGTACCGGCAAAATCTACGAGCACCTCGTCTACGACGCCGAGCAGGTACAGTACGCGACCTACGCCGCGGAGCAC
 ACCCTGACCATCAATGGCGCGAGCAAAGCGTATGCCATGACCGGTGGCGCATCGGCTACGCCGGGCGGCCCGCG
 AGGTGATTTGCCGCCATGAACGCGCTGCAATCGCAAAGCACAGCAACGCCAGCAGCGTACGCCAGTACGCCGCCCT
 CGCCGCGCTGGAACAGCACGAGGAAACCATGCGCTTCATCGACAGGGCCCGCACCCCTACCGCGAACGGCGCGAC
 CGCATCGTGGCGGGCCTCAACGCGCTGGGGTTGCCACGCCACGCCGAAGGGCCTTTTACGTGATGGCCGACA
 CCCGCGCCATTCACACCGGCAACTCGAAGCCGCCGCATCATTTCTGGATGAGGCGCAGGTGCGCGTCTGCGCGG
 CACCGATTTCCCGCGCCGGGACAGGTGCGCCTGAGCTACCGGACCAGCATGGACAACATCGAGGAAGTGCTCGCG
 CGGCTGGAAGGGTCTGTCGCGCGCTAA

> pQR1746

MTSPFRLSARAQSLKPSATVAVTSRALELQRQGLDVISMSVGEPDFDTPPHVKAAGIAAIEEGKTKYTPVSGIPEL
 REAISAKFRRENGLDYAPNAVTVTSGGKQALFNAPFALLNPGDEVLIPAPHVWSYEMVALTGAVPVTVPTTQQG
 FQLDPDALAAAITPRTRMVLINSPGNPTGAVFPPELRAVADLATQHGLMIVTDEIYEHLYDQVSIQTYAPEH
 TLTINGASKAYAMTGWIRIGYAGGPREVIAAMNALQSQSTSNASSVSQYAAALAEQHEETMRIFDRARTAYRERRD
 RIVAGLNLGLPTPTPQGAFYVMADTRAIHTDELEAARIILDEAQVAVVPGTDFAAPGQVRLSYATSMDNIEEVL
 RLEGVVRR

> pQR1748

ATGACTGACCGACCTCGTATCTCCGCACGCATCGGGCGTATCTCCGAGTCAGCGACCTGGCGGTGGACGCCAAGG
 CCAAGGCCCTGAAGCGCGTGGGCATCCCGTGATCGGCTTCGGCGCGGGGAGCCTGACTTCCCCACGCCCGACTA
 CATCGTGGAGGCAGCGGTGCGCCCTGCCGCGACTCGCGCTTCCACCGTACACCCCGCGGGAGGCCCTCCCCGAA
 CTCAAGGAAGCCATCGCGGCTAAGACGCTGCGCGACTCCGGCTACCGGGTGGAGCCGAACCAAGTCTGGTACCA
 ACGGGCGCAAGCAAGCGATCTACGAGGCGTTCGCCACGCTGCTCGATCCGGGCGACGAGGTATCGTGATCGCGCC
 CTACTGGACCACCTACCTGAATCGATCCGGCTGGCCGGAGGAACCCCGGTCTACGTGGTACCGACGAGTCCACT
 GGCTACCTGGCCACGGTTCGAGCAGTGGAGGCGGCCCGCACCGACCGCACCAAGGTGCTGCTGTTCTCCCT
 CGAACCCGACCGCGCGGTGTACTCGCCCGAGCAGTCCCGGAGATCGGCCGGTGGGCCCTCGAACACAACCTGTG
 GGTGCTCACCGACGAGATCTACGAGCACCTCGTCTACGGGGACGCCCGGTTCTCCTCGATGCCGGTGAAGTCCG
 GAACTGGCCGACCGCACCGTGGTGGTCAACGGGGTGGCCAAGACCTACGCCATGACCGGGTGGCGGGTTCGGCTGGC
 TCATCGGCCCCGTGGACGTGGTCAAGGCTGCGACCAACCTGCAGTTCGACGCCACCTCCAATGTGGCCAACGTCTC
 GCAGGCCGCGGCTCTGGCAGCGGTCTCCGGCGACCTGTGCGCGGTGGAGGAGATGAAGCAGGCCTTCGACCGCGG
 CGGCAGACCATTTGCGGATGCTCAACGAGATCCCGGTGTGGTGTGCCCGAGCCCCAGGGCGGTTCTACGCCCT
 ACCCGTGGTCAAGGAGATCCTCGCAAGGAGATCCCGGTTCAGCGTCCGACAGCTCCAGCGAGTGGCGTCTGCT
 GATCCTGGAGCACGCCAAGTTCGCGGTGGTCCCGGGCAGGCGTTCGGCACTCCGGGTACTGCGGTGTGCTTAC
 CGTGTGAGCGACCGCATCTGGTCAAGGGGTTCAGCCGGATCGCCAAGCTGCTGAGCGAAGCCACTGA

> pQR1748

MTDRPISARIGGISATLAVDAKAKALKAAGHPVIGFGAGEPDFPTPDYIVEAAVAACDRSRFHRYTPAGGLPE
 LKEAIAAKTLRDSGYRVEPNQVLVTNNGKQAIYEAFATLLDPGDEVIVIAPYWTYPESIRLAGGTPVYVVTDEST
 GYLATVEQLEAARTDRTKVLLFVSPSNPTGAVYSPEQVREIGRWALEHNLWVLTDEIYEHLYGDARFSSMPVEVP
 ELADRTVVVNGVAKTYAMTGWVWGLIGPVDVVKAAATNLQSHATSNVANVSQAALAAVSGDLSAVEEMKQAFDRR

RQTIVRMLNEIPGVVCPPEQAFYAYPSVKEILGKEIRGQRPQTSSELASLILEHAKVAVVPGEAFGTPGYLRLSY
ALSDADLVEGVSRIAKLLSEAH

> **pQR1749**

ATGGTATCCAGGAGAATATCAGAGATFCCCATATCGAAAACCATGGAACTCGACGCGAAGGCCAAAGCCCTCATAA
AAAAGGGAGAAGACGTGATCAATCTAACGGCTGGTGAGCCGGATTTTCCACACCCGGAACCCGTCGTGGAAGAAGC
GGTGAGATTTTCCAGAAAAGGAGAAGTGAATACACAGATCCTCGTGGTATCTACGAACTCAGAGAGGGTATAGCG
AAAAGGATAGGCGAGAGATACAAAAAGATACTCACCGGATCAGGTCGTGGTGACGAATGGAGCGAAACAGGCTC
TGTTCAATGCTTTCATGGCCCTTCTCGATCCCGGTGACGAAGTGATCGTGTTTTCTCCCGTCTGGGTACGTACAT
TCCTCAGATCATCTTGCTGGTGACCGGTGAACGTGGTTGAGACGTTTATGAGTAAAAATTTCCAGCCAGTCTG
GAAGAGGTGGAAGGGCTTCTTGTGGGAAAACGAAAGCCGTTCTTATCAACTCGCCGAACAATCCCCTGGTGTGG
TGTACAGAAGAGAGTTCCTTGAAGGACTTGTGAGACTTGCCAAGAAGAGGAATTTCTACATAATCAGCGACGAAGT
CTACGATFCCCCTGTTTACACGGATGAATTCACATCGATACTCGATGTTTCTGAAGGATTCGACCCGATAGTTTAC
ATAAACGGCTTCTCGAAGTCTCACTCCATGACCGGGTGGAGGGTGGGTTACCTGATATCGAGCGAAAAAGTAGCGA
CCGCTGTGTCGAAGATCCAGTCTCACACCACCTCCTGTATCAACACGGTAGCACAGTACGCCGCCTTGAAGGCTCT
GGAAGTGGACAACCTTACATGGTTCAGACCTTTAAAGAAAGAAAAATTTCTGGTGGAAAAGATTGAAAAGATG
GGTGTAAAGTTCGTGGAACCAAGAGGTGCGTCTACCTCTTTTCAAAGTCCGGGGTGACGATGTGAAATTCGTGTG
AAAGGCTCCTCGAAGAAAAGAAAGGTTGCACTCGTTCAGTTCCGATCCGCTTTTCTGAAGCTGGATTTGTGAGGCTTC
TTTTGCCACATCTATAGAAAAGACTTACGGAGGCGCTGGATAGAATTGAAGACTTCTCAATTCCTGTTGA

> **pQR1749**

MVSRRISEIPI SKTMELDAKAKALIKKGEDVINLTAGEPDFPTEPEPVVEAVRFLQKGEVKYTDPRGIYELREGIA
KRIGERYKKDISPDQVVVINGAKQALFNAMALLDPGDEVIVFSPVWVSYIPIQIILAGGTVNVVETFM SKNFQPSL
EEVEGLLVGKTKAVLINS PNNPTGVVYRREFLEGLVRLAKKRNFYIISDEVYDSLVTDEFTSILDVSEGFDRIYV
INGFSKSHSMTGWRVGYLISSEKVATAVSKIQSHTTSCINTVAQYAALKALEVDNSYMVQTFKERNFVVERLKKM
GVKFVEPEGAFYLFKVRGDDVKFCERLLEEKKVALVPGSAFLKPGFVRLSFATSIERLTEALDRIEDFLNSR

> **pQR1751**

ATGGCACCTGCAGCTGCGCCACCTGCACACCTTCGGCGAACTGGATCCGCCGCAACGCCTGTTGATGGGCCCCGGCC
CGGTCAATGCGCATCCACGCGTGCCTGCGTGCGATGGCGCCGACCTGCTTGGCCAGTTCGACCCGGAAATGACCAC
CTACATGAACGAGGTGATGGCGCTGTACCGCCCTTGTTCGGCACCCAGAACCCTGGACCTTCTGTTGTCGATGGC
ACGGCGCGCCCGCATCGAAGCCGCGCTGGTGTGCTGTTGTCAGCCGGCGACCGTGTGCTGGTGAACAATTCG
GCCGCTTCGGTTTGTTGCTGACCGAAATCCTTGGCCGGCTCGGCGCCGACGTCCACACCGTGGATGCGCCGTGGGG
CGAGGTGGTGCCTGCGCGGATGTCGAGGCGATCGCAAGCGTGGCACCCAAGCTGGTGGCCACCGTGCACGGC
GACACCTCCACCACCATGGCGCAGCCGCTCGATGGCTTAGGCGCGCTATGCCGGGCGGCCGCGCGCTGAGTTACG
TAGACGCCACAGCCACCATCGCGCGCATGGACATCGCCAGCGACCGCTGGGAGGTGGACGTGGTACC CGCGGGGCT
GCAGAAATGCCTGGGCGGGCCGCTCCGGCTCGGCGCCGATCACTGTCTCTGCCGCGCAGCGGAGGCGATCTTTGCG
CGGCGCATGTGCAACGCGGCATCGTGCGGAGGACATCGCCAACGGCAGCGGCCACGCATCGCCTCGAATTTATT
TCGACCTGGCGATGATCATGGATTACTGGTCCGACAAGCCCTCAATCACCAACCGAAGCCACACCATGCTGTA
CGGCGCGCGCAATGCGCACGCTGGCCTTTCAGGAAGCCCTGGAGGCGCGCTACGCCCGGATGCGGCTGCCGCG
CGGCGGTGACGCGCGGCTGCGCGCACTGGGGCTGGAGGTGTTTCGGCGACGATGCGCACCCGATGAGCAATGTCA
CCGGCGTGGTATCCGCGACGGCTGACAGTGAAGCAGTGGCGCGCGCATGCGCGAGGATTTCGAAATCGAGAT
CGGCACCGGCTTGGCCCGCTGCAAGGCAGGATCTGGCGCATCGGTGCGATGGGTACAACGCGATGAAGCACAAAG
GTGCTGCTACCTTGGCCGCACTGGAAGCGGTGCTGCGCGCCGAGGGCTACGCGTGCACCAAGGCCGTCGCGGTCG
AAGCCGACGCGCCGCTGGCATGCGGAGCCGGCTGCATGA

> **pQR1751**

MADPLRHLHTFGELDPQRLMLMGPVNAHPRVLRAMAADLLGQFDPEMTTYMNEVMALYRPLFGTQNRWTFVLVDG
TARAGIEAALVSLVQPGDRVLVINFRGRFLLLEITELGRLGADVHTVDAPWGEVVPPLAAIAEAIASVAPKLVATVHG
DTSTTMAQPLDGLGALCRAAGALSVDATATIGGMDIASDRWEVDVVTAGLQKCLGGPSGSAPITVSAAAAEAI FA

RRHVERGIVREDIANGSGPRIASNYFDLAMIMDYWSDKRLNHHTEATTMLYGARECARVALQEGLEARYARHAAAG
RAVSAGVRALGLEVFGDDAHRSNVTGVVI PHGVDS EAVRRRMREDFEI EIGTAFGLQGR IWRIGAMGYNAMKHK
VLLTLAALAEAVLRAEGYACTQGLAVEAARAAWHAEPAA

> **pQR1752**

ATGGGAAAGTTTCTTAAGAAACACTACATAATGGCACCTGGACCAACACCAGTCCCAAACGATATTTTAAACAGAAG
GAGCGAAGGAAACAATACACCACAGAACACCTCAGTTTGTTCATAATGGAAGAGACCCCTCGAAAGTGCAAAGTA
CATCTTTCAGACAAAACACAACGTGTACGCCTTTGCTTCCACAGGAACCTGGCGCTATGGAAGCGGCGGTGGCGAAT
CTTGTGAGCCCTGGAGACAAAGTGTATCGTGGTTGTGGCTGGAAAGTTCGGTGAAAGATGGAGAGAGCTCTGTCAAG
CTTACGGTGCTGATATCGTAGAAATCGCCCTCGAATGGGAGACGCGGTACACCTGAACAGATCGAAGAGGCTCT
CAACAAAAACCCCGATGCGAAGGTCGTCTTACCACCTACAGTGAAACATCGACGGGTACAGTCATAGACCTCGAA
GGAATTGCCAGAGTCACGAAGGAAAAAGACGTTGTCTTGTGACAGACGCTGTCAGCGCTCTTGGAGCAGAACCAC
TGAAGATGGATGAATGGGGTGTGGATCTCGTTGTGACAGGTTACAGAAAGGTTTGTATGTTACCTCCAGGACTGGC
GCTCATCTCTCAACGACAAAGCGTGGGGGCTCGTGAAAAATCCAGATCTCCAAGGTACTACTTTCGATCTGAGG
GCCTACAGGAAATCTTACCCCGACAATCCTTACACCCCGCAGTAAACATGATATACATGTTGAGAAAAGGCTCTTC
AGATGATAAAAAGGAGGACATAGAAAACGTATGGGAAAGGCACAGAATACTGGGAGACGCAACAAGAGCAGCGGT
GAAAGCACTTGGACTGGAACCTCTCGAAAAGACCGGAAACGTTGTAACAGCGTGAAAAGTGCCTGAGGCGATC
GATGGAAAACAGATTCCCAAGATCATGAGAGACAAGTACGGTGTGACCATCGCCCGTGACAGGCTAAACTCAAGG
GAAAAATATTCAGGATAGCACACCTCGGATACATGTCACCTTTCGACACCATAACTGCCATTTCCGCTCTTGAATT
AACCTTGAAGGAACTCGGTTATGAGTTCGAACTCGGAGTCCGGTGTAAAGGCAGCCGAAGCTGTCTTCGCTAAAGAA
TTCATTGGGGAGTGA

> **pQR1752**

MGKFLKKHYIMAPGPTPVPNDILTEGAKETIHHRTQPQFVS IMEETLES AKYIFQTKHN VYAFAS TGTGAMEAAVAN
LVSPGDKVIVVVVAGKFGERWRELQAYGADIVEIALEWGDVAVTPEQIEEALNKNPDAKVVFPTYSETSTGTVIDLE
GIARVTK EKDVVLVTDVAVSALGAELPKMDEWGVLDVVTGSQKGLMLPPLGLALI SLNDKAWGLVEKRSRSPRYYFDLR
AYRKSYPDNPYTPAVNMI YMLRKALQMIKEEG IENVWERHRI LGDATRAAVKALGLELLSKRPGNVVTVAVKVEPI
DGKQIPKIMRDKYGVTTIAGGQAKLKGKIFRI AHLGYMSPFDTTITAI SALELTLKELGYEFELGVGVKAAEAVFAKE
FIGE

> **pQR1755**

ATGACTCAGATTTTAAATTTTAGCGCCGGTCCAGCAATGCTGCCGGTTGAAGTACTGCGTCGTGCTGAACAGGAAT
TGTGTAATGGAAATGGCCGGGCACATCGGTTATGGAAATCAGCCACCGCAGTAAAGAGTTTATGCAGGTTGCCGC
TGAATCCGAACAGGATCTGCGTGATTTGCTGAAAATCCCTCCAACCTACAAAGTGCTCTTTTCCACGGCGGTGCT
CGTGCGCAATTCGCCGAGTGCCTTAAATCTTCTGGCGAACGCTCAACGGCCGACTACATCGACGGCGGGTATT
GGGCGCACAGCGCAATCAATGAAGCAGAAAAATACTGCACGCCTAACGTGATTGACGTGAAAATGCCCGTGGCGCA
ACTGCGTGGCATTAAAGCCGATGCGTGAATGAAAATTTGTCTGATGACGCGCGCTTTGTGCATTACTGCCGAATGAA
ACCATCGACGGTATTGCGATCGAAGAAGAGCCGGACTTTGGCGATAAAAATTTGTGGTTCGCCGACTATCTTCCAGCA
TCTGTCTCGTCGTATTGATGTCAGCCGTTACGGCGTGATCTATGCCGGTGCAGCAAAAATATCGGCCCTGCCGG
CTTGACCGCTGGTTATCGTACGTGAAGATTTGCTGGGCAAGGCCGCCGCTGAGCTGCCATCGATTCTCGATTACAG
GTTCTGGCGGACAATGACTCCATGTTTAAACACGCCACCGACCTTTGCCCTGGTACCTGTCCGGTATGGTCTTCAAAT
GGCTGAAAGAGTACGGCGTCTGGCTGAAATGAAAAACGTAACCAGGAGAAGGCTGACCTGCTGTATAGCGCGAT
TGACGGTAACGATTTCTATCGTAATGACGTTGCGGTAGCGAACCGTTCTCGCATGAATGTGCCATTCCTGTTGGCG
GATTCCTGCGCTGGATAAAGTCTTCTCGAAGAATCAGTCGCTGCAGGCTGACACGGCTGAAAGGCCATCGCGTAG
TAGCGGCATGCGTGCCCTATCTACAATGCGATGCCGTTGGAAGGCGTGAAAGTGCTGACGGAAATTTATGGCTGA
CTTCGCTCGTCGCCACGGTTGA

> **pQR1755**

MTQIFNF SAGPAML PVEVLRRAEQELCNWNLGTSVMEI SHRSKEFMQVAAESEQDLRDLK I PSNYKVL FCHGGA
RAQFAAVPLNLLGERSTADYIDGGYWAHSA INEAEKYCTPNVIDVKMRV GELRGI KPMREWKLSDDAAFVHYCPNE

TIDGIAIEEEPFDGDKIVVADYSSSILSRRIDVSRVGYIYAGAQNIGPAGLTLVIVREDLLGKARRELPISILDYQ
VLADNDSMFNTPTFFAWYLSGMVFKWLKEYGGLAEMEKRNRQEKADLLYSAIDGNDFYRNDVAVANRSRMNVPFLLA
DSALDKVFLEESVAAGLHALKGHRVVGMRASIYNAMPLEGVKVLTEFMADFARRHG

Acknowledgements

The authors are grateful to the UK Engineering and Physical Sciences Research Council (EPSRC) for financial support of this work (EP/K014897/1). Furthermore, we acknowledge the advice and support from the project Industrial Advisory Board.

Transparency document. Supporting information

Transparency document associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.05.140>.

References

- [1] M. Bawn, F. Subrizi, G.J. Lye, T.D. Sheppard, H.C. Hailes, J.M. Ward, One-pot, Two-step Transaminase and Transketolase Synthesis of L-gluco-heptulose from L-arabinose, *Enzyme Microb. Technol.* 116 (2018) 16–22.
- [2] M. Cárdenas-Fernández, M. Bawn, C. Hamley-Bennett, P.K.V. Bharat, F. Subrizi, N. Suhaili, D.P. Ward, S. Bourdin, P.A. Dalby, H.C. Hailes, P. Hewitson, S. Ignatova, C. Kontoravdi, D.J. Leak, N. Shah, T.D. Sheppard, J.M. Ward, G.J. Lye, An integrated biorefinery concept for conversion of sugar beet pulp into value-added chemicals and pharmaceutical intermediates, *Faraday Discuss.* 346 (2017) 183–190. <http://dx.doi.org/10.1039/C7FD00094D>.
- [3] F. Subrizi, M. Cárdenas-Fernández, G.J. Lye, J.M. Ward, P.A. Dalby, T.D. Sheppard, H.C. Hailes, Transketolase catalysed upgrading of L-arabinose: the one-step stereoselective synthesis of L-gluco-heptulose, *Green Chem.* 18 (2016) 3158–3165. <http://dx.doi.org/10.1039/C5GC02660A>.
- [4] T. Seliwanoff, Notiz iiber eke Fruchtzucker- reaction, *Chem. Ber.* 20 (1887) 181–182.
- [5] N.C. for B.I. National Library of Medicine (US), National Center for Biotechnology Information (NCBI), 1988. (<https://www.ncbi.nlm.nih.gov/>).
- [6] A. Bateman, M.J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-Ajee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L.G. Castro, L. Figueira, P. Garmiri, G. Georghiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M.C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuhe, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemerrier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Noupikpel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A.L. Veuthey, C.H. Wu, C.N. Arighi, L. Arminski, C. Chen, Y. Chen, J.S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D.A. Natale, K. Ross, C.R. Vinayaka, Q. Wang, Y. Wang, L.S. Yeh, J. Zhang, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.* 45 (2017) D158–D169. <http://dx.doi.org/10.1093/nar/gkw1099>.
- [7] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. Mcgettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, *Bioinformatics* 23 (2007) 2947–2948. <http://dx.doi.org/10.1093/bioinformatics/btm404>.