Electronic Supplementary Material (ESI) for Green Chemistry.
This journal is © The Royal Society of Chemistry 2021

Biosynthesis of non-animal chondroitin sulfate from methanol using genetically engineered *Pichia pastoris*

Xuerong Jin,[a,b#] Weijiao Zhang,[a,b#] Yang Wang,[a,b] Jingyu Sheng,[a,b] Ruirui Xu,[a,b] Jianghua Li,[b*] Guocheng Du,[a,b] Zhen Kang[a,b,c*]

Author information

[a]The Key Laboratory of Carbohydrate Chemistry and Biotechnology, Ministry of Education, Jiangnan University, Wuxi, China

[b]The key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi, China

[c]The Science Center for Future Foods, Jiangnan University, Wuxi, 214122, China

[#]Xuerong Jin and Weijiao Zhang contributed equally to this work

***Corresponding Authors**

Zhen Kang or Jianghua Li

zkang@jiangnan.edu.cn or lijianghua@jiangnan.edu.cn

T: (86)-510-85918307, F: (86)-510-85918309

Supplementary Information

Table S1. List of plasmids and strains used in this study

Names	Descriptions	Resources
Plasmids		
pAO815	*P. pastoris* expression vector, *AOX* promoter	Lab stock
pGAPZB	*P. pastoris* expression vector, *GAP* promoter, Zeo^r	Lab stock
pAO815-*kfoC*	pAO815 plasmid containing *kfoC*	This study
pAO815-*kfoA*	pAO815 plasmid containing *kfoA*	This study
pAO815-*kfoC*-*kfoA*	pAO815 plasmid containing *kfoC* expression cassette and *kfoA* cassette	This study
pGAPZB-*tuaD*	pGAPZB plasmid containing *tuaD*	This study
pGAPZB-*tuaD'*	pGAPZB plasmid containing *tuaD'*	This study
pAO815-*kfoC'*-*T2A*-*kfoA'*-*T2A2*-*tuaD'*	pAO815 plasmid containing *kfoC'*, *kfoA'* and *tuaD'* expression cassette which connected by T2A and T2A2 peptide	This study
pGAPZB-*kfoC'*-*T2A*-*kfoA'*-*T2A2*-*tuaD'*	pGAPZB plasmid containing *kfoC'*, *kfoA'* and *tuaD'* expression cassette which connected by T2A and T2A2 peptide	This study
pAO815-K1*egfp*	pAO815-*egfp* plasmid with Kozak sequence K1	This study
pAO815-K2*egfp*	pAO815-*egfp* plasmid with Kozak sequence K2	This study
pAO815-K3*egfp*	pAO815-*egfp* plasmid with Kozak sequence K3	This study
pAO815-K4*egfp*	pAO815-*egfp* plasmid with Kozak sequence K4	This study
pAO815-K5*egfp*	pAO815-*egfp* plasmid with Kozak sequence K5	This study
pAO815-K6*egfp*	pAO815-*egfp* plasmid with Kozak sequence K6	This study
P_x-K4*C4ST*-*egfp*	The plasmid to express C4ST fused with eGFP contained different promoters	This study
pPIC3.5K	*P. pastoris* expression vector, *AOX* promoter	Lab stock
pPIC3.5KHy	*P. pastoris* expression vector, *AOX* promoter, inserted Hyg^r cassette	This study
pPIC3.5KHy-*C4ST*	pPIC3.5KHy plasmid containing *C4ST*	This study
pPIC3.5KHy-*C4ST*-T2A3-*MET3*-T2A4-*MET14*	pPIC3.5KHy plasmid containing *C4ST*, *MET3* and *MET14*	This study
pPIC3.5KHy-*C4ST*-T2A3-*0253*-T2A4-*0667*	pPIC3.5KHy plasmid containing *C4ST*, *PAS_chr1-4_0253* and *PAS_chr3_0667*	This study
Strains		

Pp001	*P. pastoris* harboring pAO815-*kfoC-kfoA* and pGAPZB-*tuaD*	This study
Pp002	*P. pastoris* harboring pAO815-*kfoC-kfoA* and pGAPZB-*tuaD'*	This study
Pp003	*P. pastoris* harboring pAO815-*kfoC'-kfoA'* and pGAPZB-*tuaD'*	This study
Pp004	*P. pastoris* harboring pGAPZB-*kfoC'-T2A-kfoA'-T2A2-tuaD'*	This study
Pp005	*P. pastoris* harboring pAO815-*kfoC'-T2A-kfoA'-T2A2-tuaD'*	This study
Pp006	*P. pastoris* harboring pAO815-*kfoC'-T2A-kfoA'-T2A2-tuaD'* and pPIC3.5KHy-*C4ST*	This study
Pp007	*P. pastoris* harboring pAO815-*kfoC'-T2A-kfoA'-T2A2-tuaD'* and pPIC3.5KHy-*C4ST*-T2A3-*MET3*-T2A4-*MET14*	This study
Pp008	*P. pastoris* harboring pAO815-*kfoC'-T2A-kfoA'-T2A2-tuaD'* and pPIC3.5KHy-*C4ST*-T2A3-*0253*-T2A4-*0667*	This study

Table S2 DNA primers used in this study

Primers	Primer sequences (5′-3′)
kfoC-F	CTAATTATTCGAAACGAGGAATTCGCCACCATGTCTATCTTAAATCAAGCAATTAACCTG
kfoC-R	CAGTCATGTCTAAGGCGAATTCTTACAAATCGTTTTCGATTTTCTCCC
kfoA-F	CTAATTATTCGAAACGAGGAATTCGCCACCATGAACATCCTGGTCACGGGAGGCGC
kfoA-R	CAGTCATGTCTAAGGCGAATTCTTAAATGTACCCGTTAGGGTTTTTCATCTGCCAGC
tuaD-F	GAACAACTATTTCGAAACGAGGAATTCGCCACCATGAAAAAAATAGCTGTCATTGGAACAGG
tuaD-R	GCGGCCGCCGCGGCTCGAGGTACCTTATAAATTGACGCTTCCCAAGTCTTTAGCC
(815)kfoC'-F	CAACTAATTATTCGAAACGAGGAATTCGCCACCATGTCAATTTTGAATCAAGCT
kfoC'-R	TCTTCAACATCACCACAAGTCAGAAGAGAACCACGACCTTCTCCGGATCCCAAATCGTTTTCAATCTTTTCCCATGTGTACTTTC
kfoA'-F	GTTCTCTTCTGACTTGTGGTGATGTTGAAGAAAACCCAGGTCCAATGAATATTTTGGTTACTGGTGGTGCTGGTTATATTG
kfoA'-R	CTCTACGTCACCGCAAGTAAGTAGGGATCCACGACCTTCTCCGGATCCAATGTAACCATTTGGATTTTTCATTTGCCATCTC
tuaD'-F	GGATCCCTACTTACTTGCGGTGACGTAGAGGAAAACCCTGGTCCGATGAAGAAAATTGCTGTTATTGGTACTGGTTATG
(815)tuaD'-R	GGAACAGTCATGTCTAAGGCGAATTCTTACAAATTAACAGAACCCAAATCTTTTGCCAAT
(GAPZB)kfoC'-F	GAACAACTATTTCGAAACGAGGAATTCGCCACCATGTCAATTTTGAATCAAGCTATTAATTTGTACAAGAATAAGAAC
(GAPZB)tuaD'-R	CTGGCGGCCGCCGCGGCTCGAGGTACCTTACAAATTAACAGAACCCAAATCTTTTGCCAATTC
egfp-F	CTAATTATTCGAAACGAGGAATTCATGGGTAAGGGAGAAGAACTTTTCACTG
egfp-R	CAGTCATGTCTAAGGCGAATTCTTATTTGTATAGTTCATCCATGCCATGTGTAATCCC
Kozak-R	GAATTCCTCTCGAGATAGTTGTTCAATTGATTGAAATAGGG
K2-F	GCCACCATGGGTAAGGGAGAAGAACTTTTCAC
K3-F	GCCTCCATGGGTAAGGGAGAAGAACTTTTCAC
K4-F	CAAACGATGGGTAAGGGAGAAGAACTTTTCAC
K5-F	GGGACCATGGGTAAGGGAGAAGAACTTTTCAC
K6-F	GCCACCATGGGTAAGGGAGAAGAACTTTTCAC
K6-R	GCGGCCGCCTCTCGAGATAGTTGTTCAATTGATTGAAATAGGG
hpC4egfp-F	ATGGGTAAGGGAGAAGAACTTTTCACTG
hpC4egfp-R	CGTTTGGAATTCCTCGTTTCGAATAATTAGTTGTTTTTTGATC
C4egfp-F	CTAATTATTCGAAACGAGGAATTCGCCACCATGTCGGACTCAGAA

	GTCAATCAAGAAG
C4egfp-R	GAAAAGTTCTTCTCCCTTACCCATTCCGGATCCATCCAATTTCAAA TAATTAGGAACAGAATAATTGAAC
3.5KHy-F	CAACTCAGCAAAAGTTCGATGACATGGAGGCCCAGAATACCC
3.5KHy-R	CGTGACTGGGTCATGGCTGCGCAGTATAGCGACCAGCATTCACAT ACG
hp3.5K-F	CGCAGCCATGACCCAGTCAC
hp3.5K-R	ATCGAACTTTTGCTGAGTTGAAGGATCAG
C4ST-F	GATCAAAAAACAACTAATTATTCGAAGGATCCATGTCGGACTCAGA AGTCAATCAAGAAG
C4ST-R	GTCATGTCTAAGGCGAATTAATTCGCGGCCGTTAATCCAATTTCAA ATAATTAGGAACAGAATAATTGAACATCAAG
C4(T2A3)-F	GATCAAAAAACAACTAATTATTCGAAGGATCCATGTCGGACTCAGA AGTCAATCAAGAAG
C4(T2A3MET3)-R	CGTCTCCTGCTTGCTTTAACAGAGAGAAGTTCGTGGCTCCGGATC CATCCAATTTCAAATAATTAGGAACAGAATAATTG
MET3-F	CTTCTCTCTGTTAAAGCAAGCAGGAGACGTGGAAGAAAACCCCGG TCCTATGCCTGCTCCTCACGGTGGTATTCTA
MET3-R	ACGTCACCGCAAGTAAGCAAAGAACCTCTACCCTCTCCGGATCCT TAAAATACAAAAAAGCCATTGTCTTCCAAGAATAGGACAACTT
MET14-F	AGAGGTTCTTTGCTTACTTGCGGTGACGTTGAGGAAAACCCAGGT CCAATGGCTACTAATATTACTTGGCATCCAAATC
MET14-R	GGCGAATTAATTCGCGGCCGCTTACAAATGCTTACGGATGATTTTT TCACTGATTAAG
C4(T2A30253)-R	CGTCTCCTGCTTGCTTTAACAGAGAGAAGTTCGTGGCTCCGGATC CATCCAATTTCAAATAATTAGGAACAGAATAATTG
0253-F	CTTCTCTCTGTTAAAGCAAGCAGGAGACGTGGAAGAAAACCCCGG TCCTTATGCCTTCTCCTCACGGTGGTGTGCTAC
0253-R	ACGTCACCGCAAGTAAGCAAAGAACCTCTACCCTCTCCGGATGAA CTGGAAGAACCCCTGTTCTTCC
0667-F	AGAGGTTCTTTGCTTACTTGCGGTGACGTTGAGGAAAACCCAGGT CCAATGGCTACTAATATCACATGGCATG
0667-R	GGCGAATTAATTCGCGGCCGCTTCTATTTGATTAGTTTCTTCTCCA A

Note: Homologous arm sequences are shown in underlined.

Table S3 The gene sequences used in the article

Gene	sequences
kfoC'	ATGTCAATTTTGAATCAAGCTATTAATTTGTACAAGAATAAGAACTACA GACAAGCATTGTCATTGTTCGAAAAGGTTGCTGAAATCTATGATGTTTC TTGGGTTGAAGCAAACATCAAGTTGTGTCAAACTGCTTTGAATTTGTCT GAAGAAGTTGATAAGTTGAACAGAAAGGCTGTTATTGATATCGATGCT GCAACAAAGATCATGTGTTCTAACGCTAAGGCAATCTCATTGAACGAA GTTGAAAAGAATGAAATCATCTCAAAGTACAGAGAAATTACTGCTAAG AAATCTGAAAGAGCTGAATTGAAAGAAGTTGAACCAATTCCATTAGATT GGCCATCAGATTTGACTTTACCACCATTGCCAGAATCTACAAATGATTA TGTTTGGGCTGGTAAAAGAAAGGAATTGGATGATTACCCAAGAAAGCA ATTGATCATCGATGGTTTGTCTATTGTTATTCCAACATACAACAGAGCT AAGATCTTGGCAATCACTTTGGCTTGTTTGTGTAACCAAAAGACTATCT ATGATTACGAAGTTATTGTTGCTGATGATGGTTCAAAAGAAAATATTGA AGAAATTGTTAGAGAATTCGAATCTTTGTTGAACATCAAGTATGTTAGA CAAAAAGATTATGGTTACCAATTGTGTGCAGTTAGAAATTTGGGTTTGA GAGCTGCAAAGTACAACTACGTTGCAATCTTGGATTGTGATATGGCTC CAAATCCATTATGGGTTCAATCATACATGGAATTGTTAGCAGTTGATGA TAATGTTGCTTTGATCGGTCCAAGAAAGTACATCGATACTTCTAAGCAT ACATACTTGGATTTCTTGTCTCAAAAATCATTGATTAATGAAATCCCAG AAATCATCACTAACAATCAAGTTGCTGGTAAAGTTGAACAAAATAAGTC AGTTGATTGGAGAATCGAACATTTCAAGAACACTGATAATTTGAGATTA TGTAATACACCTTTTAGATTTTTCTCTGGTGGTAATGTTGCTTTTGCTA AGAAATGGTTGTTTAGAGCAGGTTGGTTTGATGAAGAATTCACTCATT GGGGTGGTGAAGATAACGAATTCGGTTACAGATTGTACAGAGAAGGT TGTTACTTTAGATCAGTTGAAGGTGCTATGGCATACCATCAAGAACCA CCAGGTAAAGAAAATGAAACTGATAGAGCTGCAGGTAAAAATATCACA GTTCAATTGTTGCAACAAAAGGTTCCATACTTCTACAGAAAGAAAGAAA AGATTGAATCTGCAACTTTGAAGAGAGTTCCATTAGTTTCTATCTATAT CCCAGCTTACAACTGTTCAAAGTACATCGTTAGATGTGTTGAATCTGC ATTGAACCAAACTATCACAGATTTGGAAGTTTGTATCTGTGATGATGGT TCAACTGATGATACATTGAGAATTTTACAAGAACATTATGCTAATCATC CAAGAGTTAGATTCATTTCTCAAAAGAATAAGGGTATTGGTTCTGCATC AAATACAGCTGTTAGATTGTGTAGAGGTTTTTACATTGGTCAATTGGAT TCTGATGATTTCTTGGAACCAGATGCTGTTGAATTGTGTTTGGATGAAT TCAGAAAGGATTTGTCATTGGCATGTGTTTACACTACAAACAGAAACAT CGATAGAGAGGGTAATTTGATCTCTAACGGTTACAACTGGCCAATATA TTCTAGAGAAAAGTTGACTTCAGCTATGATCTGTCATCATTTCAGAATG TTCACTGCTAGAGCATGGAATTTGACAGAAGGTTTTAATGAATCTATCT CAAATGCAGTTGATTACGATATGTACTTGAAGTTGTCAGAAGTTGGTC CTTTTAAACATATCAATAAGATCTGTTACAACAGAGTTTTGCATGGTGA AAACACATCTATTAAGAAATTGGATATCCAAAAGGAAAATCATTTTAAA GTTGTTAATGAATCTTTGTCAAGATTGGGTATTAAGAAATACAAATACT CTCCATTGACTAATTTGAACGAATGTAGAAAGTACACATGGGAAAAGA TTGAAAACGATTTGTAG
kfoA'	ATGAATATTTTGGTTACTGGTGGTGCTGGTTATATTGGTTCACATACAT CTTTATGTTTGTTAAATAAGGGTTACAATGTTGTTATTATTGATAATTTG

	ATTAATTCTTCATGTGAATCTATTAGAAGAATTGAATTGATTGCTAAAAA
	GAAAGTTACTTTCTACGAATTGAACATCAACAACGAAAAAGAAGTTAAT
	CAAATTTTGAAGAAACATAAATTTGATTGTATTATGCATTTTGCTGGTG
	CAAAATCAGTTGCTGAATCTTTGATTAAACCAATTTTCTATTACGATAA
	CAACGTTTCAGGTACATTGCAATTGATTAATTGTGCTATTAAAAATGAT
	GTTGCAAACTTCATTTTCTCTTCTTCTGCAACTGTTTACGGTGAATCTA
	AGATCATGCCAGTTACAGAAGATTGTCATATTGGTGGTACTTTGAACC
	CATACGGTACATCAAAGTACATCTCTGAATTGATGATTAGAGATATTGC
	TAAGAAATATTCAGATACTAATTTCTTGTGTTTAAGATACTTTAATCCAA
	CAGGTGCTCATGAATCTGGTATGATTGGTGAATCTCCAGCAGATATCC
	CATCAAATTTGGTTCCATACATCTTGCAAGTTGCTATGGGTAAATTGGA
	AAAGTTGATGGTTTTCGGTGGTGACTATCCAACTAAAGATGGTACAGG
	TGTTAGAGATTACATCCATGTTATGGATTTGGCAGAAGGTCATGTTGC
	TGCATTGTCTTATTTGTTTAGAGATAACAACACTAACTACCATGTTTTTA
	ATTTGGGTACTGGTAAAGGTTACTCTGTTTTGGAATTAGTTTCAACTTT
	CGAAAAGATTTCTGGTGTTAGAATTCCATACGAAATCGTTTCAAGAAG
	AGATGGTGACATTGCTGAATCTTGGTCATCTCCAGAAAAGGCTAATAA
	GTATTTGAACTGGAAGGCAAAGAGAGAATTGGAAACAATGTTAGAAGA
	TGCATGGAGATGGCAAATGAAAAATCCAAATGGTTACATTTAG
tuaD'	ATGAAGAAAATTGCTGTTATTGGTACTGGTTATGTTGGTTTGGTTTCTG
	GTACATGTTTTGCAGAAATCGGTAATAAGGTTGTTTGTTGTGATATCGA
	TGAATCTAAGATCAGATCATTGAAAAATGGTGTTATTCCAATCTATGAA
	CCAGGTTTGGCTGATTTGGTTGAAAAGAATGTTTTGGATCAAAGATTG
	ACTTTTACAAACGATATCCCATCTGCTATCAGAGCATCAGATATCATCT
	ATATTGCTGTTGGTACTCCAATGTCAAAAACAGGTGAAGCAGATTTGA
	CTTATGTTAAAGCTGCAGCTAAGACAATCGGTGAACATTTGAACGGTT
	ACAAAGTTATTGTTAATAAGTCAACTGTTCCAGTTGGTACTGGTAAATT
	GGTTCAATCTATCGTTCAAAAAGCATCTAAGGGTAGATACTCATTCGAT
	GTTGTTTCTAACCCAGAATTTTTGAGAGAAGGTTCTGCTATTCATGATA
	CTATGAATATGGAAAGAGCAGTTATTGGTTCTACATCACATAAAGCAG
	CTGCAATCATCGAAGAATTGCATCAACCATTCCATGCTCCAGTTATTAA
	AACTAATTTGGAATCAGCAGAAATGATCAAGTACGCTGCAAACGCATT
	TTTGGCAACAAAAATTTCTTTTATTAACGATATCGCTAATATTTGTGAAA
	GAGTTGGTGCTGATGTTTCAAAAGTTGCAGATGGTGTTGGTTTGGATT
	CTAGAATCGGTAGAAAGTTCTTGAAGGCTGGTATCGGTTTTGGTGGTT
	CATGTTTTCCAAAAGATACTACAGCATTGTTGCAAATCGCTAAGTCTGC
	AGGTTACCCTTTTAAATTGATTGAAGCTGTTATTGAAACAAATGAAAAG
	CAAAGAGTTCATATCGTTGATAAATTGTTGACTGTTATGGGTTCAGTTA
	AGGGTAGAACAATCTCTGTTTTGGGTTTGGCTTTTAAACCAAACACAA
	ACGATGTTAGATCAGCTCCAGCATTGGATATCATCCCAATGTTGCAAC
	AATTAGGTGCTCATGTTAAAGCATATGATCCAATTGCTATTCCAGAAGC
	ATCTGCAATCTTGGGTGAACAAGTTGAATACTACACTGATGTTTACGC
	TGCAATGGAAGATACTGATGCATGTTTGATCTTGACAGATTGGCCAGA
	AGTTAAGGAAATGGAATTGGTTAAGGTTAAGACATTGTTGAAGCAACC
	AGTTATTATCGATGGTAGAAATTTGTTTTCATTAGAAGAAATGCAAGCT
	GCAGGTTATATCTATCATTCAATTGGTAGACCAGCTGTTAGAGGTACT

	GAACCATCTGATAAATACTTTCCAGGTTTGCCATTAGAAGAATTGGCA AAAGATTTGGGTTCTGTTAATTTGTAG

Table S4 The sequences of 2A peptide

Names	Sequences

T2A	GAAGGTCGTGGTTCTCTTCTGACTTGTGGTGATGTTGAAGAAAACCCA GGTCCA
T2A2	GAAGGTCGTGGATCCCTACTTACTTGCGGTGACGTAGAGGAAAACCC TGGTCCG
T2A3	GAGGGTAGAGGTTCTTTGCTTACTTGCGGTGACGTTGAGGAAAACCC AGGTCCA
T2A4	GAAGGACGTGGATCCCTTTTGACCTGTGGAGATGTCGAAGAGAATCC AGGTCCT

SUPPLEMENTARY FIGURES

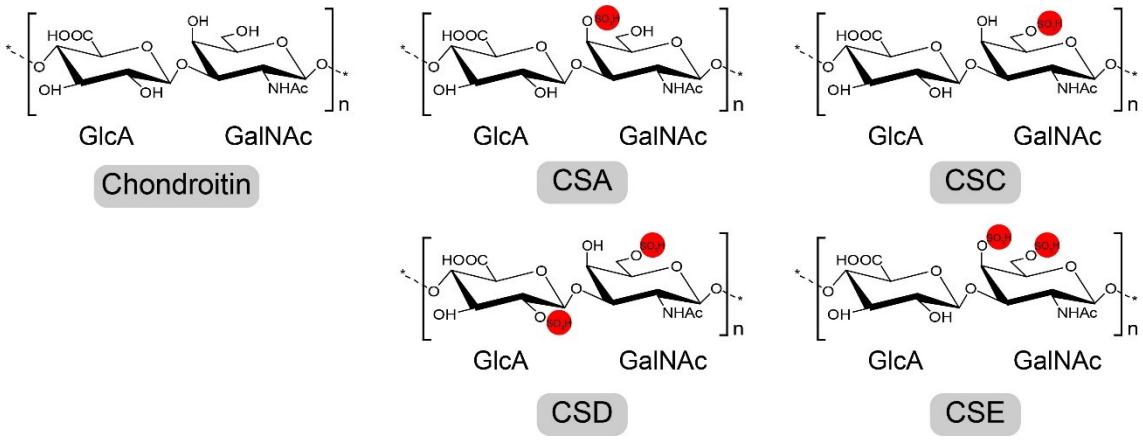

Fig. S1 Chemical structure of chondroitin sulfate with different sulfation patterns.

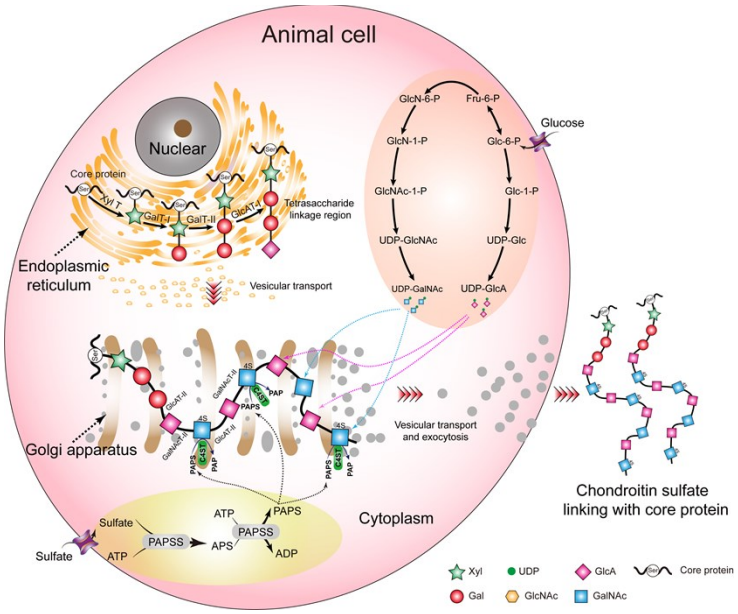

Fig. S2 The schematic diagram of the chondroitin sulfate synthesis and secretion in animal cell. XylT, xylosyltransferase; GalT-I, β1,4-galactosyltransferase-I; GalT-II, β1,3-galactosyltransferase-II; GlcAT-I, β1,3-glucuronyltransferase-I; GalNAcT-I, GalNAc transferase-I; GlcAT-II, GlcA transferase-II; GalNAcT-II, GalNAc transferase-II; Glc-6-P, glucose-6-phosphate; Glc-1-P, glucose-1-phosphate; UDP-Glc, UDP-glucose; UDP-GlcA, UDP-glucuronate; Fru-6-P, fructose-6-phosphate; GlcN-6-P, glucosamine-6-phosphate; GlcN-1-P, glucosamine-1-phosphate; GlcNAc-1-P, N-acetyl-glucosamine-1-phosphate; UDP-GlcNAc, UDP-N-acetylglucosamine; UDP-GalNAc, UDP-N-acetylgalactosamine; ATP, adenosine 5'-triphosphate; APS, adenosine 5'-phosphosulfate; ADP, adenosine 5'-diphosphate; PAPS, 3'-phosphoadenosine-5'-phosphosulfate; PAPSS, 3'-phosphoadenosine-5'-phosphosulfate synthase.

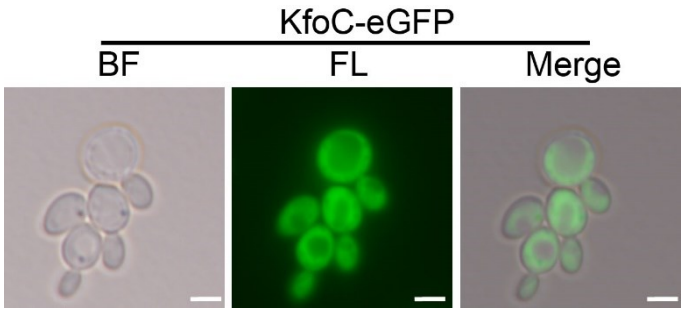

Fig. S3 Localization of chondroitin polymerase, KfoC, to indicate the position where chondroitin synthesis occurs. Scale bar = 2.5 μm. BF, bright-field microscopy; FL, fluorescence microscopy.

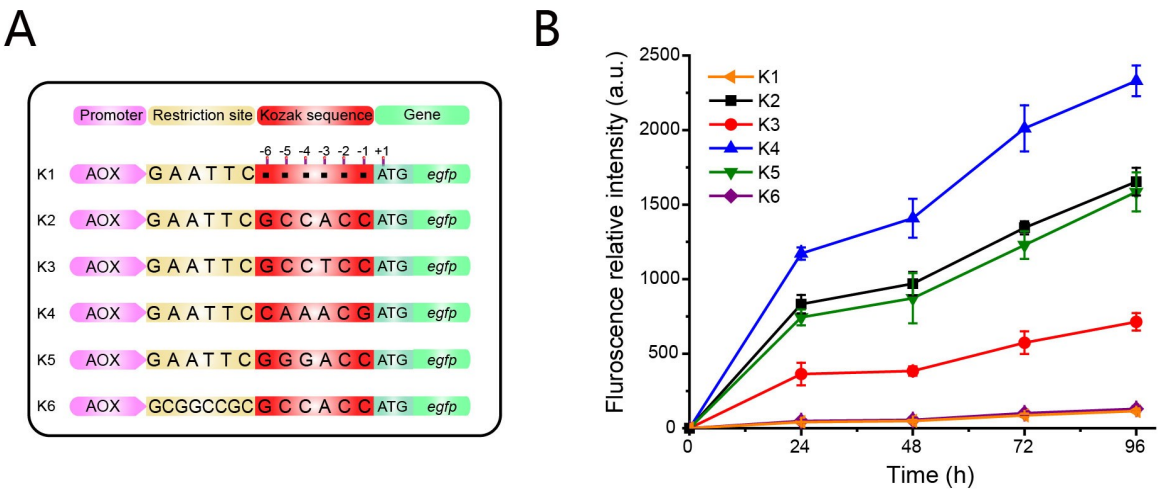

Fig. S4 Kozak sequences optimization. **(A)** Design of different kozak sequences. **(B)** The time course of the relative fluorescence intensity with different Kozak sequences.

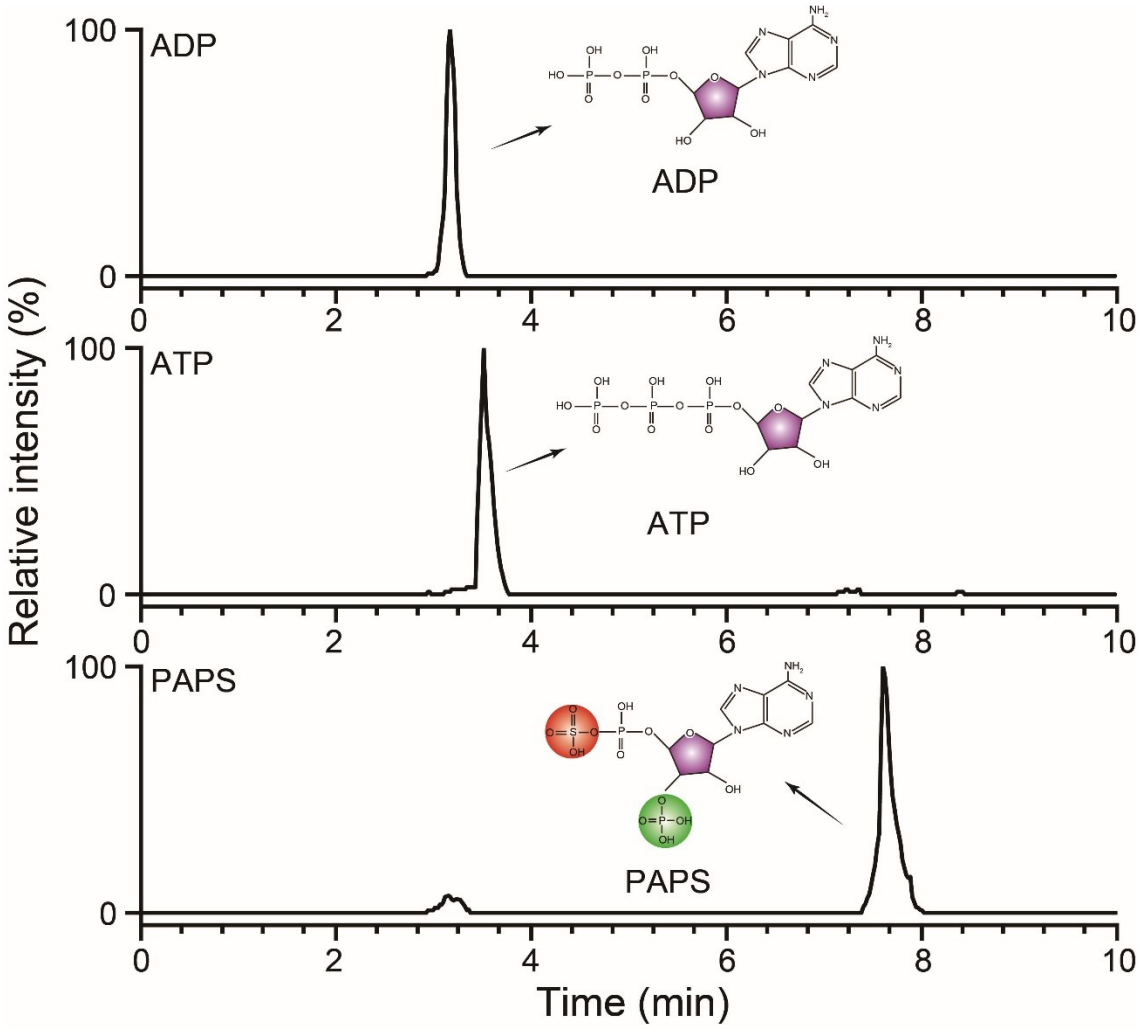

Fig. S5 The ADP, ATP and PAPS standard samples detection with LCMS-IT-TOF.

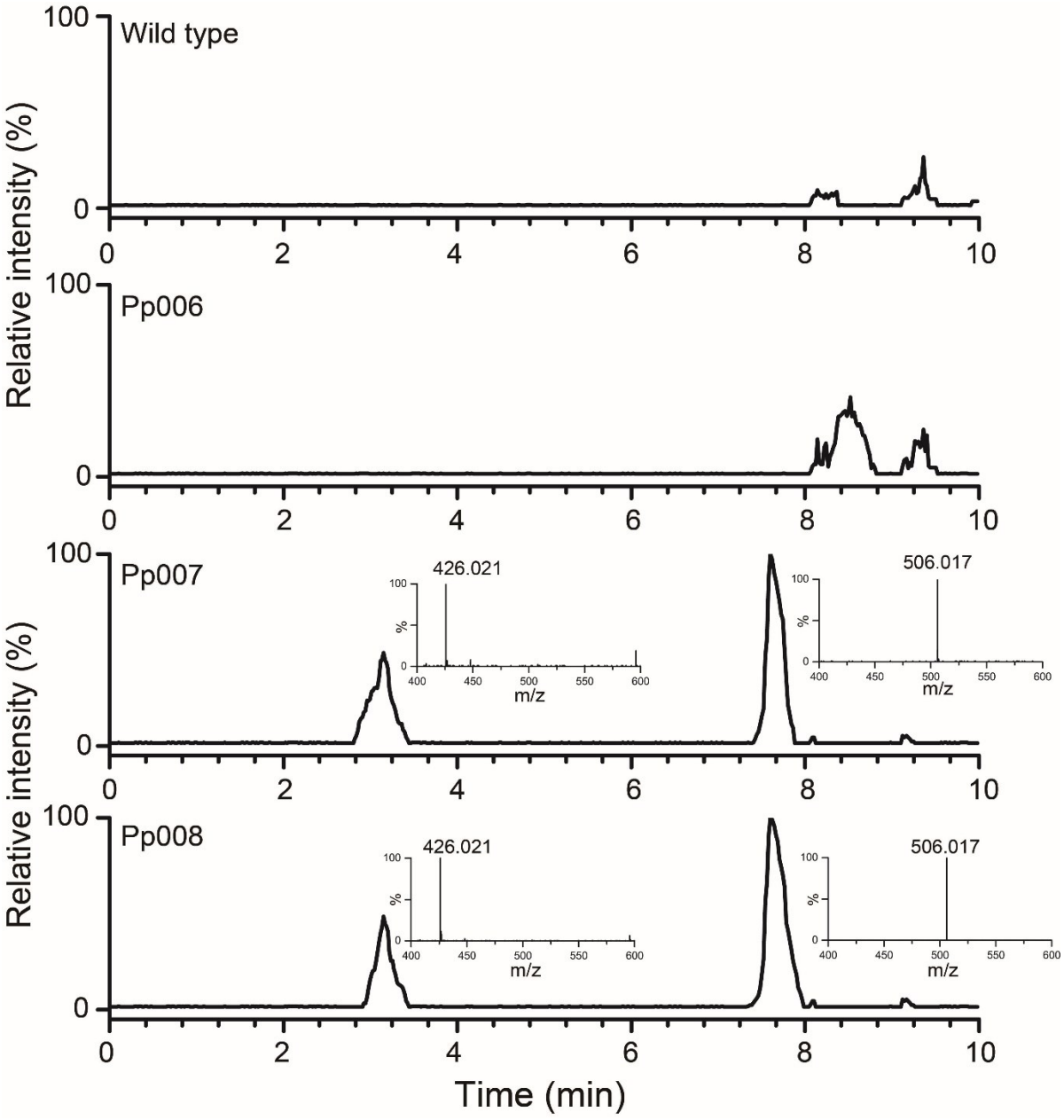

Fig. S6 The intracellular PAPS detection with LCMS-IT-TOF. The peak *m/z* 506.017 indicated the PAPS ($[M-H]^-$) and *m/z* 426.021 was the fragment pattern of PAPS ($[M-SO_3H]^-$).