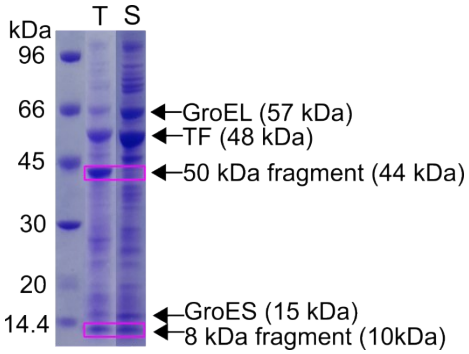
Electronic Supplementary Material (ESI) for RSC Chemical Biology.

Computational design and experimental characterisation of a stable human heparanase variant

Cassidy Whitefield[a†], Nansook Hong [a†], Joshua A. Mitchell[a] and Colin J. Jackson [a*]

[a] *Research School of Chemistry, Australian National University, Canberra, ACT, 2601, Australia.*
† Contributed equally to this work. *To whom correspondence should be addressed: colin.jackson@anu.edu.au

Supplementary Material

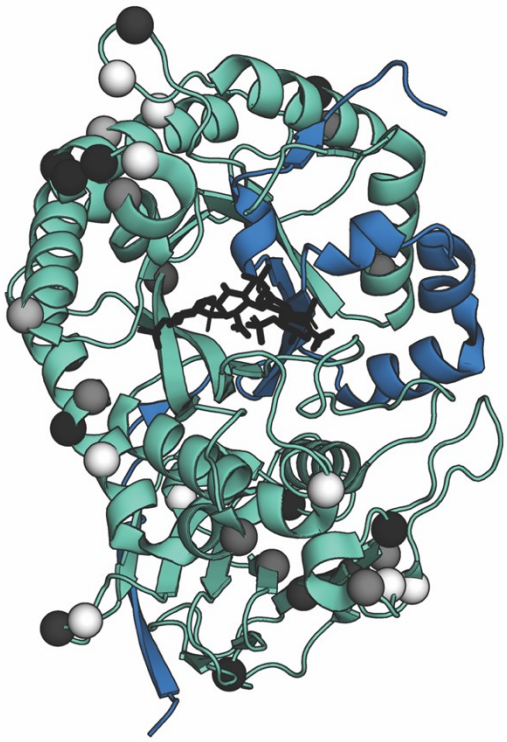


Supplementary Figure 1. Wild type human heparanase expression in *E. coli*. The 50 and 8 kDa subunits and chaperones[28] in total (annotated as T) and in soluble fraction (annotated as S) are shown with LMW protein marker (GE healthcare). 8 kDa subunit with N-terminal 6xHis-tag and 50 kDa subunit were amplified by PCR using gBlocks (IDT, **Supplementary Table 1**) and inserted into multiple cloning sites of pETDuet-1 coexpression vector (Novagen). The plasmid DNA was transformed into Shuffle T7 Express competent cells (NEB) together with chaperones in pACYC vector. Overnight seed culture from single colony was inoculated by 1% into LB media supplemented with ampicillin and chloramphenicol antibiotics. The culture was incubated at 37 °C for about 3-5 hours until the OD was about 0.8. 0.4 mM IPTG was added and further incubated for about 3 hours at 37 °C. The cell pellet was resuspended in 50 mM acetate buffer at pH 5 and lysed using BugBuster protein extraction reagent (Merck) for the solubility measurement.

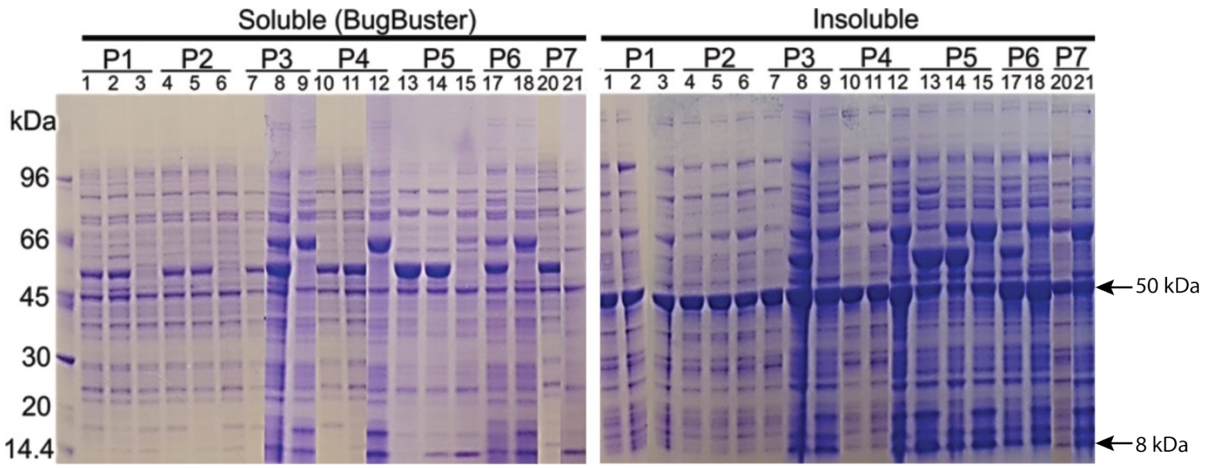
A

Design 1	1	L197G
	2	D234G
	3	Q248H
	4	I318T
	5	L354G
	6	K427D
	7	L483H
	8	L498Q
	9	A540P
Design 2	10	E244K
	11	R273G
Design 3	12	L230R
	13	S322Q
	14	H486D
	15	M512K
Design 4	16	N178K
	17	S212A
	18	R307L
	19	S530A
Design 5	20	S219D
	21	S292A
Design 6	22	A195S
	23	F327L
	24	S426Q
	25	K477Q
	26	E513P
Design 7	27	D209K
	28	F236Y
	29	N238D
	30	S240T
	31	S321Q
	32	K446R
	33	S492H
	34	S521T

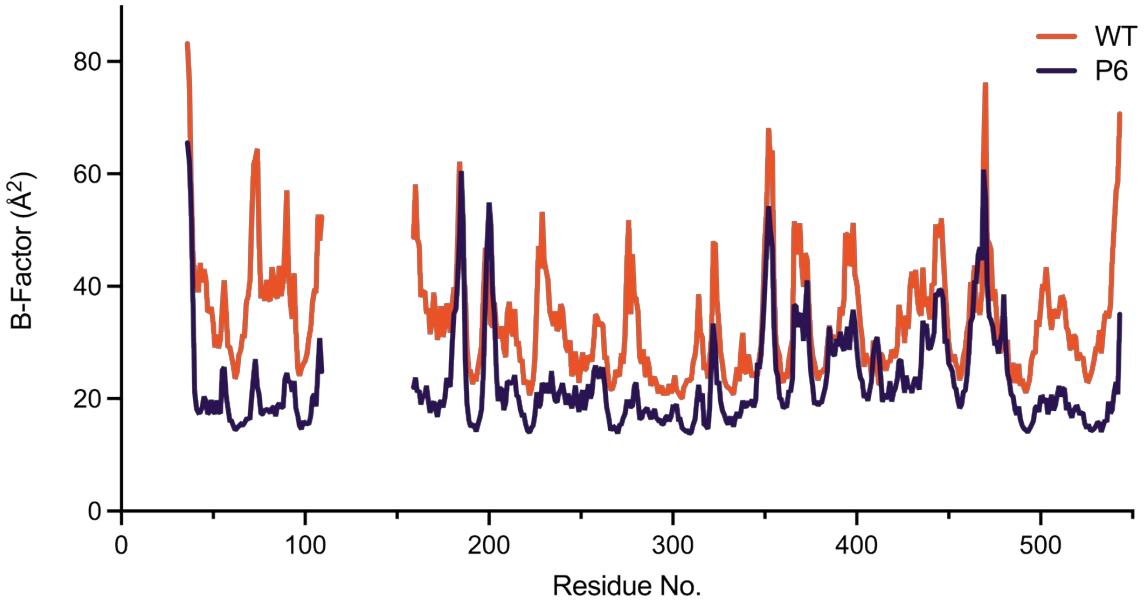
B



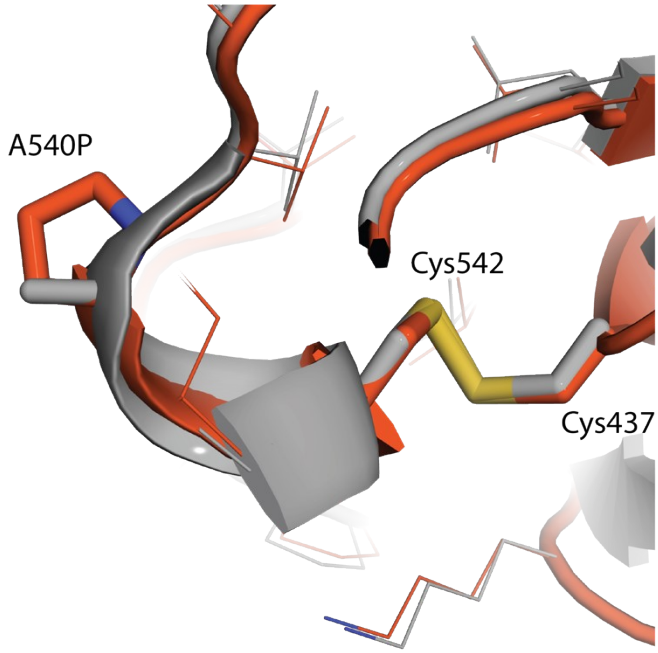
Supplementary Figure 2. Design of stable heparanase mutants from PROSS (**A**) the resulting mutations are shown for each design. (**B**) The crystal structure of heparanase (PDB ID: 5E9C) was used as template whereby 50 kDa subunit (cyan cartoon) was targeted with restriction on the residues contacting the ligand (black sticks) and the 8 kDa subunit (blue cartoon).



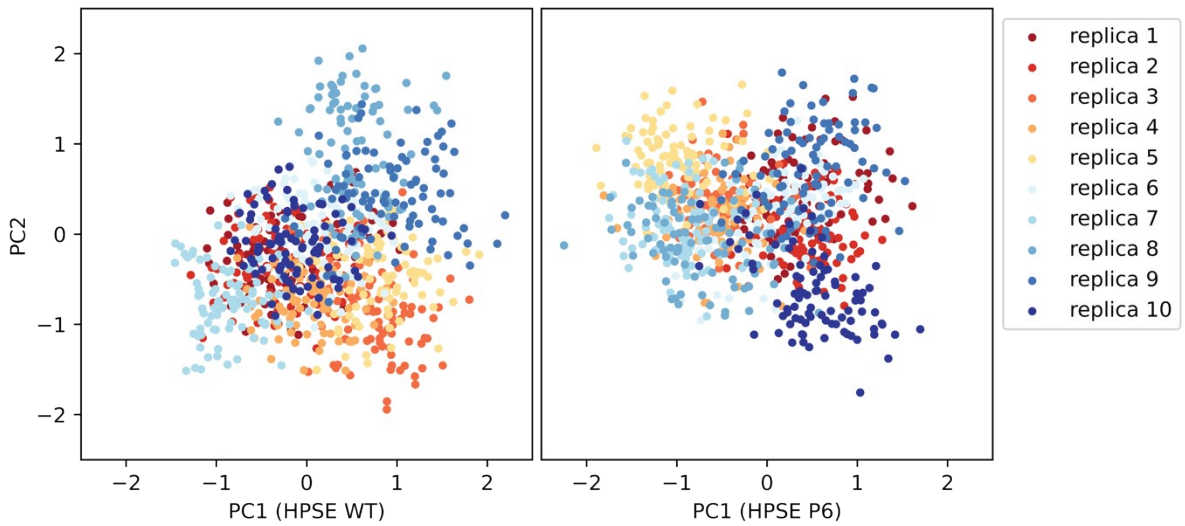
Supplementary Figure 3. Pross variant expression trials in *E. coli* SHuffle cells. P1-P7 plasmid DNA was transformed into Shuffle T7 Express competent cells (NEB) together with varying chaperones in pACYC vector. Skp+TF was tested in 1, 4, 7, 10 and 13. GroEL/ES+TF was tested in 2, 5, 8, 11, 14, 17, and 20. Skp + GroEL/ES was tetesd in 3, 6, 9, 12, 15, 18 and 21. LMW protein marker (GE healthcare) is shown on the left. Overnight seed culture from single colony was inoculated by 1% into LB media supplemented with ampicillin and chloramphenicol antibiotics. The culture was incubated at 37 °C for about 3-5 hours until the OD was about 0.8. 0.4 mM IPTG was added and further incubated for about 3 hours. The cell pellet was resuspended in 50 mM acetate buffer at pH 5 and lysed using BugBuster protein extraction reagent (Merck) for the solubility measurement.



Supplementary Figure 4: Crystallographic B-Factors of WT HSPE and HSPE P6. Comparison between the B factors from the two structures (with different Wilson B-factors and resolution) show an overall decrease in magnitude of the B factors in the Mutant P6, yet the overall trend in terms of regions of high/low B-factors is very similar.



Supplementary Figure 5: The disulfide bond between Cys542 and Cys437 that is stabilised by A540P mutation, which is in a favourable conformation is the kinked loop preceding Cys542.



Supplementary Figure 6: Principal component analysis (PCA) sampling quality for WT and P6 simulations. Sampling shows that the 10 simulations sample a large range of conformational space for both WT and mutant replica simulations, and using all replica simulations is viable in the analysis.

Supplementary Table 1. Nucleotide sequences of heparanase variants used for cloning

Name	Sequences
8 kDa subunit of wild type (gBlock)	CAGGATGTGGTCGATCTTGATTTCTTCACGCAGGAGCCTTTGCACCTTGTATCCCC CTCATTTTTAAGCGTTACCATAGACGCAAACCTTGCCACTGACCCGCGCTTCTTAA TCTTGCTTGGTAGTCCAAAGTTAAGAACGCTGGCGCGGGGGCTTAGTCCAGCATA TCTGCGCTTTGGCGGAACGAAAACAGACTTCCTGATTTTTGATCCTAAAAAGGAA
50 kDa subunit of wild type (gBlock)	AAAAAATTTAAGAACTCCACTTATAGCCGCAGTTCAGTCGACGTGCTGTACACCTT CGCGAACTGTTCGGGATTGGATTTGATATTTGGATTAAATGCATTGTTGCGCACGG CGGATCTGCAGTGGAACTCTAGTAACGCGCAATTATTGTTAGATTATTGTAGTTCG AAGGGCTACAATATATCGTGGGAATTGGGTAATGAGCCGAACAGCTTTTTGAAGAA AGCCGACATCTTTATTAATGGGTCTCAGCTGGGCGAAGATTTTATACAACTTCACA AGCTGTTACGCAAATCAACATTTAAGAACGCGAAGTTATATGGACCAGATGTTGGG CAGCCACGTAGAAAGACCGCCAAGATGCTGAAAAGCTTCCTTAAAGCAGGAGGTG AAGTGATTGACTCGGTGACCTGGCATCACTACTACTTAAACGGAAGAACAGCAACT CGTGAGGATTTCTTGAACCCGGATGTCCTTGATATATTTATTTCATCTGTACAAAAA GTCTTCCAAGTTGTAGAATCCACCAGACCTGGCAAAAAAGTGTGGTTAGGAGAGA CTTCAAGCGCTTACGGCGGTGGTGCACCTCTTTTGTCCGACACCTTCGCGGCAGG CTTCATGTGGCTGGACAAATTGGGCTTAAGCGCGCGTATGGGGATCGAAGTGGTG ATGCGGCAGGTATTCTTTGGCGCCGGGAACTATCACCTGGTCGATGAAAATTTTGA TCCTTTACCTGATTATTGGTTGTCATTACTGTTTAAAAAGTTGGTCGGGACAAAGGT CCTTATGGCTTCTGTCCAGGGGAGTAAAAGAAGAAAATTGAGAGTTTACTTGCATT GCACCAATACGGACAACCCGAGATATAAGGAAGGAGACTTGACCTTATACGCTATC AATTTGCACAATGTTACGAAATATTTGCGTTTACCTTACCCATTCTCCAACAAACAA GTTGACAAATACTTGCTGCGCCCTCTGGGTCCGCATGGCTTATTATCCAAATCGGT TCAGTTGAATGGCTTAACTCTGAAAATGGTAGATGATCAGACATTGCCACCATTGA TGGAGAAA
50 kDa subunit of design (gBlock)	ccccatcttagtatattagttaagtataagaaggagatatacatATGAAAAAATTCAAAAACTCGACGTAT AGCCGGTCTTCTGTGGATGTGCTCTATACTTTTGCGAAGTGTTCGGGCCTGGACTT AATCTTCGGCTTAAATGCACTGCTTCGGACTTCAGATGGGCAGTGGAATTCTAGCA ATGCTCAGCTCCTGCTCGATTACTGTGCCTCTAAAGGGTATAACATCGACTGGGAG TTGGGCAACGAGCCAAATAGCTTCCGTAAAAAGGCTGGGATCTTCATCAACGGGT CGCAATTAGGCAAGGACTTCATTCACCTTCACAAACTGCTCCGGAAATCGACATTT AAGAATGCGAAACTGTATGGCCCTGATGTAGGTCAACCGCGCGGGAAAACGGCCA AAATGCTTAAATCGTTCCTGAAGGCGGGCGGCGAAGTCATTGATGCAGTAACATG GCACCATTACTATTTGAATGGTCGCACCGCCACCTTAGAAGATTTCCTGAATCCGG ACGTATTGGACACGTTTATTTCTCAGGTTCAAAAGGTCTTGCAAGTTGTCGAATCG

	ACCCGGCCTGGGAAGAAAGTTTGGCTCGGGGAGACAAGTTCCGCCTATGGCGGT GGCGCTCCTGGCCTCTCAGATACCTTTGCTGCTGGTTTCATGTGGCTTGATAAACT GGGCCTCTCCGCTCGCATGGGGATCGAAGTCGTGATGCGCCAAGTATTTTTTGGC GCTGGCAACTACCACCTCGTCGACGAAAACTTCGATCCATTGCCTGACTACTGGCT GAGCCTCCTTTTCAAAAAGTTAGTTGGTACAAAGGTGTTGATGGCAAGTGTTCAGG GTCAGGATCGCCGCAAACTTCGCGTTTATCTCCATTGCACAAATACGGATAATCCT CGCTACAAAGAAGGCGACCTGACGCTCTATGCTATCAACCTCCATAACGTCACCA AGTATCTCCGCCTGCCATATCCTTTTAGTAATAAACAGGTGGATCAATATCTCTTGC GCCCTCATGGCCCTGATGGTTTACTGTCCAAGAGCGTGCAGTTGAATGGCCAGAC CCTCAAGATGGTTGATGACCAGACTTTGCCTCCTTTGAAGCCAAAACCACTGCGTC CGGGGAGCAGTCTTGGCCTGCCTGCCTTCTCCTACGCATTTTTTGTAATTCGTAAC GCAAAGGTCCCAGCCTGCATCTGATAACTCGAGTCTGGTAAAGAAACCG

Supplementary Table 2. Amino acid sequences of heparanase variants

8 kDa subunit	MGSSHHHHHHSQDPNSSSQDVVDLDFFTQEPLHLVSPSFLSVTIDANLATDPRFLIL LGSPKLRTLARGLSPAYLRFGGTKTDFLIFDPKKE*
50 kDa subunit of wild type	MKKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSS KGYNISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLLRKSTFKNAKLYGPDVGQP RRKTAKMLKSFLKAGGEVIDSVTWHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQ VVESTRPGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQ VFFGAGNYHLVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGSKRRKLRVYLHCTN TDNPRYKEGDLTLYAINLHNVTKYLRLPYPFSNKQVDKYLLRPLGPHGLLSKSVQLN GLTLKMVDDQTLPPLMEKPLRPGSSLGLPAFSYSFFVIRNAKVAACI*
50 kDa subunit of design	MKKFKNSTYSRSSVDVLYTFAKCSGLDLIFGLNALLRTSDGQWNSSNAQLLLDYCA SKGYNIDWELGNEPNSFRKKAGIFINGSQLGKDFIHLHKLLRKSTFKNAKLYGPDVG QPRGKTAKMLKSFLKAGGEVIDAVTWHHYYLNGRTATLEDFLNPDVLDTFISQVQKV LQVVESTRPGKKVWLGETSSAYGGGAPGLSDTFAAGFMWLDKLGLSARMGIEVVM RQVFFGAGNYHLVDENFDPLPDYWLSLLFKKLVGTKVLMASVQGQDRRKLRVYLH CTNTDNPRYKEGDLTLYAINLHNVTKYLRLPYPFSNKQVDQYLLRPHGPDGLLSKSV QLNGQTLKMVDDQTLPPLKPKPLRPGSSLGLPAFSYAFFVIRNAKVPACI*

Supplementary Table 3. Data collection and refinement statistics

PDB ID	7RG8
Data collection	
Space group	P 2_1 2_1 2_1
Cell dimensions	
a, *b*, *c* (Å)	59.78 76.09 124.43
α, β, γ (°)	90.00 90.00 90.00
Resolution (Å)	47.01-1.30 (1.346-1.3)
R_{merge}	0.07 (1.11)
I / σI	11.82 (0.33)
$CC_{1/2}$	0.998 (0.882)
Completeness (%)	99.78% (99.0%)
Redundancy	13.2 (11.3)
Refinement	
Resolution (Å)	47.01-1.30 (1.346-1.3)
No. reflections	139516 (13656)
R_{work} / R_{free}	0.142/ 0.164
No. atoms	
Protein	3660
Ligand/ion	11
Water	363
B-factors	
Protein	24.89
Ligand/ion	21.94
Water	41.23
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.01
Ramachandran plot	
Preferred (%)	98.68
Allowed (%)	1.32
Outliers (%)	0.00

*X-ray data were collected from single crystals. *Values in parentheses are for highest-resolution shell.