Glycomimetic Affinity-enrichment Proteomics Identifies Partners for Clinically-Utilized Iminosugar

Isa N. Cruz[†,ζ], Conor S. Barry[‡,ζ], Holger B. Kramer[§], C. Celeste Chuang[¬], Sarah Lloyd[⊥], Aarnoud C. van der Spoel[¬,¶], Frances M. Platt[¬], Min Yang[†,*], Benjamin G. Davis[‡,*]

[†] Department of Pharmaceutical & Biological Chemistry, UCL School of Pharmacy, 29/39 Brunswick Square, London, WC1N 1AX, UK.

[‡] Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford, OX1 3TA, UK.

[§] Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Rd, Oxford, OX1 3PT, UK.

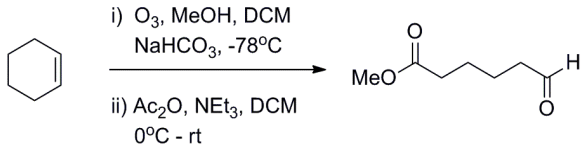
[¬] Department of Pharmacology, University of Oxford, Mansfield Road, Oxford, OX1 3QT, UK.

[⊥] MRC Prion Unit, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK.

Table of content

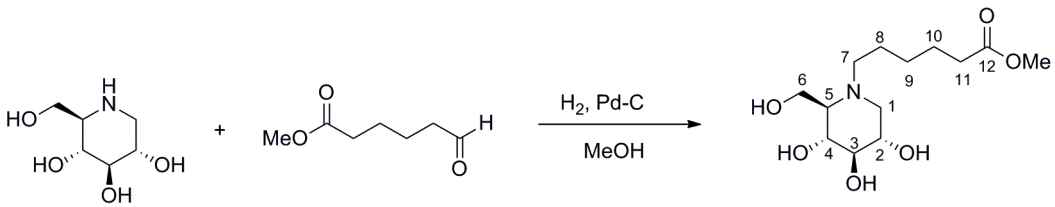
1. Synthesis of Iminosugar Affinity Resin

1.1 Synthesis of methyl 6-oxohexanoate[1]



Cyclohexene (2.53 ml, 25.0 mmol) was added to a suspension of sodium hydrogencarbonate (750 mg) in dry methanol (17 ml) and dichloromethane (83 ml) at -78 °C under a nitrogen atmosphere. Ozone was bubbled through the solution for 50 min until a pale blue colour persisted. The solution was purged until colourless to remove any excess ozone. The suspension was filtered and benzene (30 ml) added. The mixture was azeotroped to ~50 ml volume and then azeotroped again with toluene (50 ml) and dichloromethane (50 ml). The solution was diluted with dichloromethane (75 ml) and cooled to 0 °C under nitrogen. Triethylamine (5.2 ml, 37.5 mmol) and acetic anhydride (7.01 ml, 75 mmol) were added and the mixture was stirred for 30 min at 0 °C and for a further 20 h at rt. The mixture was washed with 0.1M HCl (2x 50 ml), 10% NaOH (50 ml) and water (50 ml), dried ($MgSO_4$), filtered and concentrated *in vacuo* to give a pale yellow oil (3.02 g) which was purified by column chromatography on silica gel eluting with 20% ethyl acetate/petrol to give the title compound as a colourless oil (2.60 g, 72%); δ_H (400 MHz, $CDCl_3$) 1.62-1.70 (4H, m, 3-H_2,4-H_2), 2.30-2.37 (2H, m, 2-H_2), 2.44-2.50 (2H, m, 5-H_2), 3.67 (3H, s, -OMe), 9.76 (1H, br s, 6-H); δ_C ($CDCl_3$) 21.4 (C-2), 24.3 (C-3), 33.7 (C-4), 43.4 (C-5), 52.7 (-OMe), 173.7 (C-1), 202.1 (C-6); ν_{max} (neat)/cm^{-1}) 2953, 2726 (aldehydic C-H), 1736 (C=O).

1.2 Synthesis of methyl 6-((2R,3R,4R,5S)-3,4,5-trihydroxy-2-(hydroxymethyl)piperidin-1-yl)hexanoate[2]



A solution of 1-deoxynojirimycin (50 mg, 0.31 mmol) and methyl 6-oxohexanoate (53 mg, 0.37 mmol) in methanol (5 ml) were stirred with palladium on carbon (10%w/w, 33 mg, 0.03 mmol) at room temperature under an atmosphere of hydrogen for 23 hours. Thin layer chromatography analysis (4:2:1 EtOAc/MeOH/H_2O +0.5 % NEt_3) confirmed the complete consumption of all starting material (Rf = 0.08) and formation of a single product (Rf = 0.53). The reaction flask was purged with nitrogen and the reaction mixture was filtered through a short celite pad eluting with methanol (25 ml) and 1:1 methanol/water (25 ml). The filtrate was concentrated *in vacuo* to give a yellow oil which was purified by column chromatography on silica gel eluting with 7:2:1 EtOAc/MeOH/H_2O +0.5% NEt_3 to give the title compound as a white amorphous solid (74 mg, 83%); $[\alpha]_D^{25}$ -8.1 (c=1.0, MeOH); δ_H (400 MHz, $CDCl_3$) 1.29-1.40 (2H, m, 9-H_2), 1.51-1.62 (2H, m, 8-H_2), 1.63-1.71 (2H, m, 10-H_2), 2.24-2.34 (2H, m, 1-*H*H, 5-*H*), 2.36 (2H, t, *J* 7.3, 11-H_2), 2.69 (1H, ddd, *J* 13.3, 9.2, 6.8, 7-*H*H), 2.90

(1H, ddd, *J* 13.3, 10.0, 6.1, 7-H*H*), 3.07 (1H, dd, *J* 11.3, 4.8, 1-*H*H), 3.18 (1H, t, *J* 9.2, 3-*H*), 3.40 (1H, t, *J* 9.2, 4-*H*), 3.52 (1H, ddd, *J* 10.6, 9.2, 4.9, 2-*H*), 3.67 (3H, s, -OMe), 3.89 (2H, app. d, *J* 2.6, 6-H_2); δ_C (100 MHz, $CDCl_3$) 23.7 (t, C-8), 24.8 (t, C-10), 26.9 (t, C-9), 33.6 (t, C-11), 51.0 (q, -OMe), 52.6 (t, C-7), 56.1 (t, C-1), 57.7 (t, C-6), 66.4 (d, C-5), 69.2 (d, C-2), 70.4 (d, C-4), 79.1 (d, C-3), 174.8 (s, C-12); *m/z* (ES^+) 292.17 ($[M+H]^+$, 100%, 314.16 ($[M+Na]^+$, 94%).

1.3 Preparation of DNJ-affinity Gel

Methyl 6-((2R,3R,4R,5S)-3,4,5-trihydroxy-2-(hydroxymethyl)piperidin-1-yl)hexanoate (17 mg, 0.06 mmol) was dissolved in 1M HCl (2.5 ml) and heated to 65 ˚C for 3 hours. The mixture was concentrated *in vacuo* and lyophylized from deionized water (1 ml) to give 6-((2R,3R,4R,5S)-3,4,5-trihydroxy-2-(hydroxymethyl)piperidin-1-yl)hexanoic acid **1** as a white solid (12 mg, 100%); δ_H (400 MHz, $CDCl_3$) 1.24-1.32 (2H, m, 9-H_2), 1.46-1.73 (5H, m, 8-H_2, 10-H_2, N*H*), 2.36 (2H, t, *J* 7.2, 11-H_2), 2.97 (1H, t, *J* 12.0, 1-*H*H), 3.03-3.13 (2H, m, 5-*H*, 7-*H*H), 3.26 (1H, td, *J* 12.0, 5.1, 7-*H*H), 3.39 (1H, app. t, *J* 9.4, 3-*H*), 3.46 (1H, dd, *J* 12.0, 4.8, 1-*H*H), 3.53 (1H, app. t, *J* 10.0, 4-*H*), 3.66 (1H, ddd, *J* 11.3, 6.3, 5.0, 2-*H*), 3.85 (1H, dd, *J* 13.3, 2.4, 6-*H*H), 3.97 (1H, d, *J* 13.3, 6-H*H*). Glyco-affinity probe **1** hydrochloride salt (12.4 mg, 40 µmol) was dissolved in water (1 ml) and AffiGel-102 (BioRad) (1 ml) was added. The mixture was agitated and the pH adjusted to 4.7. EDAC (10 mg, 52 µmol) was added to the mixture and the pH was adjusted to 5.0. DMSO (2.84 µL, 40 µmol) was added as an internal standard and the mixture was mixed end-over-end at room temperature for 16 hours. The modified gel was collected by gravity filtration and washed with distilled water (50 ml) to give the final immobilized glyco-affinity probe. The initial flow through was concentrated and analysed by NMR. Comparison of the integrals of the DMSO peak and one of the 6-*H*H signal in the ^{1}H NMR spectrum indicated a loading of 40.6 µmol per ml of gel.

2. DNJ-Affinity Chromatography with Ceredase

DNJ column was equilibrated with 4 x 4 mL water and 2 x 4 mL washing buffer. Ceredase™ (alglucerase; 0.16 units; Genzyme, Cambridge, MA) diluted in 3 mL washing buffer was loaded on to the column and allowed to incubate for 2 hr with gentle agitation at 4°C. Sodium acetate buffer (50 mM, 20 mL) was applied to wash the column and the flow-through collected. Miglustat buffer (3 mL) was added to the column and allowed to incubate for 2 hr at 4°C with gentle agitation. Ceredase was eluted with 20 mL miglustat buffer. The flow-through of the washing, eluting and cleaning steps was collected in separate fractions and concentrated as described above.

2.1 Ceredase Activity Assay

DNJ column matrix (50 µL) was packed in a column. Water (2 mL) and 50 mM sodium acetate (2 mL) were applied to equilibrate. Diluted Ceredase (200 µL, containing 0.016 or 0.0016 units of Ceredase) was applied to the column and allowed to drip through. The flow-through (10 µL) was incubated with Ceredase substrate buffer (10 µL; 8 mM 4-Methylumbelliferyl-β-D-glucoside, 100 mM citric acid, 4 mM β-mercaptoethanol, 1.25 mM EDTA, 0.25% (v/v) Triton X-100 and 0.25% sodium taurocholate) for 20 min at 37°C. Stop buffer (200 µL; 0.5 M sodium carbonate, pH 10.7) was added to the reaction mix. The fluorescence of the released 4-Methylumbellifer (4-MU) was read at 355 nm (excitation) and 460 nm (emission).

2.2 Wessel-Flugge Protein Precipitation

Proteins from column fractions were precipitated as described previously.[3] Protein fractions containing 50 µg protein were topped with water to 150 µL. Samples were added with 600 µL methanol and 150 µL chloroform, and vortexed. Only one phase was observed. The samples were then combined with 450 µL water, vortexed thoroughly and centrifuged at 14,000 *g* for 1 min. The upper organic phase was removed without disturbing the interphase. Methanol (450 µL) was added to the lower phase. Samples were vortexed thoroughly and centrifuged at 14,000 *g* for 3 min. The supernatants were removed and protein pellets were dried under gentle air stream.

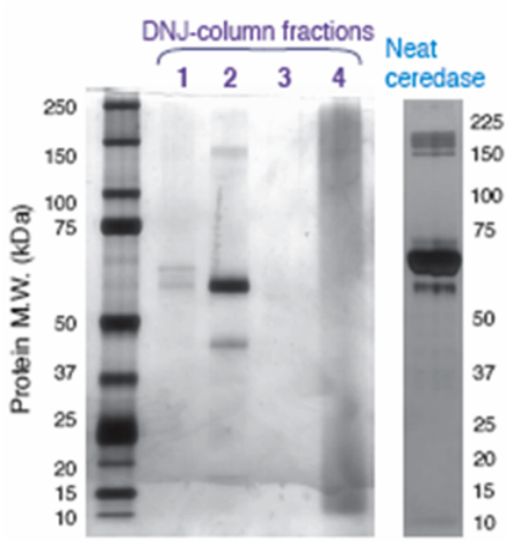


Figure 1. SDS-PAGE of Ceredase™ and fractions eluted from DNJ-affinity gel loaded with Ceredase™. DNJ-affinity gel (4 mL) was equilibrated with water and washing buffer (PBS, 0.5% NaCholate), then incubated with 0.16 U Ceredase™ for 2 hr at 4 °C. The gel column was eluted with: 1) 50 mM sodium acetate, 2) 10 mM NB-DNJ in washing buffer, and cleaned with: 1) 1 M acetic acid, 2) water, 3) 3.5 M magnesium chloride.

2.3 DNJ Affinity Chromatography with Mouse Testis Tissues

Mouse testis tissue (161.6 mg) was homogenized with a sample grinding kit (Amersham Biosciences, UK), using one grinding tube per 100 mg of tissue sample. The grinding tubes were briefly centrifuged at maximum speed to pellet the grinding resin and the liquid was removed from the grinding resin pellet. Each piece of solid tissue of up to 100 mg was placed into a 1.5 mL tube and washed with cold PBS, after which it was transferred to a grinding tube. Testis tissue was homogenized in 1 mL PBS containing 0.5% sodium cholate and 1% protease inhibitor using a pestle to thoroughly grind the sample for 15 to 60 min on ice. Tissue homogenates were lysed by 3 cycles of freeze/thawing (dry ice/37°C). In order to remove insoluble materials, homogenates were spun at 13,000 g for 20 min at 4 °C. The supernatant was collected and transferred to a clean tube. Solubilized proteins were loaded on to the affinity columns immediately.

DNJ affinity column matrix (1 mL) was loaded on to a 5-mL solid phase extraction cartridge. The column was equilibrated by 4 x 4 mL water and 2 x 4 mL washing buffer (PBS containing 0.5% sodium cholate). Solubilized proteins were added on to the column and the column was sealed. Proteins and column matrix were incubated on a rollerbank at 4°C overnight to allow maximum binding.

Washing buffer (12 x 4 mL) was applied to remove non-column-binding proteins. The flow-through was collected along with the washing fraction. Bound proteins were incubated with 4 mL miglustat buffer (PBS containing 10 mM miglustat and 0.5 % sodium cholate) for 4 hr at 4°C and competitively eluted with 11 x 4 mL miglustat buffer and collected as the eluate fraction. The column was cleaned by 5 x 4 mL 1 M acetic acid (pH 2.4), 2 x 4 mL water and 5 x 4 mL 3.5 M magnesium chloride. The flow-through was collected in the cleaning fraction. The column was stored in water supplemented with 1μM sodium azide at 4°C.

The washing, eluting, and cleaning fractions were concentrated by 15 mL centrifugal filter units (3 kDa molecular weight cut-off, Amicon-Ultra, Millipore, MA, USA) at 3,700 *g* for approximately 60 min, and further reduced to 50-100 μL by 500 μL centrifugal units (3 kDa molecular weight cut-off, Millipore) at 14,000 *g* for 20-30 min. Each fraction was added with 1% protease inhibitor and stored at -80°C.

3. Proteomics Analysis

3.1 Materials

Urea, DTT, Glycerol, bromophenol blue, iodoacetamide, agarose and acetic acid were purchased from Sigma. CHAPS, ethanol, formic acid and acetonitrile were from Fisher Scientific. Thiourea, ampholytes solution, ammonium bicarbonate were obtained from Fluka. SDS was from Sigma Aldrich and Tris was purchased from Aldrich.

3.2 Methods

3.2.1 Two-dimensional Gel Electrophoresis (2D-PAGE)

3.2.1.1 Trichloroacetic Acid Precipitation of Proteins

Proteins obtained from above procedure were mixed with 4 volumes of 20% trichloroacetic acid (TCA). The mixture was incubated on ice for 1 hr and then centrifuged at 13,000 rpm for 10 min. TCA was removed and the pellet was washed with 300 μL of 90% ice-cold acetone and centrifuged at 13,000 rpm for a further 10 min. This washing/centrifugation step was repeated to completely remove TCA and the pellet was allowed to air-dry.

3.2.1.2 First Dimension – Isoelectric Focusing (IEF)

Isoelectric focusing was performed using a Protean IEF Cell (Bio-Rad) with 7 cm ReadyStrips, pH 3 – 10 (Bio-Rad). The pellet resultant from TCA precipitation was resuspended in rehydration buffer (7 M urea, 2 M thiourea, 4% CHAPS, 20 mM DTT, 0.5% ampholytes solution pH 3 – 10) to a total volume of 125 μL. Mixtures were centrifuged at 14,000 rpm for 5 min at room temperature and the supernatant was loaded into a focusing tray. The immobilized pH gradient (IPG) strip was placed above the mixture and the liquid was allowed to distribute for 1 hr before covering the strip with mineral oil (Bio-Rad). Rehydration procedure was performed at 50 V (active rehydration) for 12 hr. The focusing was then started and carried out on a linear ramp according to the following steps: 250 V for 15 min; 4000 V for 2 hr, 4000 V until 20 000 V/h. The IPG strip was immediately used for the second dimension or stored in a -80ºC freezer.

3.2.1.3 Reduction and Alkylation

The IPG strip was washed 3 times with equilibration buffer (0.375 M Tris-HCl pH 8.8, 6 M urea, 20% glycerol, 2% SDS) and then was incubated in 55 mM DTT solution in equilibration buffer for 1 hr at room temperature with constant shaking. After incubation, the DTT solution was discarded and 100 mM iodoacetamide solution was added to the strip. The strip was then incubated in the dark for 1.5 hr at room temperature with constant shaking, after which the iodoacetamide solution was discarded. The alkylation process was stopped by adding an equal volume of rehydration buffer without DTT (7 M urea, 2 M thiourea, 4% CHAPS), followed by an incubation period of 10 min.

3.2.1.4 Second Dimension – SDS-PAGE

Each strip was transferred to the top of a precast gel, Mini-Protean TGX Precast Gels, Any kD, IPG well comb, 7 cm IPG strip (Bio-Rad). The IPG strip was fixed to the second-dimension gel with 1% low melting agarose in stacking buffer (0.5 M Tris-HCl pH 6.8) with a trace of bromophenol blue. After allowing the agarose to solidify for 5 min and filling the reservoirs with 1x Tris-Glycine-SDS running buffer, electrophoresis was started. This process was carried out using a Mini-Protean Tetra Cell System (Bio-Rad) at 150 V, until the dye front had reached the bottom of the gel.

3.2.1.5 Gel Staining — Silver Staining

The Pierce Silver Stain Kit (Thermo Scientific) was used to perform the silver staining according to the manufacturer's instructions.

3.2.1.6 Gel Imaging

Both images (Control and Enriched sample) were also obtained using EXQuest Spot Cutter (Bio-Rad, UK) and analyzed using PDQuest Advanced software version 8.0.1 (Bio-Rad, UK). The Multi-Channel Viewer tool was used to overlap the two images.

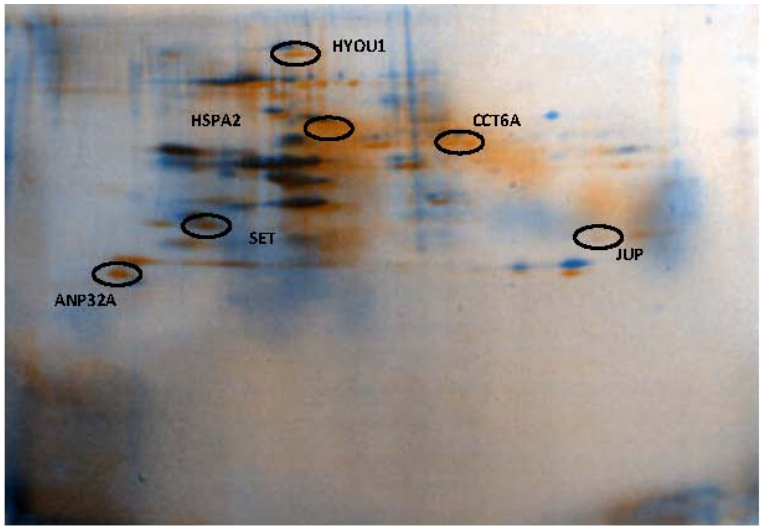


Figure 2. Overlapped gel images. Spots in the control gel are blue and spots in the enriched sample gel are in orange. Overlapped spots are shown in a darker color. The six most relevant proteins identified are highlighted in the image (HYOU1, HSPA2, CCT6A, JUP, SET, ANP32A).

3.3 Protein Identification

3.3.1 Protein Spot Cutting, Washing and Trypsin Digestion

Spots of interest were cut into 1 – 2 mm^3 gel pieces, and placed into 0.6 mL siliconized tubes. Gel pieces were rinsed twice with 190 µL of wash solution (50% methanol + 5% acetic acid) at room temperature for 3 hr and overnight, respectively. Gel pieces were then dehydrated in 190 µL of acetonitrile at room temperature for 5 min, after which the samples were dried in a vacuum centrifuge for 3 min. Sample reduction was performed with 30 µL of 10 mM DTT at room temperature for 30 min, followed by alkylation with the same volume of 100 mM iodoacetamide at room temperature for another 30 min. Samples were dehydrated again in 190 µL of acetonitrile at room temperature for 5 min and dried in a vacuum centrifuge for 3 min. Gel pieces were rehydrated in 190 µL of 100 mM ammonium bicarbonate at room temperature for 10 min and then dehydrated in acetonitrile and dried in a vacuum centrifuge another time.

Finally, samples were rehydrated on ice, for 10 min, with 20 µL of trypsin solution (20 ng/µL sequencing grade modified porcine trypsin (Promega) in 50 mM ice-cold ammonium bicarbonate pH 8.0) with occasional vortex mixing. Smples were spun down for 30 sec and excess trypsin solution was removed. An aliquot of 10 µL of 50 mM ammonium bicarbonate solution was added and proteins were digested overnight at 37ºC.

3.3.2 Peptide Extraction from Gel Pieces

After digestion, 30 µL of 50 mM ammonium bicarbonate solution were added to the tubes containing the gel pieces, which were vortexed for 10 min after which supernatants were collected and transferred to new tubes. Peptides were firstly extracted from the gel pieces with 30 µL of extraction buffer I (50% acetonitrile + 5% formic acid) for 10 min. A second extraction was performed with 30 µL of extraction buffer II (85% acetonitrile + 5% formic acid) for another 10 min. Supernatants were collected after each extraction and combined with the previous fraction. The extracts volume was reduced to <20 µL by evaporation in a vacuum centrifuge. The final dried extracts were redissolved in 5% acetonitrile + 0.1% formic acid, sonicated for 10 min, centrifuged for 5 min and transferred to mass spec vials.

Table 1. Significant proteins identified from mouse testis with putative NB-DNJ affinities.[b]

Spot	Accession Number	Protein Name	Code	Score	Mr (Da)	pI	% Coverage
73[a]	Q9JKR6	Hypoxia up-regulated protein 1	HYOU1	291	111340	5.12	68
54	Q02257	Junction plakoglobin	JUP	65	82490	5.75	37.3
77	P17156	Heat shock-related 70 kDa protein 2	HSPA2	92	69884	5.51	51
45	O35381	Acidic leucine-rich nuclear phosphoprotein 32 family member A	ANP32A	86	28691	3.99	23.9
49	Q9EQU5	Protein SET	SET	49	33358	4.22	29.1
60/61	P80317	T-complex protein 1 subunit zeta	CCT6A	45	58424	6.63	51.8
50	P70670	Nascent polypeptide-associated complex subunit alpha	NACA	104	221277	9.39	49.1
62	P80314	T-complex protein 1 subunit beta	CCT2	60	57783	5.96	13.8
56	Q059Y8	DC-STAMP domain-containing protein 1	DCST1	41	85646	9.42	37.8
71	Q8BJQ2	Ubiquitin carboxyl-terminal hydrolase 1	USP1	38	88314	5.33	41.6
77	P17879	Heat shock 70 kDa protein 1B	HSPA1B	60	70418	5.53	44.1
46	Q63ZV0	Insulinoma-associated protein 1	INSM1	46	55042	9.24	41.5
54	P01634	Ig kappa chain V-V region MOPC 21	-	40	15063	6.26	38.2

54	Q52KB6	C2 domain-containing protein 3	C2CD3	38	257281	6.52	37.8
77	Q6A068	Cell division cycle 5-related protein	CDC5L	37	92361	7.98	37.3
52	A2ARV4	Low-density lipoprotein receptor-related protein	LRP2	27	537628	4.94	18.6
54	Q8VHJ7	Peroxisome proliferator-activated receptor gamma coactivator 1	PPARGC1B	26	113773	4.92	25
45	Q8C9J3	Sperm flagellar protein 2	SPEF2	23	199477	5.65	35.6

[a] Spots in bold and italic indicate unique proteins; spots in bold indicate enriched proteins. [b] It should be noted that some membrane bound CAP glycosidases and glycosyltransferases have required bespoke separation methods (e.g. sucrose gradient for Golgi membranes)[4] and this may lead to a potential under-representation in the broad-ranging glyco-A^{e}P strategy that we have employed here.

3.4 MS/MS Analysis

LC-MS/MS analysis was performed on a Waters CapLC system coupled to the front end of a Waters Micromass Q-Tof Premier. A 63 min gradient LC method was followed: t = 0.1 min, 5% B; t = 3 min, 5% B; t = 40 min, 28% B; t = 49 min, 80% B; t = 52min, 80% B; t = 53 min, 5% B and t = 63 min, 5% B (A, 5% ACN + 0.1% formic acid (FA); B, 95% ACN = 0.1% FA). MS was monitored at 400-1700 Da and MS/MS was monitored at 50-1700 Da. Capillary was at 1.8 kv, cone voltage at 35 V and collision energy was at 35 eV.

MS/MS data were used to perform database searching using an online search engine, Mascot[5], choosing SwissProt databases to look for mouse proteins (taxonomy — *Mus musculus*). Searches were performed without restriction of protein molecular mass or pI, but with variable modifications such as carbamidomethylation of cysteines and oxidation of methionine residues. One trypsin miscleavage was allowed. Peptide and fragment mass tolerances were set to ± 0.3 Da.

3.5 LC-MS/MS Analysis

For the analysis of in-gel digested protein material, liquid chromatography was performed using an Ultimate 3000 nano-HPLC system (Dionex, Sunnyvale, CA, USA) comprising a WPS-3000 micro auto sampler, a FLM-3000 flow manager and column compartment, a UVD-3000 UV detector, an LPG-3600 dual-gradient micro-pump, and an SRD-3600 solvent rack controlled by Hystar (Bruker Daltonics, Billerica, MA, USA) and DCMS link 2.0 software. Samples were concentrated on a trapping column (Dionex, Sunnyvale, CA, USA), [300 µm i.d., 0.1 cm] at a flow rate of 20 µL/min. For the separation with a C18 Pepmap column [75 µm i.d., 15 cm], a flow rate of 250 nL/min was used as generated by a cap-flow splitter cartridge (1/1000). Peptides were eluted by the application of a 30 min multi-step gradient using solvents A (98% H_2O, 2% acetonitrile, 0.1% formic acid) and B (80% acetonitrile, 20% water, 0.1% formic acid):

Composition (% solvent B)	Run time (min)
2-10	0-3
10-25	3-18
25-50	18-30
50-90	30-30.2

The liquid chromatography was interfaced directly with a 3D high capacity ion trap mass spectrometer (amaZon; Bruker Daltonics) utilizing 10 µm i.d. distal coated SilicaTips (New Objective, Woburn, MA, USA) and nano-ESI mode. SPS parameter settings on the ion trap were tuned for a target mass of 850 *m/z*, compound stability 100% and a smart ICC target of 250,000. MS/MS analysis was initiated on a contact closure signal triggered by HyStar software (version 3.2). Up to five

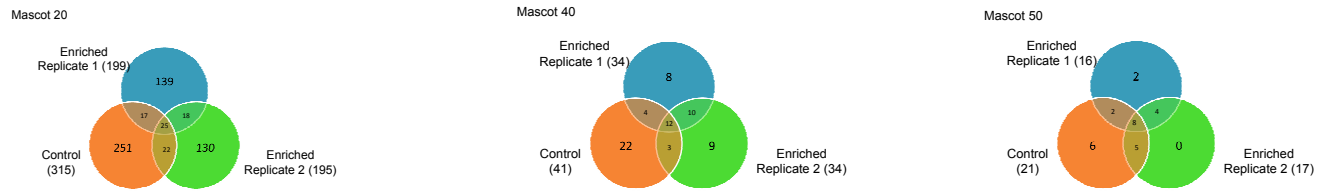
precursor ions were selected per cycle with active exclusion (0.5 min) in collision-induced dissociation (CID) mode. CID fragmentation was achieved using helium gas and a 30%–200% collision energy sweep with amplitude 1.0 (ions are ejected from the trap as soon as they fragment).

3.6 Data Processing and Database Searching

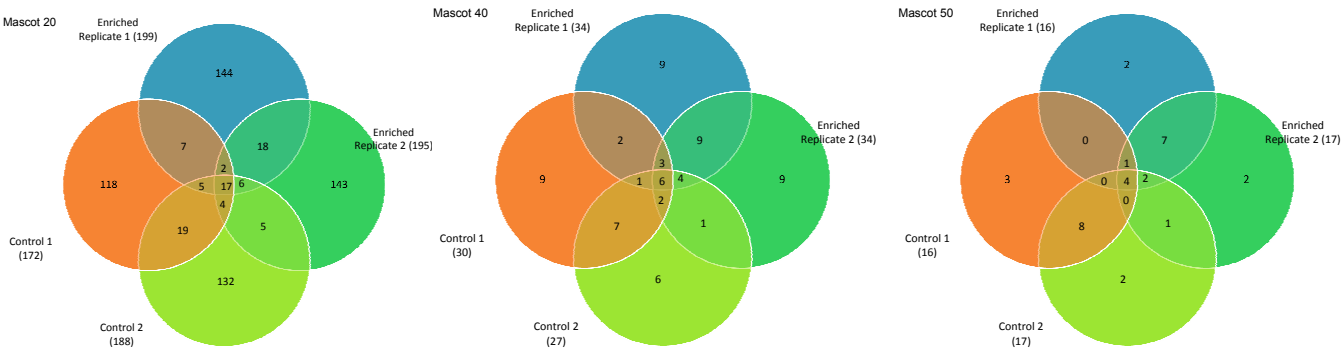
Raw LC-MS/MS data were processed and Mascot compatible files created using DataAnalysis 4.0 software (Bruker Daltonics) with the following parameters: compounds (autoMS) threshold 10,000, number of compounds 500, retention time windows 0.5 min for C18 (30 min gradient). Database searches were performed using the Mascot software (version 2.3) and the SwissProt database (v2011.02.11, number of entries 525,207) with the following parameters: 1+, 2+ and 3+ ions, peptide tolerance 0.3 Da, ^{13}C = 2, fragment tolerance 0.6 Da, missed cleavages: 2, instrument type: ESI-TRAP.

Figure 3. Venn Analysis of the outcomes at different levels of protein score thresholds and different use of either (a) 'combined exclusion' or (b) 'overlap exclusion'. Different protein scores from the Mascot software (version 2.3) lead to different outcomes in the excluded and overlapped sets of identified proteins. As the score threshold is increased, as would be expected, the agreement between sets increases.

(a)



(b)



4. Pathway Analysis

All interactions of HYOU1protein were analyzed using Pathway Studio 9.

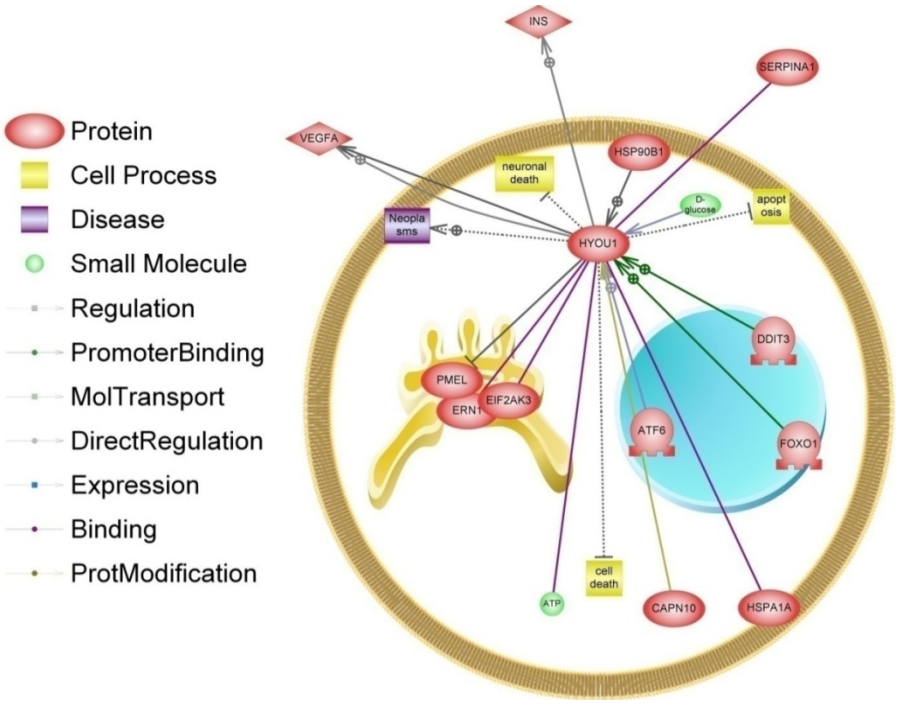


Figure 4. Pathways focussed on HYOU1

5. Single-nucleotide Polymorphism Analysis

Single-nucleotide polymorphism (SNP) analysis was applied to the genes encoding for proteins identified in Table 1 examining differences between Miglustat sensitive strain C57BL/6 and insensitive strain FVB/N.

The SNP differences between C57BL and FVB/N were downloaded from the Sanger Centre ftp site (ftp://ftp-mouse.sanger.ac.uk/REL-1206-FVBNJ/ data file: 2012-0612-snps+indels_FVBNJ_annotated.vcf). The genes, and chromosome positions that we used are listed in table 2. Annotated gene sequence were searched 5kb both upstream and downstream. No SNPs were found for *Hyou1*, *Jup*, *Anp32b* and *Set*. SNPs for *Hspa2*, *Anp32a* and *Cct6a* are in the following tables that indicate the chromosome, location, reference sequence (C57BL/6) and FVB/N sequence. There are 180 for *Cct6a* (Table 3), 57 SNPs for *HspA2* (Table 4) and 367 for *Anp32a* (Table 5). All SNP data come from the Sanger Institute sequence data and all base position numbers refer to NCBI Build 37. The accession number for each gene is given below.

Cct6a (NM_009838): SNPs in non-coding regions including 3'UTR. Six changes in the coding region; five of which do not change the corresponding amino acid. One in exon 9, Chr5 position 130299586 has the SNP designation rs13470985 and changes amino acid 348 from E (C57BL/6) to G in (FVB/N).

Hspa2 (NM_001002012): SNPs in non-coding regions including 3'UTR. Three SNPs were identified in the coding region but do not change the corresponding amino acids.

Anp32a (NM_009672) : SNPs in non-coding regions including 3'UTR.

It is possible that the non-coding changes may affect gene regulation. Since no microarray data on FVB/N testis tissue is yet available, it is not possible to compare expression levels.

Table 2 Parameters used in the SNP analysis

Gene name	chr	start	start-5000	end	end + 5000
HYOU1	9	44187573	44182573	44200452	44205452
JUP	11	100239403	100234403	100259053	100264053
HSPA2	12	77505357	77500357	77507923	77512923
ANP32A	9	62189150	62184150	62226609	62231609
ANP32B	4	46463989	46458989	46485395	46490395
CCT6A	5	130293261	130288261	130322231	130327231
SET	2	29922246	29917246	29927314	29932314

Table 3 SNP analysis of *Cct6a*

Chr	Location	Base pair	C57BL/6	FVB/N
5	upstream	130288350	A	G
5		130288539	T	A
5		130288770	ATTTTTT	ATTTTTTT
5		130288891	G	T
5		130288917	C	T
5		130289008	G	A
5		130289021	G	T
5		130289041	G	A
5		130289057	G	A
5		130289088	C	T
5		130289260	G	A
5		130289316	G	A
5		130289318	C	T
5		130289345	A	G
5		130289434	G	A
5		130289477	TAAAAAA	TAAAAAAA
5		130289494	A	G
5		130289513	C	T
5		130289583	G	A
5		130289595	A	T
5		130289732	GGTACTCGTGTAC	GGTAC
5		130289824	T	TG
5		130289902	T	C
5		130289910	C	T
5		130289932	GGAGAGAGA	GGAGA
5		130290179	T	C
5		130290180	G	C
5		130290214	T	C
5		130290294	ATTTTTTTT	ATTTTTTT
5		130290393	A	C

5	130290398	T	C
5	130290516	G	A
5	130290610	ACCCC	ACCCCC
5	130290655	T	C
5	130290690	T	G
5	130290708	A	G
5	130290889	A	G
5	130290935	A	G
5	130290954	C	T
5	130291015	ATTTTTTTT	ATTTTTT
5	130291142	G	C
5	130291148	G	A
5	130291184	G	A
5	130291220	AGAGTGAGT	AGAGT
5	130291248	C	T
5	130291342	G	C
5	130291391	A	G
5	130291458	C	T
5	130291509	G	A
5	130291638	C	T
5	130291727	A	G
5	130291940	G	A
5	130291985	T	A
5	130292137	TT	TTCT
5	130292210	GTATATATA	GTATA
5	130292232	C	T
5	130292378	A	T
5	130292563	A	C
5	130292813	GAAAAAAAAAAAA	GAAAAAAAAAA
5	130292839	T	C
5	130293048	G	C
5	130293140	G	A

5	intronic	130293530	C	G
5		130293577	C	T
5		130293597	G	T
5		130293658	C	T
5		130293666	G	C
5		130293776	ATTTTTTTTT	ATTTTTTTTTT
5		130293969	A	T
5		130294473	C	T
5		130294513	C	T
5		130294828	A	G
5		130295067	T	C
5		130295124	A	T
5		130295125	G	A
5		130295135	G	A
5		130295233	A	G
5		130295254	T	A
5		130295508	G	A
5		130295600	A	G
5	Ex4 coding gcg-gcc A152A	130295812	G	C
5	intronic	130295902	G	A
5		130295920	A	AT
5		130295964	G	A
5		130296003	C	A
5		130296126	C	T
5		130296129	A	G
5		130296183	G	C
5		130296193	A	G
5		130296245	G	A
5		130296383	A	G
5		130296730	AGG	AG
5		130296960	T	A
5		130296964	G	T

5		130297035	G	T
5		130297036	C	T
5		130297311	A	G
5	Ex6 coding act-acg T232T	130297427	T	G
5	intronic	130297489	G	A
5		130297573	G	C
5		130297580	T	A
5		130297590	C	A
5		130297595	A	G
5		130297606	A	T
5		130297623	A	G
5		130298068	A	G
5		130298094	TTTTTTTT	TTTTGTTTTT
5		130298124	T	C
5		130298140	G	A
5		130298176	CTCTT	CT
5		130298212	A	G
5		130298286	G	A
5		130298290	G	A
5		130298450	A	G
5		130298489	G	A
5		130298510	TCCCCC	TCCCC
5		130298577	G	A
5		130298596	G	A
5		130298704	T	C
5		130298771	A	G
5		130298772	A	G
5		130298944	A	G
5		130299030	C(GT)25	C(GT)11
5		130299132	C	T
5		130299349	G	T
5	**EX9 coding gag-ggg**	130299586	A	G

E348G

5	intronic	130299700	TTT(GGTT)6	TTT(GGTT)4
5		130299933	A	G
5		130300208	G	A
5		130300578	G	T
5		130320695	G	A
5	Ex11 coding tcg-tca S471S	130320911	G	A
5	Ex11 coding tcc-tcg S473S	130320917	C	G
5	intronic	130320992	GAA	GA
5		130321151	CT	C
5		130321200	G	C
5		130321356	GAAAAAAAAAAAA	GAAAAAAAA
5	Ex12 coding gct-gcc A489A	130321475	T	C
5	3UTR	130321935	CATAA	CATAATAA
5		130321967	G	A
5		130322202	T	G
5	downstream	130322330	C	G
5		130322430	ATTTTTT	ATTTTTTT
5		130322572	A	T
5		130322619	C	G
5		130322708	T	C
5		130322837	A	C
5		130322990	A	C
5		130323000	T	C
5		130323004	A	T
5		130323045	T	C
5		130323109	CTTTTTCTTTTTTTTTT	CTTTTT
5		130323147	C	T
5		130323177	G	A
5		130323301	T	C
5		130323413	A	G

5	130323429	G	C
5	130323635	C	T
5	130323708	T	G
5	130323709	T	G
5	130323710	T	G
5	130323753	A	T
5	130323768	ATTTT	ATTTTT
5	130323815	T	G
5	130324321	GTTTTTTTT	GTTTTTTTTT
5	130324544	A	G
5	130324566	C	G
5	130324841	TCACACA	TCACACACACA
5	130324937	AT	ATT
5	130325005	T	C
5	130325024	T	C
5	130325025	G	A
5	130325044	T	C
5	130325173	G	C
5	130325214	C	T
5	130325427	G	A
5	130326202	GAA	GAAA
5	130326395	A	G
5	130326618	C	T
5	130326917	C	T

Table 4 SNP analysis of *HspA2*

Chr	Location	Base pair	C57BL/6	FVB/N
12	upstream	77500511	CAAAATAAAATAAAATAAAATAAA	CAAAATAAAATAAAATAAA
12		77500802	TAA	TAAA
12		77501115	C	A
12		77501121	C	T
12		77501180	GAAAAAAAAAAA	GAAAAAAAAAAAA
12		77501825	AATATATATAT	AATATATAT
12		77502211	GTTTT	GTTT
12		77502375	C	T
12		77502483	T	C
12		77502577	ACAGC	ACAGCCAGC
12		77502666	T	A
12		77502883	T	A
12		77502964	C	T
12		77503121	TTATGTATGTATGTATGTATGT	TTATGTATGTATGTATGT
12		77503724	ATTTTTTTTTT	ATTTTTTTTT
12		77503914	CA CAAATA	
12		77504126	G	C
12		77504129	A	T

12		77504345	C	T
12		77504400	TCTACT	TCT
12		77504428	C	A
12		77504619	G	A
12		77504710	T	C
12		77504715	C	T
12		77504913	G	A
12	coding tac->tat Y184Y	77506072	C	T
12	coding gcc->gcg A494A	77507002	C	G
12	coding gtg->gtc V555V	77507155	G	C
12	3UTR	77507471	CTTTTTTTTTTTTTT	CTTTTTTTT
12	3UTR	77507742	T	G
12	downstream	77508199	C	T
12		77508240	T	A
12		77508535	T	C
12		77508611	C	A
12		77508615	G	A
12		77508637	A	G
12		77508683	T	G
12		77508921	GTTTTTTTTT	GTTTTTTT
12		77509226	TAGACA	TAGACAGACA

12	77509474	G	T
12	77509498	ATACTTAC	ATAC
12	77509792	T	C
12	77510054	C	T
12	77510303	A	T
12	77510683	G	A
12	77510745	AAGTAGTAG	AAGTAG
12	77510912	G	A
12	77511013	A	G
12	77511045	T	G
12	77511068	G	A
12	77511099	A	C
12	77511194	TAAA	TAAAA
12	77511202	TAAAAAAAAAAA	TAAAAAAAAAAAA
12	77511331	G	GT
12	77511400	T	C
12	77511490	A	G
12	77511526	T	C

Table 5 SNP analysis of *Anp32*

Chr	Location	Base pair	C57BL/6	FVB/N
9	upstream	62184340	T	C
9		62184366	C	T
9		62184540	C	T
9		62185189	C	T
9		62185595	G	A
9		62186271	G	A
9		62186357	C	A
9		62186410	G	A
9		62186445	CTTT	CTT
9		62186468	C	T
9		62186610	G	A
9		62186624	G	A
9		62186733	C	G
9		62186851	A	T
9		62186955	T	C
9		62186959	G	A
9		62187621	G	A

9		62187669	C	T
9		62187708	G	A
9		62187912	C	T
9		62187935	C	A
9		62187985	C	A
9		62188189	G	A
9		62188295	GCCCCC	GCCCC
9		62188335	A	G
9		62188685	T	C
9		62189108	T	A
9	intronic	62189712	T	C
9		62190963	A	G
9		62191285	C	G
9		62191366	C	T
9		62191805	C	A
9		62191912	TGGGGG	TGGGGGG
9		62192385	G	T
9		62192398	GCCCCCCCCCCCCC	GCCCCCCC
9		62192445	G	T
9		62192482	T	C
9		62193067	ATTTTTTTTTTT	ATTTTTTTTTT

9	62193112	A	G
9	62193413	TATCAAT	TATCAATCAAT
9	62193670	T	G
9	62193679	G	A
9	62193695	G	A
9	62193702	A	G
9	62193762	GAAA	GAA
9	62193821	T	A
9	62193876	T	A
9	62193950	G	A
9	62194090	C	T
9	62194366	A	G
9	62194504	T	C
9	62194514	A	G
9	62195010	GGGGGGGG	GGGGGGGGGG
9	62195029	T	C
9	62195218	A	T
9	62195221	C	T
9	62195492	T	C
9	62195513	T	A
9	62195570	T	TC

9	62195866	C	T
9	62195993	A	G
9	62196157	T	C
9	62196167	C	A
9	62196336	G	T
9	62196481	T	C
9	62196533	TGGGGGGGGGG	TGGGGGG
9	62196635	C	T
9	62196661	G	C
9	62196731	A	T
9	62196845	G	A
9	62197085	G	T
9	62197124	C	G
9	62197302	C	T
9	62197338	T	C
9	62197436	G	A
9	62197470	G	A
9	62197481	TCC	TCCC
9	62197602	TACACACACACACACACACACA	TACACACACACACACA
9	62197887	T	C
9	62197901	G	A

9	62198527	T	A
9	62198703	T	A
9	62198811	A	T
9	62198813	C	T
9	62199012	TAAAAATCCTTAAAAAATCCTTA	TAAAAATCCTTA
9	62199047	A	C
9	62199095	CCCGCC	CCC
9	62199111	A	G
9	62199189	TT	TTAAAAT
9	62199205	A	T
9	62199335	T	C
9	62199379	T	C
9	62199390	T	C
9	62199610	A	G
9	62199628	T	G
9	62199928	C	A
9	62199942	A	T
9	62200048	CTTTTTTT	CTTTTTT
9	62200209	G	A
9	62200361	C	T
9	62200719	C	G

9	62200940	T	G
9	62200968	T	C
9	62201005	T	C
9	62201062	AGT	A
9	62201109	CTT	CT
9	62201217	GTTTTTTTTTTTTT	GTTTTTTTTTTTT
9	62201249	C	T
9	62201323	G	T
9	62201366	A	G
9	62201452	A	G
9	62201620	A	G
9	62201631	A	C
9	62201662	C	T
9	62201879	C	A
9	62201885	A	G
9	62202063	T	C
9	62202143	G	C
9	62202152	A	G
9	62202186	A	T
9	62202343	T	C
9	62202443	TA	TATAA

9	62202451	CAA	CAAA
9	62202533	A	T
9	62202555	T	A
9	62202618	C	CCA
9	62202661	C	T
9	62202708	C	T
9	62202780	G	T
9	62202782	T	C
9	62202784	G	A
9	62202791	A	T
9	62202894	T	C
9	62202895	T	C
9	62203208	T	TA
9	62203231	T	A
9	62203477	T	C
9	62203487	A	G
9	62203489	A	C
9	62203500	C	T
9	62203513	T	C
9	62203531	G	A
9	62203680	C	T

9	62203686	T	C
9	62203789	G	A
9	62203863	C	G
9	62203929	G	A
9	62204009	G	C
9	62204148	C	T
9	62204162	AAGACCAG	AAG
9	62204499	C	T
9	62204563	A	G
9	62204579	T	C
9	62204728	C	T
9	62204768	T	G
9	62204856	CT	CTT
9	62204870	GAA	GA
9	62204928	G	A
9	62205480	C	T
9	62205496	C	T
9	62205774	CCCCCG	CCCCCGCCCCG
9	62206025	T	C
9	62206118	G	A
9	62206186	T	G

9	62206506	C	T
9	62206610	A	G
9	62206910	C	T
9	62207373	T	C
9	62207831	G	A
9	62208025	C	G
9	62208237	A	G
9	62208631	G	A
9	62208859	G	T
9	62208991	C	T
9	62209042	T	C
9	62209074	TGGGGG	TGGGG
9	62209458	A	T
9	62209461	G	T
9	62209570	G	A
9	62210012	G	T
9	62210161	G	A
9	62210551	A	T
9	62210878	A	G
9	62211031	C	T
9	62211127	C	T

9	62211185	G	A
9	62211218	G	A
9	62211326	C	T
9	62211566	G	A
9	62211680	A	G
9	62211756	G	A
9	62211829	TAAGACTACGGAAGA	TAAGA
9	62211978	G	A
9	62211997	AAAAATTAAAATTAA	AAAAATTAA
9	62212080	C	A
9	62212118	T	A
9	62212192	G	T
9	62212234	C	T
9	62212322	G	A
9	62212333	GTTTTTTTTTTTT	GTTTTTTTTTTTT
9	62212363	T	C
9	62212388	A	G
9	62212411	T	A
9	62212569	A	G
9	62212609	T	A
9	62212664	T	C

9	62212707	A	T
9	62212824	G	T
9	62213068	G	A
9	62213074	C	T
9	62213145	A	G
9	62213158	G	A
9	62213280	C	G
9	62213281	A	G
9	62213283	T	C
9	62213348	A	G
9	62213375	G	A
9	62213432	C	T
9	62213447	TCCCC	TCCC
9	62213576	T	C
9	62213658	T	C
9	62213671	G	A
9	62213765	A	T
9	62213786	A	C
9	62213935	A	G
9	62214064	G	A
9	62214114	T	C

9	62214176	A	G
9	62214219	GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	GCTCTCTCTCTCTCT
9	62214277	C	A
9	62214386	T	G
9	62214387	T	G
9	62214388	T	G
9	62214392	T	C
9	62214394	G	A
9	62214407	C	G
9	62214441	ACGCCCAAAGCTGTTGCGCCCAAAGCTGTTGCG	ACGCCCAAAGCTGTTGCG
9	62214487	G	A
9	62214691	C	T
9	62214774	A	G
9	62215037	A	G
9	62215064	A	G
9	62215297	C	G
9	62215578	C	T
9	62215634	T	A
9	62215920	GC	GCC
9	62216018	G	A
9	62216254	G	A

9	62216257	C	T
9	62216399	T	C
9	62216424	A	G
9	62216532	T	C
9	62216612	G	A
9	62216798	CT	C
9	62216837	T	C
9	62216910	G	A
9	62217038	C	T
9	62217100	A	T
9	62217144	G	A
9	62217146	C	T
9	62217174	C	T
9	62217215	T	C
9	62217404	T	C
9	62217435	T	C
9	62217437	A	G
9	62217489	A	C
9	62217539	T	G
9	62217582	G	C
9	62217588	T	C

9	62217856	A	G
9	62217911	AG	AGGGGTG
9	62217917	GGGGG	GGGGGTGGGG
9	62217969	A	G
9	62217974	G	T
9	62217999	T	A
9	62218065	C	G
9	62218068	T	C
9	62218120	TCCCCC	TCCCC
9	62218265	TCCC	TCC
9	62218283	A	G
9	62218327	G	C
9	62218347	TATGAAGCAGGGACCA	TA
9	62218391	C	G
9	62218421	T	A
9	62218423	G	T
9	62218532	A	C
9	62218707	G	A
9	62218726	TGGGG	TGGG
9	62218928	C	T
9	62218939	TCACACACACACACACACAC	TCACACACACACACAC

9	62219137	G	A
9	62219321	A	ATG
9	62219434	G	A
9	62219744	G	C
9	62219753	G	T
9	62220002	TGGGG	TGG
9	62220540	T	C
9	62220899	A	C
9	62220957	A	G
9	62220982	CTT	CT
9	62221028	CCAC	CC
9	62221045	G	A
9	62221213	G	A
9	62221503	CT	C
9	62221693	G	A
9	62222148	C	T
9	62222358	G	A
9	62222393	G	A
9	62222803	G	T
9	62222954	C	G
9	62222981	T	C

9		62223060	C	T
9		62223322	C	T
9		62223511	T	C
9		62223644	A	G
9		62224152	C	T
9		62224787	G	A
9		62224821	T	G
9		62224833	T	C
9		62225087	C	A
9	3UTR	62225954	G	T
9		62226167	CAAAAAAAA	CAAAAAAAAA
9		62226219	C	T
9		62226296	A	G
9		62226445	C	T
9		62226500	T	C
9	downstream	62226838	TGGGGGGGGGG	TGGGGGGGGGGG
9		62226862	TA(TAAA)13TAA	TA(TAAA)9TAA
9		62227338	C	T
9		62227492	TAAAAAAAAAA	TAAAAAAAAAAA
9		62228057	G	A
9		62228078	G	A

9	62228099	GGACTGACT	GGACT
9	62228557	A	G
9	62228558	A	G
9	62228649	T	A
9	62228777	T	C
9	62229110	A	G
9	62229134	G	T
9	62229194	T	C
9	62229210	T	C
9	62229372	C	T
9	62229471	C	T
9	62229688	C	T
9	62229723	AGGGGG	AGGGGGG
9	62229770	T	C
9	62229789	A	G
9	62229879	A	G
9	62229973	G	A
9	62229993	A	G
9	62230043	GGCTAG	GG
9	62230092	T	C
9	62230119	C	T

9	62230231	C	T
9	62230239	T	C
9	62230240	G	A
9	62230389	A	G
9	62230458	C	T
9	62230473	T	A
9	62230633	T	C
9	62230848	G	C
9	62231156	T	C
9	62231240	G	A
9	62231327	A	T
9	62231392	T	C
9	62231466	G	A
9	62231489	C	T

6. References

(1) Schreiber, S. L.; Claus, R. E.; Reagan, J. *Tetrahedron Lett* **1982**, *23*, 3867.
(2) Bernotas, R. C.; Ganem, B. *Biochem J* **1990**, *270*, 539.
(3) Wessel, D.; Flugge, U. I. *Anal. Biochem.* **1984**, *138*, 141.
(4) Lin, C.-H.; Lin, C.-W.; Khoo, K.-H. *Proteomics* **2008**, *8*, 475.
(5) Perkins, D.N.; Pappin, D.J.C.; Creasy, D.M.; Cottrell, J.S. *Electrophoresis* **1999**, *20*, 3551.