A new chemical probe for the detection of the cancer-linked galectin-3†

Lluis Ballell, Monique van Scherpenzeel, Katerina Buchalova, Rob M. J. Liskamp and Roland J. Pieters*

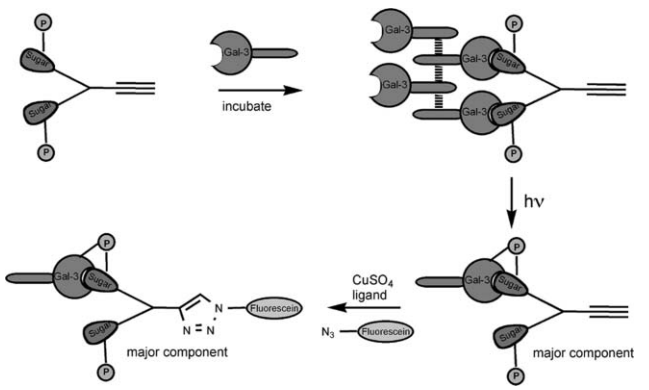
Received 31st July 2006, Accepted 12th October 2006
First published as an Advance Article on the web 30th October 2006
DOI: 10.1039/b611050a

A chemical probe was developed for the detection of the emerging cancer marker galectin-3. The probe contains a benzophenone moiety which covalently attaches itself to the protein upon binding and irradiation. Introduction of a fluorescent label *via* 'click' chemistry allows the labelled proteins to be visualized in a gel. With the probe, selective visualization of galectin-3 in protein mixtures was shown and remarkably even in cell lysates.

Introduction

The development of methods that allow the study of specific subsections of the proteome is an important research goal.[1] Current methods typically reduce the sample complexity through multidimensional separations, followed by identification and quantification by mass spectrometry. Complementary to this, synthetic molecular probes can tag a particular protein class, typically based on catalytic activity. The advantage of this path is that information about the quantities of functional proteins is obtained. Probes designed for enzymes can covalently capture their target proteins by taking advantage of their catalytic mechanisms with the use of suicide inhibitors.[2] The appeal of these probes is that the captured set of proteins is a reflection of the enzyme activity present in a certain sample, rather than the protein abundance that most other methods report on. For this reason the inhibitor-probe method is called activity-based protein profiling (ABBP).[3] However, this cannot be done for proteins that merely bind their target molecules or for enzymes for which no suicide inhibitors exist. For these cases photoreactive groups[4] have been linked to the ligands in order to attach to nearby protein residues. Such probes[3e] can thus capture proteins as a function of their binding activity within a complex sample as was shown for kinases,[5,6] metalloproteases,[7,8] HMG-CoA reductase,[9] and aspartic proteases.[10] An example of a non-catalytic protein group are the carbohydrate binding proteins: the lectins. The galectins, a group of 15 β-galactoside binding lectins, are a particularly interesting subgroup thereof due to their involvement in many health related biological processes.[11] In particular, galectin-3 expression has been correlated, with notable exceptions, with cancer aggressiveness,[12] metastasis,[13] and apoptosis.[14] In this context galectin-3 can be considered an emerging cancer marker.

We previously reported[15] on a photoaffinity probe for the profiling of galectins within samples of low protein complexity and high galectin concentration. We here report on a significant leap in development. Our focus has become the most cancer relevant galectin-3. Our approach allowed the detection of galectin-3 in more complex mixtures of proteins, including carbohydrate binding and processing ones, with detection limits in the low nanogram range per gel lane. Furthermore detection was possible in complex cell lysates. The method aims to provide an alternative to antibody based methods.[11b] In order to arrive at the improvements, multivalency was introduced into the system in order to increase galectin-3 affinity[16] and lower the detection limit.[17] Multivalency is known to be of great importance in the biological mechanism of action of the galectins which often involves cross-linking or aggregation. For galectin-3 this is particularly true as this predominantly monomeric lectin was shown to aggregate into a pentameric form of the lectin in the presence of divalent ligands.[18] A divalent lactoside, similar to the one presented here was shown to increase binding 15-fold in a heamagglutination inhibition study.[16b] The other factor favouring galectin-3 detection is the presence of an aromatic group linked to the 3′ position of the used lactose ligand. Such substituents were shown to greatly enhance the binding to galectin-3, *i.e.* over 200-fold, due to cation–π interactions of the aryl group to a protein arginine residue.[19] The detection principle and the used compounds are shown in Scheme 1 and Chart 1. The scheme shows the non-covalent binding of the probe (**1**) to a partial aggregate of galectin-3 followed by photo-labelling. In the final step a fluorescent reporter molecule (**2**) is introduced to allow visualization of the captured proteins in a gel. Such a two step protocol has advantages such as fewer artefacts due to non-specific effects of bulky labels in the crucial



Scheme 1 Depiction of the capture strategy of galectin-3 in which photocovalent attachment and chemoselective ligation are combined. The incubation of a divalent ligand and galectin-3 leads to aggregate formation and enhanced affinity.

Department of Medicinal Chemistry and Chemical Biology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, P.O. Box 80082, 3508 TB, Utrecht, The Netherlands

† Electronic supplementary information (ESI) available: Details of labeling studies and MALDI MS and ^{1}H and ^{13}C NMR spectra. See DOI: 10.1039/b611050a

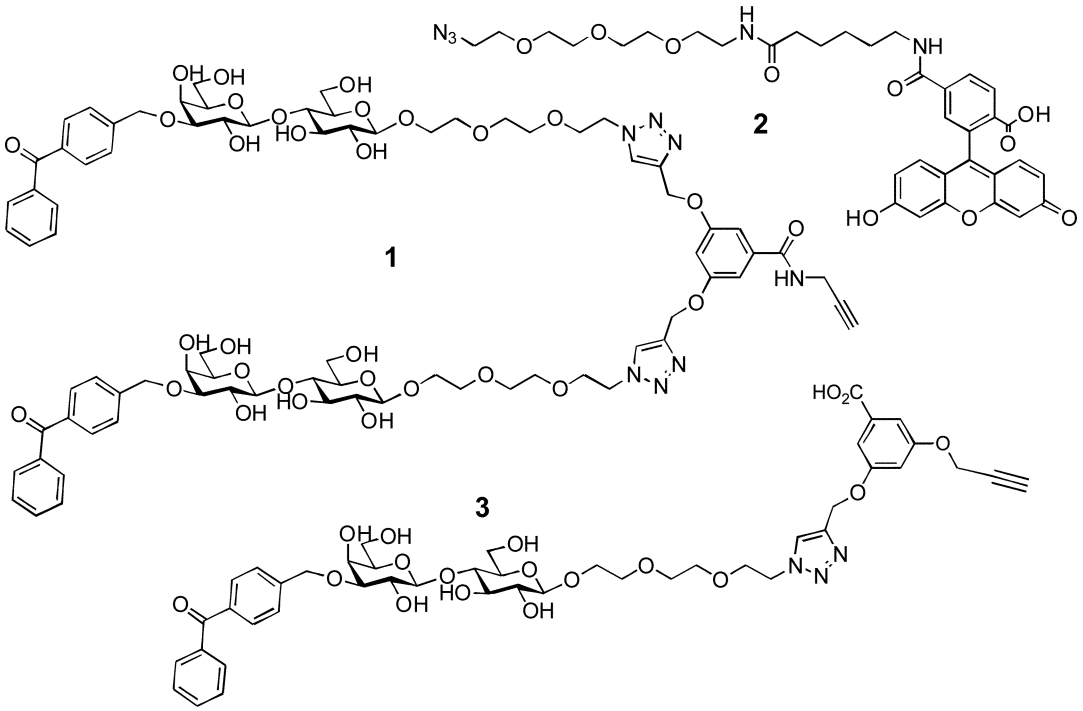
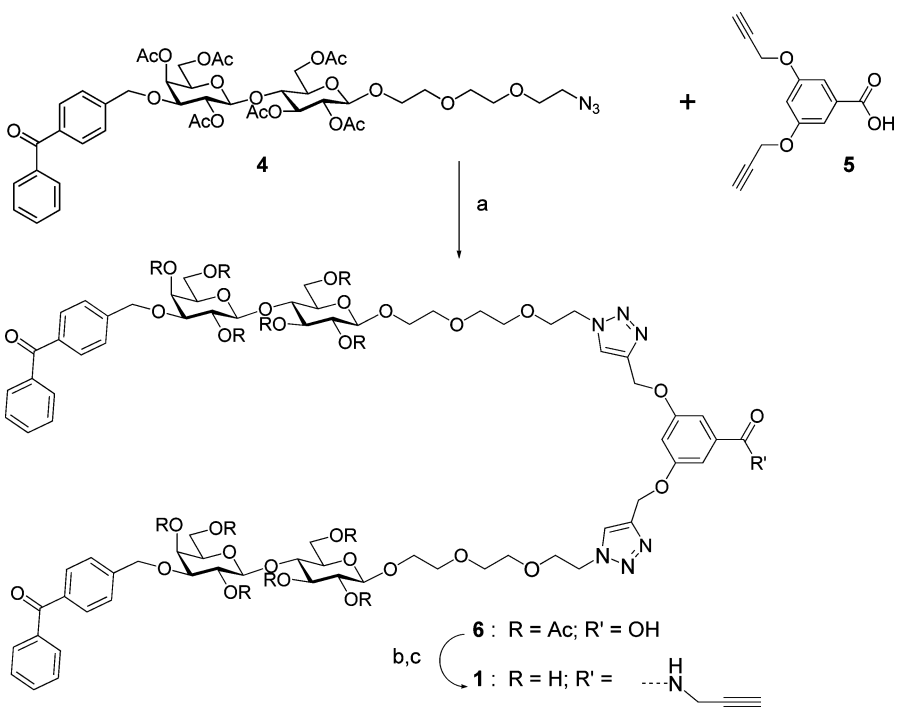


Chart 1 Structures of used probes **1** and **3** and reporter molecule **2**.

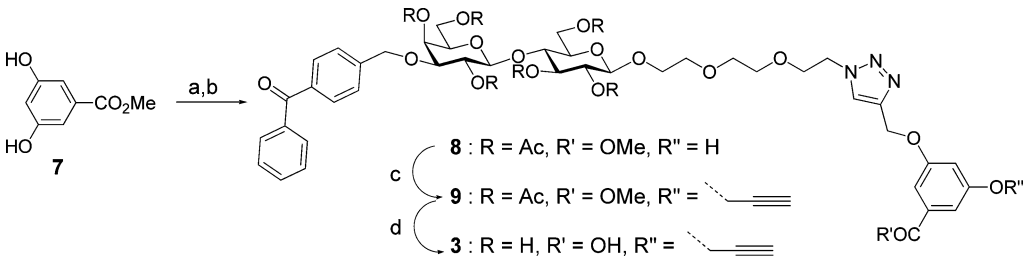
photoreaction step. The alkyne moiety on the probe allows for 'click chemistry'[20] to introduce the fluorescent reporter molecule.

Synthesis

The synthesis of probe **1** started by linking the previous probe **4**[15] to the divalent scaffold **5**[21] using 'click' chemistry (Scheme 2). Subsequently an alkyne group was introduced into the molecule by means of a BOP coupling of propargyl amine to the scaffold carboxylic acid function. Deprotecting the lactose hydroxyls yielded **1**. In order to get an idea about the magnitude of the operative multivalency effect for **1** a monovalent reference probe **3** was also prepared as shown in Scheme 3. This synthesis involved the monofunctionalization of the divalent scaffold **7**, followed by



Scheme 2 *Reagents and conditions*: a) $CuSO_4$, Na ascorbate, DMF–H_2O, MW, 20 min, 84%; b) propargyl amine, BOP, DIPEA, DMF, 14 h, 94%; c) MeONa–MeOH, quant.



Scheme 3 *Reagents and conditions*: a) propargyl bromide (1 equiv.), K_2CO_3, DMF, 16 h, 33%; b) **4**, $CuSO_4$, Na ascorbate, DMF–H_2O, MW, 20 min, 89%; c) propargyl bromide (1 equiv.), K_2CO_3, DMF, 14 h, 61%; d) MeONa–MeOH, quant.

linking of the photoligand **4**. Finally incorporation of the alkyne group and sugar deprotection yielded **3**.

Probe evaluation

To evaluate the probe, protein mixtures were prepared. The first mixture consisted of six common (marker) proteins and the second mixture contained a series of seven carbohydrate binding and processing proteins, including several galactose specific ones (protein mix I and II[22]). To these mixtures various amounts of galectin-3 were added and also various amounts of probe were applied. Following the above protocol, the fluorescent gel images obtained showed a strong and dominant band corresponding to galectin-3, indicating good selectivity of the probe in both protein mixtures (Fig. 1A). In the first mixture some minor labelling of phosphorylase B (lane 5, 94 kDa) was observed. Strikingly, in the second protein mixture the probe exhibited selectivity for galectin-3 over galectin-1, as indicated by the absence of a fluorescent band at 14 kDa in lane 8. A faint band at twice the molecular weight of galectin-3 (≈60 kDa) was seen in a sample only containing galectin-3 (lane 7), indicating the capture of two copies of this protein. In order to confirm that binding of the carbohydrate portion of the probe was a prerequisite for detection, an experiment was performed with increasing amounts of the competing lactose ligand (Fig. 1B). The experiment clearly showed that the signal of fluorescently labelled galectin-3 disappears with the addition of larger quantities of the weak lactose ligand.

The most favourable probe (**1**) concentration was determined by using a concentration range and was found to be around 5 μM (see supporting information†) with a two-fold excess of fluorescein–N_3 construct **2**. Higher concentrations of **2** gave rise to saturated fluorescence gels and thus poorly visible protein signals. With these concentrations of **1** and **2** the minimal detection limit for galectin-3 was found to be 1–5 ng which is clearly sufficient for application.[23] With the related monovalent probe **3** around 50 ng of galectin-3 could be detected, thus confirming the benefits of multivalency.

Once the optimal probe concentration and the lower detection limit were known, we proceeded to apply the methodology to an *E. coli* and a human cell line lysate to explore the probe sensitivity in a biologically complex environment (Fig. 2). The bacterial lysate was spiked with various concentrations of galectin-3. Following the protocol, fluorescent images of the gel clearly show that galectin-3 can be detected under these conditions. In the corresponding experiments with the CaCo2 lysate, which is known to express

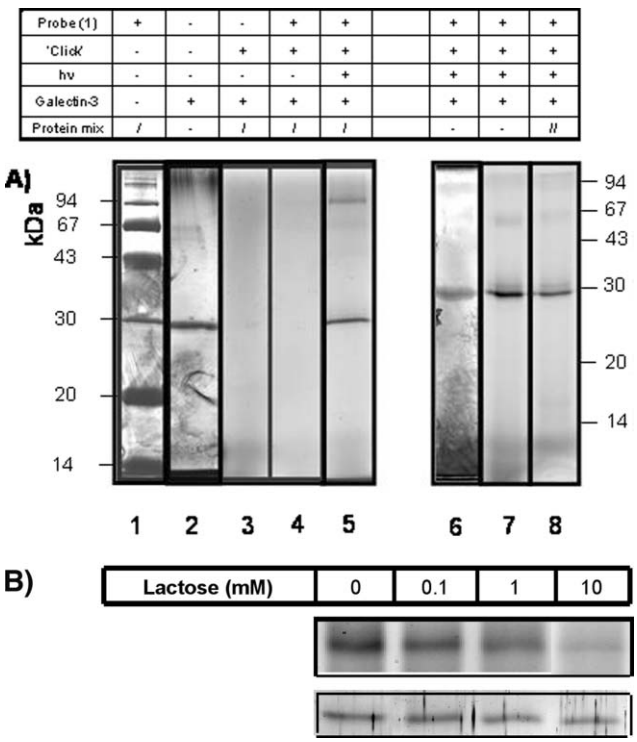


Fig. 1 (A) Evaluation of probe (**1**, 5 μM) in prepared protein mixtures. Lane 1: protein mix I[22] and lane 2: galectin-3, both silver stained. Lanes 3–5: Fluorescence images, galectin-3 (120 ng) in protein mix I (400 ng per protein). Lane 3: no probe, lane 4: no light irradiation. Lanes 6 and 7: galectin-3 (67 ng), silver stained and fluorescence image respectively; lane 8 fluorescence image of galectin-3 (40 ng) with protein mix II[22] (40 ng per protein). (B) Competition experiment between lactose and probe **1**. A sample of 50 ng galectin-3 was incubated with 10 μM divalent probe in the presence of different concentrations of lactose. Top row: fluorescent picture of labeled galectin-3. Bottom row: silver stain of the same gel, as a loading control.

galectin-3,[24] the spiked sample showed a band corresponding to the expected galectin-3 position (lane 9), but also in the non-spiked sample a band is visible (lane 10). In order to confirm the presence of endogenous galectin-3 in the CaCo2 cell line, the presumed galectin-3 band was excised from the gel, digested and analysed by MS. The analysis confirmed the presence of galectin-3. Again experiments were run with increasing amounts of competing lactose (lanes 5–8). At the higher lactose concentration (1–10 mM) the galectin-3 lane fades while non-targeted ligands

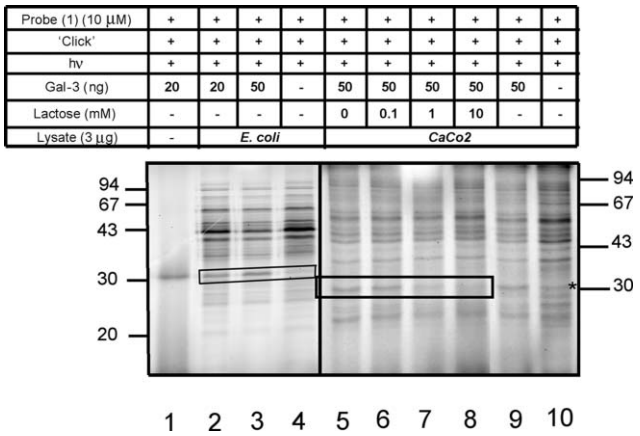


Fig. 2 Lane 1: labeling of 20 ng galectin-3. Lane 2–4: labeling of an *E. coli* lysate with different amounts of galectin-3 and with no galectin-3 added. The box indicates the galectin-3 position. Lane 9–10: fluorescent image of human colon carcinoma lysate, with and without galectin-3 added. The asterisk indicates labeling of galectin-3 present in this lysate, which was confirmed by MALDI analysis. Lane 5–8: Competition experiment between 10 μM divalent probe and increasing concentrations of lactose; 50 ng galectin-3 was added to 3 μg lysate. The box indicates the band of galectin-3, which decreases in intensity with increasing concentrations of lactose.

remained equally visible, clearly indicating that specific binding is responsible for the galectin-3 detection.

Discussion/conclusions

We here present a probe-based system for the detection of the cancer-linked galectin-3, which is considered an emerging cancer marker. The chemical probe based method, as presented here, may ultimately prove to be a valid alternative to the often used antibody based methods as a detection method for active galectin-3. However, the protein–carbohydrate interactions, governing the probe–galectin-3 binding are notoriously weak, illustrating the tremendous challenge to realize this goal. Two features of the probe presented here have allowed a significant step forwards. The first is the multivalency of the system and the second is the use of a C3′ aromatic substituent on the lactose ligand that is known to greatly enhance binding to galectin-3. With probe **1** it proved possible to strongly and selectively identify galectin-3 from a mixture of carbohydrate binding and processing proteins, including galactose binding ones and even the close cousin galectin-1. Furthermore in this study it was possible to visualize galectin-3 in spiked lysates of bacterial and human origin, despite the fact that non-targeted, non-specifically bound and likely highly abundant proteins were also labelled by the probe. Gratifyingly, in an experiment with a non-spiked caco2 cell lysate, which should contain significant amounts of endogenous galectin-3, it proved possible to excise the corresponding fluorescent band and after digestion and mass spectrometric analysis, it was unambiguously identified as galectin-3. This experiment bodes well for further development of the principle. In conclusion, we have shown the utility of a galectin-3 specific divalent photoaffinity probe for the visualization of galectin-3 in biological protein mixtures. This type of probe in combination with various reporter molecules may become a valuable tool in further deciphering the galectin-3 roles in biology and especially cancer and may ultimately aid in its diagnosis and prognosis.

Experimental

General

All reagents were purchased from commercial sources and used without further purification apart from solvents that were distilled before use. Dowex 50 × 8 (H^+ form, 20–50 mesh) from Fluka. Analytical thin layer chromatography (TLC) was performed on Merck pre-coated silica gel 60 F254 (0.25 mm) plates. Spots were visualized with UV light, ninhydrin, or Cl_2-TDM. Column chromatography was carried out using Merck Kieselgel 60 (40–63 mm). ^{1}H NMR and ^{13}C NMR were obtained on a Varian 300 MHz spectrometer. Chemical shifts are given in ppm with respect to internal TMS for ^{1}H NMR. ^{13}C NMR spectra were recorded using the attached proton test (APT) pulse sequence. Two-dimensional ^{1}H–^{1}H correlation and total correlation spectroscopy (COSY and TOCSY) and ^{1}H–^{13}C correlated heteronuclear single quantum coherence (HSQC) NMR spectra (500 MHz) were recorded at 300 K with a Varian Unity INOVA 500 spectrometer.

Low resolution ES-MS experiments were performed on a Shimadzu LCMS QP8000 system. Exact masses were measured by nanoelectrospray time-of-flight mass spectrometry on a Micromass LCToF mass spectrometer at a resolution of 5000 fwhm. Gold-coated capillaries were loaded with 1 μL of sample (concentration 20 μM) dissolved in a 1 : 1 (v/v) mixture of CH_3CN–H_2O with 0.1% formic acid. NaI or poly(ethylene glycol) (PEG) was added as internal standard. The capillary voltage was set between 1100 and 1350 V, and the cone voltage was set at 30 V.

3,5-Bis-{8-(1*H*-1,2,3-triazol-4-yl)-(3,6-dioxa)octyl-[2,4,6-tri-*O*-acetyl-3-*O*-4-methylbenzophenon-β-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-β-D-glucopyranoside]methoxy}benzoic acid (6)

To a solution of **4** (172 mg, 0.182 mmol) and **7** (10.5 mg, 0.046 mmol) in DMF (2 mL), $CuSO_4 \cdot 5H_2O$ (3.4 mg, 0.014 mmol), sodium ascorbate (5.4 mg, 0.028 mmol) and three drops of water were added. The reaction mixture was exposed to microwave irradiation for 20 min at 80 °C. TLC (DCM : MeOH, 9 : 1) and showed at this point total consumption of **5**. The mixture was concentrated and purified on silica gel (DCM : MeOH, 20 : 1) to afford **6** as a slightly yellow viscous oil (82 mg, 84%). Excess **4** was recovered by chromatography. ^{1}H NMR (500 MHz, $CDCl_3$): 1.98, 2.01, 2.03, 2.06, 2.07, 2.13 (6 × s, 6 × 6H), 3.55–3.62 (m, 16H), 3.72 (t, $J = 9.3$, 2H), 3.80–3.89 (m, 8H), 4.06–4.11 (m, 6H), 4.40 (d, $J = 8.3$, 2H), 4.43–4.45 (m, 4H), 4.51 (d, $J = 8.0$, 2H), 4.54 (t, $J = 4.9$, 4H), 4.78 (d, $J = 12.7$, 2H), 4.85 (m, 2H), 5.06 (dd, $J = 8.3$, $J = 9.8$, 2H), 5.16 (t, $J = 9.3$, 2H), 5.20 (s, 4H), 5.49 (d, $J = 2.9$, 2H), 6.85 (s, 1H), 7.31–7.33 (m, 6H), 7.47 (m, 4H), 7.58 (t, 2H, $J = 7.3$), 7.74 (d, $J = 7.8$, 4H), 7.77 (d, $J = 7.3$, 4H), 7.85 (s, 2H); ^{13}C NMR (125 MHz, $CDCl_3$): 20.9, 21.0, 21.1, 50.5, 61.6, 62.4, 65.6, 69.3, 69.6, 70.4, 70.7, 70.8, 70.9, 71.0, 71.2, 71.9, 72.9, 73.1, 76.4, 77.8, 100.8, 101.3, 107.3, 109.1, 124.6, 127.3, 128.5, 130.2, 130.4, 132.7, 137.2, 137.7, 142.3, 143.5, 159.6, 168.0, 169.5, 169.9, 170.1, 170.5, 170.7, 170.8; HR-ESI-MS calculated for $C_{101}H_{120}N_6O_{44}$ 2121.742 $[M + H]^+$ and 2143.724 $[M + Na]^+$, found 2121.730 and 2143.702.

3,5-Bis-{8-(1,2,3-triazol-4-yl)-(3,6-dioxa)octyl-[3-*O*-4-methylbenzophenone)-β-D-galactopyranosyl-(1→4)-β-D-glucopyranoside]methoxy}benzoic acid propargyl amide (1)

Carboxylic acid **6** (97 mg, 0.045 mmol) and propargylamine hydrochloride (7 mg, 0.09 mmol) were dissolved in DMF (2 mL) and the pH was adjusted to approximately 8 using DIPEA. BOP (20 mg, 0.045 mmol) was added and the pH was checked and adjusted again. The mixture was stirred overnight when TLC (DCM : MeOH, 9 : 1) indicated total conversion of the starting material. The mixture was concentrated and loaded onto a silica column for chromatography (DCM : MeOH, 20 : 1). The amide was obtained as a colorless viscous oil (91 mg; 94% yield). Part of this material compound (49 mg, 0.023 mmol) was treated with a 5 M NaOMe solution (0.2 mL) in MeOH (2 mL). The reaction was neutralized with Dowex H^+ after 3 h of stirring at room temperature. The solution was filtered off and concentrated to dryness to give to give **1** in quantitative yield. ^{1}H NMR (500 MHz, MeOD): 2.62 (s, 1H), 3.26 (m, 2H), 3.35 (m, 2H), 3.44 (dd, J = 2.9, J = 9.8, 2H), 3.50 (m, 4H), 3.57–3.64 (m, 16H), 3.67–3.84 (m, 10H), 3.87–3.92 (m, 6H), 4.09 (m, 2H), 4.13 (m, 2H), 4.28 (d, J = 7.8, 2H), 4.36 (d, J = 7.8, 2H), 4.6 (t, J = 4.9), 4.78 (d, J = 12.6, 2H), 4.86 (m, 12OH and 2H), 5.23 (s, 4H), 6.89 (s, 1H), 7.13 (d, J = 1.9, 2H), 7.53 (t, J = 7.8, 4H), 7.62–7.66 (m, 6H), 7.74–7.76 (m, 8H), 8.19 (s, 2H); ^{13}C NMR (125 MHz, CD_3OD): 29.1, 50.4, 60.8, 61.3, 61.6, 65.9, 68.6, 69.1, 70.1, 70.2, 70.6, 70.7, 71.2, 73.5, 75.1, 75.2, 75.7, 79.7, 81.6, 103.1, 103.9, 105.6, 106.8, 125.4, 127.5, 128.4, 129.8, 130.0, 132.7, 136.2, 136.6, 137.7, 143.2, 144.2, 159.8, 167.8; HR-ESI-MS calculated for $C_{80}H_{99}N_7O_{31}$ 1654.646 $[M + H]^+$ and 1676.628 $[M + Na]^+$, found 1654.641 and 1676.621.

Methyl 3-hydroxy-5-{8-(1*H*-1,2,3-triazol-4-yl)-(3,6-dioxa)octyl-[2,4,6-tri-*O*-acetyl-3-*O*-4-methylbenzophenone)-β-D-galacto-pyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-β-D-glucopyranoside]-methoxy}benzoate (8)

To a solution of methyl 2,4-dihydroxybenzoate (1 g, 5.95 mmol) in DMF (20 mL) K_2CO_3 (0.822 g, 5.9 mmol) and propargyl bromide (80% solution in toluene) (0.66 mL, 5.9 mmol) were added and the mixture was stirred for 16 h at rt. The crude product was then filtered through Hyflo SuperCell, concentrated to dryness, dissolved in EtOAc and washed 3 × with water. The organic layer was dried (K_2SO_4) and concentrated again. TLC (hexane : EtOAc, 1 : 1) showed a mixture of starting material and mono- and dialkylated products. Silica gel column chromatography (hexane : EtOAc, 3 : 1) gave monoalkylated product methyl 3-hydroxy-5-(prop-2-ynyloxy)benzoate (0.40 g; 33% yield) as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$): 2.55 (s, 1H), 3.92 (s, 3H), 4.72 (d, 2H, J = 2.4), 5.25 (s, 1H), 6.71 (d, 1H, J = 2.4), 7.2 (s, 1H), 7.24 (d, 1H, J <1); ^{13}C NMR (125 MHz, $CDCl_3$): 52.6, 56.3, 76.2, 78.2, 107.8, 108.4, 110.3, 132.4, 156.9, 158.9, 166.9; HR-ESI-MS calculated for $C_{11}H_{10}O_4$ 207.0657 $[M + H]^+$ and 229.0477 $[M + Na]^+$, found 207.0661 and 229.0482. This compound (15 mg, 0.07 mmol) and **4** (82 mg, 0.09 mmol) were dissolved in DMF (2 mL) and $CuSO_4 \cdot 5H_2O$ (2.7 mg, 0.01 mmol) and sodium ascorbate (4.3 mg, 0.022 mmol) were added together with 3 drops of water. This mixture was exposed to microwave irradiation at 80 °C for 20 min. The mixture was concentrated and purified by silica column chromatography (DCM : MeOH, 20 : 1 to 10 : 1) to afford **8** as a slightly yellow viscous oil (71 mg; 89% yield). ^{1}H NMR (500 MHz, $CDCl_3$): 1.98, 2.02, 2.03, 2.06, 2.07, 2.13 (6 × s, 6 × 3H), 3.53–3.55 (m, 8H), 3.64 (m, 1H), 3.70 (t, 1H, J = 9.3), 3.80–3.88 (m, 7H), 4.05–4.11 (m, 3H), 4.4–4.5 (m, 4H), 4.53 (t, 2H, J = 4.9), 4.77 (d, 1H, J = 12.2), 4.85 (t, 1H, J = 8.8), 5.04 (dd, 1H, J = 8.3, J = 9.3), 5.14 (t, 1H, J = 9.3), 5.2 (s, 2H), 5.49 (d, 1H, J = 2.5), 6.73 (s, 1H), 7.15 (s, 2H), 7.32 (d, 2H, J = 7.8), 7.47 (dd, 2H, J = 7.3, J = 7.8), 7.58 (t, 1H, J = 7.3), 7.72 (br s, 1OH) 7.75 (d, J = 7.8, 2H), 7.77 (d, J = 7.8, 2H), 7.87 (s, 1H); ^{13}C NMR (125 MHz, $CDCl_3$): 20.9, 21.0, 21.1, 50.6, 52.5, 61.6, 62.2, 62.3, 65.6, 69.4, 69.5, 70.4, 70.7, 70.8, 70.9, 71.0, 71.2, 71.8, 72.9, 73.1, 76.4, 77.8, 100.8, 101.4, 107.5, 107.6, 110.4, 124.7, 127.3, 128.6, 130.3, 130.5, 132.4, 132.8, 137.3, 137.7, 142.3, 143.6, 158.0, 159.5, 167.0, 169.7, 170.1, 170.2, 170.6, 170.8, 170.9; HR-ESI-MS calculated for $C_{55}H_{65}N_3O_{24}$ 1152.404 $[M + H]^+$ and 1174.386 $[M + Na]^+$, found 1152.410 and 1174.380.

Methyl 3-(prop-2-ynyl)-5-{8-(1*H*-1,2,3-triazol-4-yl)-(3,6-dioxa)-octyl-[2,4,6-tri-*O*-acetyl-3-*O*-4-methylbenzophenon)-β-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-β-D-glucopyranoside]methoxy}benzoate (9)

To a solution of compound **8** (72 mg, 0.063 mmol) and K_2CO_3 (17 mg, 0.12 mmol) in DMF (1 mL), propargyl bromide (80% solution in toluene, 14 µl, 0.12 mmol) was added. The mixture was stirred overnight. After TLC (DCM : MeOH, 9 : 1) showed completion, the mixture was loaded onto silica gel for column chromatography (DCM : MeOH, 50 : 1 to 20 : 1) to give **9** as a colorless oil (46 mg, 61%). ^{1}H NMR (500 MHz, $CDCl_3$): 2.00, 2.03, 2.04, 2.08, 2.09, 2.10 (6 × s, 6 × 3H), 2.55 (s, 1H), 3.53–3.58 (m, 9H), 3.68 (m, 1H), 3.74 (dd, J = 8.8, J = 9.8, 1H), 3.80–3.82 (m, 1H), 3.87–3.91 (m, 6H), 4.08–4.12 (m, 3H), 4.40 (d, J = 8.3, 1H), 4.40–4.46 (m, 2H), 4.53 (d, J = 8.3, 1H), 4.56 (t, J = 4.9, 2H), 4.72 (d, J = 2.4, 2H), 4.79 (d, J = 12.7, 1H), 4.87 (dd, J = 8.3, J = 9.3, 1H), 5.07 (dd, J = 8.3, J = 9.8, 1H), 5.18 (t, J = 9.7, 1H), 5.23 (s, 2H), 5.50 (d, J = 2.5), 6.84 (t, J = 2.4, 1H), 7.27 (s, 1H), 7.33–7.35 (m, 3H), 7.49 (t, J = 7.3, 2H), 7.60 (t, J = 7.3, 1H), 7.76–7.80 (m, 4H), 7.83 (s, 1H); ^{13}C NMR (125 MHz, $CDCl_3$): 20.9, 21.0, 21.1, 50.6, 52.6, 56.3, 61.6, 62.3, 62.5, 65.5, 69.3, 69.7, 70.5, 70.5, 70.8, 71.0, 71.2, 71.9, 72.9, 73.1, 76.3, 76.5, 77.9, 78.2, 100.8, 101.4, 107.5, 108.8, 109.2, 124.4, 127.3, 128.6, 130.2, 130.5, 132.5, 132.7, 137.3, 137.8, 142.2, 143.5, 158.8, 159.6, 166.7, 169.3, 169.9, 170.0, 170.5, 170.6, 170.7; HR-ESI-MS calculated for $C_{58}H_{67}N_3O_{24}$ 1190.419 $[M + H]^+$ and 1212.401 $[M + Na]^+$, found 1190.416 and 1212.399.

3-(Prop-2-ynyl)-5-bis-{8-(1,2,3-triazol-4-yl)-(3,6-dioxa)octyl-[3-*O*-4-methylbenzophenon)-β-D-galactopyranosyl-(1→4)-β-D-glucopyranoside]methoxy}benzoic acid (3)

Compound **9** (30 mg, 0.025 mmol) was dissolved in a mixture of dioxane (1.4 mL) and MeOH (0.5 mL) and 4 N NaOH (0.1 mL, Tesser's base) and the reaction was stirred for 5 h. The resulting mixture was neutralized with Dowex H^+, filtered and concentrated *in vacuo* to give **3** (21 mg, 90%). ^{1}H NMR (500 MHz, MeOD): 2.98 (t, J = 2.4, 1H), 3.26 (m, 1H), 3.41–3.43 (m, 1H), 3.44 (dd, J = 2.9, J = 9.8, 1H), 3.52–3.54 (m, 2H), 3.57–3.64 (m, 8H), 3.67–3.84 (m, 5H), 3.87–3.92 (m, 3H), 3.95–3.98 (m, 1H), 4.09 (d, J = 2.9, 1H), 4.30 (d, J = 7.8, 1H), 4.38 (d, J = 7.8, 1H), 4.60 (t, J = 4.9, 2H), 4.77 (d, J = 2.44, 2H), 4.77–4.90 (m, 6OH and 2H), 5.23

(s, 2H), 6.86 (s, 1H), 7.27 (s, 1H), 7.3 (s, 1H), 7.54 (t, $J = 7.8$, 2H), 7.63–7.67 (m, 3H), 7.76–7.78 (m, 4H), 8.16 (s, 1H). ^{13}C NMR (125 MHz, CD_3OD): 50.4, 55.8, 60.8, 61.3, 61.6, 65.9, 68.6, 69.1, 70.1, 70.2, 70.6, 70.7, 73.5, 75.1, 75.3, 75.7, 76.0, 78.3, 79.6, 81.6, 103.1, 103.9, 106.5, 108.7, 108.9, 125.3, 127.4, 128.3, 129.8, 129.9, 132.6, 134.2, 136.7, 137.8, 143.3, 144.2, 159.0, 159.5; HR-ESI-MS calculated for $C_{45}H_{53}N_3O_{18}$ 924.340 $[M + H]^+$ and 946.322 $[M + Na]^+$, found 924.341 and 946.323.

Synthesis of fluorescein–azide adduct (2)

6-[Fluorescein-5(6)-carboxamido]hexanoic acid *N*-hydroxysuccinimide ester (Sigma) (20 mg, 0.025 mmol, 75% purity) was dissolved in MeOH (0.6 mL) and 1-amino-11-azido-3,6,9-trioxaundecane (Toronto Research Chemicals) (16 mg, 0.075 mmol) was added to the stirring solution. DIPEA was then added dropwise until the solution reached pH 8–9 and the mixture was stirred overnight. The resulting solution was concentrated to dryness and loaded onto a silica gel for column chromatography (DCM : MeOH, 10 : 1) to give fluorescein-azide adduct (**2**) as a yellow solid (14 mg, 81%). ^{1}H NMR (300 MHz, CD_3OD) 8.45 (d, $J = 2.0$ Hz, 1H), 8.06 (m, 1H), 7.67 (d, $J = 2.0$ Hz, 1H), 7.33 (d, $J = 8.0$ Hz, 1H), 7.04 (m, 2H), 6.57 (m, 2H), 3.08–3.66 (m, 26H). HR-ESI-MS calculated for $C_{35}H_{39}N_5O_{10}$ 690.278, $[M + H]^+$, found 690.279.

Galectin-3 labeling studies in presence of protein mixture I

Solutions of probe **1** (10, 5, 2, 1, and 0.5 μL of a 50 μM stock solution in water : DMSO 20 : 1) were diluted to 35 μL total volume in HEPES buffer pH 7.4, 150 mM NaCl. A 5 μL aliquot of galectin-3 stock solution (human, recombinant form, carrier free, R & D Systems Europe, Abingdon, UK) (125 μg mL^{-1} in HEPES buffer) together with 10 μL of protein mix I (200 μg mL^{-1} of each protein: α-lactalbumin, trypsin inhibitor, carbonic anhydrase, ovalbumin, albumin and phosphorylase B) was added to each one of these samples and the resulting solutions were irradiated for 30 min under a 366 nm UV lamp at 4–5 cm distance at 4 °C. After photoincubation, 5 μL of ligand (100 μM in DMSO : *tert*-butyl alcohol, 1 : 4),[25] 5 μL of $CuSO_4$ (50 mM in water) and a two fold excess with respect to probe quantity of fluorescein–azide conjugate **2** (20, 10, 4, 2, and 1 μL respectively of a 50 μM solution in water) were added and the volume adjusted to 80 μL. The samples were then gently shaken for 1 h at rt and denatured (20 μL of a 10% SDS, 40% glycerol and 2% DTT solution) boiling at 95 °C for 5 min. From the total volume of 100 μL, 20 μL were loaded onto a 15% Tris-HCl gel for SDS-PAGE. After extensive washing of the developed gel (1–2 h in water) to eliminate excess dye reagent, fluorescence was detected with a Typhoon fluorescence scanner and each gel was subsequently silver stained (only the lane with 5 μL probe (5 μM) is shown in Fig. 1 (lane 5), see supporting information for the rest†).

Galectin-3 labeling studies in presence of protein mixture II

Probe **1** (5 μL of a 50 μM stock solution in water : DMSO 20 : 1) was diluted to 35 μL total volume in HEPES buffer pH 7.4, 150 mM NaCl. Decreasing amounts of galectin-3 were added to each of the samples: 5, 3, 2, 1 and 0.5 μL of galectin-3 stock solution (100 μg mL^{-1} in HEPES buffer) together with 3 μL of protein mixture II (100 μg mL^{-1} in each α-glucosidase from yeast (Serva 22828); β-galactosidase from *E. coli* (Boehringer 1510220, 105031); β-glucosidase from almonds (Sigma G4511); galectin-1 (R & D Systems 1152-GA-CF), lectin from *Arachis hypogaea* (peanut), peanut agglutinin, PNA (Sigma L0881); lectin from *Helix aspersa* (garden snail), helix aspersa agglutinin, HAA (Sigma L6635); galactosyl transferase from bovine milk (Fluka 48279) was added to each one of these samples. The resulting solutions were irradiated for 30 min under a 366 nm UV lamp at 4–5 cm distance at 4 °C. After photoincubation, 5 μL of ligand (100 μM in DMSO : *tert*-butyl alcohol, 1 : 4), 5 μL of $CuSO_4$ (50 mM in water) and a two fold excess with respect to probe quantity of fluorescein–azide conjugate **2** (10 μL of a 50 μM solution in water) were added and the volume adjusted to 100 μL. The samples were then gently shaken for 1 h at rt and denatured (50 μL of a 10% SDS, 40% glycerol and 2% DTT solution) boiling at 95 °C for 5 min. From the total volume of 150 μL, 20 μL were loaded onto a 15% Tris-HCl gel for SDS-PAGE. After extensive washing of the developed gel (1–2 h in water) to eliminate excess dye reagent, fluorescence was detected with a Typhoon fluorescence scanner and each gel was subsequently silver stained (in Fig. 1, only lane 8 is shown with 40 ng of galectin-3 *i.e.* 3 μL galectin-3, see supporting information for the rest†).

Lactose competition experiments

Stock solutions of lactose (1, 10, 100 mM) in 50 mM HEPES buffer pH 7.4, 150 mM NaCl were prepared. For a concentration range of 0.1, 1 and 10 mM, 2 μl from the lactose stock solutions were added, together with 1.2 μl galectin-3 stock solution (125 μg ml^{-1} in HEPES buffer), 4 μl probe **1** stock solution (50 μM in water : DMSO 20 : 1) and in the case of the lysate experiment 2.4 μl human cell lysate (4.3 mg mL^{-1} stock as determined by BCA protein concentration assay), to a final sample volume of 20 μL (in HEPES buffer). The samples were irradiated as described above. After photoincubation, 3 μL of fluorescein azide conjugate **2** (100 μM stock in water), 0.8 μL $CuSO_4$ and 0.8 μL TCEP (both 50 mM stock in water) were added and the final sample volume was adjusted to 40 μL. The samples were gently shaken for 1 h at rt and denatured (20 μL of a 10% SDS, 40% glycerol and 2% DTT solution) boiling at 95 °C for 5 min. From the total sample volume of 60 μL, 20 μL was loaded onto a 15% Tris-HCl gel for SDS-PAGE. After extensive washing of the developed gel (1–2 h in water), fluorescence was detected with a Typhoon fluorescence scanner and each gel was subsequently silver stained.

Labelling of galectin-3 spiked in bacterial and human cell extract by probe 1

To 40 μL cell lysate (with a protein concentration of 3 μg mL^{-1}) (or, in lane 1, 40 μL 50 mM HEPES buffer pH 7.4) 0, 1.2 or 3 μL (respectively 0, 20 or 50 ng lane^{-1}) of a 125 μg mL^{-1} stock solution of galectin-3 (in HEPES buffer pH 7.4) was added. An aliquot of 10 μL of probe **1** was added and the total volume was adjusted to 55 μL. The resulting solutions were photoactivated and submitted to conjugation to the fluorescein–azide dye **2** and 20 μL from a total sample volume of 150 μL was loaded onto a gel and visualized as described above.

Table 1 Characteristics of the identified galectin-3, notably the number of peptides used for identification of the protein, the Mascot score and the percentage sequence coverage

Protein ID	Accession NCBI	MW (Da)	pI	# peptides	Mascot score	SQ (%)
Galectin-3	gi|48145911	26057	8.60	9	346	35.3

Bacterial protein extraction

E. coli cells (BL21(DE3) from Novagen, OD600 2.2, 1.5 ml) were pelleted by centrifugation at 5000 rpm for 10 min. The cells were resuspended in 300 µl of B-Per® Reagent (Pierce, no. 78243) and vortexed for 2 min. The homogeneous mixture was centrifuged at 13000 rpm for 5 min. The supernatant was taken for the labeling studies. Protein concentrations of the solutions were determined by BCA protein concentration assay, using bovine serum albumin as a standard.

Human cell protein extraction

Colon cancer cells: 20 mL cell culture (CaCo2 cells, ATCC HTB-37) was pelleted by centrifugation (5 min 1400 rpm), the medium was removed, cells were resuspended in PBS and centrifuged again. The pellet obtained was resuspended in 1 mL lysis buffer (20 mM HEPES pH 7.4, 10 mM KCl, 1 mM $MgCl_2$, 0.5 mM DTT, 0.1% Triton X-100 (w/v), 20% glycerol (w/v), 120 mM NaCl), vortexed for 1 min and centrifuged at 13000 rpm for 5 min. The supernatant was collected and a protein concentration of 0.8 mg mL^{-1} was determined with the use of a Bradford-based protein assay.

Protein identification using mass spectrometry

After fluorescence detection, gels were silver stained and bands were manually excised. The samples were reduced with dithiothreitol (DTT) and treated with iodoacetamide to protect the cysteines. After treatment with iodoacetamide the samples were dialyzed against 50 mM ammonium bicarbonate pH 8.5 and digested with trypsin (Sigma-Aldrich) for 20 h at 37 °C. The digested protein was mixed with 5 mg mL^{-1} α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 0.1% TFA and spotted on the MALDI plate. Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) MS peptide mass fingerprints were acquired on a 4700 proteomics analyzer MALDI-TOF/TOF mass spectrometer (AB 4700 proteomics analyzer, Applied Biosystems). This instrument is equipped with a 200 Hz Nd/YAG laser operating at 355 nm. Experiments were done in a reflectron positive ion mode using delayed extraction. Typically, 2000 shots per spectrum were acquired in the MS mode and 15000 shots per spectrum in the MS/MS mode. Internal calibration was done using trypsin autodigest peaks. The spectra obtained were searched against the National Center for Biotechnology Information databases for protein identity using the Mascot search engine (Table 1).

Acknowledgements

The authors thank Stefan Harmsen and Linda Quarles van Ufford for providing CaCo2 and *E. coli* cells and Ronald van Ooijen for recording high-resolution mass spectra. This work was supported by the European Union in the form of a Marie Curie Individual Fellowship to L.B. under contract MCFI-2002-00249, and a Marie Curie Training Site Fellowship to K.B. under contract QLK2-CT-2000-60047. M.v.S. was supported by the Netherlands Proteomics Centre (NPC).

References

1 R. Aebersold and M. Mann, *Nature*, 2003, **422**, 198–207.
2 (*a*) N. Jessani and B. F. Cravatt, *Curr. Opin. Chem. Biol.*, 2004, **8**, 1–6; (*b*) M. Ichikawa and Y. Ichikawa, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 1769–1773; (*c*) D. J. Vocadlo and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2004, **43**, 5338–5342; (*d*) A. Borodovsky, H. Ovaa, W. J. N. Meester, E. S. Venanzi, M. S. Bogyo, B. G. Hekking, H. L. Ploegh, B. M. Kessler and H. S. Overkleeft, *ChemBioChem*, 2005, **6**, 287–291; (*e*) M. Bogyo, S. Verhelst, V. Bellingard-Dubouchaud, S. Toba and D. Greenbaum, *Chem. Biol.*, 2000, **7**, 27–38; (*f*) D. Greenbaum, K. F. Medzihradszky, A. Burlingame and M. Bogyo, *Chem. Biol.*, 2000, **7**, 569–581; (*g*) T. Nazif and M. Bogyo, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 2967–2972; (*h*) D. Kidd, Y. Liu and B. F. Cravatt, *Biochemistry*, 2001, **40**, 4005–4015; (*i*) G. C. Adam, B. F. Cravatt and E. J. Sorensen, *Chem. Biol.*, 2001, **8**, 81–95; B. M. Kessler, D. Tortorella, M. Altun, A. F. Kisselev, E. Fiebiger, B. G. Hekking, H. L. Ploegh and H. S. Overkleeft, *Chem. Biol.*, 2001, **8**, 913–929; (*j*) D. C. Greenbaum, W. D. Arnold, F. Lu, L. Hayrapetian, A. Baruch, J. Krumrine, S. Toba, K. Chehade and M. Bogyo, *Chem. Biol.*, 2002, **9**, 1085–1094; (*k*) G. C. Adams, E. J. Sorensen and B. F. Cravatt, *Mol. Cell. Proteomics*, 2002, **1**, 828–835; H. Ovaa, P. F. van Swieten, B. M. Kessler, M. A. Leeuwenburgh, E. Fiebiger, A. M. van den Nieuwendijk, P. J. Galardy, G. A. van der Marel, H. L. Ploegh and H. S. Overkleeft, *Angew. Chem., Int. Ed.*, 2003, **11**, 3626–3629; (*l*) A. Speers and B. F. Cravatt, *J. Am. Chem. Soc.*, 2005, **127**, 10018–10019.
3 (*a*) B. F. Cravatt and E. J. Sorensen, E. J., *Curr. Opin. Chem. Biol.*, 2004, **4**, 663–668; (*b*) G. C. Adam, E. J. Sorensen and B. F. Cravatt, *Mol. Cell. Proteomics*, 2002, **1**, 781–790; (*c*) N. Jessani and B. F. Cravatt, *Curr. Opin. Chem. Biol.*, 2004, **8**, 54–59; (*d*) S. H. L. Verhelst and M. Bogyo, *QSAR Comb. Sci.*, 2005, **24**, 261–269; (*e*) M. C. Hagenstein and N. Sewald, *J. Biotechnol.*, 2006, **124**, 56–73; (*f*) N. L. Pohl, *Curr. Opin. Chem. Biol.*, 2005, **9**, 76–81.
4 G. Dormán and G. D. Prestwich, *Trends Biotechnol.*, 2000, **18**, 64–77.
5 M. C. Hagenstein, J. H. Mussgnug, K. Lotte, R. Plessow, A. Brockhinke, O. Kruse and N. Sewald, N., *Angew. Chem., Int. Ed.*, 2003, **42**, 5635–5638.
6 S.-Y. Han, S. H. Choi, M. H. Kim, W. G. Lee, S. H. Kim, Y. K. Min and B. T. Kim, *Tetrahedron Lett.*, 2006, **47**, 2915–2919.
7 E. W. S. Chan, S. Chattopadhaya, R. C. Panicker, X. Huang and S. Q. Yao, *J. Am. Chem. Soc.*, 2004, **126**, 14435–14446.
8 A. Saghatelian, N. Jessani, A. Joseph, M. Humphrey and B. F. Cravatt, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10000–10005.
9 T. Hosoya, T. Hiramatsu, T. Ikemoto, M. Nakanishi, H. Aoyama, A. Hosoya, T. Iwata, K. Muruyama, M. Endo and M. Suzuki, *Org. Biomol. Chem.*, 2004, **2**, 637–641.
10 S. Chattopadhaya, E. W. S. Chan and S. Q. Yao, *Tetrahedron Lett.*, 2005, **46**, 4053–4056.
11 (*a*) H. Leffler, S. Carlsson, M. Hedlund, Y. Qian and F. Poirier, *Glycoconjugate J.*, 2004, **19**, 433–440; (*b*) R. J. Pieters, *ChemBioChem*, 2006, **7**, 721–728.
12 S. Califice, V. Castronovo and F. Van Den Brûle, *Int. J. Oncol.*, 2004, **25**, 983–992.
13 Y. Takenaka, T. Fukumori and A. Raz, *Glycoconjugate J.*, 2004, **19**, 543–549.
14 (*a*) S. Nakahara, N. Oka and A. Raz, *Apoptosis*, 2005, **10**, 267–275; (*b*) F.-T. Liu and G. A. Rabinovich, *Nat. Rev. Cancer*, 2005, **5**, 29–41.

15 L. Ballell, K. J. Alink, M. Slijper, C. Versluis, R. M. J. Liskamp and R. J. Pieters, *ChemBioChem*, 2005, **6**, 291–295.
16 (*a*) I. Vrasidas, S. André, P. Valentini, C. Böck, M. Lensch, H. Kaltner, R. M. J. Liskamp, H. J. Gabius and R. J. Pieters, *Org. Biomol. Chem.*, 2003, **1**, 803–810; (*b*) S. André, R. J. Pieters, I. Vrasidas, H. Kaltner, I. Kuwabara, F.-T. Liu, R. M. J. Liskamp and H. J. Gabius, *ChemBioChem*, 2001, **2**, 822–830.
17 (*a*) M. Lee, D.-W. Jung, D. Williams and I. Shin, *Org. Lett.*, 2005, **7**, 5477–5480; (*b*) G. Lauc, R. T. Lee, J. Dumi and Y. C. Lee, *Glycobiology*, 2000, **10**, 357–364.
18 N. Ahmad, H.-J. Gabius, H. Kaltner, S. Sabesan, R. Roy, B. Liu, F. Macaluso and C. F. Brewer, *J. Biol. Chem.*, 2004, **279**, 10841–10847.
19 (*a*) P. Sörme, Y. Qian, P. G. Nyholm, H. Leffler and U. J. Nilsson, *ChemBioChem*, 2002, **3**, 183–189; (*b*) P. Sörme, P. Arnoux, B. Kahl-Knutsson, H. Leffler, J. M. Rini and U. J. Nilsson, *J. Am. Chem. Soc.*, 2005, **127**, 1737–1743; (*c*) B. Salameh, H. Leffler and U. J. Nilsson, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3344–3346.
20 A. E. Speers, G. C. Adam and B. F. Cravatt, *J. Am. Chem. Soc.*, 2003, **125**, 4686–4687.
21 J. A. F. Joosten, N. T. H. Tholen, F. Ait El Maate, A. J. Brouwer, G. W. van Esse, D. T. S. Rijkers, R. M. J. Liskamp and R. J. Pieters, *Eur. J. Org. Chem.*, 2005, 3182–3185.
22 Protein mix I: α-lactalbumin, trypsin inhibitor, carbonic anhydrase, ovalbumin, albumin and phosphorylase B. Protein mix II: α-glucosidase, β-galactosidase, β-glucosidase, galectin-1, peanut agglutinin, *Helix aspersa* agglutinin, galactosyl transferase.
23 Considering that in sera of patients with metastatic gastrointenstinal carcinoma 320 ng ml^{-1} was found to be the median concentration of galectin-3. See: I. Iurisci, N. Tinari, C. Natoli, D. Angelucci, E. Cianchetti and S. Iacobelli, *Clin. Cancer Res.*, 2000, **6**, 1389–1393.
24 K. Lenaerts, E. Mariman, F. Bouwman and J. Renes, *FEBS J.*, 2005, **272**, 3350–3364.
25 Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless and M. G. Finn, *J. Am. Chem. Soc.*, 2003, **125**, 3192–3193.

