Pergamon

Bioorganic & Medicinal Chemistry 8 (2000) 1527–1535

BIOORGANIC & MEDICINAL CHEMISTRY

Controlled Site-Selective Protein Glycosylation for Precise Glycan Structure–Catalytic Activity Relationships

Benjamin G. Davis,[a,*] Richard C. Lloyd[b] and J. Bryan Jones[b,†]

[a]*Department of Chemistry, University of Durham, South Road, Durham, DH1 3LE, UK*
[b]*Department of Chemistry, University of Toronto, 80 St. George Street, Toronto, ON, Canada M5S 3H6*

Received 18 October 1999; accepted 3 February 2000

Abstract—Glycoproteins occur naturally as complex mixtures of differently glycosylated forms which are difficult to separate. To explore their individual properties, there is a need for homogeneous sources of carbohydrate–protein conjugates and this has recently prompted us to develop a novel method for the site-selective glycosylation of proteins. The potential of the method was illustrated by site-selective glycosylations of subtilisin *Bacillus lentus* (SBL) as a model protein. A representative library of mono- and disaccharide MTS reagents were synthesized from their parent carbohydrates and used to modify cysteine mutants of SBL at positions 62 in the S_2 site, 156 and 166 in the S_1 site and 217 in the S_1' site. These were the first examples of preparations of homogeneous neoglycoproteins in which both the site of glycosylation and structure of the introduced glycan were predetermined. The scope of this versatile method was expanded further through the combined use of peracetylated MTS reagents and careful pH adjustment to introduce glycans containing different numbers of acetate groups. This method provides a highly controlled and versatile route that is virtually unlimited in the scope of the sites and glycans that may be conjugated, and opens up hitherto inaccessible opportunities for the systematic determination of the properties of glycosylated proteins. This potential has been clearly demonstrated by the determination of detailed glycan structure–hydrolytic activity relationships for SBL. The 48 glycosylated CMMs formed display k_{cat}/K_M values that range from 1.1-fold higher than WT to 7-fold lower than WT. The anomeric stereochemistry of the glycans introduced modulates changes in k_{cat}/K_M upon acetylation. At positions 62 and 217 acetylation enhances the activity of α-glycosylated CMMs but decreases that of β-glycosylated. This trend is reversed at position 166 where, in contrast, acetylation enhances the k_{cat}/K_Ms of β-glycosylated CMMs but decreases those of α-glycosylated. Consistent with its surface exposed nature changes at position 156 are more modest, but still allow control of activity, particularly through glycosylation with disaccharide lactose.

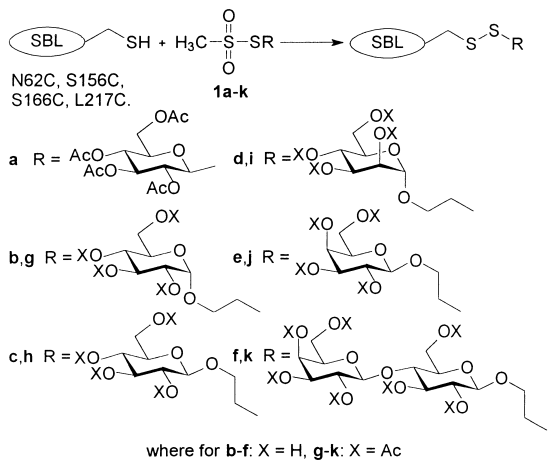
Introduction

In addition to their critical role in communication events,[1–5] glycosylated proteins display altered catalytic properties.[6–13] For example, carbohydrate–protease conjugates show greater stabilities at high temperatures[14] and in organic solvents[15,16] and catalyze high yielding peptide syntheses.[15,17] In addition, the functionality found in carbohydrates may be used to introduce novel environments that selectively recognize certain functional groups. For example, polyalkylated cyclodextrins have been used as ionophores for the selective complexation of ammonium salts of biologically important bases such as dopamine and imipramine.[18,19] In spite of these positive indications, a systematic study of the effects of protein glycosylation has not been made. Until recently this has been hampered by the lack of suitably glycan- or regio-specific methods for the glycosylation of proteins. This need for sources of pure, precisely-controlled carbohydrate–protein conjugates recently prompted us to develop a novel method for the site-selective glycosylation of proteins, which would allow us to address our goal of controlling the specificity and activity of enzymes through glycosylation.[20]

Site-directed mutagenesis combined with chemical modification has permitted us to realize this goal and provides a general method that allows both regio- and glycan-selective glycosylation of proteins. This method is rapid, utilizes reagents that may be prepared in a facile manner and is unlimited in the scope of sites and glycans that may be conjugated. The strategy involves the introduction of cysteine at preselected positions and then reaction of its thiol residue with glycomethanethiosulfonate reagents (Scheme 1). Methanethiosulfonate (MTS) reagents react specifically and quantitatively with

*Corresponding author. Tel.: +44-191-374-3136; fax: +44-191-384-4737; e-mail: ben.davis@durham.ac.uk
†Second corresponding author. Tel.: +1-416-978-3589; fax: +1-416-978-1553; e-mail: jbjones@alchemy.chem.utoronto.ca

0968-0896/00/$ - see front matter
PII: S0968-0896(00)00083-3



Scheme 1.

thiols[21,22] and allow the controlled formation of neutral disulfide linkages. Recently, we have successfully used the representative serine protease subtilisin *Bacillus lentus* (SBL) as our vehicle and continue to do so here.[23–29] SBL is an ideal model protein for evaluating the validity of this strategy as it does not contain a natural cysteine and is not naturally glycosylated. This site-selective glycosylation procedure is also ideal for evaluating glycosylated enzymes as new catalysts[6] since it is powerful enough to introduce internally-oriented glycans within the active site and it yields pure glycoprotein products.

Four SBL sites at different locations and of different characteristics were selected for mutation to cysteine. S156 of the S_1[30] pocket is a surface-exposed residue that permits the introduction of externally-disposed glycans mirroring those found naturally in glycoproteins.[31] In contrast, N62 in the S_2 pocket, S166 in the S_1 pocket, and L217 in the S_1' pocket have side chains which are internally oriented. These tested the applicability of the method for introducing sugars at hindered locations and allowed us to investigate the influence of glycosylation upon catalysis.[6] Broad applicability with respect to the sugar moiety was evaluated by using the representative library of protected and deprotected, mono- and disaccharide methanethiosulfonates **1a**–**k**. Two types of glycosylating reagents, the untethered methanethiosulfonate **1a** and the ethyl-tethered methanethiosulfonates **1b**–**k**, were prepared, as described previously,[20] from their parent carbohydrates in good to excellent yields. The preparation of these reagents in fully protected **1a**,**g**–**k** and deprotected **1b**–**f** forms also allowed the effects of increased steric bulk and hydrophobicity to be assessed.

Results and Discussion

Site-specific glycosylation

The glyco-MTS reagents **1a**–**k** were reacted with the chosen cysteine mutants SBL-N62C, -S156C, -S166C and -L217C in aqueous buffer. These reactions were rapid and quantitative, as judged by monitoring of changes in specific activity and by titration of free thiols with Ellman's reagent.[32] The glycosylated chemically modified mutants (CMMs) were purified by size-exclusion chromatography and dialysis, and their structures (Table 1) were confirmed by rigorous ES–MS analyses (± 7 Da). The CMMs each appeared as a single band on non-denaturing gradient PAGE, thereby establishing their high purities. The active enzyme concentration of the resulting CMM solutions was determined by active site titration with α-toluenesulfonyl fluoride (PMSF) using a fluoride ion-sensitive electrode.[33] In all cases, modification with the fully deprotected reagents **1b**–**f** led to site-specific glycosylations and the formation of single glycoforms. Furthermore, modification with the protected MTS reagents **1a**,**g**–**k** gave products with controllable levels of acetylation. Through adjustment of pH and appropriate selection of the glycosylation site, we were able to prepare differently acetylated glycoforms of SBL. This ability to modulate the level of acetylation through pH-control vastly expands the structural variety of glyco-CMMs that can be conveniently accessed and we probed its scope through the reaction of **1a** with SBL-N62C, -S156C, -S166C and -L217C (Scheme 2).

The extent of deacetylation during modification is highly site-dependent. Modification of L217C with reagent **1a** at pH 9.5 was accompanied by complete in situ deacetylation, and the sole product was the fully deprotected glucosylated-SBL, L217C-S-β-Glc. In contrast, treatment of N62C, S156C and S166C with **1a** at pH 9.5 yielded only fully acetylated products, N62C-S-**a**, S156C-S-**a** and S166C-S-**a**, respectively. To examine the effects of pH upon deacetylation, the reaction of L217C with **1a** was chosen. At pH 7.5 and 5.5 the products retained two and three acetate groups, forming L217C-S-β-Glc$(Ac)_2$ and L217C-S-β-Glc$(Ac)_3$, respectively. In all cases, complete integrity of the site selectivity was retained.

We attribute this valuable site-dependent deacetylation to a novel intramolecular SBL-catalyzed process. Although acetate esters are moderately chemically labile in aqueous solution at pH 9.5 they are not at either pH 7.5 or 5.5.[34] The striking differences in behavior during modification between L217C and the three other mutants N62C, S156C and S166C under identical reaction conditions discounts the possibility of deacetylation prior to modification. In addition, it should be noted that position 217 bears an internally-oriented side chain and that modification of surface exposed position 156 showed no sign of deacetylation. This observation discounts both the possibility of either in situ chemical deacetylation or intermolecular enzymatic deacetylation. Furthermore, this ability of SBL to intramolecularly deacetylate was confirmed by the reaction of L217C-S-β-Glc$(Ac)_3$ at pH 9.5. Incubation of L217C-S-β-Glc$(Ac)_3$ under standard modification reaction conditions, but without reagent **1a**, gave L217C-S-β-Glc as the sole product (Scheme 2).

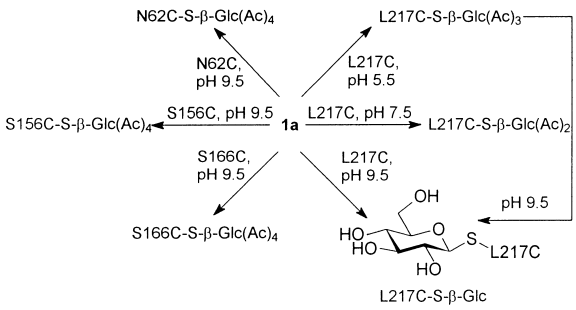
The enormous potential of this method was demonstrated by the preparation of a small library of differently acetylated glycosylated CMMs through the reaction of

Table 1. Properties and kinetic parameters[a] of modified enzymes

Entry	Reactant enzyme	Pocket	MTS reagent	Reaction pH	Product(s)	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} mM^{-1})
1	SBL-WT	—	—	—	—	153±4	0.73±0.05	209±15
2	N62C	S_2	—	—	—	174±9	1.90±0.20	92±11
3			**1a**	9.5	N62C-S-**a**[b]	67.9±3.5	0.52±0.07	130.6±18.8
4			**1b**	6.5	N62C-S-**b**[b]	135.3±3.5	0.94±0.05	143.9±8.5
5			**1c**	6.5	N62C-S-**c**[b]	132.7±4.0	1.25±0.08	106.2±7.5
6			**1d**	6.5	N62C-S-**d**[b]	132.9±3.1	1.04±0.05	127.8±6.8
7			**1e**	6.5	N62C-S-**e**[b]	119.3±3.6	0.99±0.07	120.5±9.3
8			**1f**	6.5	N62C-S-**f**[b]	129.8±2.4	1.04±0.04	124.8±5.3
9			**1g**	5.5	N62C-S-**g**[b]	120.0±2.7	0.52±0.03	230.8±14.3
10			**1h**	6.5	N62C-S-Et-β-Glc$(Ac)_2$[c]	87.7±4.2	1.63± 0.15	53.8±5.8
11			**1h**	5.5	N62C-S-Et-β-Glc$(Ac)_3$[c]	100.3±3.5	1.86±0.12	53.9±4.0
12			**1i**	5.5	N62C-S-**i**[b]	123.0±1.6	1.05±0.03	117.1±3.7
13			**1j**	5.5	N62C-S-Et-β-Gal$(Ac)_3$[c]	103.4±4.3	2.36±0.17	43.8±3.6
14			**1k**	5.5	N62C-S-**k**[b]	64.9±1.5	0.88±0.05	73.8±4.5
15	L217C	S_1'	—	—	—	41±1	0.80±0.04	51±3
16			**1a**	9.5	L217C-S-β-Glc[b]	27.7±0.4	0.79±0.03	35.1±1.4
17			**1a**	7.5	L217C-S-β-Glc$(Ac)_2$[c]	44.9±2.0	0.44±0.06	102.0±14.6
18			**1a**	5.5	L217C-S-β-Glc$(Ac)_3$[c]	36.3±0.8	0.36±0.03	100.8±8.7
19			**1b**	6.5	L217C-S-**b**[b]	57.8±0.6	0.67±0.02	86.3±2.7
20			**1c**	6.5	L217C-S-**c**[b]	50.6±0.9	0.67±0.03	75.5±3.6
21			**1d**	6.5	L217C-S-**d**[b]	62.0±1.3	0.55±0.03	112.7±6.6
22			**1e**	6.5	L217C-S-**e**[b]	46.2±0.8	0.63±0.08	73.3±3.7
23			**1f**	6.5	L217C-S-**f**[b]	30.4±0.6	0.46±0.03	66.1±4.5
24			**1g**	5.5	L217C-S-Et-α-Glc$(Ac)_2$[c]	72.7±3.1	0.73±0.08	99.6±11.8
25			**1h**	5.5	L217C-S-Et-β-Glc$(Ac)_3$[c]	29.4±0.8	0.93±0.06	31.6±2.2
26			**1i**	5.5	L217C-S-Et-α-Man$(Ac)_3$[c]	97.8±2.4	0.59±0.04	165.8±12.0
27			**1j**	5.5	L217C-S-Et-β-Gal$(Ac)_3$[c]	39.2±0.8	1.17±0.05	33.5±1.6
28			**1k**	5.5	L217C-S-Et-Lac$(Ac)_6$[c]	27.1±0.6	0.69±0.04	39.3±2.4
29	S156C	S_1	—	—	—	125±4	0.85±0.06	147±11
30			**1a**	9.5	S156C-S-**a**[b]	54.8±1.3	0.70±0.04	78.3±4.8
31			**1b**	6.5	S156C-S-**b**[b]	77.0±1.2	0.84±0.03	91.7±3.6
32			**1c**	6.5	S156C-S-**c**[b]	76.6±1.7	0.73±0.04	104.9±6.2
33			**1d**	6.5	S156C-S-**d**[b]	88.6±2.8	0.79±0.06	112.2±9.2
34			**1e**	6.5	S156C-S-**e**[b]	78.9±1.9	0.89±0.04	89.7±4.4
35			**1f**	6.5	S156C-S-**f**[b]	63.6±1.4	0.89±0.05	71.8±4.3
36			**1g**	5.5	S156C-S-**g**[b]	43.6±0.8	0.78±0.04	55.9±3.0
37			**1h**	5.5	S156C-S-**h**[b]	64.0±1.3	0.72±0.04	88.9±5.2
38			**1i**	5.5	S156C-S-**i**[b]	60.3±0.9	0.71±0.03	84.9±3.8
39			**1j**	5.5	S156C-S-**j**[b]	51.9±0.6	0.61±0.02	85.1±3.0
40			**1k**	5.5	S156C-S-**k**[b]	53.6±0.8	0.79±0.03	67.4±2.8
41	S166C	S_1	—	—	—	42±1	0.50±0.05	84±9
42			**1a**	9.5	S166C-S-**a**[b]	33.8±1.3	0.66±0.06	51.2±5.0
43			**1b**	6.5	S166C-S-**b**[b]	81.9±1.1	1.14±0.03	71.8±2.1
44			**1c**	6.5	S166C-S-**c**[b]	67.0±2.2	0.99±0.07	67.6±5.3
45			**1d**	6.5	S166C-S-**d**[b]	76.5±2.0	1.17±0.07	65.4±4.3
46			**1e**	6.5	S166C-S-**e**[b]	62.2±1.4	1.08±0.05	57.6±3.0
47			**1f**	6.5	S166C-S-**f**[b]	58.2±1.2	1.02±0.04	57.1± 2.5
48			**1g**	5.5	S166C-S-Et-α-Glc$(Ac)_3$[c]	31.0±0.8	0.77±0.05	40.3±2.8
49			**1h**	6.5	S166C-S-Et-β-Glc$(Ac)_2$[c] S166C-S-Et-β-Glc$(Ac)_3$[d]	95.0±2.1	0.87±0.05	109.2±6.7
50			**1h**	5.5	S166C-S-Et-β-Glc$(Ac)_2$[e] S166C-S-**h**[e]	72.9±1.7	0.65±0.04	112.2±7.4
51			**1i**	5.5	S166C-S-Et-α-Man$(Ac)_3$[c]	67.7±1.9	1.64±0.09	41.3±2.5
52			**1j**	5.5	S166C-S-Et-β-Gal$(Ac)_3$[d]	65.1±0.9	0.80±0.03	81.3±3.3
53			**1k**	5.5	S166C-S-Et-Lac$(Ac)_5$[d]	67.4±1.6	1.65±0.07	40.8±2.0

[a]Michaelis–Menten constants were measured at 25 °C according to the initial rates method in 0.1 M Tris–HCl buffer at pH 8.6, 0.005% Tween 80, 1% DMSO, suc-AAPF-*p*NA as the substrate.
[b]Single species.
[c]Single product mass by ES–MS.
[d]Major component.
[e]Minor component.

SBL-N62C, -S156C, -S166C and -L217C with MTS reagents **1g**–**k**. Using the pH-activity profiles of WT and CMMs of SBL[35] as a guide, pH 5.5 and 6.5 were chosen to minimize deacetylation. Typically the specific activity of SBL and its CMMs drops sharply below pH 7.5 to levels that at pH 5.5 are 5–20% those at optimal pH (8.5–9.5).[35] As expected, this drop in hydrolytic activity was reflected in the products of these modifications with **1g**–**k**, which in all cases retained two or more acetate groups.



Scheme 2.

For example, at pH 5.5 all reactions of SBL-L217C and -S166C created singly deacetylated CMMs, with the exception of the pure dideacetylated glyco-CMMs L217C-S-Et-α-Glc(Ac)$_2$ and S166C-S-Lac(Ac)$_5$. The formation of the latter may reflect the presence of two primary acetates in disaccharidic MTS reagent **1k**. Primary acetate groups are typically more labile than secondary acetate groups under conditions of intermolecular enzymatic deacetylation.[36] In contrast, all reactions of SBL-S156C gave only the fully acetylated CMMs, S156C-S-**g**–**k**. This uniform lack of deacetylation observed for surface exposed glycans at position 156 is consistent with an intramolecular enzyme-catalyzed mechanism requiring internally-oriented acetate groups. Interestingly, the deacetylation reactions of SBL-N62C were also determined by the anomeric configuration of **1g**–**k**: α-MTS reagents **1g**,**i** gave products that retained all acetate groups while β-MTS reagents **1h**,**j**,**k** were monodeacetylated. The range of accessible acetylated glyco-CMMs was further extended through selected modifications at pH 6.5. For example, at position 62 it allowed the introduction of diacetylated β-glucose, forming N62C-S-Et-β-Glc(Ac)$_2$ in place of the triacetylated N62C-S-Et-β-Glc(Ac)$_3$ formed at pH 5.5. The range of acetylation at position 166 was similarly expanded through, for example, the formation of S166C-S-Et-β-Glc(Ac)$_2$ at pH 6.5 in place of S166C-S-Et-β-Glc(Ac)$_3$ at pH 5.5.

Glycan structure–hydrolytic activity relationships

The effects of glycosylation upon SBL were assessed by the determination of k_{cat} and K_M for the hydrolysis of succinyl-AAPF-*p*-nitroanilide (Suc-AAPF-*p*NA) at pH 8.6. The kinetic parameters of the 48 CMMs generated are compared with those of WT and unmodified mutants in Table 1. The excellently selective and controlled method shown in Scheme 1 allowed the introduction of structurally related monosaccharides, D-glucose, D-galactose and D-mannose, in addition to the more sterically bulky disaccharide lactose. From the resulting glycosylated CMMs a detailed and precise set of structure–activity relationships (SARs) was generated (Figs 1–3).

At position 62, in the S_2 pocket, the 2.3-fold reduction in k_{cat}/K_M caused by mutation to cysteine is partially restored by glycosylation (Fig. 1(a)). The introduction of deprotected ethyl-tethered α- or β-glucose, β-galactose

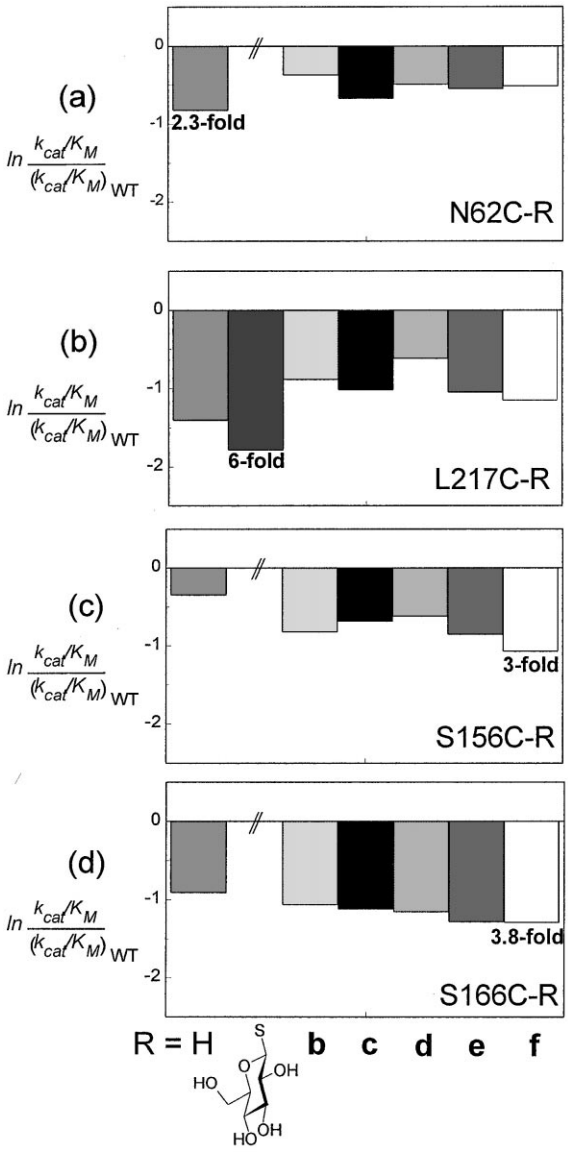


Figure 1. Deprotected glycan structure–proteolytic activity SARs of SBL cysteine mutants and glycosylated CMMs relative to WT. A break in the axis indicates that the value was not determined. (a) At position 62, glycosylation partially restores the decrease in k_{cat}/K_M caused by mutation to cysteine (R = H). (b) At position 217, the effects of mutation and modification are more dramatic. The 4-fold decrease in activity caused by mutation (R = H) is amplified to 6-fold lower than WT by glycosylation with untethered S-β-Glc, but reduced to around 2.5-fold lower than WT by glycosylation with ethyl-tethered glycans (**b**–**f**). (c) At position 156, an arced variation in activity reaches a 3-fold lower than WT minimum k_{cat}/K_M at bulky lacto-CMM S156C-S-**f**. (d) At position 166, the 2.5-fold decrease in k_{cat}/K_M caused by mutation is amplified by glycosylation. k_{cat}/K_M decreases monotonically from S166C (R = H) to a value that is 3.8-fold lower than WT for S166C-S-**f**.

or α-mannose to N62C increased k_{cat}/K_M and formed N62C-S-**b**–**e** with k_{cat}/K_Ms 1.5- to 2-fold lower than WT. Despite its steric bulk and high hydrophilicity disaccharide-containing lacto-CMM N62C-S-**f** also shows higher activity than N62C with k_{cat}/K_M only 1.7-fold lower than WT.

The effects of the mutation of position 217 in the S_1' pocket are intrinsically more dramatic as indicated by a

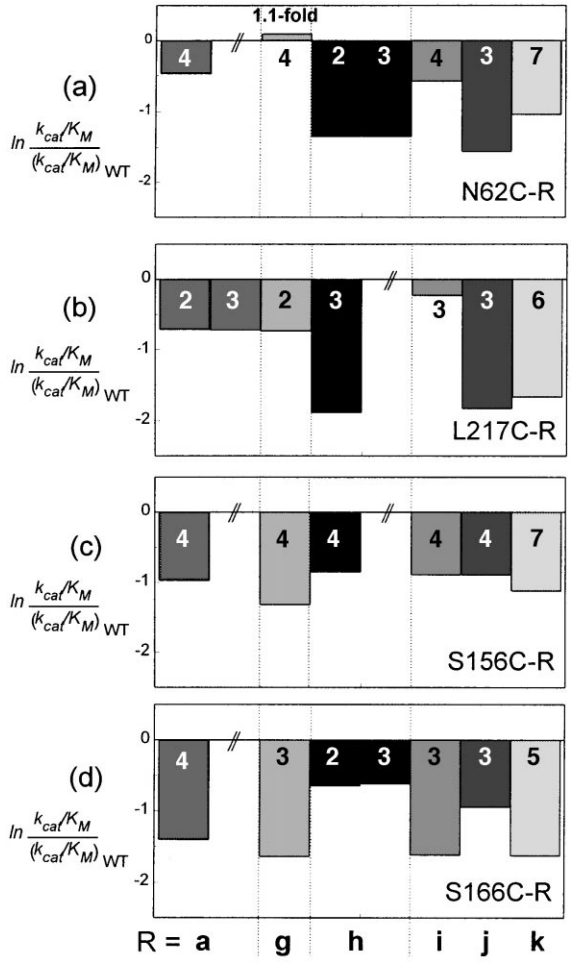


Figure 2. Acetylated glycan structure–proteolytic activity SARs of glycosylated CMMs relative to WT: the broad range of acetylation that is accessible expands the relationships that may be probed. For each glycan the number of acetate groups present is indicated by a label on the corresponding bar. A break in the axis indicates that the value was not determined. At positions 62 (a), 217 (b) and 166 (d) an interesting alternating trend in activity is observed as a result of the opposite effects of acetylation upon k_{cat}/K_M according to anomeric stereochemistry (see Fig. 3). Gratifyingly, this results in a k_{cat}/K_M for N62C-S-**g** that is 1.1-fold higher than WT. At position 156 (c) variations are slight and this is consistent with its surface exposed orientation.

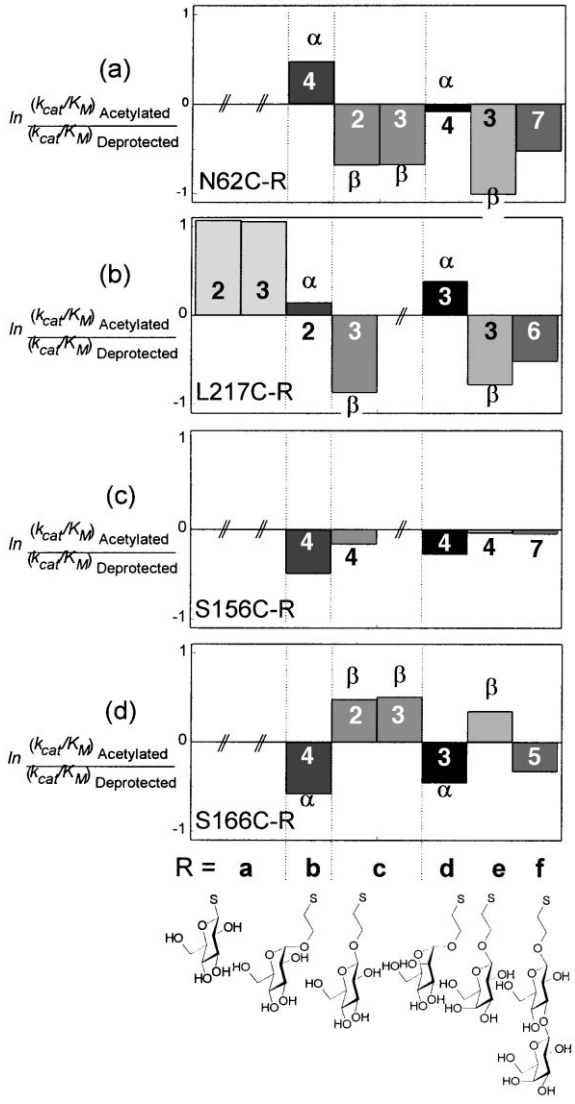


Figure 3. Variation in proteolytic activity of glycosylated CMMs of SBL upon acetylation of glycans. Comparison of the activity of acetylated with fully deprotected CMMs shows that, at positions 62 (a) and 217 (b), acetylation enhances the activity of α-tethered CMMs but decreases that of β-tethered. In contrast, at position 166 (d) acetylation decreases the activity of α-tethered CMMs but increases that of β-tethered. Consistent with its surface exposed orientation changes at position 156 (c) are modest. For each glycan the number of acetate groups present is indicated by a label on the corresponding bar. A break in the axis indicates that the value was not determined.

value of k_{cat}/K_M for L217C that is 4-fold lower than WT (Fig. 1(b)). The introduction of deprotected untethered glucose, forming L217C-S-β-Glc, lowered k_{cat}/K_M further to 6-fold lower than WT. In contrast, glycosylation of position 217 with deprotected ethyl-tethered MTS reagents **1b**–**f** restored activity and k_{cat}/K_Ms for L217C-S-**b**–**f** are similar to each other in the range 2.5- to 3.1-fold lower than WT. This striking difference between tethered L217C-S-**b**–**f** and untethered L217C-S-β-Glc illustrates that SBL tolerates the replacement of hydrophobic Leu with highly hydrophilic carbohydrate moieties when they are linked by a hydrophobic ethyl spacer group better than directly-linked Cys-S-β-Glc. This may indicate that a structural requirement for efficient amidase activity is a closely-bound hydrophobic residue in the S_1' subsite of SBL and contrasts sharply with the excellent enhancement of esterase activity caused by the same Cys-S-β-Glc substitution.[37]

Mutation of position 156 in the S_1 pocket to cysteine causes a 1.4-fold drop in k_{cat}/K_M (Fig. 1(c)). Subsequent introduction of deprotected S-Et-α-Glc, side chain **b**, results in a k_{cat}/K_M for S156C-S-**b** that is 2.3-fold lower than WT. From S156C-S-**b** to -**f** k_{cat}/K_Ms vary in an arced manner peaking at 1.9-fold lower than WT for S156C-S-**d** and then decreasing monotonically to a k_{cat}/K_M for S156C-S-**f** that is 3-fold lower than WT. The similar K_M values of these S156 CMMs to those of SBL-WT are indicative of these modifications having

little effect upon ground state binding and are consistent with the surface exposed orientation of the S156 side chain.

At position 166, in the S_1 pocket, the 2.5-fold decrease in k_{cat}/K_M caused by mutation to cysteine is amplified by modification with **1b** and leads to a k_{cat}/K_M value 3-fold lower than WT for S166C-S-**b**. From S166C-S-**b** to -**f** k_{cat}/K_M decreases monotonically to a k_{cat}/K_M for S166C-S-**f**, in which the S_1 binding site is occupied by the sterically bulky disaccharide lactose, that is 3.8-fold lower than WT.

Kinetic effects of glycosylation with acetylated carbohydrates

The enormous potential of the controlled site-selective glycosylation approach depicted in Scheme 1 is illustrated by the great variety of changes in k_{cat}/K_M that are caused by the introduction of acetylated side chains **a**,**g**–**k** to SBL. These dramatic changes contrast with the slight variations found for deprotected side chains **b**–**f**. For example, at position 62 an alternating decrease–increase pattern is observed (Fig. 2(a)). Gratifyingly, this results in a k_{cat}/K_M for tetraacetylated α-gluco-CMM N62C-S-**g** that is 1.1-fold higher than WT. Similar alternating patterns are also seen at positions 217 (Fig. 2(b)) and 166 (Fig. 2(d)).

To examine the cause of these interesting variations, the k_{cat}/K_Ms of acetylated glycosylated CMMs were compared with those for deprotected glycosylated CMMs with the same glycan structure and stereochemistry (Fig. 3). This separated the effects of acetylation from the effects of glycosylation and allowed the underlying effects of modification to be dissected. It is clear from Figure 3 that the anomeric stereochemistry of the acetylated glycans modulates k_{cat}/K_M.

For example, at position 62 (Fig. 3(a)) comparison of N62C-S-**b**,**c** with N62C-S-**g** and N62C-S-Et-β-Glc(Ac)$_{2,3}$ shows that increasing the number of acetate groups from zero to four, from N62C-S-**b** to N62C-S-**d**, increases k_{cat}/K_M 1.6-fold for the α-gluco side-chain **b**. In contrast, increasing the number of acetate groups from zero to two or three, from N62C-S-**c** to N62C-S-Et-β-Glc(Ac)$_{2\ or\ 3}$, is detrimental for the β-gluco side-chain **c** and leads to a 2-fold decrease. Similarly, N62C-S-Et-β-Gal(Ac)$_3$ displays a distinctly lower k_{cat}/K_M than N62C-S-**e** that is 5-fold lower than WT. These dramatic changes in k_{cat}/K_M upon acetylation are manifested largely through increased or decreased ground state binding, of which the most striking example is a K_M for N62C-S-Et-β-Gal(Ac)$_3$ that is 2.4-fold higher than the corresponding deprotected galacto-CMM N62C-S-**e**.

At position 217, control of the level of acetylation through pH, as shown in Scheme 2, had allowed the introduction at position 217 of untethered β-D-glucose bearing zero, two and three acetate groups. As Figure 3(b) illustrates, the addition of two or three acetate groups restores k_{cat}/K_M from 6-fold lower than WT for L217C-S-β-Glc to 2-fold lower than WT for L217C-S-β-Glc(Ac)$_2$ or L217C-S-β-Glc(Ac)$_3$. Again this shows that acetylation allows fine-tuning of activity and interestingly parallels increases in the esterase k_{cat}/K_Ms of these CMMs.[37]

The same trend in k_{cat}/K_M is observed for the L217C ethyl-tethered CMMs as at position 62: acetylation is beneficial to α-tethered but detrimental to β-tethered CMMs. For example, increasing the number of acetate groups in the α-linked glucose moiety from zero to two, i.e., from L217C-S-**b** to L217C-S-Et-α-Glc(Ac)$_2$, increases k_{cat}/K_M to 2-fold lower than WT. In contrast, increasing the number of acetate groups in the epimeric β-linked moiety from zero to three, i.e., from L217C-S-**c** to L217C-S-Et-β-Glc(Ac)$_3$, halves k_{cat}/K_M to 6-fold lower than WT. Similarly, the k_{cat}/K_M of α-linked L217C-S-Et-α-Man(Ac)$_3$ is 1.5-fold higher than the corresponding deprotected L217C-S-**d**, while β-linked L217C-S-Et-β-Gal(Ac)$_3$ is 2-fold lower than the corresponding deprotected L217C-S-**e**.

Consistent with its surface exposed orientation, the changes at position 156 caused by acetylation are slight and no variation with anomeric stereochemistry is seen (Fig. 3(c)). Increasing the number of acetate groups from zero to four, from S156C-S-**b**–**e** to S156C-S-**g**–**j**, decreases k_{cat}/K_M by 1.05- to 1.6-fold. The most sterically bulky lactose side-chain (-**k**) gives rise to the lowest k_{cat}/K_M at position 156, 3.2-fold lower than WT, and indicates that even at the surface of SBL the introduction of sterically bulky groups still allows tailoring of k_{cat}/K_M.

At position 166, the effects of increased acetylation are, as at positions 62 and 217, modulated by glycan anomeric configuration. However, the direction of these increases and decreases is reversed: acetylation is beneficial to β-tethered but detrimental to α-tethered CMMs. For example, the α-tethered S166C-S-Et-α-Glc(Ac)$_3$ has a 1.8-fold lower k_{cat}/K_M value than the corresponding deprotected S166C-S-**b**, while β-linked CMMs S166-S-Et-β-Glc(Ac)$_{2,3}$ have 1.6-fold higher k_{cat}/K_M values than the corresponding fully deprotected S166-S-**c**. Again, these variations are largely manifested through changes in ground state binding. For example, K_M increases 1.4-fold from S166C-S-Et-α-Man (-**d**) to S166C-S-Et-α-Man(Ac)$_3$.

It should be noted that the changes in activity of the lacto-CMMs upon acetylation fall largely outside of these trends and at all four positions acetylation of the bulky, disaccharidic side-chain **f** causes a general decrease in k_{cat}/K_Ms. For example, at position 62 heptaacetylation, from N62C-S-**f** to N62C-S-**k**, results in a lowering of k_{cat}/K_M to a value that is 3-fold lower than WT. Interestingly, despite the greater steric bulk of side chain **k** this drop is a consequence of a lower k_{cat}, 2-fold lower than N62C-**f**, rather than a higher K_M. In fact, the K_M value of N62C-S-**k** is 1.3-fold lower than N62C-S-**f**. Similarly, L217C-S-Et-Lac(Ac)$_6$ and S166C-S-Et-Lac(Ac)$_5$ have 1.7-fold and 1.4-fold lower k_{cat}/K_Ms than the corresponding deprotected CMMs, respectively.

Conclusions

In summary, we have exploited the strategy of site-directed mutagenesis combined with chemical modification for the site-selective glycosylation of SBL. This method is general, versatile and allowed the preparation of pure glycoforms which constitute the first examples of regio- and glycan-specific protein glycosylation at predetermined sites. Careful control of a novel SBL-catalyzed intramolecular deacetylation greatly expanded the scope of this method and through reaction of SBL-N62C, -S156C, -S166C and -L217C with peracetylated MTS reagents **1a**,**g**–**k** allowed the introduction of glycans with precisely modulated levels of acetylation.

The glycosylated CMMs formed display k_{cat}/K_M values that range from 1.1-fold higher than WT to 7-fold lower than WT. Without the use of a highly selective glycosylation technique the determination of the precise trends described in this paper would have been unachievable and variations caused by previous non-specific glycosylations could only be interpreted in a general manner.[6] We have demonstrated in this work that subtle differences in carbohydrate structure may be used to fine tune the activity of SBL. For example, the anomeric stereochemistry of the glycans introduced modulates changes in k_{cat}/K_M upon acetylation. At positions 62 and 217 acetylation enhances the activity of α-tethered CMMs but decreases that of β-tethered. This trend is reversed at position 166 where, in contrast, acetylation enhances the k_{cat}/K_Ms of β-tethered CMMs but decreases those of α-tethered. Consistent with its surface exposed nature, changes at position 156 are more modest, but still allow control of activity particularly through glycosylation with disaccharide lactose. These results illustrate the great potential for tailoring activity through the correct choice of glycan and glycosylation site. The exciting opportunities that are created by this method may be readily extended to the tailoring of esterase activity and this further benefit has been explored in the following paper.[37]

The ability of the glycosylation method we have described to glycosylate the binding pockets of SBL also creates opportunities to broaden its substrate specificity. For instance, an array of hydrogen bonding hydroxyl groups may broaden its specificity towards hydrogen bonding substrates such as glycosylated amino acids. Subtilisins have been elegantly used to catalyze the synthesis of glycopeptides.[38,39] However the natural specificity of these enzymes has limited these peptide ligations to those in which the glycosylated residues are typically at least one residue distant (P_2,P_3... or P_2',P_3'...) from the amide bond formed. For example, while ligation of Z-Gly-OBz with H-Gly-Ser(Ac_3 GlcNAcβ)-NH_2 was successful, no yield of product was obtained with H-Ser(Ac_3GlcNAcβ)-NH_2.[38] The introduction of sugars to the S_1 and S_1' subsites as hydrogen bonding groups demonstrated here may enhance the specificity of proteases towards hydrophilic substrates.

Furthermore, by choosing carbohydrate attachments that differ from each other at only one stereocenter, SARs may be determined by examining changes in activity as the nature of the sugar side-chain is varied. For example, the effect of inverting stereocenters in the order C-4→C-1→C-2 can be determined using CMMs in the series **e**→**c**→**b**→**d**. While the current illustrations have been with SBL as a protein example, the method is clearly amenable to the glycosylation of any protein and is without limitation with respect to the sites and to the glycans that may be conjugated. It will therefore allow the introduction of any therapeutically important carbohydrate recognition determinant, of which the β-D-galactopyranosyl moiety of **e** and **f** that represents a ligand of the hepatic asialoglycoprotein receptor[1] is just one example. Accordingly, a further valuable benefit of this approach will be the ability to target proteins, including enzymes, to carbohydrate-binding proteins.

Experimental

Mutants of subtilisin *B. lentus* (SBL) were generated, and WT and mutant enzymes purified as described previously.[23,24] MTS reagents **1a**–**k** were prepared as described previously.[20]

General procedure for modification of SBL mutants stored as flash-frozen solutions

A 1.25 mL frozen aliquot of the mutant enzyme (N62C, L217C or S166C) containing approximately 25 mg of enzyme was thawed and added to 1.25 mL of Modifying Buffer (see below) in a polypropylene test-tube. To this solution was added 100 μL of a 0.2 M glyco-MTS reagent, prepared as described previously,[20] solution (**1a**,**g**–**k** in MeCN, **1b**–**f** in water). The mixture was sealed, vortexed and placed on an end-over-end rotator at room temperature. Completion of modification was determined by a specific activity assay using succinyl-AlaAlaProPhe-*p*-nitroanilide ($\epsilon_{410} = 8800\ M^{-1}\ cm^{-1}$)[40] as substrate in 0.1 M Tris–HCl buffer containing 0.005% Tween 80, 1% DMSO, pH 8.6, showing constant activity and by titration with Ellman's reagent[32] ($\epsilon_{412} = 13{,}600\ M^{-1}\ cm^{-1}$) showing no free thiol present in solution, a further 50 μL of the modifying reagent solution was added and the mixture placed back on the end-over-end rotator for a further 10 min. The reaction was poured onto a pre-packed, pre-equilibrated G-25 Sephadex® PD10 column and eluted with 3.5 mL of Quench Buffer (see below). The eluant was dialysed at 4 °C against 10 mM MES, 1 mM $CaCl_2$, pH 5.8 (2 × 1 L, 2 × 45 min). The resulting dialysate was flash frozen in liquid nitrogen and stored at −18 C.

Modifying Buffer: pH 9.5: 140 mM CHES, 2 mM $CaCl_2$; pH 7.5: 140 mM HEPES, 2 mM $CaCl_2$; pH 6.5: 140 mM MES, 2 mM $CaCl_2$; pH 5.5: 140 mM MES, 2 mM $CaCl_2$. Quench Buffer: reactions at pH 7.5–9.5: 5 mM MES 1 mM $CaCl_2$, pH 6.5; reactions at pH 5.5: 5 mM MES 1 mM $CaCl_2$, pH 5.5.

The free thiol content of all CMMs, was determined spectrophotometrically by titration with Ellman's reagent[32] in phosphate buffer (0.25 M, pH 8.0). In all

cases no free thiol was detected. Modified enzymes were analyzed by non-denaturing gradient (8–25%) gels at pH 4.2, run towards the cathode, on the Pharmacia Phast-system and appeared as a single band. Prior to ES–MS analysis CMMs were purified by FPLC (BioRad, Biologic System) on a Source 15 RPC matrix (17-0727-20 from Pharmacia) with 5% acetonitrile, 0.01% TFA as the running buffer and eluted with 80% acetonitrile, 0.01% TFA in a one step gradient. MS m/z (ES–MS): N62C-S-**a** calcd 27049, found 27051; N62C-S-**b** calcd 26925, found 26928; N62C-S-**c** calcd 26925, found 26928; N62C-S-**d** calcd 26925, found 26925; N62C-S-**e** calcd 26925, found 26925; N62C-S-**f** calcd 27087, found 27087; N62C-S-**g** calcd 27093, found 27096; N62C-S-Et-β-Glc(Ac)$_2$ calcd 27009, found 27015; N62C-S-Et-β-Glc(Ac)$_3$ calcd 27051, found 27053; N62C-S-**i** calcd 27093, found 27098; N62C-S-Et-β-Gal(Ac)$_3$ calcd 27051, found 27051; N62C-S-**k** calcd 27381, found 27386; L217C-S-β-Glc calcd 26882, found 26879; L217C-S-β-Glc(Ac)$_2$ calcd 26966, found 26962; L217C-S-β-Glc(Ac)$_3$ calcd 27008, found 27006; L217C-S-**b** calcd 26926, found 26928; L217C-S-**c** calcd 26926, found 26925; L217C-S-**d** calcd 26926, found 26925; L217C-S-**e** calcd 26926, found 26928; L217C-S-**f** calcd 27088, found 27087; L217C-S-Et-α-Glc(Ac)$_2$ calcd 27010, found 27012; L217C-S-Et-β-Glc(Ac)$_3$ calcd 27052, found 27056; L217C-S-Et-α-Man(Ac)$_3$ calcd 27052, found 27056; L217C-S-Et-β-Gal(Ac)$_3$ calcd 27052, found 27053; L217C-S-Et-Lac(Ac)$_6$ calcd 27340, found 27342; S166C-S-**a** calcd 27076, found 27080; S166C-S-**b** calcd 26952, found 26955; S166C-S-**c** calcd 26952, found 26950; S166C-S-**d** calcd 26952, found 26952; S166C-S-**e** calcd 26952, found 26952; S166C-S-**f** calcd 27114, found 27112; S166C-S-Et-α-Glc(Ac)$_3$ calcd 27078, found 27078; S166C-S-Et-β-Glc(Ac)$_2$ calcd 27036, found 27040; S166C-S-Et-β-Glc(Ac)$_3$ (major) with S166C-S-**h** (minor) and S166C-S-Et-β-Glc(Ac)$_2$ (minor) calcd 27078 (major), 27120 (minor), 27036 (minor), found 27081 (major), 27121 (minor), 27036 (minor); S166C-S-Et-α-Man(Ac)$_3$ calcd 27078, found 27085; S166C-S-Et-β-Gal(Ac)$_3$ calcd 27078, found 27079; S166C-S-Et-Lac(Ac)$_5$ calcd 27324, found 27331.

General procedure for modification of SBL mutants stored as lyophilized powders

This procedure was used for S156C, which was stored as a lyophilized powder to prevent dimerization. Into a polypropylene test tube was weighed about 25–30 mg of lyophilized S156C. This was dissolved in the following modifying buffers (2.5 mL).

Modifying buffer: pH 9.5: 70 mM CHES, 2 mM $CaCl_2$; pH 7.5: 70 mM HEPES, 2 mM $CaCl_2$; pH 6.5: 70 mM MES, 2 mM $CaCl_2$; pH 5.5: 70 mM MES, 2 mM $CaCl_2$.

Glyco-MTS reagent was added and the reaction then proceeded as for the other mutants, using the appropriate quench buffer. MS m/z (ES–MS): S156C-S-**a** calcd 27076, found 27079; S156C-S-**b** calcd 26952, found 26955; S156C-S-**c** calcd 26952, found 26952; S156C-S-**d** calcd 26952, found 26952; S156C-S-**e** calcd 26952, found 26952; S156C-S-**f** calcd 27114, found 27115; S156C-S-**g** calcd 27120, found 27123; S156C-S-**h** calcd 27120, found 27122; S156C-S-**i** calcd 27120, found 27123; S156C-S-**j** calcd 27120, found 27120; S156C-S-**k** calcd 27408, found 27411.

Incubation of L217C-S-β-Glc(Ac)$_3$ at pH 9.5

The general procedure for modification of SBL mutants stored as flash-frozen solutions was used to incubate 1.26 mg of L217C-S-β-Glc(Ac)$_3$ as a 0.5 mL aliquot in the absence of MTS reagent for 2 h to give L217C-S-β-Glc as the sole product. MS m/z (ES–MS): L217C-S-β-Glc calcd 26882, found 26885.

Active site titrations

The active enzyme concentration was determined as previously described[33] by monitoring fluoride release upon enzyme reaction with α-toluenesulfonyl fluoride (PMSF) as measured by a fluoride ion-sensitive electrode (Orion Research 96-09). The active enzyme concentration determined in this way was used to calculate k_{cat} values for each CMM.

Kinetic measurements

Michaelis–Menten constants were measured at 25(±0.2)°C by curve fitting (GraFit® 3.03) of the initial rate data determined at nine concentrations (0.125–3.0 mM) of succinyl-AAPF-*p*NA substrate in 0.1 M Tris–HCl buffer containing 0.005% Tween 80, 1% DMSO, pH 8.6 ($\epsilon_{410} = 8800\ M^{-1}\ cm^{-1}$).[40]

Acknowledgements

We acknowledge the generous financial support of the Natural Sciences and Engineering Research Council of Canada and Genencor International, Inc. and the University of Durham, and we thank Dr. Rick Bott for helpful discussions.

References

1. Sharon, N.; Lis, H. *Essays Biochem*. **1995**, *30*, 59.
2. Lasky, L. A. *Annu. Rev. Biochem*. **1995**, *64*, 113.
3. Weis, W. I.; Drickamer, K. *Annu. Rev. Biochem*. **1996**, *65*, 441.
4. Varki, A. *Glycobiology* **1993**, *3*, 97.
5. Dwek, R. A. *Chem. Rev*. **1996**, *96*, 683.
6. Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33*, 17.
7. Kohen, A.; Jonsson, T.; Klinman, J. P. *Biochemistry* **1997**, *36*, 2603.
8. Yanez, E.; Carmona, T. A.; Tiemblo, M.; Jimenez, A.; Fernandez-Lobato, M. *Biochem. J*. **1998**, *329*, 65.
9. For previous examples of the effects of non-specific chemical glycosylation upon enzyme activity see: Levashov, A. V.; Rariy, R. V.; Martinek, K.; Klyachko, N. L. *FEBS Lett*. **1993**, *336*, 385 and refs 10–13.
10. Longo, M. A.; Combes, D. *FEBS Lett*. **1995**, *375*, 63.
11. Longo, M. A.; Combes, D. *J. Mol. Catal. B* **1997**, *2*, 281.

12. Baek, W. O.; Vijayalakshmi, M. A. *Biochim. Biophys. Acta* **1997**, *1336*, 394.
13. Jiang, K. Y.; Pitiot, O.; Anissimova, M.; Adenier, H.; Vijayalakshmi, M. A. *Biochim. Biophys. Acta* **1999**, *1433*, 198.
14. Hill, T. G.; Wang, P.; Huston, M. E.; Wartchow, C. A.; Oehler, L. M.; Smith, M. B.; Bednarski, M. D.; Callstrom, M. R. *Tetrahedron Lett*. **1991**, *32*, 6823.
15. Wang, P.; Hill, T. G.; Wartchow, C. A.; Huston, M. E.; Oehler, L. M.; Smith, M. B.; Bednarski, M. D.; Callstrom, M. R. *J. Am. Chem. Soc*. **1992**, *114*, 378.
16. Wartchow, C. A.; Wang, P.; Bednarski, M. D.; Callstrom, M. R. *J. Org. Chem*. **1995**, *60*, 2216.
17. Wang, P.; Hill, T. G.; Bednarski, M. D.; Callstrom, M. R. *Tetrahedron Lett*. **1991**, *32*, 6827.
18. Parker, D.; Kataky, R. *J. Chem. Soc., Chem. Commun*. **1997**, 141
19. Parker, D.; Kataky, R. *Pure Appl. Chem*. **1996**, *68*, 1219.
20. Davis, B. G.; Lloyd, R. C.; Jones, J. B. *J. Org. Chem*. **1998**, *63*, 9614.
21. Kenyon, G. L.; Bruice, T. W. *Methods Enzymol*. **1977**, *47*, 407.
22. Wynn, R.; Richards, F. M. *Methods Enzymol*. **1995**, *251*, 351.
23. Stabile, M. R.; Lai, W. G.; DeSantis, G.; Gold, M.; Jones, J. B.; Mitchinson, C.; Bott, R. R.; Graycar, T. P.; Liu, C.-C. *Bioorg. Med. Chem. Lett*. **1996**, *6*, 2501.
24. Berglund, P.; DeSantis, G.; Stabile, M. R.; Shang, X.; Gold, M.; Bott, R. R.; Graycar, T. P.; Lau, T. H.; Mitchinson, C.; Jones, J. B. *J. Am. Chem. Soc*. **1997**, *119*, 5265.
25. DeSantis, G.; Berglund, P.; Stabile, M. R.; Gold, M.; Jones, J. B. *Biochemistry* **1998**, *37*, 5968.
26. Plettner, E.; Khumtaveeporn, K.; Shang, X.; Jones, J. B. *Bioorg. Med. Chem. Lett*. **1998**, *8*, 2291.
27. Dickman, M.; Lloyd, R. C.; Jones, J. B. *Tetrahedron: Asymm*. **1998**, *9*, 4099.
28. Jones, J. B.; Desantis, G. *Acc. Chem. Res*. **1999**, *32*, 99.
29. Plettner, E.; DeSantis, G.; Stabile, M. R.; Jones, J. B. *J. Am. Chem. Soc*. **1999**, *121*, 4977.
30. Nomenclature of Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun*. **1967,** *27*, 157.
31. Fukuda, M.; Hindsgaul, O., Eds. *Molecular Glycobiology*. Oxford: Oxford University, 1994.
32. Ellman, G. L.; Courtney, K. D.; Andres, V.; Featherstone, R. M. *Biochem. Pharmacol*. **1961**, *7*, 88.
33. Hsia, C. Y.; Ganshaw, G.; Paech, C.; Murray, C. *J. Anal. Biochem*. **1996**, *242*, 221.
34. Greene, T. W.; Wuts, P. G. M. *Protective Groups in Organic Synthesis*, 3rd ed.; New York: Wiley, 1999.
35. Desantis, G.; Jones, J. B. *J. Am. Chem. Soc*. **1998**, *120*, 8582.
36. Bashir, N. B.; Phythian, S. J.; Reason, A. J.; Roberts, S. *J. Chem. Soc., Perkin Trans. 1* **1995**, 2203.
37. Lloyd, R. C.; Davis, B. G.; Jones, J. B. *Bioorg. Med. Chem*. **2000**, *8*, 1537.
38. Witte, K.; Seitz, O.; Wong, C.-H. *J. Am. Chem. Soc*. **1998**, *120*, 1979.
39. Wong, C.-H.; Schuster, M.; Wang, P.; Sears, P. *J. Am. Chem. Soc*. **1993**, *115*, 5893.
40. Bonneau, P. R.; Graycar, T. P.; Estell, D. A.; Jones, J. B. *J. Am. Chem. Soc*. **1991**, *119*, 1026.