Green Solvents Derived from Glycerol Modified Classical Regioselectivity in the Enzymatic Synthesis of Disaccharides with Biolacta β-Galactosidase

María Pérez-Sánchez,[a] **Manuel Sandoval,**[a] **Alvaro Cortés-Cabrera,** [b] **Héctor García-Marín,** [c] **José V. Sinisterra,**[d] **José I. García** [c] **and María J. Hernaiz**[a,d*]

Green solvents made from glycerol change the classical regioselectivity of Biolacta N° 5 β-galactosidase, from β(1→4) toward β(1→6) linkages when 2M concentration was used. In order to explain these results, the non-proteic compounds present in Biolacta preparation were separated by precipitation with ammonium sulfate and the remaining protein extract was used to set reactions with appropriate organic solvents to find that the regioselectivity towards the β(1→6) isomer is retained. According to proteomic analysis, a 98% homology between *Streptococcus pneumoniae* and Biolacta β-galactosidase preparation was found. With these data, molecular modelling was done and predicts a tridimensional interaction in the enzyme active site with the donor (GlcNAc) and the water-solvent mixture that explains this phenomenon.

Introduction

In recent years, Green Chemistry has become an area of great interest for developing more environmentally acceptable processes. According to the Green Chemistry principles,[1-3] biocatalysis and solvents play an important role in green processes, since they provide a valuable alternative to classic organic chemistry.[4] First, enzymes offer suitable tools for industrial reactions, which can be carried out under milder conditions, without using of heavy metals and with a great control over chemo-, regio- and stereoselectivity using appropriate enzymes. Second, there are numerous potential advantages in employing enzymes in organic solvents, such as increased solubility of nonpolar substrates, shifting of thermodynamic equilibria to favor synthesis over hydrolysis, decrease of water-dependent side reactions, enhanced thermal stability of enzymes, and elimination of microbial contamination, among others.[5] On the other hand, the presence of organic solvents in enzymatic reactions can alter the activity and specificity of the enzyme. There are numerous examples in the literature wherein solvents physical properties such as dielectric constant, dipole moment and hydrophobicity are related to various effects on enzyme activity, specificity and enantioselectivity.[6-9]

Nowadays, a new generation of solvents has been developed, whose search has become an area of interest. Currently, water, supercritical fluids (SCFs), fluorous solvents, and solvents from renewable sources are considered green solvents and have been used in biocatalysis.[10, 11] In that sense glycerol is a green solvent that has received attention, especially because it is a by-product of biodiesel production, and hence it can be obtained from renewable sources. Glycerol is an environmentally friendly, non-toxic, biodegradable solvent. It has already been used in biocatalytic reactions, and its potential for these reactions demostrated.[12-14]

On the other hand, oligosaccharides are involved in a wide range of biological processes including bacterial and viral infection, cancer metastasis, the blood-clotting cascade and many other crucial intercellular recognition events.[15-18] As the understanding of these biological functions increases, the need for practical synthetic procedures of oligosaccharides in large quantities has become a major subject. Organic chemical methods for obtaining them have been developed,[19-21] but they involve several elaborate protection and deprotection procedures. Glycosyl hydrolases (glycosidases) can be used to synthesize oligosaccharides in a kinetically controlled reaction, where a glycosyl donor is used to transfer its glycosyl residue to a sugar acceptor present in the reaction medium. In spite of the increased amount of work carried out with glycosyl hydrolases, their main drawback is the lack of regioselectivity, which limits their use for synthetic purposes. In an important contribution to glycosidase-catalyzed oligosaccharide synthesis, Usui *et al.*[22] reported for the first time the use of a β-galactosidase from *Bacillus circulans* (Biolacta) to synthesize some β-D-(1→4) galactosyl disaccharides bearing a GlcNAc residue at the reducing end. Some β-D-(1→6) linkages were produced as well, but to the best of our knowledge, it was the first time that a preparative scale β-D-galactosyltransfer has been shown to occur preferentially at the O-4 position using a galactosyl hydrolase. Also it has been shown that the use of the thioethyl glycoside of *N*-acetylglucosamine as acceptor and *p*-nitrophenyl β-D-galactoside as donor gave complete selectivity for 1→4 transfer[23-26]. In this work, we present the first study of a change of regioselectivity in the enzymatic synthesis of disaccharides with Biolacta β-galactosidase when green solvents derived from glycerol are used.

Results and discussion

The benchmark reaction used throughout the work has been the transglycosilation reaction of *N*-acetyl glucosamine with *p*-nitrophenyl-β-D-galactopyranoside, catalyzed by Biolacta N° 5 β-galactosidase (Scheme 1). This reaction can afford Gal-β-(1→4)-GlcNAc as mayor product and Gal-β-(1→6)-GlcNAc as minor

product.
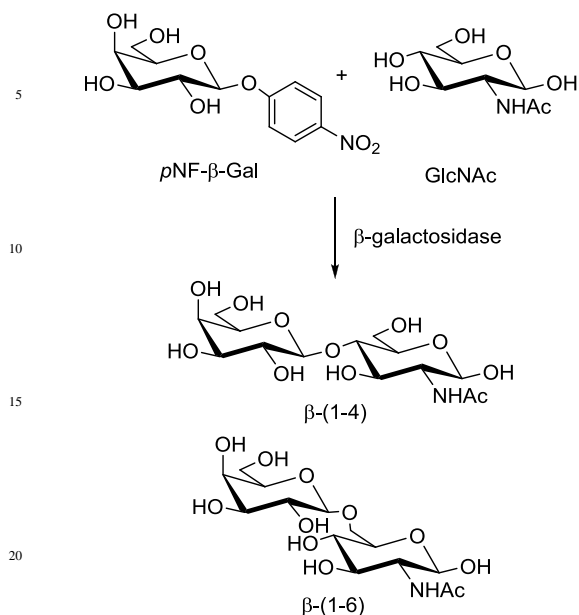
**Scheme 1.** General scheme of transglycosilation reaction catalyzed by Biolacta N º 5 β-galactosidase.

Synthesis and toxicity issues of glycerol-derived solvents.

Glycerol-based solvents were synthesized using the same procedures previously described.[27] The complete listing of the solvents used in this work is given in Figure 1. Some physico-chemical and solvatochromic parameters of the solvents used in this work are gathered in Table 1. As can be seen, the structural diversity of the glycerol derivatives results in a wide range of hydrophobicities, covering four orders of magnitude. Bearing this in mind, a solvent set was selected, since hydrophobicity could be a key factor in solvent-enzyme substrate interactions.

Also, dielectric constants and dipole moments cover a broad range of values, as well as the Dimroth–Reichardt solvatochromic parameter, E_T^N.[28] The latter is mainly a measure of the hydrogen-bond donor ability of the solvent, a feature that could also be of importance for solvent-substrate and/or solvent-enzyme interactions. Some of the glycerol-derived solvents used, namely those bearing fluorinated chains, display simultaneously both high hydrophobicity and hydrogen-bond donor ability, a rather unusual combination in conventional organic solvents. Miscibility with water, another important issue, also depends on solvent structure. Thus, solvents **1–3** are completely miscible with water, **5** and **8** are able to dissolve an appreciable amount of water. On the other hand, triethers (**6** and **9**) and highly fluorinated alcohols (**10** and **11**) display a very limited compatibility with water.[27]

Concerning toxicity issues, there are some experimental data pointing to glycerol-derived solvents as being low hazardous. For instance, toxicities of glycerol-based solvents like **1**, **8** and the closely related 1,3-diethoxy-2-propanol (LD$_{50}$ > 3000 mg/kg, oral in mice)[32] and 1,2,3-triethoxypropane (LD$_{50}$ = 660 mg/kg, intraperitoneal in mice)[32] are very low, in some cases even lower than that of butanol (LD$_{50}$ = 2680 mg/kg, oral in mice). Fluorinated solvents will be probably more toxic than aliphatic

ones, but lacking in experimental data for them it would be reasonable to take the toxicities of compounds produced in their probable decomposition, like trifluoroethanol (LD$_{50}$ = 366 mg/kg, oral in mice),[33] trifluoroacetic acid (LD$_{lo}$ = 150 mg/kg, intraperitoneal in mice)[33] or 2,2,2-trifluoroethyl ethyl ether (LD$_{50}$ = 5100 mg/kg, intraperitoneal in rats),[34] as referents of the toxicity of these kinds of solvents.
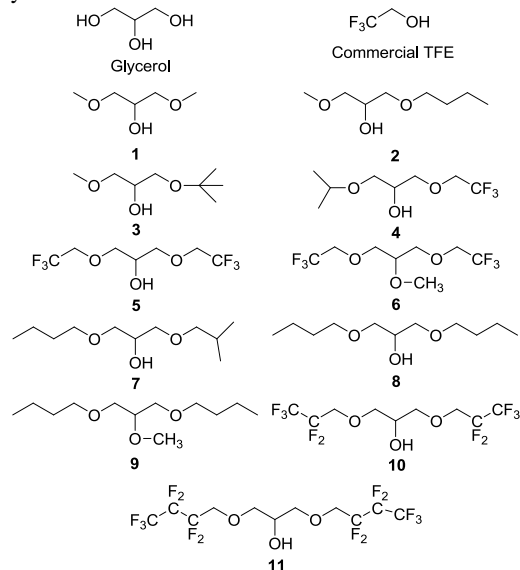
Fig. 1 Structure of different solvents derived from glycerol employed in the transglycosilation reaction with Biolacta β-galactosidase.

Moreover using these glycerol derivatives allow to reduce some other hazardous aspects. For instance, they have low volatility,[27] so the concentration in the air using them is lower than using other conventional solvents. Albeit toxicities of these solvents are not well known at this moment it is clear they are not extremely toxic, meeting one of the key requirements for green applications.

Table 1. Transglycosylation yields (%) obtained with commercial Biolacta β-galactosidase in presence of green solvents.

Medium	log P[a]	μ (D)[a]	ε[a]	$E_T^{N a}$	Gal	Disaccharide β-(1→4)	Disaccharide β-(1→6)
Buffer	0.00	1.85			-	84	17
Glycerol	−1.33	-			35	14	9
TFE	0.71	2.64[b]	27.1[c]	0.90[d]	9	73	-
1	−0.60	3.00	13.0	0.61	-	15	85
2	0.14	2.40	7.0	0.48	-	23	77
3	0.27	2.36	6.6	0.44	-	13	87
4	1.14	3.36	12.0	0.59	-	-	93
5	1.42	4.10	14.6	0.70	-	-	100
6	1.71	4.20	13.7	0.55	79	-	21
7	1.93	2.30	5.2	0.46	-	17	83
8	2.07	2.40	5.6	0.45	-	-	92
9	2.48	2.20	4.7	0.15	-	21	71
10	2.56	3.94	10.4	0.70	57	-	42
11	3.77	3.90	8.6	0.07	100	-	-

[a] Values from ref. 27, unless otherwise indicated. [b] Value taken from ref. 30. [c] Value taken from ref.29. [d] Value taken from ref.31.

Enzymatic synthesis in presence of green solvents with commercial and semipurified Biolacta β–galactosidase.

Powder of Biolacta Nº5 contains about 11% proteins in its composition. After the concentration process we achieved 71% of proteins in solid sample, which means we reduce non-proteic components from the crude preparation. Commercial (C) and semipurified (S) enzyme showed a similar specific activity (4.0–4.2 U mg^{-1} respectively). SDS-PAGE showed the same gel profile elution for the two preparations mentioned above (Fig. 2).
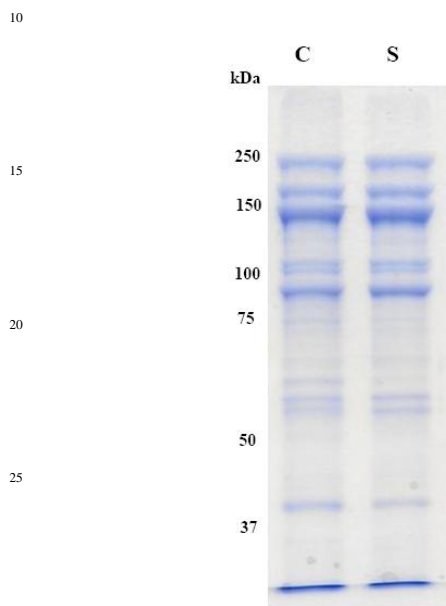
Fig. 2. Electrophoresis gel (7.5%) with different samples obtained from Biolacta Nº 5 (S, semipurified; C, commercial).

According to SDS-PAGE results C and S preparation gave about 11 protein bands in the range of 250–37 kDa, but there are three important bands at 230, 145 and 86 KDa that ~~is~~ agree with previous publication reporting three isoforms in the Biolacta β–galactosidase preparation.[35] These suggests that these proteins may be assembly in many subunits.(?) ver carta referees First, the regioselectivity showed by this enzyme was investigated when the reaction was carried out in 50 mM citrate/phosphate buffer at pH 5.0 (Table 1). Biolacta β-galactosidase gave a mixture of regioisomers, the major β-(1→4) linker (84 % conversion) and minor product β-(1→6) linker (17 %).

Second, hydrolytic activity of this β-D-galactosidase was measured in different concentrations of solvent **1**. The best result was achieved in 2 M concentration of that solvent (Fig. 3). Therefore, transglycosylation reactions catalyzed by commercial preparation of the Biolacta β-galactosidase were carried out in buffer media and green solvent (2 M)-buffer mixture (see experimental section). Results from these reactions are shown in Table 1.

The synthetic activity of the β-galactosidase was examined, and a change in the regioselectivity of the reaction was confirmed. In presence of solvents **1**, **4**, **5**, **8** and **10** the regioisomer β-(1→6) was formed as the sole product, the best yields being obtained in the presence of **4** (93 %), **5** (100 %) and **8** (92%), while with the rest of the solvents tested the main product formed was β-(1→6) with yields ranging from 71% to 85 % and β-(1→4) as the minor product (13–23 %). This change in the regioselectivity had also been noticed in previous works by our group in the reaction catalyzed by Biolacta β-galactosidase, when *N,N*-dimetylamide-type solvents were employed.[36]
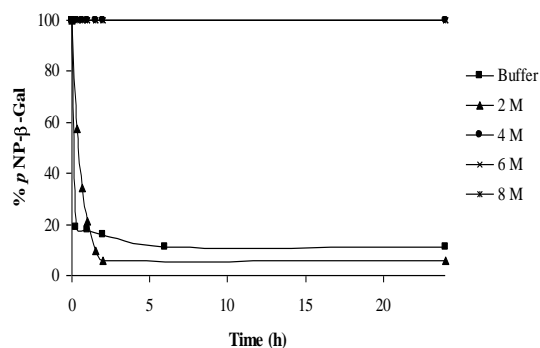
Fig. 3 Determination of optimal concentration of solvent 1 for the enzyme activity of Biolacta β-galactosidase at 30ºC (followed by hydrolysis of *p*NP-β-Gal).

In fact, the change of regioselectivity cannot be easily related to the available physical parameters (dipole moment, dielectric constant or log P). In principle, it could be due to either solvent-enzyme or solvent-substrate interactions, as well as to the presence of impurities in the enzyme.

In order to explain the changes in regioselectivity observed in reactions with some solvents and commercial enzyme, most of non-proteic substances present in Biolacta were removed by precipitation with $(NH_4)_2SO_4$. The semipurified Biolacta β-galactosidase was used with the best green solvents, selected accordingly with the previous results obtained with the commercial enzyme. Table 2 shows the yields obtained in different media for these reactions.

Table 2. Transglycosylation yields (%) obtained with semipurified Biolacta β-galactosidase in the presence of glycerol-derived solvents.

Medium	Hydrolysis (galactose)	Synthesis β(1→4)	Synthesis β(1→6)
Buffer	-	90	10
4	6	6	88
5	14	11	75
8	9	9	82

Using semipurified enzyme we observed that solvent **4** leads to the best yields for Galβ(1→6)GlcNAc product. This result represents a very important regioselectivity change in comparison with the natural behavior of this enzyme that shows a high tendency to give Galβ(1→4)GlcNAc product.[24-26]
Using commercial or semipurified enzyme as catalyst in buffer medium usually gives ride to yields about 80% or 90% β(1→4) and 17 or 10% β(1→6) product (See Table 1 and 2). Finally, and taking into account that this reaction was performed after elimination of non-proteic substances present in Biolacta, one can conclude that those substances were not related to regioselectivity

changes observed during transglycosylation reactions and could be related to molecular interactions between enzyme, solvent and substrates.

Molecular modeling studies

In order to analyze the proteic nature of the commercial Biolacta β-galactosidase a spot from the SDS-PAGE (approximately 86 kDa) was selected according to the molecular mass expected for the β-galactosidase for Biolacta crude extract[35] and the and the mass spectrometry analysis of the protein spots was done. Results from the proteomic analysis showed 98% homology with *Streptococcus pneumoniae* SP9-BS68 β-galactosidase, which posses a 247 kDa molecular mass. This information could explain the synthetic behavior of Biolacta enzymes with high preference to Gal-β(1→4)GlcNAc synthesis instead of the classical *B. circulans* tendency toward β(1→3) linkages[37, 38] that does not occur in Biolacta preparation.

The *S. pneumoniae* β-galactosidase have proved to be more selective for catabolic degradation of Gal-β(1→4)GlcNAc than lactose because its role in infection processes.[39, 40] This feature is also present in Biolacta crude extract and suggests that both could be indeed the same enzyme. These data also support the idea that protein conglomerates are present in Biolacta preparation and they are cleaved by denaturing conditions of SDS-PAGE.

In order to come up with a rational explanation and to confirm our hypothesis about the influence of the solvent properties in reaction stereoselectivity in the glycosidic linkage formation, molecular modeling studies were carried out.

A 3D β-galactosidase *S. pneumoniae* model was build and the molecules implicated in the mechanism docked, then, MD simulations were run in two different solvent conditions and the results compared.

The active site was located using a combination of sequence and structure alignment and blind docking. The enzyme model revealed that the active center could be similar to *E. coli* lacZ. A simple superimposition of the structure of a monomer and the model yielded the identification of the catalytic pair (Glu-564 and Glu-645) and the chemical behavior of each of them. Residue Glu-645 corresponds to Glu-537 in lacZ which is accepted as the nucleophile,[41] while Glu-564 was identified as the acid catalyst. This fact is also consistent with the first docking results.

It was observed that the compound tended to be located near the Glu-645 residue, packed over the Trp-708 residue. This kind of interaction has been deeply investigated[42] and it is identified as crucial for other galactosidases. Additional interactions with residues His-450, His-484, Asn-563 and Glu-716 contribute to stabilize the molecule as seen in Fig. 4.

After modeling the glycosyl-enzyme, GlcNAc was docked in the active site. The best docking solution was used as a starting point for MD simulations. Regarding to the preferred stereochemistry of β-galactosidases, distances between anomeric carbon (C1) and, 4 and 6 hydroxyl groups (O4 and O6) of the acceptor molecule, were shown to be valid indicators of the carbon atom in the best position to be attacked and correlated with the regiospecificity of various enzyme used in the enzymatic synthesis of carbohydrates.[43]

With this aim, the distances between the C1 and O4 and O6 groups and between oxygen atoms of Glu-564 and protons of each hydroxyl group were measured. Special attention was paid for structural information in the active site, particularly for differential binding mode between simulations that could explain the change in the stereoselectivity. Average distances are presented in table 3 together with RMSD of alpha carbons.

Table 3. Average distances (in Å) between anomeric carbon (C1) and hydroxyl groups (O4 and O6) and RMSD of the simulation.

	C1-O6	C1-O4	COO⁻-HO-C6	COO⁻-HO-C4	RMSD C_α (Å)
Water	5.76[a]	5.26[a]	5.73[a]	5.81[a]	2.27
5-Water	4.79	7.89	7.42	8.93	1.93

[a] After change of position.

Although the docked solution showed a clear preference for O6, not compatible with the experimental product, during the MD simulation the distances from the anomeric carbon to O4 and O6 changed at 700 ps, indicating a variation in the configuration of the acceptor. Indeed, a different binding mode can be observed, in which the acceptor is found deeper in the active center, interacting mainly with Tyr-624 and Lys-664.

This change could be a consequence of a pushing effect of the solvent (water) due to the hydrophobic nature of the acetyl chain, which can be found interacting with Tyr-624. The distance C1-O4 was found to be more favorable; nevertheless, O6 is still close enough to allow both linkages as determined experimentally.
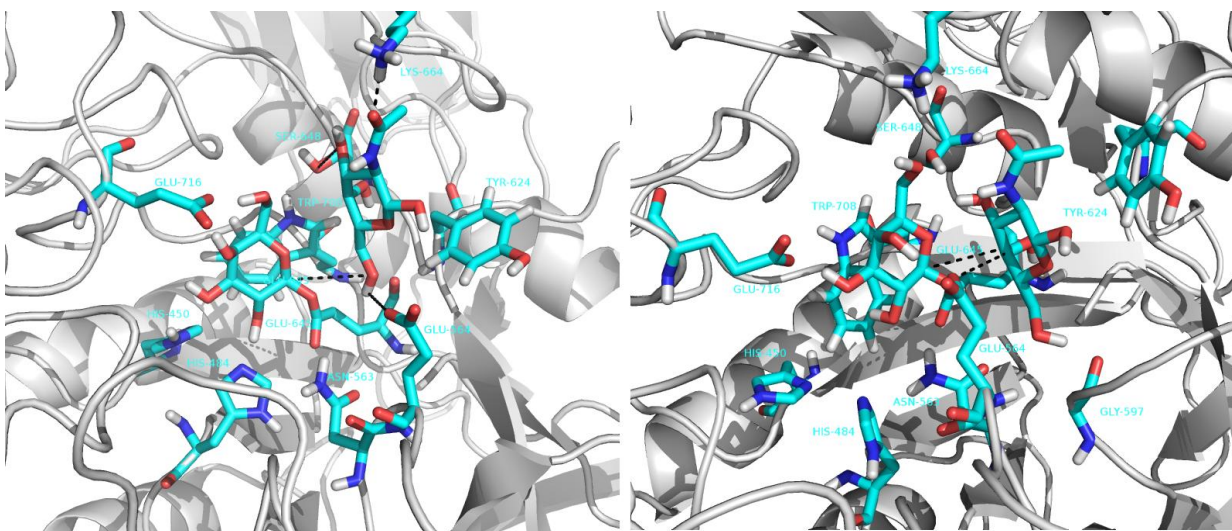
Fig. 4. Binding modes: *Shallow* mode in **5**-water simulation at left, *deep* mode in water simulation on the right.

In the case of solvent **5**-water mixture, the initial docked solution was found to be coincident with the experimental product. The acceptor exhibited during the entire simulation a clear preference for O6, supporting the formation of the $\beta(1\rightarrow6)$ linkage. However, a key interaction present in this complex, the residue Ser-648, is found to be interacting with O4 hydroxyl of the acceptor. This interaction is not present in the water simulation because of the *deep* binding mode of the molecule and it could contribute to the stabilization of the complex. In this case, the acetyl chain is not pushing deeper the acceptor. Apparently, the interaction with the solvent is not as unfavorable as in the case of water only, and the molecule adopts a *shallow* binding mode with essentially the same interactions of the *deep* binding mode (Tyr-624 and Lys-664) with the additional interaction with Ser-648.

The distances measured during MD simulations are correlated with the experimental results, in which the formation of $\beta(1\rightarrow6)$ linkage is expected in **5**-water mixture, while the $\beta(1\rightarrow4)$ linkage is the main product in water with some $\beta(1\rightarrow6)$ linked molecules. In the superimposition of both solutions (water and **5**-water mixture, Fig. 5) can be appreciated that $O4_{wat}$ and $O6_{5\text{-wat}}$ are located in the same position, at a good distance from C1 and the catalytic Glu-564.

The role of the acceptor in the stereochemical changes in carbohydrate enzymatic synthesis is of great importance as it has been previously reported.[38] Particularly, some groups attached to the acceptor could help to get the substrate molecule closer to the catalytic residues and induce a change in the sugar conformation. This is the case of the 2-acetamido group,[38] which greatly influences the stereoselectivity of the overall reaction. In our system, the interaction between the acetyl chain and the solvent triggers a binding mode change in which the molecule may reach a deep position in the active center, improving the O4/O6 groups accessibility, and therefore, the stereochemistry. The differences between solvents may be explained by the different size and characteristics of the derivatives that may be the final responsible of the interaction with the acceptor.
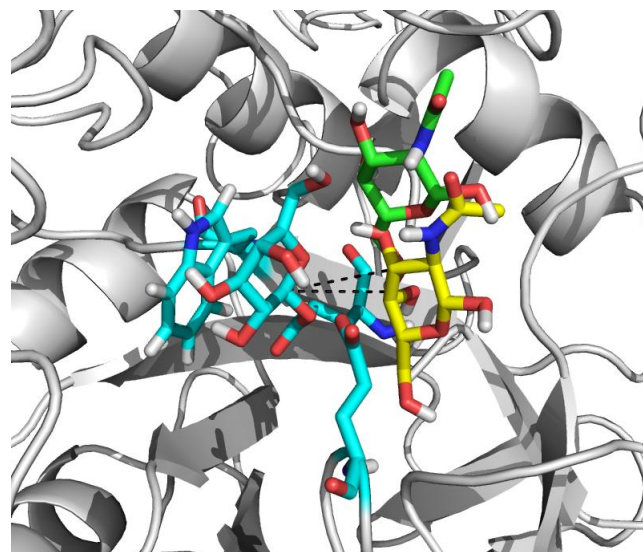
Fig. 5 Superimposition of the two binding modes: shallow (in green) and deep (in yellow). Catalytic residues are shown in light blue. Superimposition of O6 (shallow) and O4 (deep) can be observed.

Conclusions

In this work we determined a change in classical regioselectivity of Biolacta β-D-galactosidase when reaction is performed in the presence of glycerol-based solvents. These solvents promote the $\beta(1\rightarrow6)$ linkages between donor and acceptor instead of $\beta(1\rightarrow4)$. This reactivity change is maintained even after the Biolacta non protein components are removed, which means the change is mainly related to enzyme-solvent interactions. The proteomic analysis showed a high similarity between *S. pneumoniae* β-D-galactosidase and one of Biolacta proteins, which could explain the tendency toward $\beta(1\rightarrow4)$ linkages present in both enzymes. Using *S. pneumoniae* protein structure as a basis for understanding the solvent-enzyme interaction, we undertook a

molecular modeling study and found that the three-dimensional arrangement between GlcNAc and the water-solvent (**5**) mixture in the active site of the enzyme, favors the β(1→6) linkage. This disposition is also stabilized by the OH interaction between Ser-648 and O4 from GlcNAc, which also explains the reduced proportion of β(1→4) linkage product.

Experimental

Materials. *p*-nitrophenol (*p*NP), *p*-nitrophenyl-β-D-galactopyranoside (*p*NP-β-Gal), *N*-acetyl-D-glucosamine (GlcNAc), Gal-β(1→4)GlcNAc, Gal-β(1→6)GlcNAc, D-(+)-galactose, ammonium sulfate and bovine serum albumin (BSA), were purchased from Sigma Aldrich. Biolacta Nº5 β-galactosidase from *B. circulans* was received from Daiwa Kasei, Japan. Dye reagent for protein determination was purchased from Bio Rad. All other chemicals were from analytical grade.
UV-visible spectra were recorded on a UV-2401 PC Shimadzu. HPLC Agilent 1100 with UV-vis detector using Mediterraneasea18 15 cm x 0.46 5 mm column (Teknokroma) with water: acetonitrile (75:25) as a mobile phase at a flow of 0.7 mL min⁻¹. HPLC Jasco with an evaporative light scattering detector (ELSD) using NH$_2$P50-4E amino column (Asahipak, Japan) with acetonitrile: water (80:20) as a mobile phase at a flow of 1.0 mL min⁻¹. NMR spectra were recorded on Bruker AV 500 MHz spectrometers. The structures of the enzymatically synthesized disaccharides were assigned by ¹H, ¹³C and 2-D NMR (see Supporting Information for details).

Enzyme assays. Hydrolytic activity was determined by spectrophotometrical quantification of *p*NP liberated from *p*NP-β-Gal in citrate/phosphate buffer 50 mM, pH 5.0, using both continuous and discontinuous methods. *Continuous method:* hydrolysis of *p*NP-β-Gal (5 mM) in a 1 μL cell by measuring the increase in absorbance at 410 nm during 2.5 minutes at 37ºC.[44] *Discontinuous method:* hydrolysis of 200 μL of *p*NP-β-Gal (1.0 mM) for 10 minutes at 37ºC and then reaction was stopped by adding 1.00 mL of Na$_2$CO$_3$ 0.20 M.[44,36] Each experimental assay was determined at least three samples with standard deviation under 5% of the average of samples. Protein concentration was determined by Bradford method[45] following the manufacturer's protocol (Bio-Rad) and bovine serum albumin as the standard. One enzymatic unit was defined as the amount of protein that hydrolyses 1 μmol of *p*-nitrophenyl-β-D-galactopyranoside per minute under the conditions described before.

Semipurification of enzymes. Semipurified Biolacta β-galactosidase was obtained as a precipitate by centrifugation after the addition of ammonium sulfate, (55% saturation at 0ºC) to a solution of 5% m/v from commercial enzyme in citrate/phosphate buffer 50 mM, pH 5.0,. The pellet was resuspended in in citrate/phosphate buffer 50 mM, pH 5.0.

Proteomic Analysis. In order to identify proteic nature of the β-galactosidases present in the commercial Biolacta preparation, a spot from the SDS-PAGE[46] (7,5%) was selected according to the molecular relative mass for this enzyme described by Vetere and Paoleti.[42] Sample was sent to Proteomic Unit Services at Universidad Complutense. The protein was identified by peptide mass fingerprints and MS/MS sequencing analyses. Homology search of the sequences was obtained by Blast (http://www.ncbi.nlm.nih.gov/BLAST).

Glycerol-derived solvents. Glycerol-based solvents were synthesized using the same procedures previously described.[27] The alcoholysis of either epichlorohydrin (symmetrical derivatives) or the appropriate glycidol ether (unsymmetrical derivatives), gave the corresponding 1,3-dialkoxy-2-propanols. *O*-methylation of some selected targets leads to the corresponding 1,2,3-trialkoxypropanes. All the solvents were purified by distillation and used in high pure form. The complete listing of the solvents used in this work is given in Fig 1.

Effect of solvent molarity on enzyme activity. In order to determine the molarity of green solvent required for the transglycosylation reactions we measured the hydrolytic rates of the crude enzyme using solvent **1** in the following concentrations: 2, 4, 6 and 8 M. The hydrolysis measurements were performed by quantification of the *p*-nitrophenyl-β-D-galactopiranoside by UV-visible detector on HPLC.

Transglycosylation reactions. Transglycosylation reactions were carried out using commercial preparations and semipurified enzymes. To begin with, solvents from **1** to **11** were employed with a commercial crude extract of the enzyme. Then, the best solvents were chosen according to best performance and used to carry out reactions with semipurified enzymes in order to determine the effect of non-proteic substances present in Biolacta preparation in reaction yields.
A 0.17 M solution *p*NP-β-Gal (donor) 0.85 M of GlcNAc (acceptor) in 1 mL of green solvent (2 M)-buffer mixture was pre-equilibrated to 30 ℃. Afterwards, 155 μmol min⁻¹ (U) of Biolacta β-galactosidase were added to the reaction mixture. Reaction was monitored by HPLC UV-visible, and final products obtained were analyzed by HPLC-ELSD. The reaction was stopped by heating to 100 ℃ for 5 min.
The reaction mixture was then directly loaded onto a carbon-celite column. The column was first eluted with water (200 mL) and then with a linear gradient of 0% to 15% (v/v) of ethanol. Solvents were eliminated and disaccharides were dissolved in D$_2$O to be characterized by H¹ and C¹³ NMR spectroscopy.

Molecular modeling studies. Upon information available, molecular modeling studies were carried out to explain interactions between glycerol-based solvents and substrates as well as solvents-enzyme interactions.
Homology modeling. According to the results obtained with the peptide mass fingerprint, *Streptococcus pneumoniae* β-galactosidase was selected as the starting point for the molecular modeling studies. The 3D model of the domain with catalytic activity of this β-galactosidase (residues 150–979) was built by homology modeling based on available crystal structures of homologous proteins. The complete amino acid sequence of the target protein was retrieved from NCBI protein sequence database (Accession No. gi|148992245) in FASTA format. The NCBI Basic Local Alignment Search Tool (BLAST),[47] for sequence similarities was used for searching the crystal structures of the closest homologues available in the Brookhaven Protein

Data Bank (PDB). The results revealed β-galactosidase from *Bacteroides vulgatus* (PDB ID: 3GM8), with a resolution of 2.4 Å, was a suitable template (81% of homology in the selected zone and E-value 2e-104).

5 The sequences of β-galactosidases from *B. vulgatus, B. fragilis, B. thetaiotaomicron* and *E. coli,* which were the most similar to that of *S. pneumoniae*, were used for a pairwise alignment employing the Multiple Sequence Comparison by Log-Expectation (MUSCLE) tool.[48]

10 Homology modeling of the target was performed by two automated homology modeling servers: SWISS-MODEL[49] and 3D-JIGSAW[50] using as template the protein 3GM8. The first was found the best model, and was selected after performing a the 1000 steps of steepest descent (SD) energy minimization using 15 the GROMOS 96 force field and GROMACS 4.0.3 package,[51] to regularize the protein structure geometry.

The quality of the model was analyzed with the PROCHECK[52] program to calculate the main-chain torsional angles (Ramachandran plot) and it was verified using the on-line 20 software ERRAT[53] and VERIFY_3D[54] at the UCLA-DOE Structure Analysis and Verification Server (http://www.doe-mbi.ucla.edu/Services/SV/). Most of the analyzed parameters had statistical values that were in the range of those expected for a naturally folded protein and therefore, the model was considered 25 validated.

Active site search. The absence of a crystal structure of the complex forced us to deduce the location of the active center by means of conservation analysis of residues, blind docking and structure superimposition.

30 A comparative analysis of the catalytic residues of resolved structures (lacZ *E. coli*) yielded as a result a perfect conservation of two glutamic acid residues, Glu-564 and Glu-645, corresponding with residues 461 (acid catalyst) and 537 (nucleophile) in the lacZ of *E. coli*.[41] The structure of the chain A 35 of lacZ (PDB ID: 1PX3) was superimposed on the model using PyMOL. A very good superimposition of the analyzed residues could be observed depicting a typical active center of β-galactosidase in which the two glutamic residues were placed at a distance of 5.5 Å, the usual distance for this class of enzymes.

40 Finally, a blind docking using the whole structure as the grid was performed, and the results confirmed the hypothetic active site.

Molecular docking. As in previous work a two[43] dockings approach was followed. A *p*-nitrophenol-β-galactopiranose was initially docked into the active center to determine to the position 45 of the molecule to simulate the glycosyl-enzyme intermediate. With this aim, we selected the best pose according the geometrical (close enough to the nucleophile Glu-645) and energetic criteria (favorable energy) and set up a covalent link between the Glu-645 residue and the galactose.

50 Regarding the second docking, the glycosyl-enzyme was used to perform a docking with the *N*-acetyl-glucosamine (GlcNAc), the best solution was selected taking into account energetic criteria only, and it was taken as starting structure for the subsequent molecular dynamics studies.

55 All docking procedures were carried out with Autodock 4 software,[55] using a 60x60x60 grid points box around Glu-645 residue, to embrace all residues implicated in the mechanism, and with a grid spacing of 0.375 Å. A Lamarckian genetic algorithm with the standard parameters was selected.

60 *Solvent parameterization.* One of the experimental solvents, namely 1,3-bis(2,2,2-trifluoroethoxy)propan-2-ol (**5**) was used to solvate the protein and simulate the system in order to model the change in the stereoselectivity induced by the solvent and get insights into the molecular details of this process. Due to the lack 65 of parameters for **5** for the GROMOS 96 43a1 force field, a parameterization of this molecule was carried out. The structure was optimized within the density functional theory (DFT) framework, at the B3LYP/6-31G** level, using the Massively Parallel Quantum Chemistry package (MPQC).[56] All the 70 parameters (bond lengths, angles and dihedrals) and charges were extracted from the calculations or from the standard force field.[57]

To check the derived parameters, a test system was set up formed by a cubic box of 216 molecules of **5**. This box was equilibrated 300 ps at 298K and 1 atm. Following, a production simulation of 75 5 ns was carried out at the same NpT conditions recording position and energy values every 1 ps. Finally, density was calculated and compared with the experimental values,[58] confirming a good agreement (ρ_{exp}=1330 Kg m^{-3};[29] ρ_{calc}=1360 Kg m^{-3}).

80 *Molecular dynamics.* Molecular dynamics simulations were set up with the best solution of docking and the glycosyl-enzyme intermediate. Parameters adopted in GROMOS 96 43a1 were used for the protein, while the parameters required for the galactose-glutamic residue were derived in a consistent manner 85 with the force field. GlcNAc parameters were extracted from Dundee PRODRG 2.5 Server (beta) (http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg_beta) and the charges were properly checked.

An initial minimization of the complex was performed with 500 90 steps of SD followed by 1500 steps of Polak-Ribiere Conjugate Gradients (CG). The minimized complex was solvated in two different cubic boxes, fulfilling the requirement of a minimum distance of 1.2 nm between any atom of the complex and the faces of the box. The first box was filled with SPC water 95 molecules to reproduce the usual experimental conditions for the enzyme, while the second one was filled with a mixture of SPC water molecules and **5** molecules trying to reproduce the conditions for the stereochemistry change presented in this work. Then, for each box, a new minimization procedure of 500 steps of 100 SD and 3000 steps of CG was followed. The systems were equilibrated 100 ps at 298K (NVT conditions) and another 100 ps at 298K and 1 atm (NPT conditions). Finally, 1 ns production simulations were carried out.

Notes and references

aDepartment of Organic and Pharmaceutical Chemistry, Faculty of Pharmacy, Complutense University of Madrid, Campus de Moncloa, 28040 Madrid, Spain. Fax: (+) 34 9139 41822; Tel: (+) 34 91139 41821

5 *b Unidad de Bioinformática. Centro de Biología Molecular "Severo Ochoa" (CBMSO), CSIC, Universidad Autónoma de Madrid (UAM), C/ Nicolás Cabrera 1, 28049, Madrid. Spain.*

c Instituto de Ciencia de Materiales de Aragón, CSIC-Universidad de Zaragoza. C/ Pedro Cerbuna, 12, E-50009, Zaragoza. Spain.

10 *d Servicio de Interacciones Moleculares. (Pz/ Ramón y Cajal s/n. 28040 Madrid) y Servicio de Biotransformaciones Industriales (C/ Santiago Grisolía. 28760 Tres Cantos). Parque Científico de Madrid. Spain*

∗ mjhernai@farm.ucm.es (María J. Hernáiz)

15 1. P. T. W. Anastas, J. C.;, *Green Chemistry: Theory and Practice*, New York, 1998.

2. P. T. Anastas, *Chem. Revi.*, 2007, **107**, 2167-2168.

3. I. T. Horvath, P. T. Anastas, *Chem. Rev.*, 2007, **107**, 2169-2173.

4. R. W. Kates, W. C. Clark, R. Corell, J. M. Hall, C. C. Jaeger, I. 20 Lowe, J. J. McCarthy, H. J. Schellnhuber, B. Bolin, N. M. Dickson, S. Faucheux, G. C. Gallopin, A. Grubler, B. Huntley, J. Jager, N. S. Jodha, R. E. Kasperson, A. Mabogunje, P. Matson, H. Mooney, B. Moore, T. O'Riordan and U. Svedin, *Science*, 2001, **292**, 641-642.

5. J. S. Dordick, *Enzyme Microbiol. Technol.*, 1989, **11**, 194-211.

25 6. S. V. Kamat, E. J. Beckman and A. J. Russell, *J. Am. Chem. Soc.*, 1993, **115**, 8845-8846.

7. K. Ryu and J. S. Dordick, *Biochemistry*, **1992**, 31, 2588-2598.

8. S. Tawaki and A. M. Klibanov, *J. Am. Chem. Soc.*, 1992, **114**, 1882-1884.

30 9. C. R. Wescott and A. M. Klibanov, *J. Am. Chem. Soc.*, 1993, **115**, 1629-1631.

10. M. J. Hernaiz, A. R. Alcántara, J. I. García and J. V. Sinisterra, *Chem. Eur. J.*, 2010, **16**, 9422 - 9437.

11. M. S. Pérez, J. V. Sinisterra; M. J. Hernáiz, *Curr. Org. Chem.*, 2010, 35 **14**, 2366-2383.

12. A. Wolfson, C. Dlugy and Y. Shotland, *Environ. Chem. Lett.*, 2007, **5**, 67-71.

13. A. Wolfson, C. Dlugy, D. Tavor, J. Blumenfeld and Y. Shotland, *Tetrahedron: Asymmetry*, 2006, **17**, 2043-2045.

40 14. L. Andrade, L. Piovan and M. D. Pasquini, *Tetrahedron: Asymmetry*, 2009, **20**, 1521-1525.

15. M. E. C. Caines, H. Zhu, M. Vuckovic, L. M. Willis, S. G. Withers, W. W. Wakarchuk and N. C. J. Strynadka, *J. Biol. Chem.*, 2008, **283**, 31279-31283.

45 16. H. Shirato, S. Ogawa, H. Ito, T. Sato, A. Kameyama, H. Narimatsu, Z. Xiaofan, T. Miyamura, T. Wakita, K. Ishii and N. Takeda, *J. Virol.*, 2008, **82**, 10756-10767.

17. G. F. Springer, *Science*, 1984, **224**, 1198-1206.

18. G. F. Springer, P. R. Desai, W. Wise, S. C. Carlstedt, H. Tegtmeyer, 50 R. Stein and E. F. Scanlon, *Immunol Ser*, 1990, **53**, 587-612.

19. R. R. Schmidt and E. Rücker, *Tetrahedron Lett.*, 1980, **21**, 1421-1424.

20. R. R. Schmidt, *Angew. Chem. Int. Ed. Engl.*, 1986, **25**, 212-235.

21. H. Paulsen, *Angew. Chem. Int. Ed. Engl.*, 1982, **21**, 155-173.

55 22. T. Usui, S. Kubota and H. Ohi, *Carbohydr. Res.*, 1983, **244**, 315-323.

23. M. J. Hernaiz and D. H. G. Crout, *J. Mol. Catal. B. Enzym.*, 2000, **10**, 403-408.

24. A. Vetere and S. Paoletti, *Biochem. Biophys. Res. Commun.*, 1996, **219**, 6-13.

60 25. K. Sakai, R. Katsumi, H. Ohi, T. Usui and Y. Ishido, *J. Carbohydr. Chem.*, 1992, **11**, 553 - 565.

26. T. Usui, S. Morimoto, Y. Hayakawa, M. Kawaguchi, T. Murata, Y. Matahira and Y. Nishida, *Carbohydr. Res.*, 1996, **285**, 29-39.

27. J. I. García, H. García-Marín, J. A. Mayoral and P. Pérez, *Green* 65 *Chem.*, 2010, **12**, 426-434.

28. C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*, 2003.

29. D.-P. Hong, M. Hoshino, R. Kuboi and Y. Goto, *J. Am. Chem. Soc.*, 1999, **121**, 8427-8433.

70 30. A. M. P. Mainar, J.; García, J. I.; Royo, F. M.; Urieta, J. S., *J. Chem. Soc. Faraday Trans. *, 1998, **94**, 3595-3599.

31. C. Laurence, P. Nicolet and C. Reichardt, *Bull. Soc. Chim. Fr.*, 1987, 1001-1005.

32. J. Wagner and H. Grill, ed. US. Pat., 3888994., U.S.A, 1975.

75 33. C. Blake D. A., H.F., Rozman R. S., Meyer, F. J., *Toxicol. Appl. Pharm.*, 1969, **15**.

34. D. A. D. M. J. Murphy, L. S. Kaminsky, *Toxicol. Appl. Pharm.*, 1983, **71**.

35. J. Bigorra, V. Fabry, M. Perez, J. V. Sinisterra and M. J. Hernaiz. 80 European Patent EP: 08011388, 2008.

36. A. Vetere and S. Paoletti, *Biochim. Biophys. Acta - General Subjects*, 1998, **1380**, 223-231.

37. Y. Ito and T. Sasaki, *Bios. Biotech. Biochem.*, 1997, **61**, 1270-1276.

38. X. X. Zeng, R. Yoshino, T. Murata, K. Ajisaka and T. Usui, 85 *Carbohydr. Res.*, 2000, **325**, 120-131.

39. D. Zahner and R. Hakenbeck, *J. Bacteriol.*, 2000, **182**, 5919-5921.

40. R. Zeleny, F. Altmann and W. Praznik, *Anal. Biochem.*, 1997, **246**, 96-101.

41. D. H. Juers, T. D. Heightman, A. Vasella, J. D. McCarter, L. 90 Mackenzie, S. G. Withers and B. W. Matthews, *Biochemistry*, 2001, **40**, 14781-14794.

42. V. Spiwok, P. Lipovová, T. Skálová, E. Buchtelová, J. Hasek and B. Králová, *Carbohydr. Res.*, 2004, **339**, 2275-2280.

43. N. Brás, P. Fernandes and M. Ramos, *Theor. Chem. Acc.*, 2009, **122**, 95 283-296.

44. L. Fourage, M. Helbert, P. Nicolet and B. Colas, *Anal.Biochem.*, 1999, **270**, 184-185.

45. M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248-254.

46. U. K. Laemmli, *Nature*, 1970, **227**, 680-685.

100 47. T. Tatusova and T. Madden, *FEMS Microbiol. Lett.*, 1999, **174**.

48. R. C. Edgar, *Nucl. Acids Res.*, 2004, **32**, 1792-1797.

49. T. Schwede, J. R. Kopp, N. Guex and M. C. Peitsch, *Nucl. Acids Res.,* 2003, **31**, 3381-3385.

50. P. Bates, L. Kelley, R. MacCallum and M. Sternberg, *Proteins:* 105 *Struct. Funct. Bioinf.*, 2001, **45**, 39-46.

51. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4.**, 435-447.

52. R. Laskowski, M. MacArthur, D. Moss and J. M. Thornton, *J. Appl. Cryst.*, 1993, **26**, 283-291.

110 53. C. Colovos and T. Yeates, *Prot. Sci.*, 1993, **2**, 1511-1519.

54. D. Eisenberg, R. Lüthy and J. Bowie, *Methods in enzymology, 1997, 277, 396.*, 1997, 277- 396.

55. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785-2791.
56. C. Janssen, I. Nielsen, M. Leininger, E. Valeev and E. Seidl.
57. R. Suardiaz, M. Maestre, E. Suárez and C. Pérez, *J. Mol. Struct: THEOCHEM*, 2006, **778**, 21-25.
58. C. Resende and G. F. L Prado, *J. Mol. Struct: THEOCHEM*, 2007, **847**, 93-100.

5

10