Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 339 (2004) 961–966
Carbohydrate RESEARCH

The conformations of the O-specific polysaccharides of *Shigella dysenteriae* type 4 and *Escherichia coli* O159 studied with molecular mechanics (MM3) filtered systematic search

Jimmy Rosen, Armin Robobi and Per-Georg Nyholm*

Department of Medical Biochemistry/Centre for Structural Biology, Göteborg University, Medicinaregatan 7B, S405 30 Göteborg, Sweden

Received 6 July 2003; accepted 19 November 2003

Abstract—The branched O-antigens of *Escherichia coli* O159 and *Shigella dysenteriae* type 4 are structurally related and are known to show cross-reactivity with antibodies. In the present study, conformational analyses were performed on these two O-antigens using molecular mechanics MM3(96) with filtered systematic search. The results show very strong steric restrictions for the trisaccharide at the branch point of the *E. coli* O159 antigen, especially for the β-D-GlcNAc-(1 → 3)-β-D-GlcNAc linkage of the main chain. For the type 4 O-antigen the calculations show essentially a single conformation with respect to the α-D-GlcNAc-(1 → 3)-α-D-GlcNAc linkage of the main chain and three different favoured conformations for the fucose branch. Consecutive repeating units of the *S. dysenteriae* type 4 and *E. coli* O159 O-antigens form linear extended chains with significant flexibility between the branches. Comparative calculations carried out with the SWEET server indicate that our method of filtered systematic search is a superior method in the case of branched, constrained oligosaccharides. Based on the results of the MM3 calculations, we propose that the common epitope explaining the cross-reactivity comprises the fucose branch, the downstream GlcNAc and part of the uronic acid.

Keywords: Conformation; Molecular mechanics; O-Antigen; Epitope; *Shigella*

1. Introduction

Escherichia coli O159 is a serogroup among the ETEC strains, which gives rise to diarrhoea in infants in developing countries.[1] The structure of the O-antigen of *E. coli* O159 was recently determined based on NMR analysis and degradation studies.[2] It was found that the *E. coli* O159 structure is closely related to the O-antigen of *Shigella dysenteriae*[3] type 4. This serotype of *Shigella* is known to be a minor etiologic agent for shigellosis.[4] As shown in Figure 1, there are four differences between the structures of the O-antigens of *E. coli* O159 and *S. dysenteriae* type 4: the anomeric configuration of two consecutive D-GlcNAc residues in the main chain, the configuration at C-4 of a downstream hexuronic acid and the O-acetylation of the fucose branch of the type 4 antigen. Conformational studies on the two O-antigens have so far not been reported in the literature. The O-antigens of *S. dysenteriae* type 1 and type 2 were recently studied with molecular mechanics (MM3) using an approach with filtered systematic search.[5,6] In the present

E. coli O159

α-L-Fuc*p*
|
4
[-3)-β-D-Glc*p*NAc-(1-3)-β-D-Glc*p*NAc-(1-4)-α-D-Gal*p*A-(1-3)-α-L-Fuc*p*-(1-]

S. dysenteriae type 4:

O-Ac-3/4-α-L-Fuc*p*
|
4
[-3)-α-D-Glc*p*NAc (1-3)-α-D-Glc*p*NAc-(1-4)-α-D-Glc*p*A-(1-3)-α-L-Fuc*p*-(1-]

Figure 1. Structural formulae of the repeating units of the O-specific polysaccharides of *E. coli* O159 and *S. dysenteriae* type 4.

* Corresponding author. Fax: +46-31-823758;
e-mail: nyholm@medkem.gu.se

0008-6215/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2003.11.018

study we have used the same method to investigate the favoured conformations of the O-antigens of *E. coli* O159 and *S. dysenteriae* type 4.

2. Methods

Structural formulae of the repeating units of the O-antigens of *E. coli* 159 and *S. dysenteriae* type 4 are shown in Figure 1. The *O*-acetyl group of the *S. dysenteriae* type 4 antigen was excluded from the calculations since the position of this substitution of the fucose branch is not known. The notation used for the torsion angles of the glycosidic linkages is $\phi = $ O5–C1–O-1–CX, $\psi = $ C-1–O-1–CX–C(X+1) with $X = 3$ and 4 in the studied $1\rightarrow3$ and $1\rightarrow4$ glycosidic linkages, respectively.

Initial calculations were carried out with the GLYCAN program,[5] which allows rapid screening of the conformational space of medium sized oligosaccharides. However, GLYCAN calculations indicated that the branches in both the studied O-antigens were highly constrained and it was considered necessary to perform full minimizations with molecular mechanics. The MM3(96) program (Allinger, University of Georgia) was used to calculate relaxed potential energy ϕ/ψ maps for the di- and trisaccharide moieties forming the branching points of the two repeating units. The disaccharides were calculated with driver option 4, which ensures that the starting conformation at each ϕ/ψ point is recalculated with correct geometry for the rings and the pendant groups.[7] Different combinations of starting conformations for the C-6 hydroxymethylene groups were considered as described by French et al.[7] and the lowest energy value obtained at each ϕ/ψ point was used in the generation of adiabatic potential energy ϕ/ψ maps with a resolution of 15°. In the case of the trisaccharide fragments four-dimensional MM3 driver calculations were carried out for the ϕ/ψ torsions of the constituent glycosidic linkages. The filtered input data for these calculations was generated with an in-house developed program selecting ϕ/ψ points within a range of 12 kcal/mol from the global minimum of the adiabatic energy map of the corresponding disaccharide units.[5,6] In the case of MM3 calculations on branch trisaccharides the two favoured rotamers of the hydroxymethylene group and the *r* orientation of the hydroxyl groups were investigated forming a total of 167,479 conformations in the case of type 4 and 251,165 in the case of *E. coli* O159. All the MM3 calculations were performed with a dielectric constant ε of 80 to simulate an aqueous medium. The carboxyl group of the uronic acid was treated as a deprotonated group using the preliminary parameters developed by Lii and Allinger for the charged carboxylate oxygen.[8] The calculations were carried out on five computers with dual AMD 2200+ processors using the Linux version of MM3(96). The total CPU time required for this study was about 360 h. Visualization of molecular models including the generation of Connolly surfaces and calculations of Poisson–Boltzmann electrostatic surfaces was performed with Sybyl version 6.8 (Tripos Inc.).

3. Results and discussion

3.1. Systematic search with MM3

The MM3 adiabatic energy maps of the glycosidic linkages contained in the *E. coli* O159 O-antigen are shown in Figure 2. It is apparent that the β-D-GlcNAc-$(1\rightarrow3)$-β-D-GlcNAc glycosidic linkage has a global minimum at $\phi/\psi = -75°/135°$ in the disaccharide and the conformation about this linkage is more restricted in the case of the branch trisaccharide. The restrictions are primarily due to close contacts between the *N*-acetyl groups of the GlcNAc residues. The adiabatic map of the α-L-Fuc-$(1\rightarrow4)$-β-D-GlcNAc disaccharide shows a wide energy valley with an energy minimum at $\phi/\psi = -90°/-165°$. The energy map is in excellent agreement with earlier MM3 results on the same disaccharide obtained by Imberty et al.[9] In the trisaccharide however, the adiabatic map of the α-L-Fuc-$(1\rightarrow4)$-β-D-GlcNAc linkage shows a well defined energy minimum dislocated towards $\phi/\psi = -75°/-105°$. Thus, the steric interference from the upstream β-D-GlcNAc residue causes a significant displacement of the global minimum found for the corresponding disaccharide. The β-D-GlcNAc-$(1\rightarrow4)$-α-D-GalA disaccharide shows a major energy minimum at $\phi/\psi = -75°/150°$. The α-D-GalA-$(1\rightarrow3)$-α-L-Fuc disaccharide shows a minimum energy trough at $\phi = 90°$ extending in the ψ-dimension from $-165°$ to $-45°$. The α-L-Fuc-$(1\rightarrow3)$-β-D-GlcNAc disaccharide similarly shows a low energy valley at $\phi = -90°$ extending over a wide range of ψ-values.

The MM3 adiabatic energy maps of the disaccharide moieties present in the O-antigen of *S. dysenteriae* type 4 are shown in the lower row in Figure 3. The α-D-glycosidic linkage of α-D-GlcNAc-$(1\rightarrow3)$-α-D-GlcNAc shows two energy minima, the global minimum at $\phi/\psi = 90°/90°$ and a secondary minimum at $\phi/\psi = 105°/165°$. These minima are separated by an energy barrier similar to what was found for the α-D-Gal-$(1\rightarrow3)$-α-D-GlcNAc disaccharide moiety of the *S. dysenteriae* type 1 antigen[5] and the α-D-GalNAc-$(1\rightarrow3)$-α-D-GalNAc sequence in the O-antigen of *S. dysenteriae* type 2.[6] In all three cases the barrier is due to the collision of the hydroxymethylene group of the upstream residue with the *N*-acetyl group of the downstream residue in the disaccharide moiety. The global minimum energy ϕ/ψ region in the case of the α-D-GlcNAc-$(1\rightarrow3)$-α-D-GlcNAc disaccharide of the type 4

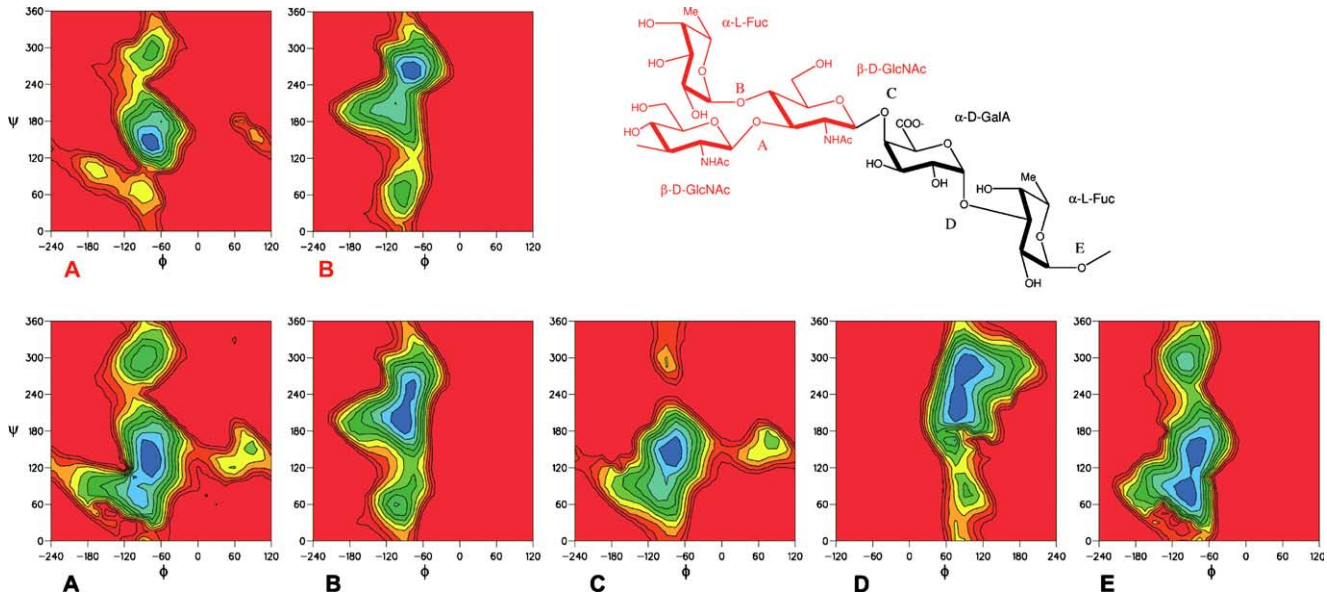

Figure 2. MM3 adiabatic energy maps for the glycosidic linkages of the different disaccharide moieties (lower row) as well as the trisaccharide branch (upper row) contained within the structure of the *E. coli* O159 O-antigen. The labels for the linkages are defined in the schematic drawing in the upper right corner. Both the β-D-GlcNAc-(1→3)-β-D-GlcNAc and the α-L-Fuc-(1→4)-β-D-GlcNAc linkages are significantly restricted in the trisaccharide structure. The starting geometry at the point −105°/105° in the contour plot for β-D-GlcNAc-(1→3)-β-D-GlcNAc was manually modified (±20° on the H2–C2–N2–H torsion angle) to alleviate clashing of the *N*-acetyl groups. Energy contours are shown at each kcal/mol starting from the global energy minimum (blue).

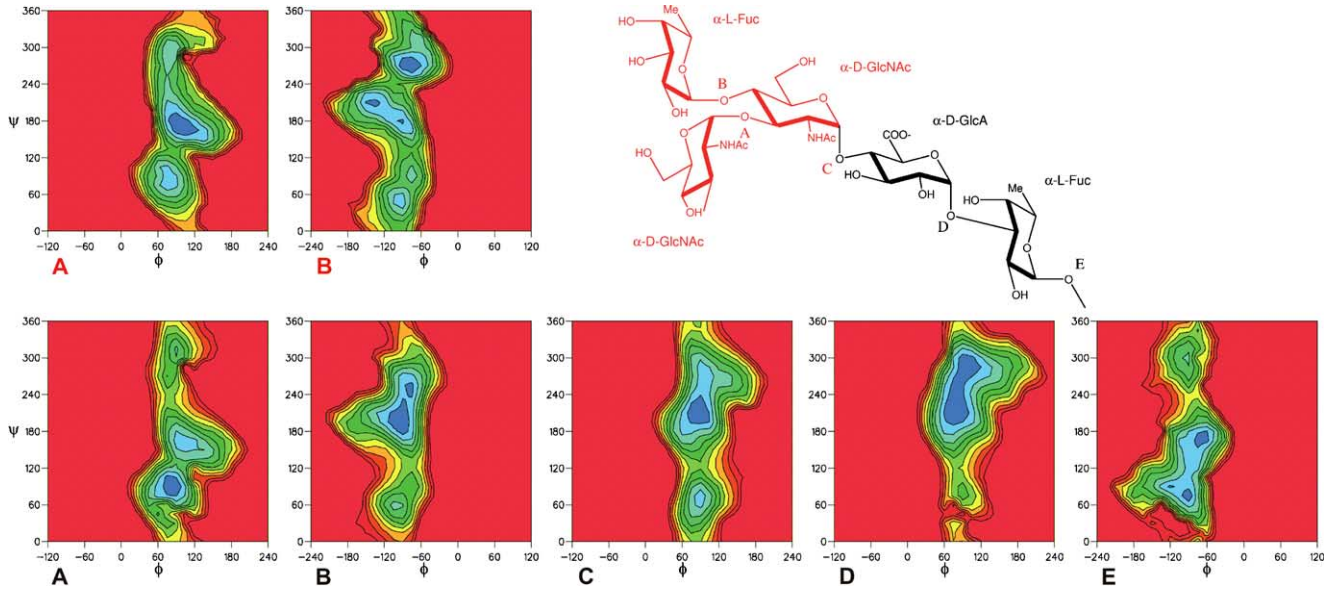

Figure 3. MM3 adiabatic energy maps for the glycosidic linkages of the different disaccharide moieties (lower row) as well as in the branch trisaccharide (upper row) of the repeating unit of the *S. dysenteriae* type 4 O-antigen. For the α-D-GlcNAc-(1→3)-α-D-GlcNAc linkage there are two energy minima separated by a barrier. In the disaccharide the minimum at 90°/90° is preferred whereas the minimum at 105°/165° is preferred in the trisaccharide. For the α-L-Fuc-(1→4)-β-D-GlcNAc linkage there is a trough in the case of the disaccharide but three distinct minima in the trisaccharide. The α-D-GlcNAc-(1→4)-α-D-GlcA linkage is fairly constrained already in the disaccharide. The α-D-GlcA-(1→3)-α-L-Fuc and α-L-Fuc-(1→4)-β-D-GlcNAc linkages are more flexible. Labelling, energy contours and colouring are as in Figure 2.

antigen is in agreement with the favoured conformations of the mentioned α-D-(1→3) linkages in *S. dysenteriae* type 1 and 2. Furthermore, this conformation is supported by experimental data on the conformation of the α-D-GalNAc-(1→3)-β-D-GalNAc disaccharide moiety studied by NMR and X-ray methods in the Forssman antigen.[10,11] The α-L-Fuc-(1→4)-α-D-GlcNAc disaccharide contained in the branch shows a low energy

region at $\phi = -90°$ extending from $\psi = 180°$ to $-90°$. This is in agreement with the results obtained for the structurally related disaccharide α-L-Fuc-(1→4)-β-D-GlcNAc (Fig. 2B, lower panel). The α-D-GlcNAc-(1→4)-α-D-GlcA disaccharide shows a major energy minimum at $\phi/\psi = 90°/-150°$. The α-D-GlcA-(1→3)-α-L-Fuc disaccharide shows a minimum energy trough at $\phi = 75°$ extending in the ψ-dimension from $-165°$ to $-60°$. The α-L-Fuc-(1→3)-α-D-GlcNAc disaccharide shows two minima at $\phi = -90°/75°$ and $\phi = -75°/165°$.

In the branch trisaccharide of the type 4 antigen, the α-D-GlcNAc-(1→3)-α-D-GlcNAc linkage (Fig. 3A, upper row) shows two different energy minima but the relative energy ranking of these minima is opposite from the situation with the disaccharide. For the trisaccharide, the global minimum of this linkage is at $\phi/\psi = 105°/165°$ with the secondary minimum at $\phi/\psi = 75°/75°$. Apparently, this shift of the energy order of the minima for this linkage is due to interference from the adjacent α-L-Fuc residue. The adiabatic map of the α-L-Fuc-(1→4)-α-D-GlcNAc linkage in the trisaccharide branch shows three major energy minima, one at $\phi/\psi = -75°/-90°$ and the others at $-135°/-150°$ and $-90°/180°$. There is a significant energy barrier (about 3 kcal/mol) between the former minimum and the two latter. This barrier is apparently caused by the presence of the α-D-GlcNAc residue since the energy map of the α-L-Fuc-(1→4)-α-D-GlcNAc disaccharide linkage shows a single wide energy trough in this region.

3.2. Comparative calculations with SWEET

Calculations with the SWEET server[12] on the branched trisaccharides showed that the results from SWEET depend on whether the branch was entered above or below the main chain. In the case of *S. dysenteriae* type 4 two SWEET runs with different ordering of the input resulted in two widely different conformations for the α-D-GlcNAc-(1→3)-α-D-GlcNAc linkage, roughly corresponding to the lowest and second lowest minimum obtained with our MM3(96) calculations. The ϕ/ψ values obtained with SWEET are listed in Figure 4. It thus appears that SWEET has problems in performing a thorough conformational search in the case of branched structures. We conclude that the present version of SWEET should be used with caution in the case of sterically restricted oligosaccharides, for example, structures with vicinal branches.

3.3. Modelling of the repeating units

For each of the studied O-antigens a model of the repeating unit and one additional downstream residue was built using the data from the search on the branch trisaccharides and the favoured conformations of the disaccharides for the other glycosidic linkages. The final minimum energy conformations obtained after MM3 minimization are shown in Figure 5. Through visual inspection in Sybyl of the favoured conformations of the type 4 structure, it was observed that the *O*-acetyl group at the 3 or 4-position of the fucose branch points away from the rest of the structure and is unlikely to have significant effects on the overall conformation of the repeating unit. Based on the conformations of the repeating units, models of four consecutive repeating units were generated (Fig. 6). These sequences showed extended, almost linear conformations although the *E. coli* O159 antigen shows rather sharp bends at each branch point. It should be observed that there is a considerable flexibility in these structures in between the branches.

***E. coli* O159:**

α-L-Fuc*p* **-72°/-104°**, *-75°/-88°* or *-76°/-109°*

-87°/74° *-57°/-170°* or -71°/137° | **-76°/149°** -166°/93° or *-72°/128°* | **-91°/149°**, *-81°/120°* or *-78°/131°* | **89°/-73°**, *106°/-153°* or *71°/-100°*

4

[-3)-β-D-Glc*p*NAc-(1-3)-β-D-Glc*p*NAc-(1-4)-α-D-Gal*p*A-(1-3)-α-L-Fuc*p*-(1-]

***S. dysenteriae* type 4:**

α-L-Fuc*p* **-72°/-89°**, *-112°/-79°* or *-69°/-89°*

-87°/74°, *-68°/168°* or *-57°/-169°* | **89°/179°**, *76°/96°* or 145°/153° | **89°/-150°**, *148°/-106°* or *136°/-102°* | **74°/-148°**, *101°/-148°* or *102°/-150°*

4

[-3)-α-D-Glc*p*NAc-(1-3)-α-D-Glc*p*NAc-(1-4)-α-D-Glc*p*A-(1-3)-α-L-Fuc*p*-(1-]

Figure 4. Values in bold show the ϕ/ψ minimum energy conformations of the repeating units of the O-antigens of *E. coli* O159 and *S. dysenteriae* type 4 obtained by MM3 starting from the global minima found in the systematic search. Values in italics refer to the results obtained with the SWEET server. Two values are given since the SWEET server reports different values depending on whether the fucose branch is entered above or below the main chain.

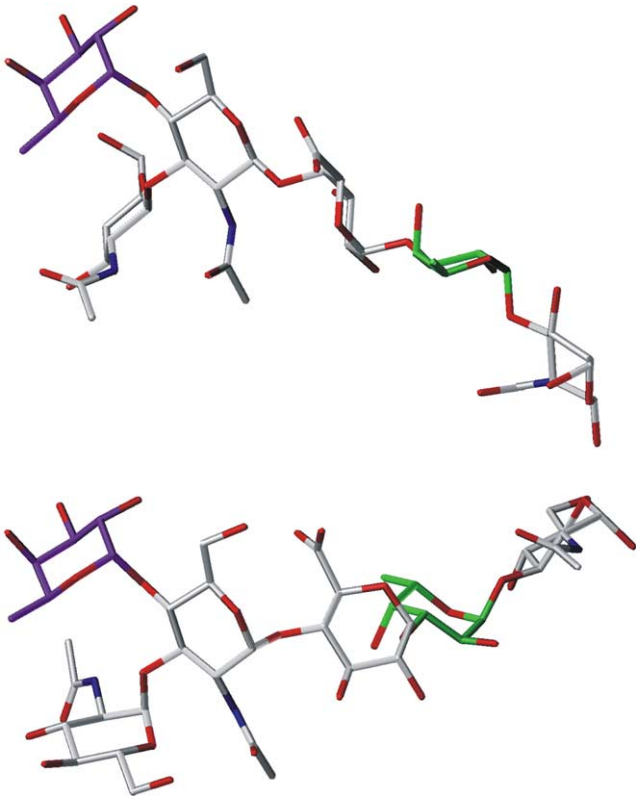

Figure 5. Minimum energy conformations of one repeating unit and one additional downstream residue of the O-antigens of *E. coli* O159 (top) and of *S. dysenteriae* type 4 (bottom). The carbons of the fucose branch are shown in violet, while the carbons of the downstream fucose are shown in green, respectively.

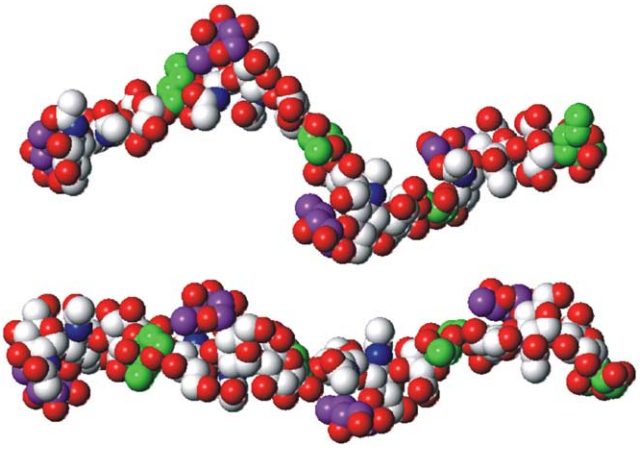

Figure 6. Models of a sequence of four repeating units of the O-antigens of *E. coli* O159 (top) and *S. dysenteriae* type 4 (bottom). The conformations are based on the minimum energy conformations of the respective repeating unit. Colouring as in Figure 5.

3.4. Common epitopes and cross-reactivity

Both the studied O-antigens have α-L-Fuc branches (1→4) linked to GlcNAc and it appeared possible that this moiety gives rise to a common epitope, which can explain the observed cross-reactivity.[2] In fact the global minimum energy conformation of the *S. dysenteriae* type 4 antigen has the α-L-Fuc-(1→4)-GlcNAc branch in the ϕ/ψ conformation −75°/−90°, which is similar to what it observed for the fucose branch in the *E. coli*

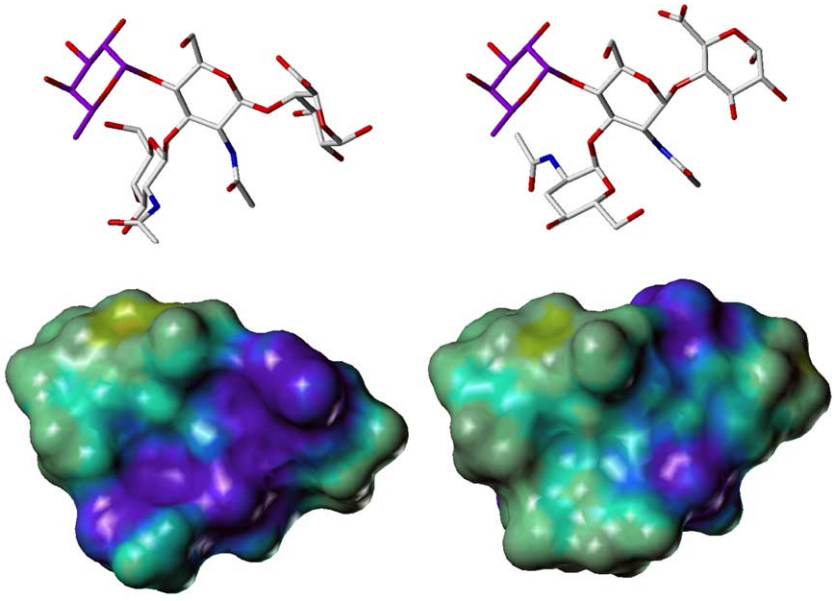

Figure 7. Top: Stick models of the global minimum energy conformations of the branch tetrasaccharides of *E. coli* O159 (left) and *S. dysenteriae* type 4 (right). Bottom: Connolly surfaces with colour coding according to the Poisson–Boltzmann electrostatic potential for the tetrasaccharides shown above. Blue indicates negative potential while brown indicates positive values. It is suggested that a common epitope is formed by the fucose branch, part of the adjacent GlcNAc and the carboxyl group of the downstream uronic acid explaining the cross-reactivity.

O159 antigen. Molecular surfaces of the global minimum energy conformations of the two branch tetrasaccharides were generated and colour coded with respect to electrostatic potential (Fig. 7). It appears that there is a common epitope comprising the fucose branch, the downstream GlcNAc and also the carboxylate group of the uronic acid, despite the fact that this latter residue is linked in different configurations in the two antigens. The flexibility for the fucose branch in the case of the *S. dysenteriae* type 4 antigen (Fig. 3) could of course limit the occurrence of the suggested epitope. Since our MM3 results regarding the *S. dysenteriae* type 4 antigen suggest the presence of several conformations, especially for the fucose branch, experimental studies will be required to determine the populations of different conformations of potential relevance for the proposed epitope in this antigen.

Acknowledgements

This work was supported by a grant from the Swedish Medical Research Council (006). Rickard Nordström at Biognos AB, Göteborg, is acknowledged for proofreading the manuscript.

References

1. Nataro, J. P.; Kaper, J. B. *Clin. Microbiol. Rev.* **1998**, *11*, 142–201.
2. Linneborg, M.; Weintraub, A.; Widmalm, G. *Eur. J. Biochem.* **1999**, *266*, 246–251.
3. Dmitriev, B. A.; L'Vov, V. L.; Kochetkov, N. K.; Hofman, I. L. *Bioorg. Chim.* **1977**, *3*, 1226–1233.
4. Kotloff, K. L.; Winickoff, J. P.; Ivanoff, B.; Clemens, J. D.; Swerdlow, D. L.; Sansonetti, P. J.; Adak, G. K.; Levine, M. M. *Bull. WHO* **1999**, *77*, 651–666.
5. Nyholm, P. G.; Mulard, L. A.; Miller, C. E.; Lew, T.; Olin, R.; Glaudemans, C. P. J. *Glycobiology* **2001**, *11*, 945–955.
6. Rosen, J.; Robobi, A.; Nyholm, P.-G. *Carbohydr. Res.* **2002**, *337*, 1633–1640.
7. French, A. D.; Tran, V. H.; Perez, S. In *Computer Modelling of Carbohydrate Molecules*. ACS Symp. Ser. 430; American Chemical Society: Washington D.C., 1990. p 191–212.
8. Lii, J. H.; Allinger, N. L. *J. Comp. Chem.* **1999**, *12*, 186–199.
9. Imberty, A.; Mikros, E.; Koča, J.; Mollicone, R.; Oriol, R.; Perez, S. *Glycoconjugate J.* **1995**, *12*, 331–349.
10. Grönberg, G.; Nilsson, U.; Bock, K.; Magnusson, G. *Carbohydr. Res.* **1994**, *257*, 35–54.
11. Hamelryck, T. W.; Loris, R.; Bouckeart, J.; Dao-Thi, M. H.; Strecker, G.; Imberty, A.; Fernandez, E.; Wyns, L.; Etzler, M. *J. Mol. Biol.* **1999**, *286*, 1161–1177.
12. Bohne, A.; Lang, E.; von der Lieth, C. W. *J. Mol. Model.* **1998**, *4*, 33–43.