Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 340 (2005) 1059–1064
Carbohydrate RESEARCH

The use of a genetic algorithm search for molecular mechanics (MM3)-based conformational analysis of oligosaccharides

Abraham Nahmany,[a,†] Francesco Strino,[a,†] Jimmy Rosen,[a] Graham J. L. Kemp[b] and Per-Georg Nyholm[a,*]

[a]*Department of Medical Biochemistry, Göteborg University, Medicinaregatan 7B, Box 440, SE-405 30 Göteborg, Sweden*
[b]*Department of Computing Science, Chalmers University of Technology, SE-412 96 Göteborg, Sweden*

Received 24 September 2004; accepted 27 December 2004

Dedicated to Professor David A. Brant

Abstract—We have implemented a system called GLYGAL that can perform conformational searches on oligosaccharides using several different genetic algorithm (GA) search methods. The searches are performed in the torsion angle conformational space, considering both the primary glycosidic linkages as well as the pendant groups (C-5–C-6 and hydroxyl groups) where energy calculations are performed using the MM3(96) force field. The system includes a graphical user interface for setting calculation parameters and incorporates a 3D molecular viewer. The system was tested using dozens of structures and we present two case studies for two previously investigated O-specific oligosaccharides of the *Shigella dysenteriae* type 2 and 4. The results obtained using GLYGAL show a significant reduction in the number of structures that need to be sampled in order to find the best conformation, as compared to filtered systematic search.

Keywords: Genetic algorithms; Molecular mechanics; Oligosaccharide; O-antigen; *Shigella dysenteriae*

1. Introduction

Saccharides linked to proteins and lipids cover a large fraction of the surface area of most cells. Many of these saccharides are involved in specific recognition processes. To understand their biological function in detail it is necessary to have information about their three dimensional (3D) structure. Knowledge about the 3D structure of oligosaccharides also has medical applications, for example, in the design of vaccines targeted at surface saccharides of bacteria.[1–4]

The two main experimental techniques for determining the 3D structures of molecules—NMR spectroscopy and X-ray crystallography—are often difficult to apply to oligosaccharides. X-Ray is problematic since it is difficult to obtain crystals of suitable quality. NMR determination is difficult due to the paucity of well defined NOEs. An alternative approach for oligosaccharide conformational analysis is to search the space of possible conformations using computational methods to find favorable low energy conformations.

The O-specific oligosaccharide 3D structure of *Shigella dysenteriae* type 1 has been the subject of experimental as well as computational studies.[3,5] An earlier study by Rosen et al. used filtered systematic search to predict favorable conformations of the O-specific oligosaccharide of *S. dysenteriae* type 2 and *S. dysenteriae* type 4.[6,7]

A major problem with computational conformational analysis of oligosaccharides is achieving a good trade-off between the sampling of the conformational space and the required computation time.

Here we present the results obtained using the newly developed GLYGAL program coupled to MM3,[8]

* Corresponding author. Tel.: +46 (0)31 773 34 54; fax: +46 (0)31 823758; e-mail: nyholm@medkem.gu.se
† Contributed equally to this paper.

0008-6215/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2004.12.037

concerning *S. dysenteriae* type 2 and *S. dysenteriae* type 4 and a comparison with the results obtained by filtered systematic search.

2. Methods

The basic ideas for predicting oligosaccharide conformations using a standard genetic algorithm[9–11] are:

1. Initial population of randomly generated conformations ('individuals' or 'chromosomes') with respect to the torsion angles of the glycosidic linkages and pendant groups ('genes'). Each such 'chromosome' is represented as a vector of real-numbers for the torsion angles. The torsions of the rings are not randomized.
2. Evaluation using molecular mechanics MM3 as fitness function. Selection is carried out using the roulette wheel method, where individuals are evaluated as a function of their conformational energy.
3. Standard genetic operators like mutation and crossover are used to generate offspring.
4. Termination criteria are satisfied either after a fixed number of generations or when no improvement has occurred during several generations.

The major goal was to turn these ideas into easy-to-use software for oligosaccharide conformational search. The software which was developed was called GLYGAL. The program implements three different GAs for the purpose of oligosaccharide conformational search: standard GA, parallel GA, and an evolutionary programming algorithm.[9] All the GAs implemented can be used with local minimization and propagation of the minimized geometry to the progeny, namely Lamarckian GA.[9,12] Some default GA parameters, such as population size, number of iterations, etc. are suggested to the user and those can easily be set using the GLYGAL graphical user interface.

GLYGAL also requires a template file containing the coordinates. Pdb, xyz and MM3 files are currently supported. The following steps describe the course of events of the conformational search:

1. The torsion angles to be modified in the search are assigned automatically.
2. The template structure file is copied as many times as the size of the initial population of the GA. The copied structures are modified by torsion-angle adjustments to create the first randomly generated population of the GA.
3. The files are then sent to the MM3 program for evaluation and local minimization, and the results (i.e., energies and geometries) are sent back to GLYGAL.
4. The structures to be manipulated by the genetic algorithm operators are selected randomly using a roulette wheel method.
5. Genetic algorithm operators, such as mutation and crossover are performed on the structures and the next generation of structures is created. In the case of parallel GA a migration operator is also involved.
6. The structures are evaluated and termination criteria are checked. If not fulfilled, the process will resume from step 3.

As mentioned, GLYGAL uses the MM3 (96)[8,13] forcefield program for the energy evaluation and local minimization. The MM3 calculations are distributed on a Linux cluster (Csol Hoborg) of five nodes, each with dual 2200+ AMD processors.

One of the main problems with the existing methods for oligosaccharide modeling is the need of manual preprocessing, that is, to identify the torsion angles to be searched. In GLYGAL we solved this problem by developing an algorithm for identifying the torsion angles automatically. This algorithm creates a connection matrix containing the torsion angles for each glycosidic linkage and, on the diagonal, a ring vector containing all the torsion angles needed to identify the position of all atoms within a certain ring.

The pendant groups, that is, the C-5–C-6 and the hydroxyl groups, respectively, are included in the GA search just as the primary torsions of the glycosidic linkages. Since these minor torsions have less impact on the energy of the structure their sampling can be weighted accordingly, to achieve a thorough but efficient sampling of the conformational space. The systematic search[3,6,7] uses the simplification that the hydroxyls align in chains of weak hydrogen bonds[14,15] to reduce the number of dimensions to sample. This is not necessary when using GA search, as the algorithmic complexity scales well with higher dimensionality. It is thus possible to sample the hydroxyls individually within our computational capacity, which is to be preferred.[16]

The pendant groups add significantly to the number of sample dimensions. However, since the GA search scales well to increasing dimensionality it does not introduce too much of a computational burden. To further reduce the computational load and increase the sampling efficiency, the search space can be divided into layers and be successively refined as the search progresses. Another method, available in GLYGAL, is to attach different sampling 'weights' to the search dimensions, reflecting their impact on the conformational energy. Each individual (or 'chromosome') is represented as a vector of torsion angles of the glycosidic linkages and pendant groups ('genes'). For example, a vector representing a trisaccharide will contain five genes: one gene for each glycosidic linkage and one gene for the pendant groups of each residue. The weight values indicate the probability of the