Pergamon

Bioorganic & Medicinal Chemistry Letters 12 (2002) 1031–1034

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

Optimization of the Enzymatic Synthesis of *O*-Glycan Core 2 Structure by Use of a Genetic Algorithm

Christoph Hoh, Gregor Dudziak and Andreas Liese*

Forschungszentrum Jülich GmbH, Institut für Biotechnologie 2, D-52425 Jülich, Germany

Received 9 November 2001; accepted 31 January 2002

Abstract—The enzymatic synthesis of Gal-β1,3[GlcNAc-β1,6]-GalNAc-α1-OBn (core 2-Bn) using a multi-enzyme system consisting of a β-galactosidase (EC 3.2.1.23) from bovine testes and a recombinant core 2 β1,6-GlcNAc transferase (C2GnT, EC 2.4.1.102) was empirically optimized by the use of a genetic algorithm. After variation of seven relevant parameters and performance of 56 experiments, two local maxima regarding the selection criterias could be found after four generations of optimization. The selectivity of core 2-Bn formation showed values up to 90%.

Core 2-based *O*-glycan structures play a key role in mucin-type *O*-glycosylation patterns.[1–4] These branched oligosaccharides attached to the respective protein backbone are involved in a variety of different biological processes.[5] The core 2 trisaccharide is biosynthesized by the enzyme family of the β1,6-GlcNAc transferases (C2GnT). Three homologous C2GnTs have been cloned and characterized so far.[6–8] Since the acceptor substrate of all these enzymes for synthesis of the core 2 trisaccharide is the core 1 [Gal-β1,3-GalNAc-α1-OR] disaccharide, several approaches have been made for its synthesis.[9–11] Despite the very recent cloning and expression of the cDNA of the appropriate β1,3-galactosyltransferase,[12] the use of a β-galactosidase from bovine testes proved to be a potent alternative method for generating this structure. Several approaches using this enzyme effectively have been made including the combined use of the β-galactosidase with a sialyltransferase in order to avoid product hydrolysis[13,14] or the use of oversaturated solutions of the donor molecule.[15]

In earlier works we could show that the GlcNAc-β1,6-GalNAc-α1-OBn structure (core 6-Bn, **4**) can be synthesized by combination of the β-galactosidase from bovine testes and a recombinant C2GnT[16] in yields >90% starting from GalNAc-α1-OBn (**1**) as a model compound for serine/threonine linked *O*-glycans.[17–19] However, the core 6 disaccharide is the product of core 2 trisaccharide hydrolysis occuring under these reaction conditions.

The intention of our work was to optimize this multi-enzyme system in that way that the secondary hydrolysis of the core 2-Bn by the β-galactosidase is discriminated and that formation of the undesired core 6-Bn is avoided. Figure 1 shows the simplified reaction scheme.

In the first step, the core 1-Bn (**2**) is formed by transfer of the galactose residue from lactose (donor) on the GalNAc-α1-OBn acceptor (**1**). This step is catalyzed by the β-galactosidase. The core 2-Bn (**3**) is then formed by transfer of the *N*-acetylglucosamine (GlcNAc) residue from UDP-GlcNAc to the core 1-Bn (**2**) by C2GnT. Eventual transglycosylation products other than core 1-Bn formed by β-galactosidase catalysis (e.g., 1,4-linked disaccharides) are not accepted by C2GnT. Undesired secondary hydrolysis of the core 1-Bn and the core 2-Bn are catalyzed by the β-galactosidase to form the core 6-Bn (**4**) and the GalNAc-α1-OBn (**1**), respectively. Calf intestine alkaline phosphatase (CIAP; EC 3.1.3.1) was added as a third enzyme in order to overcome inhibition of C2GnT by UDP.

Because of the high complexity of the reaction system due to many cross inhibitions and the poor regioselectivity of the β-1,3 linkage by the β-galactosidase,[9,10] we decided to optimize this synthesis by the use of a genetic algorithm.

*Corresponding author. Tel.: +49-2461-616044; fax: +49-2461-613870; e-mail: a.liese@fz-juelich.de

0960-894X/02/$ - see front matter
PII: S0960-894X(02)00087-2

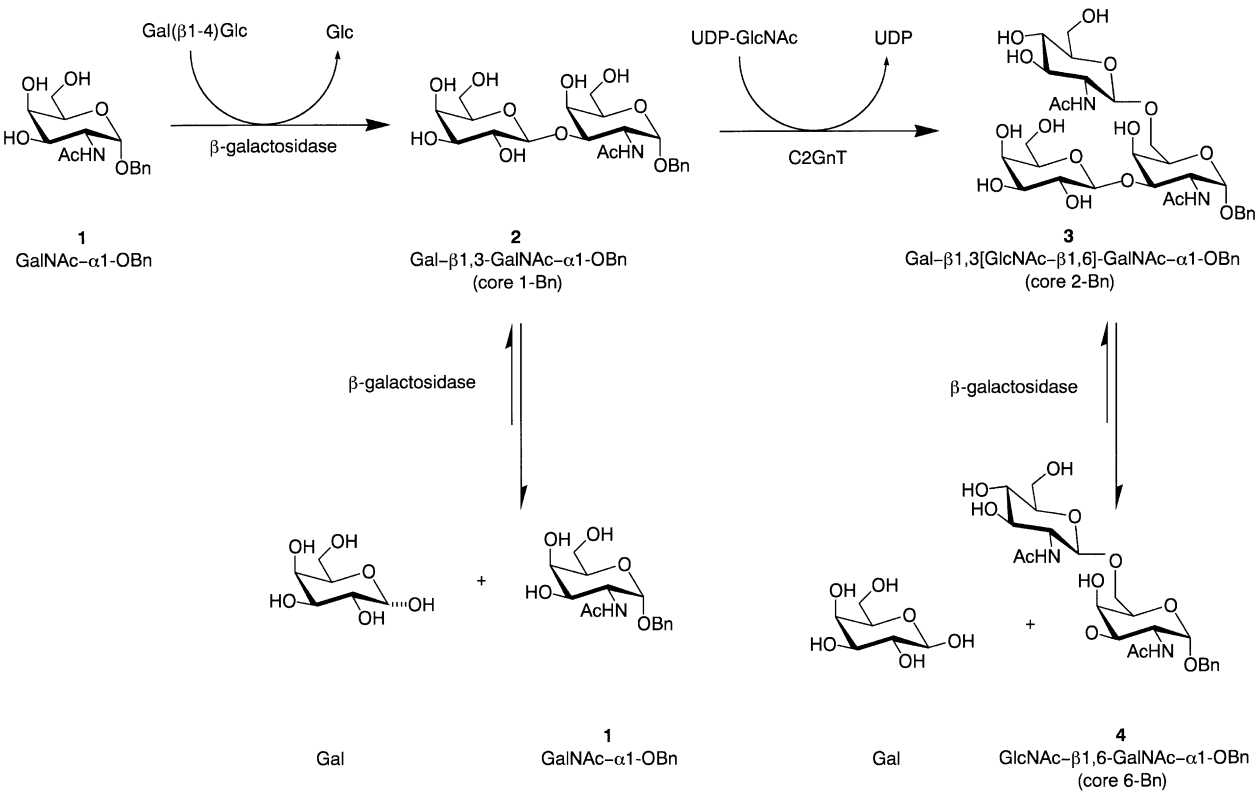

Figure 1. Simplified reaction scheme of the synthesis of *O*-glycan core structures (cores 1, 2 and 6).

Genetic algorithms (GAs) are non-model based 'black-box' optimization methods with the ability to search a large parameter space in a highly directed way.[20,21] The major strengths of GAs are that optimization is possible without initial guesses per se, that no derivatives of the objective function are needed and the fact that GA optimization is independent from the choice of the starting point. This makes it a candidate for experimental optimization of complex reaction or fermentation systems.[22]

The motivation of this work was to find reaction conditions suitable to generate the desired product in high concentration and with high selectivity. Therefore, we chose the following two selection criterias: (1) maximum core 2-Bn concentration and (2) high selectivity of core 2-Bn formation expressed as the ratio of the concentrations of core 2-Bn to core 6-Bn at the point of the maximum product concentration.

Table 1. Parameters of optimization and their range of variation

Parameter	Dimension	Range of variation	Stepwidth[a]
GalNAc-Bn	mM	10–40	10 (4)
Lactose	mM	100–800	100 (8)
UDP-GlcNAc	mM	10–40	5 (7)
β-galactosidase	mU/mL[b]	250–1000	125 (7)
C2GnT	mU/mL[b]	50–250	50 (5)
pH	—	5.0–7.5	0.5 (6)
Temperature	°C	10–50	10 (5)

[a]The values given in parenthesis indicate the number of the values possible for each parameter. The product of these values states the number of total possible experiments: $4\times7\times8\times7\times5\times6\times5=235{,}200$.
[b]Enzyme activities refer to the respective standard assay.

In total, seven parameters were varied throughout the four generations of optimization. The respective parameters and their range of variation are summarized in Table 1.

As shown in Figure 2, a set of parameters (substrate concentrations, enzyme activities, pH and temperature) is varied for each generation (input). The experiments of each generation are carried out in parallel. After each batch reaction (representing the 'black box') the obtained results regarding the selection criterias (output) are then judged and the genetic algorithm and gives origin to new improved parameters. The first set of parameter values was chosen randomly considering the respective range of parameters. GAs are based on three

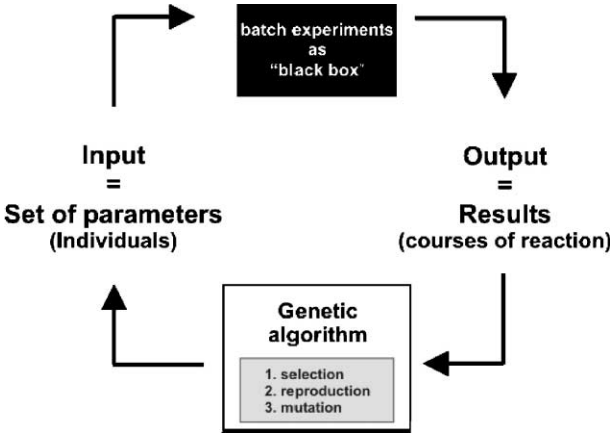

Figure 2. Principle of optimization by the use of a genetic algorithm.

major processes:[18] (1) selection, (2) reproduction and (3) mutation.

All processes regarding the GA were realized with the GALOP software, developed in-house.[23] The best individual for each generation was taken over, without change, into the next generation. Each generation consisted of 14 experiments (individuals), double the amount of parameters to be varied.

All compounds of interest were quantified by HPLC analysis (conditions: Aminex HPX-87-H column, 6 mM H_2SO_4, 0.6 mL/min, $T = 65\,^{\circ}C$, UV-detection at $\lambda = 195$ nm, typical retention times: GalNAc-α1-OBn (**1**) = 33.5 min, core 1-Bn (**2**) = 19.1 min, core 2-Bn (**3**) = 13.5 min, core 6-Bn (**4**) = 20.4 min). The yield of each core structure (product) was calculated using the following formula: yield = concentration of (**2**), (**3**) or (**4**) divided by the sum of (**1**), (**2**), (**3**) and (**4**).

After four generations of optimization in a total of 56 from 235,200 possible experiments, the genetic algorithm found two local maxima regarding the selection criterias.

As shown in Figure 3, a maximum concentration of the core 2-Bn trisaccharide of 8 mM was achieved. Although this product concentration corresponds to a yield of only 27%, the selectivity of the core 2-Bn formation at this point was 55%. Compared to Dudziak et al.[17] the concentration of core 2-Bn could be doubled under these conditions.

The individual showing the highest selectivity of up to 90% for core 2-Bn formation generated the desired trisaccharide in a maximum concentration of 2.1 mM, corresponding to 74% selectivity at the point of maximum product concentration and a yield of 21% (Fig. 4).

Depending on the selection criteria, different values of the relevant parameters turned out to be ideal. Regarding the criteria of high product concentration, a reaction

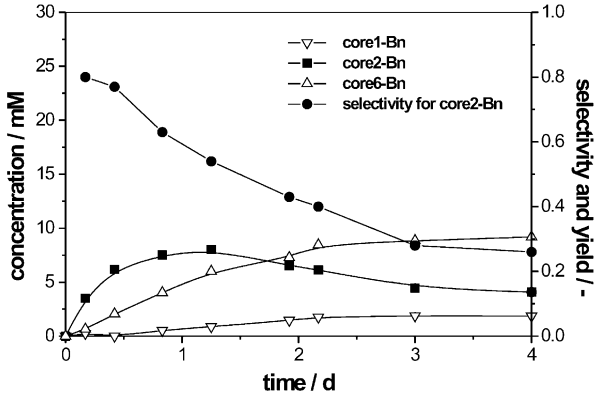

Figure 3. Best individual after four generations of optimization regarding the criteria of maximum product concentration. For each core structure the concentration and the respective yield is plotted. Conditions: 0.5 M Lactose, 30 mM GalNAc-Bn, 30 mM UDP-GlcNAc, 10 mM $MgCl_2$, 250 mU/mL β-galactosidase, 150 mU/mL C2GnT, 10 U/mL calf intestine alkaline phosphatase, 0.5 mg/mL BSA, MES buffer 50 mM, pH 6.5, 30 °C.

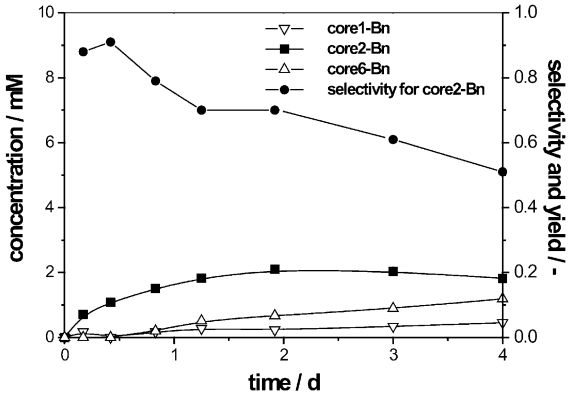

Figure 4. Best individual after four generations of optimization regarding the criteria of maximum ratio core 2-Bn/core 6-Bn at the point of maximum core 2-Bn concentration. For each core structure the concentration and the respective yield is plotted. Conditions: 0.6 M Lactose, 10 mM GalNAc-Bn, 10 mM UDP-GlcNAc, 10 mM $MgCl_2$, 375 mU/mL β-galactosidase, 150 mU/mL C2GnT, 10 U/mL calf intestine alkaline phosphatase, 0.5 mg/mL BSA, MES buffer 50 mM, pH 6.5, 10 °C.

temperature of 30 °C is optimal. In contrast, high selectivity for core 2-Bn concentration exclusively could be found at 10 °C, although low temperatures lowered product concentration drastically.

For both criterias of selection, a pH value of 6.5 was found to be best. Compared to the system of Dudziak et al.[17] suitable for generating the core 6-Bn disaccharide, this value obviously marks the optimum condition for core 2-Bn formation considering both activity and stability of the β-galactosidase and C2GnT.

Not surprisingly, high lactose concentrations were found to be elementary for high product concentration and high product selectivity. The donor molecule of the β-galactosidase for generating the core 1-Bn disaccharide must be present in high excess, since hydrolysis of lactose is much faster than the respective transglycosylation reaction yielding the substrate for C2GnT.

To generate high product concentrations *and* high selectivities for core 2-Bn formation, the optimal C2GnT concentration was found to be 150 mU/mL. This finding is rather unexpected since one might predict high amounts of C2GnT for fast and selective core 2-Bn synthesis. However, the highest enzyme level allowed for GA was set to 250 mU/mL, but this enzyme concentration proved to be inappropriate for effective core 2-Bn synthesis.

Since a system generating selectively the desired trisaccharide core 2-Bn will have to discriminate the secondary hydrolysis of core 6-Bn and core 1-Bn respectively, the optimal β-galactosidase activity is expected to be rather low. In fact, the best two individuals regarding the selection criterias (Figs. 3 and 4) both showed values for the β-galactosidase at the very lower limit of this parameter, namely 250 and 375 mU/mL, respectively.

In summary, we could show that optimization of a complex multi-enzyme system generating different *O*-glycan core structures can be achieved by using the method of a genetic algorithm. The selectivity for the core 2-Bn trisaccharide formation at the point of maximum product concentration has been doubled from 34[17] to 74% (Fig. 4).

Acknowledgements

The recombinant CHO cells expressing the soluble mouse C2GnT were kindly provided by Prof. Dr. E. G. Berger. The authors would like to thank Dr. Detlev Eisenkrätzer for his kind help during C2GnT purification and Prof. Dr. C. Wandrey for his ongoing generous support.

References and Notes

1. Brockhausen, I. In *Glycoproteins*; Montreuil, J., Schachter, H., Vliegenthart, J. F. G., Eds., Elsevier Science, Amsterdam: 1995; p 201.
2. Wilkins, P. P.; McEver, R. P.; Cummings, R. D. *J. Biol. Chem.* **1996**, *271*, 18732.
3. Hemmerich, S.; Lefflers, H.; Rosen, S. D. *J. Biol. Chem.* **1995**, *270*, 12035.
4. Snapp, K. R.; Heitzig, C. E.; Ellies, L. G.; Marth, J. D.; Kansas, G. S. *Blood* **2001**, *97*, 3806.
5. Varki, A. *Glycobiology* **1993**, *3*, 97.
6. Bierhuizen, M. F. A.; Fukuda, M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9326.
7. Yeh, J. C.; Ong, E.; Fukuda, M. *J. Biol. Chem.* **1999**, *274*, 3215.
8. Schwientek, T.; Yeh, J. C.; Levery, S. B.; Keck, B.; Merkx, G.; van Kessel, A. G.; Fukuda, M.; Clausen, H. *J. Biol. Chem.* **2000**, *275*, 11106.
9. Hedbys, L.; Johansson, E.; Mosbach, K.; Larsson, P. O. *Carbohydr. Res.* **1989**, *186*, 217.
10. Fernández-Mayoralas, A. *Topics Curr. Chem.* **1997**, *186*, 1.
11. Suzuki, K.; Fujimoto, H.; Ito, Y.; Sasaki, T.; Ajisaka, K. *Tetrahedron Lett.* **1997**, *38*, 1211.
12. Canfield, W. M.; Cummings, R. D.; Ju, T. US Patent WO 01/44478 A2, 2001.
13. Gambert, U.; Thiem, J. *Eur. J. Org. Chem.* **1999**, *1*, 107.
14. Kren, V.; Thiem, J. *Angew. Chem.* **1995**, *107*, 979.
15. Lio, G. R.; Thiem, J. *Carbohydr. Res.* **1999**, *317*, 180.
16. Zeng, S.; Dinter, D.; Eisenkrätzer, D.; Biselli, M.; Wandrey, C.; Berger, E. G. *Biochem. Biophys. Res. Commun.* **1997**, *237*, 653.
17. Dudziak, G.; Zeng, S.; Berger, E. G.; Gallego, R. G.; Kamerling, J. P.; Kragl, U.; Wandrey, C. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2595.
18. Dudziak, G.; Bézay, N.; Schwientek, T.; Clausen, H.; Kunz, H.; Liese, A. *Tetrahedron* **2000**, *56*, 5865.
19. Bézay, N.; Dudziak, G.; Liese, A.; Kunz, H. *Angew. Chem. Intl. Ed.* **2001**, *40*, 2292.
20. Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison Wesley: Reading, MA, 1989.
21. Holland, J. H. *Sci. Am.* **1992**, *267*, 44.
22. Weuster-Botz, D.; Pramatarova, V.; Spassov, G.; Wandrey, C. *J. Chem. Tech. Biotechnol.* **1995**, *64*, 563.
23. Software: Möllney, M.; Freyer, S.; Wiechert, W.; Weuster-Botz, D. *GALOP (Genetic Algorithms for the Optimization of Processes)*, version 2.20; Institute of Biotechnology: Research Center Jülich, Germany, 1993.