Using a Galactose Library for Exploration of a Novel Hydrophobic Pocket in the Receptor Binding Site of the Escherichia coli Heat-labile Enterotoxin*

Wendy E. Minke‡, Feng Hong‡, Christophe L. M. J. Verlinde, Wim G. J. Hol§, and Erkang Fan¶

From the Department of Biological Structure, the Biomolecular Structure Center, and the ¶Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195

© 1999 by The American Society for Biochemistry and Molecular Biology, Inc.

(Received for publication, June 28, 1999, and in revised form, August 10, 1999)

The binding of the B subunits of Escherichia coli heat-labile enterotoxin (LT) to epithelial cells lining the intestines is a critical step for the toxin to invade the host. This mechanism suggests that molecules which possess high affinity to the receptor binding site of the toxin would be good leads for the development of therapeutics against LT. The natural receptor for LT is the complex ganglioside GM1, which has galactose as its terminal sugar. A chemical library targeting a novel hydrophobic pocket in the receptor binding site of LT was constructed based on galactose derivatives and screened for high affinity to the receptor binding site of LT. This screening identified compounds that have 2–3 orders of magnitude higher affinity toward the receptor binding site of LT than the parent compound, galactose. The present findings will pave the way for developing simple and easily synthesizable molecules, instead of complex oligosaccharides, as drugs and/or prophylactics against LT-caused disease.

Escherichia coli: heat-labile enterotoxin (LT)1 is the causative agent of traveler’s diarrhea and is also responsible for the death of hundreds of thousands of children each year in developing countries (1, 2). There is currently no prophylaxis and only a very labor-intensive therapy against diarrhea caused by LT. Therefore, the discovery of potent inhibitors which can block the function of LT is very important for the development of effective drugs for the prevention and treatment of LT-caused diarrhea. The toxin is a heterohexamer and is assembled in the periplasm of the bacterium with one A subunit which contains its enzymatic active site and five identical B subunits which are responsible for target cell recognition. The mechanism of action by LT can be separated into several steps. After release from the bacterium, LT first binds to epithelial cells lining the intestines through its B pentameric subunits. The A subunit of LT is then translocated across the cell membrane. Inside the intestinal cell, the A subunit of the toxin modifies the a subunit of the trimeric protein Gs so that Gs loses its GTPase activity and remains constitutively in its GTP-bound state (3). This in turn causes a continuous stimulation of adenylate cyclase. The resulting elevated levels of cyclic AMP in the cell lead to massive loss of fluid and ions from the cell, the characteristic pathology of enterotoxigenic disease. Based on this mechanism, one approach of developing drugs against LT would be to create molecules which have high affinity for the receptor binding site of LT and can therefore block the binding of LT to the epithelial cells. The natural cell surface receptor for LT is ganglioside GM1 (Gal-β1,3-GalNAc-β1,4-(NeuAc-α2,3)-Gal-β1,4-Glc-β1,1-ceramide). The oligosaccharide part of GM1 (GM1-OS) is responsible for binding to LT. Because of the high cost of synthesizing large amounts of complex oligosaccharide such as GM1-OS, the use of GM1-OS itself or its close derivatives as drugs against LT would not be economically feasible. Therefore, the discovery of simple and easily synthesizable small molecules which can effectively compete with GM1-OS for binding to the receptor binding site of LT would be of great value.

The three-dimensional structure of LT in its apo state is well known (4). In addition, the crystal structure of the complex between GM1-OS and cholera toxin (CT), a closely related toxin with 80% sequence identity to LT, is also known (5). Structurally, both LT and CT have essentially identical binding sites for their natural receptor GM1 (5). Therefore, the two structures (LT and CT) can be used interchangeably in the design of easily synthesizable small molecules which can inhibit the receptor binding process of LT. The crystallographically determined structure of cholera toxin bound to its natural receptor (CT-GM1-OS (5)) reveals which regions in the toxin can be exploited in drug design. The two terminal sugars, galactose and sialic acid, make the most contributions to the binding of GM1 to the toxin, and therefore their binding pockets are the primary targets. Binding of the terminal galactose group is very specific, because all hydroxyls are involved in multiple hydrogen bonds with the protein, and substituting this galactose moiety may be extremely difficult. The sialic acid moiety on the other hand is less intimately bound by the protein. The sugar ring of the sialic acid makes hydrophobic interactions with Tyr-12, whereas the carboxylic acid, the hydroxyl, and the N-acetyl substituents form hydrogen bonds with the protein backbone. The glycerol tail is only involved in water-mediated hydrogen bonding. However, close reexamination of the structure showed that the glycerol tail is very close to a hydrophobic pocket formed by Ile-58 and Lys-34 of the neighboring B subunit (see Fig. 1A), although no significant hydrophobic interactions are present. Subsequently, we used the program GRID with a hydrophobic probe (6) to explore hydrophobic binding...
sites in LT. The results suggest that the proposed pocket is indeed favorable for hydrophobic binding determinants (Fig. 1B) in CT as well as in LT.

The proposed hydrophobic target region formed by Ile-58 and Lys-34 has never been occupied in any of the other toxin structures complexed to different galactose derivatives (7–9). Therefore, we wanted to explore this region by screening ligands composed of galactose, a linker, and a hydrophobic group for their ability to inhibit toxin binding to GM1. Galactose was used as an anchor to direct binding to the GM1 pocket, which is needed to prevent possible unspecific binding of the hydrophobic group to other hydrophobic regions on the protein. The linker was necessary to bridge the distance between the galactose anchor and the target site for the hydrophobic group, the Ile-58/Lys-34 pocket.

A library of compounds was synthesized to obtain a large and diverse selection of ligands with a hydrophobic group linked to the galactose anchor. For this library three different anchors were used, each of which had a different substituent on C1 of the galactose structure (11). The galactose rigid including the C1-N or C1-O dihedral in its optimal position. For library I, this optimal position has an H1-C1-N1-C7 dihedral of 20°, because in the Cambridge Structural Data Base (12) five structures of pyranoses which are D-sugars, and 60° for b-D-sugars (13). The ligands were energy-minimized using a conjugate gradient algorithm and the CFF91 force-field until the r.m.s. derivative was smaller than 0.001 kcal mol\(^{-1}\) Å\(^{-1}\). The library compounds were built starting from \(n = 1\), until \(n = 20\) Å.

**ACD3D Data Base Computer Screening**—A two-dimensional substructure search was conducted in the Available Chemicals Directory, ACD3D 97.2 (Molecular Design Ltd., San Leandro, CA). The software used for searching was the program ISIS (Molecular Design Ltd.) (14).

**Docking III3J Using SAS**—For flexible docking of III3J to the rigid compounds were synthesized. These library compounds were tested for their ability to inhibit binding of LT to gangliosides. Two of these compounds appeared to be the galactose derivatives with the highest affinity for the receptor binding site observed so far.

**EXPERIMENTAL PROCEDURES**

**GRID**—The program GRID, Version 16 (6), was used to indicate energetically favorable binding pockets in the receptor binding site of LT for a hydrophobic group. To find such favorable binding pockets, a hydrophobic probe was used with the default parameters to determine the interaction energies with the protein on a grid with 0.5-Å spacing. The grid was placed over the 2.2-Å resolution structure of LT-galactose (11) with the waters and galactose omitted while the grid covered a volume of 23 × 20 × 20 Å.

**Modeling the Linker Length**—Gal-\(\beta\)NHCO-(CH\(_2\))\(_n\)-Ph, Gal-\(\alpha\)O-(CH\(_2\))\(_n\)-NHCO-(CH\(_2\))\(_n\)-Ph, and Gal-\(\beta\)O-(CH\(_2\))\(_n\)-NHCO-(CH\(_2\))\(_n\)-Ph were built in InsightII (Version 97.0, Biosym/MSI) to represent, respectively, libraries I, II, and III. The galactose in binding site D of the LT-galactose structure (11) was used as the starting point from which the ligands were extended. The dihedrals of these compounds were adjusted manually to reach the Ile-58/Lys-34 pocket, leaving the galactose rigid including the C1-N or C1-O dihedral in its optimal position. For library I, this optimal position has an H1-C1-N1-C7 dihedral angle of 20°, because in the Cambridge Structural Data Base (12) five structures of pyranoses which are \(N\)-acyetylated at C1 have an average dihedral of 20° ± 15°. For libraries II and III, the glyclosidic torsion prefers the exo-anomeric conformation; H1-C1-O1-C7 = −60° ± 60° for \(\alpha\)-sugars, and 60° ± 60° for \(\beta\)-sugars (13). The ligands were energy-minimized using a conjugate gradient algorithm and the CFF91 force-field until the r.m.s. derivative was smaller than 0.001 kcal mol\(^{-1}\) Å\(^{-1}\). The library compounds were built starting from \(n = 1\), until \(n = 20\) Å.

**ACD3D Data Base Computer Screening**—A two-dimensional substructure search was conducted in the Available Chemicals Directory, ACD3D 97.2 (Molecular Design Ltd., San Leandro, CA). The software used for searching was the program ISIS (Molecular Design Ltd.) (14).

**Docking III3J Using SAS**—For flexible docking of III3J to the rigid

---

2 F. Hong and E. Fan, submitted for publication.
target site in pLT the program SAS (25) (Stochastic Approximation with Smoothing) was used. The protein coordinates used were prepared as follows: (i) the coordinates of pLT complexed to galactose (11) were taken from the Protein Data Bank, (ii) galactose was deleted from the target site in pLT the program SAS (25) (Stochastic Approximation with Smoothing), (iii) the positions of the hydrogens on hetero atoms were taken from the Protein Data Bank, (iv) hydrogens on hetero atoms were added with InsightII, and (v) charges were assigned using the CFF91 forcefield. Affinity grids of the protein were calculated using the AUTOGRID program supplied with the AUTODOCK 2.2 package (17).

The ligand was built in InsightII, subsequently energy-minimized using a conjugate gradient algorithm and the CFF91 forcefield until the target site in pLT the program SAS (25) (Stochastic Approximation with Smoothing) was used. The protein coordinates used were prepared as follows: (i) the coordinates of pLT complexed to galactose (11) were taken from the Protein Data Bank, (ii) galactose was deleted from the coordinate set, (iii) the positions of the hydrogens on hetero atoms were determined using the HB2NET option in the program WHATIF (16), (iv) other hydrogens were added with InsightII, and (v) charges were assigned using the CFF91 forcefield. Affinity grids of the protein were calculated using the AUTOGRID program supplied with the AUTODOCK 2.2 package (17).

r.m.s. derivative was smaller than 0.001 kcal mol⁻¹ Å⁻¹, and prepared for flexible docking using the program add_hydrogens supplied with the SAS package (25).

The default parameters were used for running the docking program SAS. The ligand was docked 50 times, each time doing 150 cycles of 6000 iterations. After running SAS, the solutions were clustered in groups with r.m.s. deviations lower than 1.0 Å. The clusters were ranked by the lowest energy representative of each cluster.

**Protein and Chemicals—**Porcine LT holotoxin was expressed from plasmid pIOPT-LT in E. coli strain MC1061 (18). The protein was purified using the protocol of Uesaka et al. (19), except that the clarified cell lysate was subjected to a 30% ammonium sulfate precipitation in 20 mM Tris-HCl buffer (pH 7.5) prior to affinity chromatography on immobilized β-galactose. The purity was estimated to be at least 95% on the basis of silver-stained SDS-polyacrylamide gel electrophoresis gels. Anti-LT B monoclonal antibody mAb 118–87 was a kind gift from Dr. T. Hirst (University of Bristol, Bristol, UK). Commercially obtained assay materials were: IgG horseradish peroxidase conjugate (Roche Molecular Biochemicals) and GD1b (Matreya, Pleasant Gap, PA). Commercially obtained acids for library synthesis were: 6-maleimidocaproic acid, 3-cyclohexylpropionic acid, and cyclohexanecarboxylic acid (Fluka, Milwaukee, WI); 3-phthalimidopropionic acid and 7-phenylheptanoic acid (Calbiochem, La Jolla, CA); β-N-indolepropionic acid (Platzl & Bauer Inc., Waterbury, CT); and all other acids were from Aldrich.

**Synthesis—**Fig. 2A depicts a short schematic of the synthesis, which uses trialkylphosphine-mediated amide formation (20). The detailed library synthesis protocol will be published elsewhere. All library compounds screened gave satisfactory mass spectrosopy results and were judged to be ~95% pure by thin layer chromatography or high performance liquid chromatography before use.

**LT GD1b ELISA—**The LT GD1b ELISA was performed as described by Minke et al. (21). Test samples consisted of 0.2 μg/mL porcine LT toxin preincubated with the library compound for 2 h at room temperature. For initial screening, the test samples were diluted in a 0.1% (w/v) bovine serum albumin/phosphate-buffered saline solution (saline solution containing 150 mM NaCl and 10 mM potassium phosphate at pH 7.2), and subsequently the solutions were filtered over a 0.22-μm membrane. For determination of the IC₅₀ values, the test samples were prepared in 10% dimethyl sulfoxide (Me₃SO), 0.1% bovine serum albumin/phosphate-buffered saline.

All experiments were carried out in quadruplicate and validated against a concentration gradient of 0, 0.1, 0.2, and 0.3 μg/mL toxin. IC₅₀ values were calculated from at least five different concentrations of competitive ligand by nonlinear regression as described previously (22) with the statistical package S-PLUS (Mathsoft, Inc., Cambridge, MA). For the final IC₅₀ values, the IC₅₀ was determined at least two times in independent experiments. The reported IC₅₀ of a ligand is the weighted average (weight = 1/estimated S.D.²) of the IC₅₀ values from its different determinations.

**RESULTS AND DISCUSSION**

The binding pocket of the glycerol tail of GM1 suggested the presence of a hydrophobic site in LT and CT formed by Ile-58 and Lys-34, which was supported by the fact that a hydropho-
bic probe had very high affinity in this region as determined by the program GRID (Fig. 1). To explore this binding pocket, we decided to construct a library from galactose and carboxylate building blocks (Fig. 2A). Manual docking procedures were used to determine the minimal length of the methylene linker of the carboxylate building block: a phenyl ring was positioned in the Ile-58/Lys-34 pocket of LT while the phenyl was attached to galactose in the galactose binding site through a methylene linker joined by an amide bond. For sub-library I (Gal-βNHCO-(CH₂)ₙ-Ph) at least three methylene units were needed to bridge the distance between the galactose-C1-NH and the hydrophobic pocket. For sub-libraries II and III (Gal-O-(CH₂)₂-NHCO-(CH₂)ₙ-Ph, α for sub-library II and β for sub-library III), at least three and two methylene units were needed, respectively, to reach the Ile-58/Lys-34 pocket. Therefore, it was decided to use carboxylates with two to six methylene units attached to a hydrophobic group as building blocks for a galactose library to screen for ligands that inhibit LT binding to GD1b by making use of the hydrophobic pocket.

To select appropriate building blocks for library synthesis, the Available Chemicals Directory was searched for the substructure depicted in Fig. 2B, which yielded 485 carboxylate containing molecules. These compounds were screened for chemical properties, and the following substituents were excluded: additional carboxylates, free alcohols, free amines/imines, free thiols, aldehydes, borines, seleniums, furans, thienyls, and radicals. In addition, the following groups were rejected: porphyrins, steroids, and microperoxidases. Of the remaining 242 compounds we rejected the doubles, in which we considered methyls and halogens to be identical. Also, we discarded compounds without a price listing or compounds that cost more than $100. This led to a group of 58 compounds of which a subgroup of 33 compounds was selected based on diversity. Of these 33 acids that were ordered, only 24 were actually available and used in the library synthesis. The diversity of this library for the design of future inhibitors. First of all, the effect of the linker length should be discussed. As was predicted, a linker of two methylene units is too short for libraries I and III, whereas a linker of one methylene unit is too long and can lead to solubility problems; these compounds are marked with an asterisk in Fig. 3. The following compounds showed more than 50% inhibition and were subsequently tested at 0.5 mM: I2J (I2J denotes the sub-library I compound with a linker of two methylene units and substituent 2), I3J, I4C, II2B, II2C, II2F, II2G, II3J, II4B, II4C, and II6A. At 0.5 mM only galactose library compounds II3J and II6A showed significant inhibition in the LT ELISA, respectively, 100% and 60%. Subsequently these compounds were synthesized at larger scale to confirm their inhibitory power and to determine the IC₅₀. Unfortunately, the results with compound II6A from the larger scale do not confirm with the earlier, screening results: the IC₅₀ is only 3.2 ± 0.1 mM. The better result obtained during the initial screening may have been because of impurities in the smaller scale library synthesis. However, compound II3J is indeed an extremely good inhibitor with an IC₅₀ of 0.23 ± 0.04 mM (Fig. 4A), which means that the affinity is 300-fold higher than that of galactose (21). Because II3J is such a good inhibitor and compound III3J was only tested at an unknown and very low concentration, we also decided to resynthesize library compound III3J at a larger scale and test it in an appropriate solvent. The presence of 10% Me₂SO in the III3J sample solutions did not have any effect on the IC₅₀ of III3J (Fig. 4A), and in addition, III3J dissolved completely in the presence of 10% Me₂SO. Therefore, the IC₅₀ of III3J was determined in the presence of 10% Me₂SO and turns out to be 40 ± 2 μM (Fig. 4B). The IC₅₀ of III3J is even 5-fold better than its α-isomer in library II or in other words it is 1500-fold better than galactose.

Besides identifying several tight binding inhibitors of LT binding to GD1b, another motivation for designing, synthesizing and screening a galactose library was to analyze characteristics of this library for the design of future inhibitors. First of all, the effect of the linker length should be discussed. As was predicted, a linker of two methylene units is too short for libraries I and III (compounds I2J and III2J being exceptions, probably because of the presence of a second phenyl group). However, for the compounds with longer linkers in libraries I and III and for library II, there is not a clear correlation between linker length and inhibition of LT binding to GD1b. Second, library II distinctly contains better inhibitors (Fig. 3), although the best compound is from library III. This confirms our earlier findings that α anomers of galactose derivatives in general are better inhibitors than their β anomers (21). To conclude, library II contains relatively the most potent inhibitors, and even compounds with the smallest linker in this library perform well in the ELISA. Therefore, it is probably...
wise to expand this library in the future and also include compounds in library II with only one methylene group as the linker.

Of course our immediate goal is to follow up the hits of this successful library screen. To design the next round of inhibitors based on the very potent inhibitor III3J, we need to know its exact binding mode. Initially, we tried to expedite the design process by predicting the binding mode of III3J in LT using the docking program SAS. However, as shown in Fig. 5, no single binding mode was preferred. Instead, of the 50 solutions suggested by the program, only 7 were within 1 Å of another solution. Therefore, we are in the process of crystallizing the LT-III3J complex to obtain a better model for the binding mode of III3J in LT.

Acknowledgments—We thank Dr. David Diller for early release of the program SAS and Dr. Peter Goodford for providing GRID. And we are very grateful to Dr. Tim Hirst for providing us with anti-LTB.

REFERENCES
1. Wanke, C. A., Lima, A. A., and Guerrant, R. L. (1987) *Baillieres Clin. Gastroenterol.* 1, 335–359
2. Holmgren, J., and Svennerholm, A. M. (1992) *Gastroenterol. Clin. N. Am.* 21, 283–302
3. Spangler, B. D. (1992) *Microbiol. Rev.* 56, 622–647
4. Sixma, T. K., Pronk, S. E., Kalk, K. H., Wartna, E. S., van Zanten, B. A., Witholt, B., and Hol, W. G. J. (1991) *Nature* 351, 371–377
5. Merritt, E. A., Kuhn, P., Sarfaty, S., Erbe, J. L., Holmes, R. K., and Hol, W. G. J. (1998) *J. Mol. Biol.* 282, 1043–1059
6. Goodford, P. J. (1985) *J. Med. Chem.* 28, 849–857
7. Sixma, T. K., Pronk, S. E., Kalk, K. H., van Zanten, B. A., Berghuis, A. M., and Hol, W. G. J. (1992) *Nature* 355, 561–564
8. Van den Akker, F., Feil, I. K., Roach, C., Plat, A. A., Merritt, E. A., and Hol, W. G. J. (1997) *Protein Sci.* 6, 2644–2649
9. Merritt, E. A., Sarfaty, S., Feil, I. K., and Hol, W. G. J. (1997) *Structure* 5, 1485–1499
10. Chernyak, A. Y., Sharma, G. V. M., Kononov, L. O., Krishna, P. R., Levinsky, A. B., Kochetkov, N. K., and Rama Rao, A. V. (1992) *Carbohydr. Res.* 223, 303–309
11. Merritt, E. A., Sixma, T. K., Kalk, K. H., Van Zanten, B. A., and Hol, W. G. J. (1994) *Mol. Microbiol.* 13, 745–753
12. Allen, F., and Kennard, O. (1993) *Chem. Des. Automata. News* 8, 1 and 31–37
13. Tvaroska, I., and Bleha, T. (1989) *Adv. Carb. Chem. Biochem.* 47, 45–123
14. Mook, T. E., Henry, D. R., Ozbekak, A. G., and Alamgir, M. (1994) *J. Chem. Info. Comput. Sci.* 34, 184–189
15. Apparu, M., Blanc-Muesser, M., Defaye, J., and Driguez, H. (1981) *Can. J. Chem.* 59, 314–320
16. Hooft, R. W., Sander, C., and Vriend, G. (1996) *Proteins* 26, 363–376
17. Goodsell, D. S., and Olson, A. J. (1990) *Proteins* 8, 195–202
18. Feil, I. K., Reddy, R., de Haan, L., Merritt, E. A., Van den Akker, F., Storm, D. R., and Hol, W. G. J. (1996) *Mol. Microbiol.* 20, 823–832
19. Uesaka, Y., Otsuka, Y., Lin, Z., Yamashita, S., Yamaoka, J., Kurazono, H., and Takeda, Y. (1994) *Microbial. Pathogen.* 16, 71–76
20. Minke, W. E., Roach, C., Hol, W. G. J., and Verlinde, C. L. M. J. (1999) *Biochemistry* 38, 5684–5692
21. Bowen, W. P., and Jerman, J. C. (1995) *Trends Pharmacol. Sci.* 16, 413–417
22. Sugii, S., and Tsuji, T. (1989) *Can. J. Microbiol.* 35, 670–673
23. Susaki, H., Suzuki, K., Ikeda, M., Yamada, H., and Watanabe, H. (1994) *Chem. Pharm. Bull.* 42, 2090–2096
24. Diller, D. J., and Verlinde, C. L. M. J. (1999) *J. Comput. Chem., in press*