High-Affinity Pentavalent Ligands of *Escherichia coli* Heat-Labile Enterotoxin by Modular Structure-Based Design

Erkang Fan,* Zhongsheng Zhang, Wendy E. Minke, Zheng Hou, Christophe L. M. J. Verlinde, and Wim G. J. Hol†

Department of Biological Structure Biomolecular Structure Center and Howard Hughes Medical Institute University of Washington, Box 357742 Seattle, Washington 98195

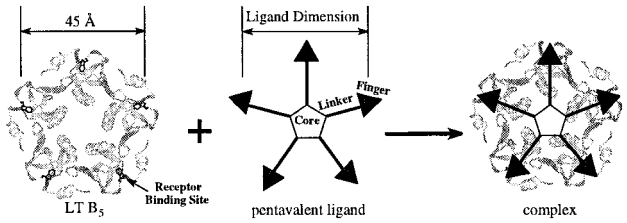
Received September 20, 1999

High-affinity protein ligands have wide applications in disease diagnosis, prevention, and treatment. A promising strategy to arrive at such ligands is the creation of multivalent compounds that bind to a multivalent target.[1] Current approaches include the synthesis of bivalent ligands, in which two copies of a ligand are joined by a flexible linker;[2] and more generally, the attachment of a larger number of monovalent ligands onto a selected backbone. Examples of backbones include polymers/oligomers,[3] membranes,[4] and dendrimers.[5] In these approaches structure-based information on the spatial arrangement of the target's multiple binding sites does not usually enter the design process. Therefore, the obtained ligands may not be ideal for maximizing the interaction with the target. Here, we present a novel approach toward high-affinity multivalent ligands: a modular design that incorporates structural information of the multiple target sites.

Our work focuses on an ideal target model: the heat-labile enterotoxin (LT) from *Escherichia coli*. LT, a close relative of cholera toxin, is a member of the AB_5 family of bacterial toxins which also includes shiga toxin, shiga-like toxin, and pertussis toxin.[6] The symmetrical arrangement of the five B subunits of LT presents five identical carbohydrate binding sites that recognize ganglioside GM1 headgroups protruding from cells of the gastrointestinal lumen. This LT-carbohydrate interaction has been determined in atomic detail,[7] and provides the basis for the structure-based design of pentavalent ligands.

The large dimensions of LT make the design and synthesis of a structurally complementary pentavalent ligand a major challenge. As shown in Scheme 1, distances between the toxin's nonadjacent binding sites are 45 Å. In our modular design, the large pentavalent ligand is divided into three modules: a semirigid

Scheme 1



"core" that can adopt a 5-fold symmetric configuration and provides a foundation for the generation of structural complementarity for the overall pentavalent ligand to LT, "linkers" that project in the direction of the receptor binding sites, and "fingers" that fit snugly into the binding sites (Scheme 1). With efficient chemistry to connect these modules, assembly of the large pentavalent ligands is synthetically feasible, as will be shown.

In this report, we demonstrate the power of a modular synthesis procedure, which allowed us to explore in detail the effects of linker length on affinity. For the core, we chose acylated pentacyclen **5**. Force-field calculations show that this molecule can adopt a conformation close to 5-fold symmetry. We used 1-β-amidated D-galactose **1** as the finger. D-galactose is a terminal sugar unit of LT's natural receptor GM1. It interacts very specifically with the toxin via defined hydrogen bonds and a carbohydrate against tryptophan stacking.[7a] Galactose and its derivatives with substitutions at C1 have been observed to bind in the same manner in LT's receptor binding site.[7] Thus, **1** would be expected to be a good mimic of galactose while providing functional groups for full ligand assembly.

The success of this design, however, depends critically on finding suitable linker modules. Rigid linkers complementary to the protein surface[8] would be ideal but are difficult to achieve in one design step. Here we present a solution to this problem, based on readily available flexible linkers. This raises the following questions: can a large gain in affinity be obtained with flexible linkers, and how long do the linkers need to be? To answer these questions, we have chosen to span a large range of linker lengths using the commercially available 4,7,10-trioxa-1,13-tridecanediamine **3** as the basic unit of the linkers. Longer linkers in our pentavalent ligands consist of up to four units of **3**.

The modular synthesis of our full ligands is shown in Scheme 2. Although each long linker could be synthesized separately from **3**, it is more economical to perform the stepwise coupling of each unit of **3** to the core−linker assembly. As an alternative to the HPLC purification of each reaction intermediate, we also developed purification protocols based on C18 Sep-Pak cartridges[9] which can handle far greater sample load than a typical research-lab preparative HPLC setup. The squaric acid diester mediated coupling reaction[10] was very clean and efficient. There was no partially derivatized product detectable at each reaction step. Typical isolated yields for each coupling product were around 80∼95%. To the best of our knowledge, this represents one of the first efficient syntheses of pure single-species protein ligands with large molecular weights (4−8 kDa).

After obtaining a series of pentavalent ligands with various linker lengths (**10**−**13**), we tested their ability to inhibit the binding of LT B pentamer (LT-B_5) to ganglioside, using the identical ELISA protocol published previously[11] (Table 1). Each

† Howard Hughes Medical Institute, University of Washington.

(1) Mammen, M.; Choi, S.-K.; Whitesides, G. M. *Angew. Chem., Int. Ed.* **1998**, *37*, 2754−2794.

(2) (a) Kramer, R. H.; Karpen, J. W. *Nature* **1998**, *395*, 710−713. (b) Profit, A. A.; Lee, T. R.; Lawrence, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 280−283. (c) Glick, G. D.; Toogood, P. L.; Wiley: D. C.; Skehel, J. J.; Knowles, J. R. *J. Biol. Chem.* **1991**, *266*, 23660−23669.

(3) (a) Kanai, M.; Mortell, K. H.; Kiessling, L. L. *J. Am. Chem. Soc.* **1997**, *119*, 9931−9932. (b) Mammen, M.; Dahmann, G.; Whitesides, G. M. *J. Med. Chem.* **1995**, *38*, 4179−4190. (c) Connolly, D. T.; Townsend, R. R.; Kawaguchi, K.; Bell, W. R.; Lee, Y. C. *J. Biol. Chem.* **1982**, *257*, 939−945. (d) Lee, T. C. *FASEB J.* **1992**, *6*, 3193−3200.

(4) (a) Kingery-Wood, J. E.; Williams, K. W.; Sigal, G. B.; Whitesides, G. M. *J. Am. Chem. Soc.* **1992**, *114*, 7303−7305. (b) Spevak, W.; Nagy, J. O.; Charych, D. H.; Schaefer, M. E.; Gilbert, J. H. *J. Am. Chem. Soc.* **1993**, *115*, 1146−1147.

(5) (a) Thompson, J. P.; Schengrund, C. L. *Glycoconjugate J.* **1997**, *14*, 837−845. (b) Roy, R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 692−702. (c) Kitov, P. I.; Sadowska, J. M.; Bundle, D. R.; Read, R. R.; Armstrong G. D. *Abstracts of the XV Interactional Symposium on Glycoconjugates*, Tokyo, 1999; pp S44.

(6) Merritt, E. A.; Hol, W. G. J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 165−171.

(7) (a) Sixma, T. K.; Pronk, S. E.; Kalk, K. H.; Van Zanten, B. A.; Berghuis, A. M.; Hol, W. G. J. *Nature* **1992**, *355*, 561−564. (b) Van den Akker, F.; Feil, I. K.; Roach, C.; Platas, A. A.; Merritt, E. A.; Hol, W. G. J. *Protein Sci.* **1997**, *6*, 2644−2649. (c) Merritt, E. A.; Sarfaty, S.; Feil, I. K.; Hol, W. G. J. *Structure* **1997**, *5*, 1485−1499.

(8) Park, H. S.; Lin, Q.; Hamilton, A. D. *J. Am. Chem. Soc.* **1999**, *121*, 8−13.

(9) *Waters Sep-Pak cartridge application bibliography*, 5th ed.; McDonald, P. D., Ed.; Waters: Milford, MA, 1991.

(10) Tietze, L. F.; Arlt, M.; Beller, M.; Glusenkamp, K.-H.; Jahde, E.; Rajewsky, F. *Chem. Ber.* **1991**, *124*, 1215−1221.

10.1021/ja993388a CCC: $19.00
Published on Web 03/03/2000

Scheme 2

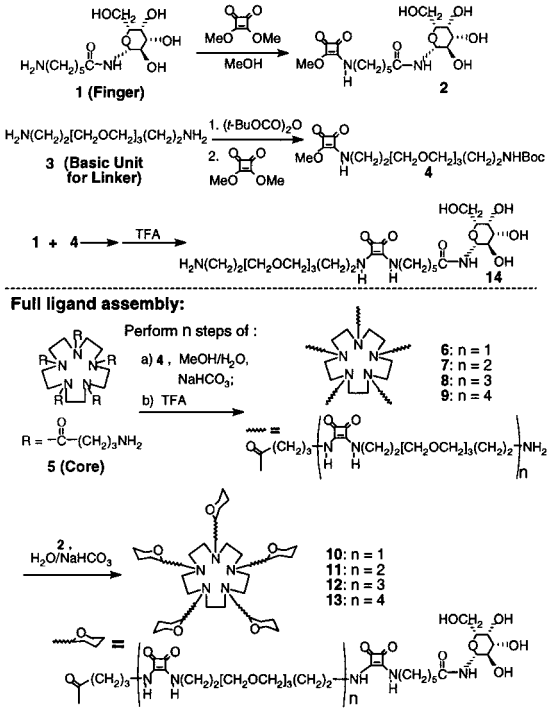


Table 1

ligand	IC_{50} (μM)	gain over galactose
galactose	58000 ± 8000	1
14	5000 ± 200	11
10	242 ± 91	240
11	16 ± 8	3600
12	6 ± 4	10 000
13	0.56 ± 0.06	104 000
GM1-OS	0.01 ± 0.01	5 800 000

IC_{50} was the standard-deviation-weighted average of two or more independent quadruplicate determinations.

The results in Table 1 clearly show that our structure-based design of pentavalent ligands leads to very significant affinity gains compared to the monovalent ligand. The best pentavalent ligand **13** shows an IC_{50} that is 10^5-fold better than galactose, the molecular moiety mostly responsible for the affinity of our fingers to LT. The IC_{50} of **13** approaches that of GM1-OS, the oligosaccharide portion of LT's natural receptor ganglioside GM1.[11] To further dissect the effect of pentavalent linkage rather than that of linker structure, we synthesized compound **14** consisting of both the finger and the linker modules. Compared to **14**, ligand **13** shows more than 10^4-fold gain in affinity, or 2000-fold on a valency-corrected basis. In comparison, recently reported GM1-OS derivatized dendrimers only exhibited a few 100-fold affinity increase per finger compared to GM1-OS by itself,[5a] which underlines the efficiency of our approach. To demonstrate that the improvements in IC_{50}'s by our pentavalent ligands over galactose is most likely due to the formation of a 1:1 complex in solution, rather than due to the effect of cross-linking target protein in solution,[2c, 3c, 5c] we have performed dynamic light scattering (DLS) studies on ligand−protein complexes. Solutions of LT-B_5 at micromolar concentrations in the presence of 0.5, 1, or 2 equiv of ligand **13** showed no evidence of aggregation.[12] This clearly ruled out the cross-linking effect by our pentavalent ligand and is supportive for the formation of a 1:1 complex in solution between **13** and LT-B_5.

In addition to demonstrating a large affinity gain by our pentavalent ligands over the monovalent ligand, it is exciting to note that in this first systematic study of the effects of flexible linker length on pentavalent ligand affinity, our results agree very well with the notion brought up by Kramer and Karpen[2a] for flexible bivalent ligands. That is: in order to obtain a large gain in affinity, the ligand dimensions should match that of the binding site distribution in the target (Scheme 1). Most importantly, this match should not be based on the extended conformation of the flexible linkers, rather, the linkers' *effective* lengths should be considered. As a guideline for determining the effective dimensions of our pentavalent ligands, we followed the equation used by Kramer and Karpen, which is based on a polymer model described by Knoll and Hermans.[13] Hence, the effective dimensions of our ligands are taken to be proportional to the square root of the molecular weights between two fingers. As shown in Figure 1, even though all ligands' extended dimensions are far greater than the distance between nonadjacent binding sites in LT, the effective dimensions of ligands **10**−**12** are less than the nonadjacent binding site distance in LT. Therefore, we see a steady drop in IC_{50}'s as the ligand's effective dimensions increase in ligands **10**−**13**. In ligand **13**, where the ligand's effect dimensions have the best match of LT's binding site distribution in this series of ligands, we obtained the largest gain in affinity to LT over galactose. Although it would be interesting to see the affinities of ligands with effective linker lengths substantially larger than the distances between the target binding sites, the low solubilities of the synthetic intermediates prevented us from obtaining such ligands. It is, however, worth noting that according to the findings of Kramer and Karpen,[2a] ligands with too long a linker length would have less affinity toward the target protein than ligands whose effective dimensions match the distribution of binding sites on the target protein.

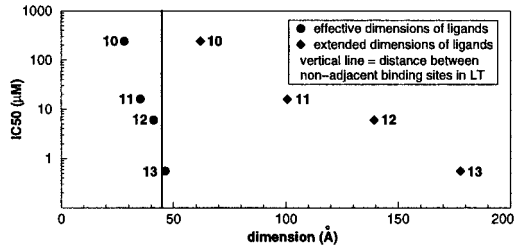


Figure 1. Plot of IC_{50}'s versus the dimensions of ligands.

In summary, our modular approach has allowed for efficient synthesis of large molecular weight protein ligands and, for the first time, a systematic study of the effects of flexible-linker lengths on the affinities of multivalent ligands. The design of multivalent ligands on the basis of structural information appears more powerful in gaining ligand affinity than conventional approaches based on structurally less well-defined supports. Further study of multivalent ligands based on our approach, such as using higher affinity fingers[11] and incorporating protein surface recognition[8] into linkers, would provide even more insight into multivalent protein−ligand interactions.

Acknowledgment. We are grateful to Dr. Tim Hirst for providing antibodies used in the ELISA and Dr. Ethan Merritt for helpful discussions. This work is supported by the School of Medicine (to E.F.), and grants from the National Institute of Health (GM54618 to C.V. and AI34501 to W.H.).

Supporting Information Available: Preparation of compounds and DLS experiments (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA993388A

(11) Minke, W. E.; Roach, C.; Hol, W. G. J.; Verlinde, C. L. M. J. *Biochemistry* **1999**, *38*, 5684−5692.

(12) The DLS-measured molecular weights for LT-B_5 in solution alone at 3.9 and 7.8 μM are 54 and 52 kD, respectively (theoretical value: 59 kD). In the presence of various amounts of **13**, the measured molecular weights varied from 53 to 57 kD. As a positive control for protein aggregation, a solution of LT-B_5 at 7.8 μM in the presence of the aggregating reagent chlorophenolred-β-D-galactopyranoside showed a molecular weight of >400 kD. The detailed experimental data is provided in the Supporting Information.

(13) Knoll, D.; Hermans, J. *J. Biol. Chem.* **1983**, *258*, 5710−5715.