J|A|C|S
ARTICLES
Published on Web 10/09/2002

Solution and Crystallographic Studies of Branched Multivalent Ligands that Inhibit the Receptor-Binding of Cholera Toxin

Zhongsheng Zhang, Ethan A. Merritt, Misol Ahn, Claudia Roach, Zheng Hou, Christophe L. M. J. Verlinde, Wim G. J. Hol,*,† and Erkang Fan*

Contribution from the Biomolecular Structure Center and Departments of Biochemistry and Biological Structure, University of Washington, Seattle, Washington 98195, and Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815

Received July 8, 2002

Abstract: The structure-based design of multivalent ligands offers an attractive strategy toward high affinity protein inhibitors. The spatial arrangement of the receptor-binding sites of cholera toxin, the causative agent of the severe diarrheal disease cholera and a member of the AB_5 bacterial toxin family, provides the opportunity of designing branched multivalent ligands with 5-fold symmetry. Our modular synthesis enabled the construction of a family of complex ligands with five flexible arms each ending with a bivalent ligand. The largest of these ligands has a molecular weight of 10.6 kDa. These ligands are capable of simultaneously binding to two toxin B pentamer molecules with high affinity, thus blocking the receptor-binding process of cholera toxin. A more than million-fold improvement over the monovalent ligand in inhibitory power was achieved with the best branched decavalent ligand. This is better than the improvement observed earlier for the corresponding nonbranched pentavalent ligand. Dynamic light scattering studies demonstrate the formation of concentration-dependent unique 1:1 and 1:2 ligand/toxin complexes in solution with no sign of nonspecific aggregation. This is in complete agreement with a crystal structure of the branched multivalent ligand/toxin B pentamer complex solved at 1.45 Å resolution that shows the specific 1:2 ligand/toxin complex formation in the solid state. These results reiterate the power of the structure-based design of multivalent protein ligands as a general strategy for achieving high affinity and potent inhibition.

Introduction

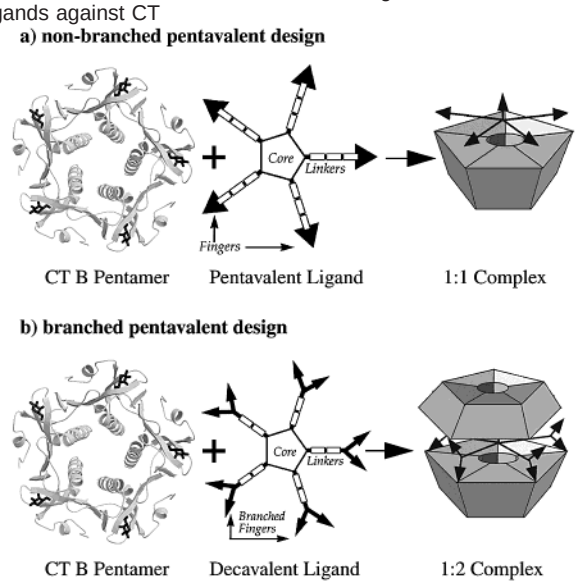
Cholera toxin (CT), the causative agent of the deadly diarrheal disease cholera, is a member of the AB_5 family of bacterial toxins.[1] After the toxin's assembly in the periplasm of *Vibrio cholerae* and secretion into the lumen of the gut, a critical step follows in which five identical receptor-binding sites on the toxin's B pentamer mediate binding to the epithelial cell surface of the human host through specific interaction with ganglioside GM1. This receptor-binding process is an attractive target for the design of prophylactics against the acute result of infection by *V. cholerae*. Because of the involvement of the multimeric B subunits of CT, receptor binding by cholera toxin is also an attractive model system for exploring the design of high affinity multivalent antagonists. An essentially identical mode of binding to GM1 on human intestine is exhibited by *Escherichia coli* heat-labile enterotoxin (LT), the causative agent of traveler's diarrhea, which is closely related to CT with 80% sequence identity for both the A and B subunits of the toxin. We are actively engaged in the design of multivalent receptor-binding antagonists against CT and LT, with a particular focus on structure-based designs that exploit the 5-fold symmetry of the binding sites on the B pentamer.

The design and synthesis of multivalent ligands represent a unique method toward high affinity protein inhibitors for drug development and a variety of other biologically related applications.[2−5] Traditional methods of synthesizing multivalent ligands in general rely on attaching multiple copies of a monovalent ligand onto a flexible generic backbone, such as a linear polymer.[2] In such cases, the resulting multivalent ligand can adopt a variety of conformations to occupy multiple binding sites of its target, and the affinity gained by multivalent presentation is due primarily to an increase in the effective local concentration of the monovalent ligand. For highly multimeric ligands, it is sometimes difficult to distinguish ligand-mediated aggregation of the protein from an actual gain in intrinsic affinity.[6] In contrast, in a structure-based design approach, one aims to arrive at multivalent ligands with geometries and dimensions that match the spatial distributions of binding sites on the target macromolecule so that the ligand would bind to

* E-mail erkang@u.washington.edu (Fan) or hol@gouda.bmsc.washington.edu (Hol).
† Howard Hughes Medical Institute.

(1) Merritt, E. A.; Hol, W. G. J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 165−171.
(2) Mammen, M.; Choi, S. K.; Whitesides, G. M. *Angew. Chem., Int. Ed.* **1998**, *37*, 2755−2794.
(3) Kiessling, L. L.; Strong, L. E.; Gestwicki, J. E. *Annu. Rep. Med. Chem.* **2000**, *35*, 321−330.
(4) Lee, R. T.; Lee, Y. C. *Glycoconjugate J.* **2000**, *17*, 543−551.
(5) Borman, S. *Chem. Eng. News* **2000**, *78*, 48−53.
(6) Dimick, S. M.; Powell, S. C.; McMahon, S. A.; Moothoo, D. N.; Naismaith, J. H.; Toone, E. J. *J. Am. Chem. Soc.* **1999**, *121*, 10286−10296.

Scheme 1. Structure-Based Modular Design of Multivalent Ligands against CT



its target through a single mode. In an ideal case, the structure-based design of multivalent ligands would also lead to ligands whose backbone or linker portion also interacts favorably with the target, thereby further enhancing affinity to the target beyond gains due to multivalency on generic backbones. In recent years, the structure-based design of multivalent ligands has attracted increasing attention[7−12] with some of the most successful examples developed for targeting AB_5 toxins.[8,9]

While several groups have investigated the multivalent inhibition of AB_5 toxins,[13−15] structure-based designs of multivalent ligands have so far mainly been reported by two groups. The Bundle group designed a decavalent inhibitor for Shiga-like toxins based on an asymmetric scaffold.[8] Because each B subunit of Shiga-like toxin's B pentamer contains three receptor-binding sites of varying affinities, this presented the opportunity to block up to 10 of these 15 total sites upon binding a pentavalent cluster of bivalent ligands that extend from a glucose core with five arms, as intended in the report by the Bundle group.[8] The resulting decavalent ligand was shown to have several million-fold gains in affinity over the monovalent trisaccharide. However, a crystal structure revealed a surprising yet very intriguing complex, where two toxin B pentamers sandwiched a single inhibitor, with the single highest affinity binding site occupied on each monomer of the two toxin pentamers.[8] In our studies on CT and LT, a modular synthetic approach is combined with structure-based design (Scheme 1a) to investigate the effect of optimizing the effective dimensions of a symmetric pentavalent inhibitor.[9] We have shown that the highest affinity gain by the pentavalent ligand over that of its monovalent counterpart was realized when the effective dimension of the pentavalent ligand closely matched the binding site distribution on the toxin.[9] This confirmed the concept first put forward by Kramer and Karpen, who used the effective length of linear flexible linkers for divalent ligand design.[7]

A close examination of the surprising 1:2 ligand/toxin complex revealed by the Bundle group and our pentavalent ligand/CT complex raises several questions: will a branched multivalent ligand designed to simultaneously bind two toxin B pentamers be a general approach for multivalent design (Scheme 1b), and will such ligands perform equally well or better than nonbranched multivalent ligands? CT is a simpler system than Shiga-like toxin in which to answer these questions because CT contains only one receptor-binding site per subunit on the toxin B pentamer. This facilitates the characterization of ligand−toxin interaction. In this report, we evaluate a series of branched decavalent ligands that can bind to the CT B pentamer in a 1:2 fashion and compare results for the branched multivalent ligands with a series of nonbranched pentavalent ligands from our previous study.[9] We also investigate dynamic light scattering (DLS) as a tool to probe higher order ligand−toxin interactions in solution. A high resolution crystal structure of a 10.6-kDa branched inhibitor bound to two CT B pentamers is described. This structure of a 1:2 ligand/toxin complex is in complete agreement with the information obtained from solution DLS studies.

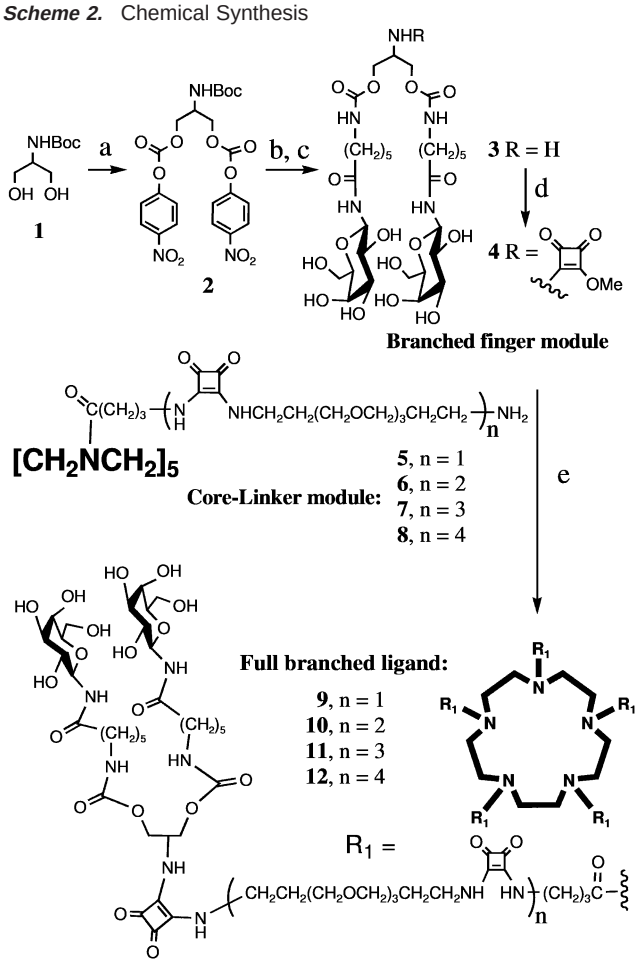
Results and Discussion

To facilitate a direct comparison of branched multivalent ligands studied in this report with nonbranched ligands reported previously, we decided to use the identical galactose-based finger module for ligand assembly (Scheme 2). The branching was achieved by attaching two finger molecules (*N*-(ϵ-aminocaproyl)-β-D-galactopyranosylamine) to a protected 2-amino-1,3-propanediol (**1**) through a carbamide linkage at the 1,3 position, while the 2-amino group of **1** provides a coupling point for full ligand assembly. The resulting branched finger module **4** was readily coupled to a series of core−linker modules (**5**−**8**) using our established modular synthesis protocols[9] to yield sufficient quantities of branched full multivalent ligands (**9**−**12**).

All the full decavalent ligands (**9**−**12**) were tested for their ability to inhibit the CT B pentamer binding to ganglioside using a competitive inhibition assay.[16] The IC_{50}'s are listed in Table 1. From Table 1, it is very clear that all branched multivalent ligands exhibit substantial gain in inhibiting CT receptor binding relative to the monovalent parent inhibitor, galactose. In our previous study,[9] we demonstrated that nonbranched pentavalent ligand with four linker units had the highest affinity toward the toxin B pentamer. We attributed this to the fact that the ligand's effective dimension was estimated to be very close to the binding site distribution at the bottom face of the toxin B pentamer. The correlation between the inhibitory power and the multivalent ligand's effective dimension is confirmed in this series of branched multivalent ligands. This is evident by a larger gain in inhibitory power, compared to that of galactose, by ligand **12** (four linker units) than by ligand **11** (three linker units).

(7) Kramer, R. H.; Karpen, J. W. *Nature* **1998**, *395*, 710−713.
(8) Kitov, P. I.; Sadowska, J. M.; Mulvey, G.; Armstrong, G. D.; Ling, H.; Pannu, N. S.; Read, R. J.; Bundle, D. R. *Nature* **2000**, *403*, 669−672.
(9) Fan, E.; Zhang, Z.; Minke, W. E.; Hou, Z.; Verlinde, C. L. M. J.; Hol, W. G. J. *J. Am. Chem. Soc.* **2000**, *122*, 2663−2664.
(10) Loidl, G.; Groll, M.; Musiol, H. J.; Huber, R.; Moroder, L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5418−5422.
(11) Loidl, G.; Musiol, H. J.; Groll, M.; Huber, R.; Moroder, L. *J. Pept. Sci.* **2000**, *6*, 36−46.
(12) Schaschke, N.; Matschiner, G.; Zettl, F.; Marquardt, U.; Bergner, A.; Bode, W.; Sommerhoff, C. P.; Moroder, L. *Chem. Biol.* **2001**, *8*, 313−327.
(13) Thompson, J. P.; Schengrund, C.-L. *Glycoconjugate J.* **1997**, *14*, 837−845.
(14) Thompson, J. P.; Schengrund, C. L. *Biochem. Pharmacol.* **1998**, *56*, 591−597.
(15) Gargano, J. M.; Ngo, T.; Kim, J. Y.; Acheson, D. W. K.; Lees, W. J. *J. Am. Chem. Soc.* **2001**, *123*, 12909−12910.
(16) Minke, W. E.; Roach, C.; Hol, W. G. J.; Verlinde, C. L. M. J. *Biochemistry* **1999**, *38*, 5684−5692.

Scheme 2. Chemical Synthesis



(a) 4-Nitrophenyl chloroformate, DIPEA, CH_2Cl_2, 60%; (b) *N*-(ϵ-aminopropyl)-β-D-galactopyranosylamine, pyridine, 82%; (c) 1:1 CH_2Cl_2/TFA; (d) dimethyl squarate, MeOH/H_2O, 60%; (e) $NaHCO_3$, MeOH/H_2O, 78−85%.

It appears that all the branched multivalent ligands, except ligand **9**, are significantly better inhibitors than the corresponding nonbranched pentavalent ligands (Table 1). There is up to 1 order of magnitude gain in inhibitory power when going from a nonbranched pentavalent ligand to a corresponding branched ligand. These measurements of IC_{50} demonstrated the high affinity of branched multivalent ligands toward the CT B pentamer. It is, therefore, clear that a branched multivalent approach toward high affinity protein ligands should be generally applicable as a follow-up to a nonbranched approach. However, the IC_{50} measurements obviously could neither reveal if 1:1 or 1:2 ligand/protein complexes were formed in solution nor provide insight into what might be responsible for the improved IC_{50} over a nonbranched multivalent approach. To further study the solution characteristics of a branched multivalent ligand in complex with CT B pentamer(s), we used DLS to probe the complex formation in solution.

DLS is a powerful tool for detecting in solution the size and polydispersity of solutes. Although DLS has been widely used for the characterization of synthetic polymers and biomacromolecules, as well as protein−protein interactions, few have used the technique to study synthetic ligand−protein interactions. Because of DLS's high sensitivity in detection of molecular aggregation, we first used DLS in our previous study[9] to prove that our general structure-based design of multivalent ligands leads to compounds that inhibit toxin binding through specific 1:1 complex formation because a ligand-mediated protein aggregation model is ruled out based on DLS results. The current series of branched multivalent ligands, which are designed to be capable of forming 1:2 ligand/protein complexes, provides the opportunity to observe the formation of higher-order discrete multivalent ligand/protein complexes, while at the same time ruling out any concerns of ligand-mediated protein aggregation behavior that has been seen in other multivalent ligand systems.[6]

The results of DLS studies using the best ligand in this series, **12**, with the CT B pentamer are collectively shown in Figure 1. Figure 1A shows clearly that, in the solution of a **12**/CT B pentamer mixture, there is no sign of the formation of large aggregates. Rather, only a discrete complex was formed. A positive control for ligand-mediated aggregation was performed using a CT B pentamer with a known aggregating ligand chlorophenolred-β-D-galactopyranoside.[9] This mixture only gave a stable DLS signal for the first few minutes after mixing and always produced visible precipitate later. The signal from such a sample gave a very large sized aggregate, as shown in Figure 1A, and the measured hydrodynamic radius (Rh) differed from run to run. In contrast, the DLS analysis of the **12**/CT B pentamer system only showed narrowly distributed solution species that are between the sizes of a single CT B pentamer and a double CT B pentamer (3−4 nm in Rh). A more detailed measurement of complex formation is shown in Figure 1B. In this study, the concentration of the CT B pentamer was fixed at 12.0 μM, and the concentration of ligand **12** was titrated in a range of 2.0−24.0 μM. The aim here is to use the molar ratio method[17] to identify the most favorable complex species in solution. As Figure 1B shows, the peak of Rh values was reached when the CT B pentamer molar ratio is around 0.66, indicative of the formation of a 1:2 **12**/CT B pentamer complex. In addition, the change of measured Rh values from 3.1 nm of CT B pentamer alone to a peak of 4.1 nm corresponds to a 1.9-fold change in molecular weight.[18] This observed change almost exactly corresponds to the expected ratio of the molecular weights of a 1:2 complex between one **12** molecule and two CT B pentamers and that of a 1:1 complex of **12** and CT B pentamer.

The combination of DLS (Figure 1B) and ELISA data (Table 1) at this stage allows us to arrive at an estimate of the K_d for the association of ligand **12** toward the CT B pentamer. The DLS results, as shown in Figure 1B, suggest that two possible association modes exist for the **12**/CT B pentamer in solution. When the toxin B pentamer is in excess, the major binding mode is a 1:2 **12**/toxin ternary complex. When excess ligand is present (low mole ratio side of Figure 1B), the binding shifts toward the 1:1 **12**/toxin binary complex. Since in ELISA assays ligand **12** is always in large excess relative to the CT B pentamer, DLS suggests that the IC_{50} measured by ELISA is the true reflection of the affinity for the 1:1 ligand/toxin association. By the same ELISA assay, galactose showed an IC_{50} value of 38.5 mM (Table 1). The dissociation constant of galactose to CT B pentamer was reported to be about 40 mM.[19] This suggests

(17) Yoe, J. H.; Jones, A. L. *Ind. Eng. Chem., Anal. Ed.* **1944**, *16*, 111−115.
(18) According to the equation MW = (1.68 × Rh)∧2.3398, established for the DLS instrument by the manufacturer.
(19) Mertz, J. A.; McCann, J. A.; Picking, W. D. *Biochem. Biophys. Res. Commun.* **1996**, *226*, 140−144.

Table 1. IC_{50} Values for Branched Multivalent Ligands (**9**−**12**) and Comparison with Nonbranched Pentavalent Ligands

linker units of multivalent ligands (*n* in Scheme 2)	IC_{50} values (μM) of branched (penta-difinger) ligands (**9**–**12**)	gain by branched multivalent ligands over galactose	IC_{50} values (μM) of nonbranched (penta-monofinger) ligands[a]	gain by nonbranched multivalent ligands over galactose[a]
monovalent galactose	38 500	1	58 000	1
$n = 1$	191 (**9**)	200	242	240
$n = 2$	5.3 (**10**)	7 000	16	3600
$n = 3$	1.4 (**11**)	28 000	6	9700
$n = 4$	0.040 (**12**)	∼1 000 000	0.56	∼100 000

[a] Reported previously,[9] measured using an ELISA assay against the LT B pentamer binding to ganglioside. It was shown previously that one can directly compare results obtained from the LT-based ELISA assay with those of the CT-based assay used in this study.[16]

that the apparent K_d for the 1:1 **12**/CT B pentamer complex would be about 40 nM on the basis of the ELISA results listed in Table 1. This affinity is at the same level of the binding of CT's natural receptor GM1 toward the CT B pentamer, which was reported to be 50 nM.[19]

Ideally, one would like to use the DLS data from Figure 1B to also obtain the affinity of the 1:2 complex. However, the intrinsic properties of DLS measurement prevent such an analysis. Since DLS is sensitive to both the concentration and the size of a solution species, at any given concentration, the signal is dominated by the larger molecular species in solution. This is evident in Figure 1B, as the drop-off of Rh from the peak is slower at the high molar ratio side of the CT B pentamer when compared to that of the low molar ratio side for CT B pentamer, indicating that even when excess CT B pentamer is present, the measured signal is biased toward the 1:2 complex relative to the 1:1 complex or the unbound CT B pentamer. To alleviate this problem and obtain an estimate of the strength of the 1:2 complex formation for **12**/CT B pentamer, we performed a competition experiment using DLS in the presence of a known ligand of the CT B pentamer. We chose to use the well characterized *m*-nitrophenyl-α-D-galactopyranoside (MNPG), which has a K_d in the range of 100−200 μM toward the CT B pentamer (the K_d of MNPG to the LT B pentamer is 175 μM,[20] and it is known that MNPG exhibits the same inhibitory power against receptor binding by either LT or CT[16]). The result from the competition experiment is shown in Figure 1C. Although one still cannot obtain a precise K_d value by fitting the curve because of the intrinsic bias toward high molecular weight species by the DLS experiment (as evident by the nonsigmoidal curve in Figure 1C), we can estimate a limit of K_d for the formation of the 1:2 **12**/CT B pentamer complex. It is clear from the data that a few thousand times higher concentration of MNPG, relative to the concentration of ligand **12**, is needed to completely disrupt the formation of the 1:2 **12**/CT B pentamer complex. On the basis of the dissociation constant of MNPG toward the CT B pentamer (100−200 μM), we can conclude that the stability of the 1:2 **12**/toxin ternary complex is very high and that the complex has a dissociation constant roughly in the high tens and low hundreds nanomolar range.

The combined ELISA and DLS results also help us to distinguish between two different entropic contributions to multivalent inhibition. One primary entropic effect is due to an increase of the local concentration of finger molecules when one molecule of **12** first binds to one CT B pentamer. Each arm of **12** terminates in two finger molecules capable of binding, however, only one of which is needed to fill a binding site on CT in the 1:1 complex. As discussed above, the IC_{50} value obtained is a close representation of the affinity constant for the 1:1 complex formation. Therefore, one can conclude that the improvement in IC_{50} by a branched decavalent ligand over a nonbranched pentavalent ligand, as measured in this study (Table 1), is primarily due to the entropic effect of an increase of local ligand concentration.

The other major entropic effect comes when a second toxin B pentamer binds to an existing 1:1 ligand/toxin binary complex to yield a 1:2 ternary complex. In this case, the five remaining unbound fingers of **12** have been preorganized into an approximate pentagonal arrangement matching that of the binding sites on the second toxin B pentamer. This preordering should reduce the entropic cost of binding the second CT B pentamer relative to binding the first CT B pentamer. However, when both fingers on each arm of ligand **12** are involved in toxin binding, the perfection of branching point design becomes relevant as well. For 1:1 ligand/toxin binding, the branching point design may be less important, since the whole linker portion is designed to remain flexible in order to keep the overall loss of entropy small upon complex formation. In this case, the unbound finger on each ligand arm can still retain most of its free movement. However, upon association of the second toxin B pentamer, the branching design will be important because unfavorable entropic effect may be introduced when both fingers on each ligand arm are assuming a specific conformation. The combined ELISA and DLS studies for ligand **12** do suggest that the association for the second toxin B pentamer is less favorable than the association of the first B pentamer. This is evident from the rapid shift from a 1:2 to a 1:1 **12**/toxin complex seen in Figure 1B on the low mole ration side, as well as by the estimation of K_d values for both the 1:1 (∼40 nM) and 1:2 (low hundreds of nanomolar) **12**/CT B pentamer complex formation discussed earlier. All this information seems to suggest that the current design of the branching is less than ideal for optimal 1:2 lignad/toxin complex formation, but the exact cause for the unfavorable entropic effect can be discussed in more depth on the basis of structural data.

Further characterization of the association between ligand **12** and the CT B pentamer was carried out by cocrystallizing the ligand and protein molecules. Crystals of the **12**/CT B pentamer complex appear to be isomorphous to those of the GM1 pentasaccharide (GM1-OS)/CT B pentamer complex[21,22] and those of the CT B pentamer complexed with a pentavalent inhibitor.[23] In all three cases, the crystals diffract to near-atomic resolution. Refinement of the present structure against 1.45-Å

(20) Pickens, J. C.; Merritt, E. A.; Ahn, M.; Verlinde, C. L. M. J.; Hol, W. G. J.; Fan, E. *Chem. Biol.* **2002**, *9*, 215−224.

(21) Merritt, E. A.; Sarfaty, S.; van den Akker, F.; L'hoir, C.; Martial, J. A.; Hol, W. G. J. *Protein Sci.* **1994**, *3*, 166−175.

(22) Merritt, E. A.; Kuhn, P.; Sarfaty, S.; Erbe, J. L.; Holmes, R. K.; Hol, W. G. J. *J. Mol. Biol.* **1998**, *282*, 1043−1059.

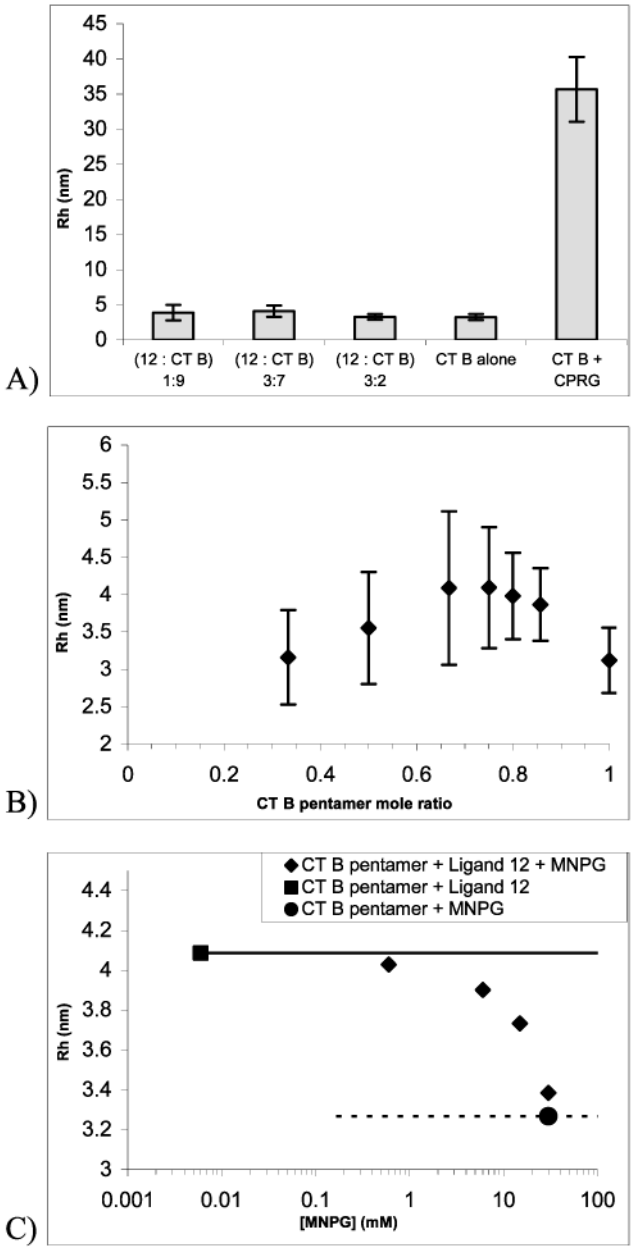


Figure 1. DLS results for the complexation of branched multivalent ligand **12** and CT B pentamer in solution. In all diagrams, data points are measured peak positions. Error bars in parts A and B are the polydispersities of the sample and are removed in part C for clarity. (A) Hydrodynamic radii measured by DLS at fixed total (CT B pentamer + ligand **12**) concentration (28.0 μM) with various ligand/CT B pentamer ratios. The positive control for large random aggregation was measured using 28.0 μM CT B pentamer with 0.5 mM chlorophenolred-β-D-galactopyranoside (CPRG). (B) Hydrodynamic radii measured by DLS at fixed CT B pentamer concentration (12.0 μM), with varying amounts of ligand **12** (2.0−24.0 μM). In the presence of excess ligand (low CT B pentamer mole ratio), the predominant species contributing to the measured Rh is the 1:1 ligand/toxin complex. In the presence of excess toxin pentamer (high mole ratio), the predominant species contributing to the measured Rh is the free toxin B pentamer. At a mole ratio of 0.66, the predominant species is the 1:2 ligand/toxin complex. (C) Hydrodynamic radii measured by DLS with fixed concentrations of CT B pentamer (12.0 μM) and, if present, 6.0 μM ligand **12**. The concentrations of MNPG, if present, were ranging from 0.6 to 30.0 mM. The top solid horizontal line represents the boundary for the 1:2 **12**/CT B pentamer complex, and the bottom dashed horizontal line represents the boundary for a single CT B pentamer.

data yielded an excellent crystallographic model (Table 2) that reveals the key binding interactions of the double finger bivalent

Table 2. Crystallographic Data

data collection			
resolution (highest shell)	50−1.45 Å (1.51−1.45 Å)		
unique data measured	81475 (5338)		
completeness	92% (54%)		
R_{merge} on intensities	0.048 (0.306)		
data used in refinement			
reflections (working set)	77338		
reflections (R_{free} set)	4085		
cutoffs	25−1.45 Å		
model			
R	0.125		
R_{free}	0.164		
protein atoms	4095	$\langle B_{iso}\rangle = 17$ Å^2	$\langle A\rangle = 0.44$
ligand atoms	198	$\langle B_{iso}\rangle = 27$ Å^2	$\langle A\rangle = 0.66$
water molecules	674	$\langle B_{iso}\rangle = 35$ Å^2	$\langle A\rangle = 0.53$
stereochemistry			
rms nonideality bond lengths	0.018 Å		
rms nonideality 1−3 lengths	0.035 Å		
overall coordinate ESU (Cruickshank DPI)	0.07 Å		
overall coordinate ESU (maximum likelihood)	0.06 Å		

group terminating each of the ligand's five arms (Figure 2). The ligand is observed to bind two toxin B pentamers simultaneously, forming a "sandwich" similar to that observed by Kitov et al.[8] for the binding of a smaller decavalent ligand to two B pentamers of the Shiga-like toxin SLT-1. This confirms the 1:2 ligand/protein binding mode observed in the DLS study described above. The asymmetric unit of the present **12**/CT B pentamer crystal structure contains one toxin pentamer and half of one molecule of the inhibitor **12**. The two toxin pentamers of the sandwich are, thus, crystallographically equivalent. The central portion of the ligand is necessarily in violation of this strict crystallographic symmetry and is, in any case, expected to be relatively flexible; thus, it is disordered and not visible in the crystal structure. However, the divalent fingers at the tip of each arm of the ligand obey the crystallographic symmetry relating the two toxin pentamers which they bridge.

The five crystallographically independent receptor-binding sites are each observed to be occupied by one of the 10 β-D-galactose moieties in ligand **12**. The binding mode of this sugar is, as expected, identical to that seen for the terminal galactose of the native receptor saccharide GM1-OS[21] and for many smaller receptor-binding inhibitors derived from β-D-galactose.[20,24,25] The amide-linked $(CH_2)_4$ moiety attached at the galactose C1 atom lies in roughly the same volume of space as the GalNAc residue of GM1-OS, while the remaining well-ordered portion of the ligand **12** does not closely approach either of the two toxin pentamers that it bridges (Figures 2 and 3).

The observed binding mode of the linking group attached to the galactose moiety is in distinct contrast to the binding mode observed previously for a pentavalent compound on the basis of a tighter binding monovalent inhibitor (MNPG) containing the anomeric form α-D-galactoside.[23] In the case of the pentavalent inhibitor, the α-anomeric linkage allows the ligand to conform much more closely to the surface of the protein and

(23) Merritt, E. A.; Zhang, Z.; Pickens, J. C.; Ahn, M.; Hol, W. G. J.; Fan, E. *J. Am. Chem. Soc.* **2002**, *124*, 8818−8824.
(24) Sixma, T. K.; Pronk, S. E.; Kalk, K. H.; Van Zanten, B. A. M.; Kingma, J.; Witholt, B.; Hol, W. G. J. *Nature* **1992**, *355*, 561−564.
(25) Merritt, E. A.; Sarfaty, S.; Feil, I. K.; Hol, W. G. J. *Structure* **1997**, *5*, 1485−1499.

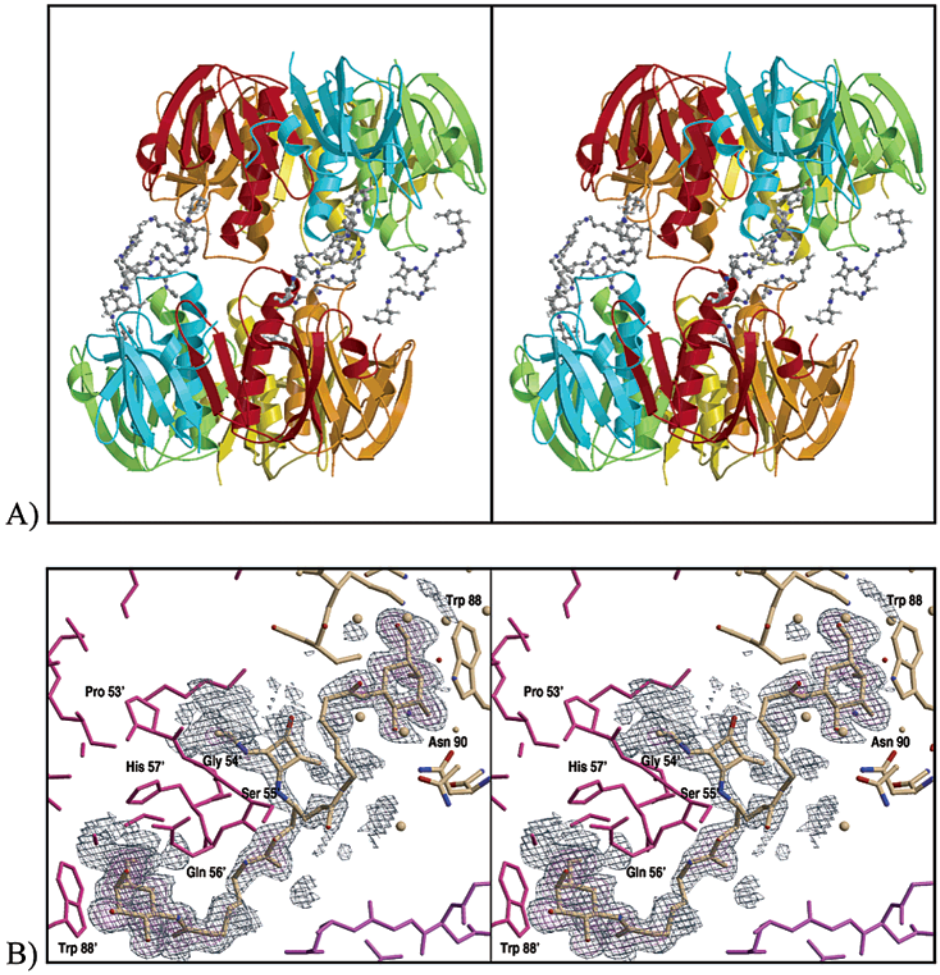


Figure 2. (A) Stereo view of the 1:2 complex between ligand **12** and the CT B pentamer as observed crystallographically. (B) Closer look of two fingers of one arm of ligand **12** in the crystal structure (stereo). The final crystallographic model for one of the five bivalent end groups of ligand **12** is shown superimposed onto difference electron density from σ_A-weighted (mFo−Fc) difference map calculated prior to inclusion of the ligand in structure refinement. Map contours are drawn at 1.5 and 3.0 σ.

the atoms of the first modular linking group lie in the portion of the receptor-binding site occupied by a sialic acid residue in the GM1-OS/CT B pentamer complex. In the branched multivalent ligand inhibitor **12**, because of the β-D-galactose finger, the chain leading to the branch point is pointing away from the face of the toxin B pentamer (Figure 3). Both binding modes leave substantial portions of the receptor-binding site unoccupied, suggesting that it will be possible to further improve the multivalent affinity by optimizing the single site affinity for both α- and β-anomeric galactose derivatives.

The **12**/CT B pentamer structure suggests that another avenue of ligand optimization may also be possible. As discussed earlier, the entropic effect upon the binding of the second toxin B pentamer seems to be less favorable than the binding to the first B pentamer, despite a favorable preorganization of five unbound fingers after the formation of a 1:1 complex. The structural data now offer the insight to this cause. The central core of ligand **12** and the larger portion of the linking groups attached to it are highly flexible. This causes the effective length of the linkers to grow only as the square root of the linear atom count (see Fan et al.[9] or Kramer and Karpen[7] and references therein) and also accounts for the lack of an ordered state implied by the lack of electron density for these regions in the crystal structure. Since this portion of the ligand is flexible in solution and remains so after binding, the concomitant entropy loss upon complex formation must be relatively small. In contrast to this, the bivalent fragment at the end of each linker arm is presumably also quite flexible in solution but can only bridge toxin pentamers by assuming an extended conformation. This extended conformation is well-ordered in the crystal structure (Figure 2). This ordering of an extended conformation upon binding presumably acts to make formation of the toxin/**12**/toxin sandwich less favorable in free energy than it would otherwise be, despite an overall favorable change in entropy, as discussed earlier. Reducing the flexibility in this divalent branching fragment or optimizing the branch on the basis of structural data such that the branch interacts favorably with the toxin's surface should, therefore, increase the effective binding affinity of the decavalent ligand, even if the long linkers joining these terminal fragments to the central core remain flexible.

In summary, this study confirms that the structure-based design of branched multivalent ligands, which are capable of binding to more than one target protein molecule, is a general approach toward high-affinity protein ligands. The current study also reiterates the success of our modular synthetic method, showing that it can efficiently lead to multivalent ligands with effective dimensions matching the binding site distribution of the target protein. We have also shown that dynamic light

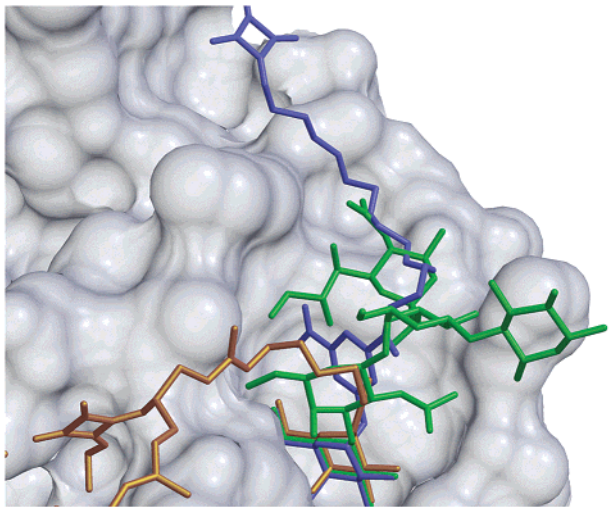


Figure 3. Receptor-binding site of cholera toxin, showing the binding modes of the natural receptor GM1 and two competitive multivalent ligands. The terminal pentasaccharide of the natural receptor, ganglioside GM1, is shown in green.[22] One arm of a pentavalent ligand based on *m*-nitrophenyl-α-D-galactose is shown in blue.[23] One tip of one arm of the decavalent ligand **12** as seen in the present structure is shown in yellow. Crystal structures of these three toxin/ligand complexes are all isomophous. Both GM1 and ligand **12** contain the *β*-anomer of galactose, whereas the pentavalent ligand contains the α-anomer. Neither of the multivalent ligands occupies the full GM1-binding site, suggesting that there is scope for further optimization to improve the single site affinity.

scattering measurement not only permits the unequivocal proof that these multivalent ligands bind to their target protein through intrinsic affinity because of the absence of large random aggregates but also can identify discrete ligand/protein complexes in solution. DLS can also aid in the analysis of complex stability and should be a complementary tool to other analytical techniques such as calorimetry titration[26] or other methods.[27,28] X-ray crystallography provides solid support to solution studies, demonstrating the power of combining different biophysical techniques. In addition, with more and higher resolution crystal structures available for structure-based multivalent ligand design,[8,10,12,23] one should expect eventually the arrival of ultrahigh affinity multivalent ligands that are geometrically and chemically complementary to their macromolecular targets. For CT and LT related diarrheal diseases, the prospect of using such ultrahigh affinity multivalent ligands as therapeutic intervention is promising. Because the site of biological action is at the epithelial cell surfaces in the intestinal lumen, a fully optimized multivalent ligand against CT or LT receptor binding should be orally delivered and ideally would not be transported into the bloodstream. This is in contrast to the more typical requirement that a drug enter the bloodstream and be delivered to intracellular targets elsewhere in the body. The molecular weights and physical dimensions of the optimized multivalent ligands will be large, which make those ligands unlikely to enter or circulate in the bloodstream, but eminently suited for use as prophylaxis against CT and LT.

Experimental Section

Synthesis. (A) General. Commercially available reagents and common solvents (reagent grade or HPLC grade) for synthesis and separation were used without further purification as purchased. HPLC purification was performed on an Agilent 1100 quaternary pump system with a variable wavelength detector. The C18 preparative column was purchased from Vydac (21 × 250 mm^2, 10−15 μm). ^{1}H NMR spectra were recorded at 300 MHz on a Bruker AC-300 instrument, while mass spectra were obtained from a Bruker Esquire 3000 or an Agilent 1100 MSD-Trap SL electrospray ion trap mass spectrometer.

(B) 2-Amino-1,3-propanediol Di[5-(*β*-D-galactosylcarbamoyl)-pentylcarbamoylacid ester] (3). A solution of 1.58 g of 4-nitrophenyl chloroformate in a small amount of dry CH_2Cl_2 was added to an ice-cold solution of 0.5 g (2.62 mmol) of *N*-Boc-2-amino-1,3-propanediol (compound **1**)[29] in 20 mL of dry CH_2Cl_2 with 228 μL DIPEA. After the mixture was stirred for 1 h, the ice bath was removed and the reaction mixture was stirred overnight at rt. Solvent removal and subsequent chromatography on silica gel (20% hexane/CH_2Cl_2) gave the activated dicarbonate **2** in (0.823 g) 60% yield. ^{1}H NMR ($CDCl_3$) δ 8.30 (d, 4H), 7.38 (d, 4H), 5.05 (broad, 1H), 4.50 (m, 5H), 1.50 (s, 9H). ESI-MS *m*/*z* 544.2 $(M + Na)^+$.

Compound **2** (0.16 g) was dissolved in THF and added dropwise to a solution containing 0.20 g of *N*-(*ϵ*-aminocaproyl)-*β*-D-galactopyranosylamine in 60 mL of water and 0.5 mL of pyridine. After the solution was stirred overnight at rt, the solvent was removed and the residue was dissolved in water. The pH of the aqueous solution was adjusted to 2 by the addition of 1% TFA. The resulting solution was subjected to HPLC purification (Solvent A: 0.1% TFA in water. Solvent B: CH_3CN. Gradient: 0% B for 10 min, and then 0−40% B over 40 min.) to give Boc-protected **3** (0.21 g) in 82% yield. ESI-MS *m*/*z* 828.7 $(M + H)^+$. Then 0.21 g of Boc-protected **3** was deprotected by treatment with 5 mL of 1:1 CH_2Cl_2/TFA for 5 min and used for the next step without further purification. ^{1}H NMR (D_2O) δ 4.23 (br, 4H), 3.94 (s, br, 2H), 3.56−3.79 (m, 13H), 3.09 (br, 4H), 2.29 (m, br, 4H), 1.60 (m, br, 4H), 1.48 (m, br, 4H), 1.32 (m, br, 4H). ESI-MS *m*/*z* 728.6 $(M + H)^+$.

(C) General Protocol for Full Ligand Assembly. After evaporation of solvent, compound **3** from the above deprotection reaction was dissolved in 10 mL of water and treated with NaOH and $NaHCO_3$ to a neutral pH. This was mixed with a solution of dimethyl squarate (71 mg) in 10 mL of MeOH. The solution was stirred at rt overnight to produce activated divalent finger module **4**. After removal of MeOH and adjusting pH to 2 by TFA, the solution was subjected to HPLC separation (Solvent A: 0.1% TFA in water. Solvent B: CH_3CN. Gradient: 0% B for 10 min, 0−8% B over 20 min, and 8−15% B over 20 min.) to produce pure branched finger module **4** [60% from **3** based on UV absorbance ($\epsilon_{250} = 1.818 \times 10^4$ M^{-1} cm^{-1}), ESI-MS *m*/*z* 860.5 $(M + Na)^+$]. Subsequently, compound **4** and a core−linker module (**5**−**8**)[9] were mixed in 10 mL of 1:1 MeOH/water. The solution was adjusted to pH ≈ 9 by 0.2 M $NaHCO_3$. This solution was stirred at rt for 2 days. HPLC purification (Solvent A: 0.1% TFA in water. Solvent B: CH_3CN. Gradient: 5% B for 3 min, 5−10% B over 7 min, and 10−35% B over 20 min) gave the final branched multivalent ligand (**9**−**12**).

Ligand **9** was prepared from 25.4 μmol of **4** and 1.30 μmol of **5** in 78% yield. ESI-MS *m*/*z*: 1233.0 $(M + H)^{5+}$, 1540.8 $(M + H)^{4+}$, 2053.9 $(M + H)^{3+}$.

Ligand **10** was prepared from 25.4 μmol of **4** and 0.80 μmol of **6** in 81% yield. ESI-MS *m*/*z*: 1094.0 $(M + H)^{7+}$, 1276.2 $(M + H)^{6+}$, 1531.1 $(M + H)^{5+}$, 1913.5 $(M + H)^{4+}$.

Ligand **11** was prepared from 25.4 μmol of **4** and 0.83 μmol of **7** in 85% yield. ESI-MS *m*/*z*: 915.1 $(M + H)^{10+}$, 1016.9 $(M + H)^{9+}$, 1143.9 $(M + H)^{8+}$, 1307.0 $(M + H)^{7+}$, 1524.6 $(M + H)^{6+}$, 1829.1 $(M + H)^{5+}$.

Ligand **12** was prepared from 25.4 μmol of **4** and 2.08 μmol of **8** in 83% yield. ESI-MS *m*/*z*: 1064.9 $(M + H)^{10+}$, 1183.1 $(M + H)^{9+}$, 1330.7 $(M + H)^{8+}$, 1520.4 $(M + H)^{7+}$, 1773.7 $(M + H)^{6+}$.

(26) Dam, T. K.; Roy, R.; Das, S. K.; Oscarson, S.; Brewer, C. F. *Glycobiology* **2000**, *10*, 56.
(27) Gestwicki, J. E.; Cairo, C. W.; Mann, D. A.; Owen, R. M.; Kiessling, L. L. *Anal. Biochem.* **2002**, *305*, 149−155.
(28) Kitova, E. N.; Kitov, P. I.; Bundle, D. R.; Klassen, J. S. *Glycobiology* **2001**, *11*, 605−611.

IC_{50} Measurements. The CT GD1b enzyme-linked adhesion assay was performed as previously reported.[16] Test samples consisted of 6 ng/mL CT B pentamer−horseradish peroxidase conjugate (Sigma-Aldrich, Milwaukee, WI) preincubated with desired ligands for 2 h at rt. At least two independent experiments for each inhibitor were carried out in triplicates and validated against a concentration gradient of 0, 1.5, 3, 6, and 12 ng/mL toxin peroxidase conjugate. IC_{50} values were calculated from triple data sets of at least seven different concentrations of competitive ligands by nonlinear regression using the Prism software (version 3.0, GraphPad Software, Inc.) and reported as an average of two or three independent determinations (Table 1).

Dynamic Light Scattering (DLS). Protein or ligand in phosphate buffered saline (150 mM NaCl/10mM phosphate, pH 7.2) was individually filtered through inorganic membrane filters (Whatman, Anodisc13, 0.02 μm) into separate vials. Then various portions of protein and ligand were mixed and, without further filtration, transferred into a sample cell for measurement. DLS measurement was done on a DynaPro99 instrument (Protein Solutions Inc.) illuminated by a 25-mW, 832.8-nm wavelength, solid-state laser at 25 °C. Data analysis was performed using the dynamics software (Version 5.25.44) provided with the instrument. Inverse Laplace Transform ("regularization fit") analysis was used to find the mean and standard deviation (polydispersity) of the hydrodynamic radius (Rh) distribution for the molecule/complex species in solution.

Protein Production. Cholera toxin B pentamer for crystallization experiments was expressed in *E. coli* and purified by galactose affinity chromatography essentially as described.[30]

Crystallization and Structure Determination. Crystals of CT B pentamer complexed with ligand **12** grew from sitting drops consisting of 2 μL of protein at 5.0 mg/mL in 100 mM TrisHCl at pH 7.5, 2 μL of 0.25 mM ligand **12** in the same buffer, and 2 μL of well buffer containing 50 mM NaCl, 100 mM TrisHCl at pH 7.5, and 40% PEG 1000. Crystals were isomorphous to the previously determined complex of CT B pentamer with the GM1 pentasaccharide (GM1-OS).[21]

X-ray intensities from a single flash-frozen crystal were measured using a wavelength of 0.9789 Å at the APS Structural Biology Center beamline 19 BM. The data were integrated and scaled using programs HKL2000 and TRUNCATE.[31,32] Isotropic refinement of the protein and 497 well-ordered water molecules not in the region of the receptor-binding site was carried out in REFMAC version 4[33] using data to 1.5 Å. This yielded residuals $R = 0.188$ and $R_{free} = 0.216$. Clear electron density was present in σ_A weighted difference maps (mFo−Fc) for the "double finger" portion of the inhibitor at all five binding sites. The 1:2 complex of ligand/toxin was found to span two asymmetric units across a crystallographic 2-fold axis. This necessarily means that the central portions of the full ligand **12** are crystallographically disordered, as the core structure is not itself 2-fold symmetric. However, the double finger terminating each arm of the ligand can, and does, conform to the crystallographic 2-fold, although two alternate conformations are found for two of the five crystallographically unique arms. This portion of the ligand was added to the crystallographic model, and refinement was continued in alternation with manual inspection and refitting of the model. Addition of individual atomic displacement parameters U^{ij} to the model and inclusion of data to 1.45 Å brought the residuals to $R = 0.133$ and $R_{free} = 0.171$. Additional discrete water molecules were identified, as well as partial oxidation or misformation of the Cys9−Cys89 disulfide linkages as previously found in the isomorphous GM1-OS/CT B pentamer structure.[22] The final model yielded residuals $R = 0.125$ and $R_{free} = 0.164$, with statistical properties described in Table 2.

Model-fitting, placement, and real-space refinement of ligand and water molecules were carried out using XFIT.[34] The choice of restraint weights for U^{ij} parameters was guided by analysis of the overall distribution of anisotropy using PARVATI,[35] and the anisotropic treatment of water molecules was additionally restrained toward isotropy using a local modification to REFMAC.[36] The resulting mean value of atomic anisotropy, $\langle A \rangle$, is given in Table 2 for protein, ligand, and solvent atoms. No noncrystallographic symmetry restraints were used during refinement. Figures were produced using XFIT,[34] MSMS,[37] and RASTER3D.[38] Structure factors and the refined model have been deposited with the PDB (accession code 1MD2).

Acknowledgment. This work was supported by the NIH (AI44954, AI34501, GM62617, and GM54618). Use of the Argonne National Laboratory Structural Biology Center beamline 19BM at the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract No. W-31-109-ENG-38.

JA027584K

(29) Benoist, E.; Loussouarn, A.; Remaud, P.; Chatal, J.-F.; Gestin, J.-F. *Synthesis* **1998**, *8*, 1113−1118.
(30) Minke, W. E.; Pickens, J.; Merritt, E. A.; Fan, E.; Verlinde, C. L. M. J.; Hol, W. G. J. *Acta Crystallogr.* **2000**, *D56*, 795−804.
(31) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307−326.
(32) Collaborative Computational Project Number 4. *Acta Crystallogr.* **1994**, *D50*, 760−763.
(33) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr.* **1997**, *D53*, 240−255.
(34) McRee, D. *Practical Protein Crystallography*; Academic Press: San Diego, CA, 1993.
(35) Merritt, E. A. *Acta Crystallogr.* **1999**, *D55*, 1109−1117.
(36) Murshudov, G.; Lebedev, A.; Vagin, A. A.; Wilson, K. S.; Dodson, E. J. *Acta Crystallogr.* **1999**, *D55*, 247−255.
(37) Sanner, M. F.; Spehner, J.-C.; Olson, A. J. *Biopolymers* **1996**, *38*, 305−320.
(38) Merritt, E. A.; Bacon, D. J. *Methods Enzymol.* **1997**, *277*, 505−524.