# Development of small peptides recognizing a monosaccharide by combinatorial chemistry

**Naoki Sugimoto,*[ab] Daisuke Miyoshi[a] and Jin Zou[ac]**

[a] *Department of Chemistry, Faculty of Science, Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe 658-8501, Japan. E-mail: sugimoto@konan-u.ac.jp; Fax: +81-78-435-2539; Phone: +81-78-435-2497*

[b] *High Technology Research Center, Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe 658-8501, Japan*

[c] *Department of Biological Engineering, School of Chemical Engineering, Heibei University of Technology, Tiaujin 300130, People's Republic of China*

***Received (in Cambridge, UK) 18th August 2000, Accepted 9th October 2000***
***First published as an Advance Article on the web 7th November 2000***

**Small peptides selected from a library of 62 000 chemically synthesized peptides were able to recognize a monosaccharide specifically and efficiently using both sandwiching and hydrogen bonding interactions, much like sugar-binding proteins (lectins) with the appropriate orientation of amino acid side chains.**

Protein–oligosaccharide interactions play a key role in the control of various normal and pathological processes.[1,2] Because of their importance, there has been considerable drive toward the better understanding of the saccharide (carbohydrate) binding specificity of proteins (lectins). Progress has been made using X-ray crystallography, isothermal titration microcalorimetry and site-directed mutagenesis experiments.[3–6] These numerous studies have found diverse mechanisms for saccharide recognition by proteins that have evolved independently but that share some key features. High selectivity for a saccharide is achieved through a combination of hydrogen bonding to the saccharide hydroxy groups and van der Waals packing of the hydrophobic sugar face against aromatic amino acid side chains forming what is called sandwiching interaction.[7,8] Since the binding of saccharides by proteins involves a variety of recognition processes and the proteins themselves exhibit considerable structural diversity, it is difficult to obtain detailed information using thermodynamic and kinetic parameters resulting from the analysis of a large lectin–oligosaccharide complex. Establishing a model system is one of the essential methods to reveal function and mechanism of the specific interactions quantitatively, but there is no good model system that includes the two binding interactions found between proteins and saccharides.

Here, a small oligopeptide–monosaccharide complex was developed as a model system for wild type protein–saccharide complexes which includes the hydrogen bonding and sandwiching recognition mechanisms. Firstly, in order to investigate the relationship between peptide conformation and sandwiching interaction ability, we synthesized alanine-based oligopeptides with one or two tryptophan residues to test the sandwiching interaction between the monosaccharide and the two tryptophan indole side chains, KW (S *i*+0), KWW (S *i*+1), KWAW (S *i*+2), KWAAW (S *i*+3), KWAAAW (S *i*+4), KWAAAAW (S *i*+5) and KWAAAAAW (S *i*+6).† Fig. 1 shows relative fluorescence intensities ($\Delta F$) of these peptides in the presence ($F$) and absence ($F^0$) of D-galactose, $\Delta F = (F - F^0) \times 100/F^0$.‡ Because $\Delta F$ shows saccharide sandwiching ability of the oligopeptide,[8] these results indicate that the oligopeptide is able to recognize the monosaccharide effectively with W residues at *i* and *i*+4. Aoyama and co-workers showed only two residual peptides, W–W, can recognize maltodextrin *via* sugar–bisindole interaction,[12] as the result of the (S *i*+1) peptide. Our results show that the (S *i*+4) peptide is capable of recognizing a monosaccharide more efficiently than the (S *i*+1) peptide and that the aromatic residues at *i* and *i*+4 have a critical role in this recognition. Previously, tryptophan and porphyrin binding peptides were obtained using combinatorial chemistry and phage display, respectively.[9–11] These peptides also had two aromatic residues and recognized each target molecule using only the sandwiching interaction with these two aromatic residues. Therefore, if both the hydrogen bonding and sandwiching interaction can be introduced suitably in one peptide, more efficient and specific recognition of a monosaccharide could be acquired, even in a small oligopeptide.

We have attempted to put both of the binding modes in the appropriate orientation into a small peptide using combinatorial chemistry. It was shown that the two aromatic residues at *i* and *i*+4 had a critical role for the sandwiching interaction by the fluorescence measurements of (S *i*+0 ~ 6) peptides. Therefore, using a basic split and mix method on a solid phase,[12] a biased combinatorial peptide library with Aro-X-X-X-Aro, where Aro is an L-aromatic amino acid and X is a natural L-amino acid (except C), was synthesized to have sandwiching interaction with *i* and *i*+4 aromatic residues. D-Erythrose ($HOCH_2$-$[CH(OH)]_2CHO$, Sigma–Aldrich Co.) was chosen as a target monosaccharide because D–erythrose is a light yellow syrup and can be followed with visible light.[13] The peptide library and D-erythrose were mixed, gently shaken overnight in a buffer containing 10 mM NaCl, 10 mM $Na_2HPO_4$, and 1 mM EDTA (pH 7.0) at rt then washed with buffer solution to eliminate any non-specific interactions. It was confirmed that the background signal arising from the non-specific interaction between D-erythrose and the beads in the selection procedure is low enough to select specific colored beads. Using a microscope, 15 beads were then selected from a library theoretically containing

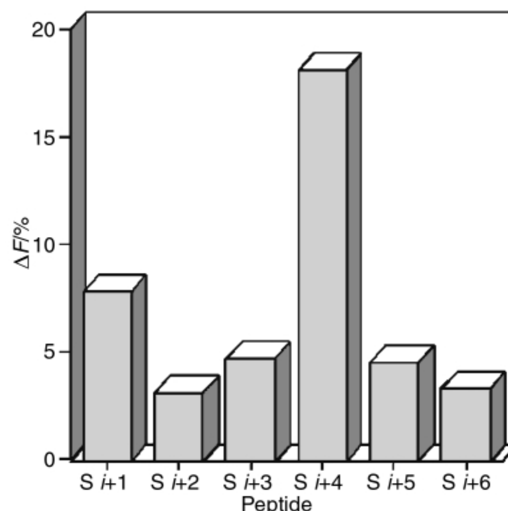


**Fig. 1** Ratio of fluorescence intensities ($\Delta F$) of 5 μM peptides at 348 nm and 1.0 °C. All experiments were carried out in a buffer of 10 mM NaCl, 1 mM $Na_2HPO_4$ and 1 mM EDTA (pH 7.0). The excitation wavelength for Trp is 278 nm and detected at 348 nm.

DOI: 10.1039/b006714h



**This journal is © The Royal Society of Chemistry 2000**

**Table 1** Binding constants and free energies of peptide–saccharide complexes at 1 °C

| Peptide sequence | $K_a/10^4$ M[a] | | $-\Delta G^0$/kcal mol$^{-1}$ [b] | |
|---|---|---|---|---|
| | D-Erythrose | D-Galactose | D-Erythrose | D-Galactose |
| WGDEY | 35.0 ± 2.9 | 5.22 ± 0.42 | 6.95 ± 0.05 | 5.91 ± 0.04 |
| WADEF | 26.5 ± 2.3 | 3.81 ± 0.27 | 6.80 ± 0.05 | 5.75 ± 0.04 |
| KWAAAW | 0.41 ± 0.06 | 3.89 ± 0.13 | 4.53 ± 0.06 | 5.76 ± 0.02 |

[a] $K_a$ values were estimated by the curve fitting procedure described in the text. [b] $-\Delta G^0$ values were determined with the equation $\Delta G^0 = -RT\ln K_a$

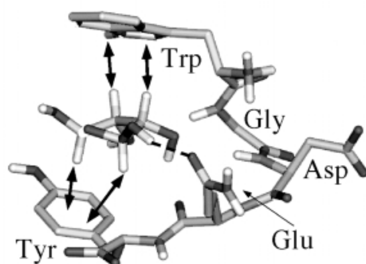


**Fig. 2** Minimized energy structure of WGDEY–D-erythrose. Lines and dotted lines indicate CH-π and hydrogen-bonding interaction, respectively.

62 000 kinds of peptides. Ten of the 15 selected peptides had a common motif: WNNNF/T where two of the three N residues were negatively charged amino acids (D or E).

To confirm the highly selective monosaccharide binding by motif peptides and to investigate the motif peptide–D-erythrose interaction, association constants ($K_a$) of motif peptides for monosaccharides were estimated from fluorescence intensity change by a curve fitting procedure described previously,[14] using the same buffer as the selection experiments. Table 1 shows the $K_a$ of WGDEY and WADEF with D-erythrose and a reference sugar (D-galactose), respectively, and the values of the (S $i$+4) peptide with sugars as references. The difference observed between motif peptides-D-erythrose and motif peptide-D-galactose suggests that motif peptides can recognize a target molecule selectively. Affinity of the lectin for a monosaccharide is usually weak with association constants in the millimolar range, yet it is often highly selective.[2] On the other hand, affinity constants for binding of a simple oligosaccharide to most lectins are estimated to be $10^3$–$10^6$ M$^{-1}$).[15] Therefore, the $K_a$ value of the motif peptide-D-erythrose is large enough for a monosaccharide recognition protein. The difference between $\Delta G°$ of motif peptides-D-erythrose (−7.0 and −6.8 kcal mol$^{-1}$) and reference peptide-D-galactose (−4.5 kcal mol$^{-1}$) is about −2.5 kcal mol$^{-1}$. The net contributions of an intramolecular hydrogen bond between protein side chains and an intermolecular hydrogen bond between protein and water were estimated to be −2.3 and −1.2 kcal mol$^{-1}$, respectively.[16] Furthermore, hydrogen bonds in proteins and α-helical peptides were estimated to be −1.3 and −1.0 kcal mol$^{-1}$, respectively.[17,18] Therefore, one or two appropriate hydrogen bonding interactions were successfully introduced into the selected peptides-D-erythrose interaction using combinatorial chemistry. Fig. 2 shows the minimized energy structure of WGDEY–D-erythrose complex§ which suggests that the peptide sandwiches D-erythrose with W and Y and that a hydrogen bond is formed between the hydroxy group of D-erythrose and a polar side chain of the peptide. Both thermodynamic and molecular modelling results indicate that the motif peptides can bind D-erythrose with the two binding manners as observed in the wild-type saccharide binding proteins. These results also show that all ($i$ ~ $i$+4) residues have critical roles for specific monosaccharide recognition. The $i$ and $i$+4 aromatic residues are necessary to form a sandwiching interaction, and the acidic residues at $i$+1, $i$+2 and $i$+3 are well suited for hydrogen-bonding interactions. Other selected small residues such as A and G are also important to fit peptide structure for the multi-point saccharide recognition. These delicate residue arrangements were achieved with a combination of rational design studies and a biased combinatorial library.

In summary, the smallest model system of saccharide-binding protein with both hydrogen bonding and the sandwiching interaction was developed using a combination of rational design and combinatorial chemistry. This system is very useful to obtain quantitative parameters of general saccharide–protein interactions.

We thank Dr Angel Maki, Department of Chemistry, Stanford University, for critical reading of the manuscript and helpful comments. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan, and Grants from 'Research for the Future' Program of the Japan Society for the Promotion of Science.

## Notes and references

† All peptides used in this study were synthesized by the Fmoc procedure on a solid support of Fmoc-NH-SAL resin then purified by HPLC. These peptides were named as S $i$+$x$ ($x$ = 0–6) by means of short oligopeptides which have one or two Trp residues at $i$ and $i$+$x$. The peptide sequences were confirmed by peptide sequencing.

‡ All fluorescence spectra were measured from 250 to 350 nm and excited at 278 nm in a 1 cm path length cuvette in the presence and absence of D-Gal at 1.0 °C. All experiments were done in a buffer of 10 mM NaCl, 1 mM $Na_2HPO_4$ and 1 mM $EDTA_4$ (pH 7.0). Peptide and saccharide concentrations were 5 μM and 50 mM, respectively.

§ The minimized energy structure was obtained after molecular calculation using QUANTA 96/CHARMM 23.2 on a Silicon Graphics Indigo² workstation running IRIX 5.3. The energy of the complex was minimized by the parameter of Newton-Raphson and VERLET with an adopted basis set. Dynamics simulation of the minimized energy structure was performed to obtain various conformations as initial coordinates for the calculation. The VERLET algorithm with a time step of 1 fs was used in the dynamics simulation. The system was heated from 0 to 1000 K for the conformational search in the MD simulation.


1 N. Sharon and H. Lis, *Science*, 1989, **246**, 227.
2 L. A. Lasky, *Science*, 1992, **258**, 964.
3 F. A. Quiocho, D. K. Wilson and N. K. Vyas, *Nature*, 1989, **340**, 404.
4 B. W. Sigurskjold and D. R. Bundle, *J. Biol. Chem.*, 1992, **267**, 8371.
5 M. H. Du, U. Spohr and R. U. Lemieux, *Glycoconj. J.*, 1994, **11**, 443.
6 W. I. Weis and K. Drickamer, *Annu. Rev. Biochem.*, 1996, **65**, 441.
7 J. M. Rini, K. D. Hardman, H. Einspahr, F. L. Suddath and J. P. Carver, *J. Biol. Chem.*, 1993, **268**, 10 126.
8 J. Otsuki, K. Kobayashi, H. Toi and Y. Aoyama, *Tetrahedron Lett.*, 1993, **34**, 1945.
9 N. Sugimoto and S. Nakano, *Chem. Lett.*, 1997, 939.
10 N. Sugimoto, H. Kazuta, J. Zou and D. Miyoshi, *Chem. Commun.*, 1999, 677.
11 J. Kawakami, T. Kitano and N. Sugimoto, *Chem., Commun.*, 1999, 1765.
12 K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski and R. J. Knapp, *Nature*, 1991, **354**, 82.
13 K. Sonogashira and M. Nakagawa, *Bull. Chem. Soc. Jpn.*, 1972, **45**, 2616.
14 S. Nakano and N. Sugimoto, *Bull. Chem. Soc. Jpn.*, 1998, **71**, 2205.
15 D. K. Mandal, N. Kishore and C. F. Brewer, *Biochemistry*, 1994, **33**, 1149.
16 K. Tanaka, Y. Yamagata, J. Funahashi, Y. Hioki, S. Kuramitsu and K. Yutani, *Biochemistry*, 1999, **38**, 12 698.
17 J. Fernandez-Recio, A. Romero and J. Sancho, *J. Mol. Biol.*, 1999, **290**, 319.
18 B. J. Stapley and A. J. Doig, *J. Mol. Biol.*, 1997, **272**, 465.