This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
Trivalent ligands for CXCR4 bearing polyproline linkers show specific recognition for cells with increased CXCR4 expression

Wataru Nomura, Taisuke Koseki, Nami Ohashi, Takaaki Mizuguchi and Hirokazu Tamamura*

The assembly status of G protein-coupled receptors (GPCR) on the cellular surface is of interest because the multimerization of GPCR could play pivotal roles in cellular functions. A bivalent ligand with polyproline linkers for CXCR4 has been shown to serve as "molecular ruler" as a result of the rigid structure of polyproline helices. To expand the utility of the ligands with rigid linkers and explore the possible multimeric forms of GPCR, trivalent ligands with polyproline helices were newly designed and synthesized. The binding affinities of the trivalent ligands for CXCR4 suggested that the ligands recognize the dimeric form of CXCR4 on the cellular surface. The fluorescent imaging and analysis by flow cytometry revealed that the ligand with 9-proline linkers binds to CXCR4 with remarkable specificity. The results of the present study suggest that the ligand design with rigid linkers is useful in the multimeric form, but, the design for the trivalent ligands requires different strategic approaches.

Results and discussion

The rigidity of polyproline helix has been proved to be useful for linker structure of multimerized ligands because the ligand moieties are maintained at constant distance at the terminus of the helix structure. In the design of trivalent ligands, the polyprolines were linked to the tripodal scaffold via flexible azidohexanoic acid moieties. The flexibility at the part of the polyprolines was conjugated to tripropargylamine utilizing properties. This technique with a rigid linker could also be extended to other GPCRs or GPCRs forming heterodimers. The solution of the X-ray crystal structure of GPCR has seen remarkable progress, and the number structures of GPCRs is increasing but the structure in the crystal might not reflect the dynamic state of the receptors. When compared to structural analysis by x-ray crystallography, analysis with the help of ligands might have advantages in an understanding of the dynamic states of receptors on the cellular surface. In addition, it is possible that GPCRs could exist as a multimeric complex on the cell surface. Currently, the relationships between the functions of GPCRs and their multimerization states have not been well analyzed. To analyze the possible form of multimerization, the development of linkers or templates for ligands is required. In this report, we describe the design and synthesis of a novel three-branched linker for a CXCR4-specific ligand, FC131, which is a downsized cyclic pentapeptide derived from polyphemusin II. The possible multimerization states of GPCR were assessed with multimer-ligands with various linker lengths.

Introduction

Multimerization of receptors plays pivotal roles in cellular functions such as signaling. Currently, up to 60% of drugs target molecules located at the cell surface, and half of the 60% are GPCRs. With GPCRs, dimerization of the receptors has been suggested as key to the internalization and activation of signaling pathways. Thus, it is important to analyze the multimerization states of the receptors. With the recent advances of the techniques for the crystallization of the GPCRs, the number of reported X-ray structures of GPCRs has increased. The bivalent ligands utilizing the polyproline helix have been viewed as an alternative approach with which to elucidate the state of GPCRs on the cell membrane. Polyprolines have a relatively rigid structure, and this makes it possible to determine the distance between the ends of a polyproline helix. In this way, evaluation of the distances between the ligand binding sites on a GPCR on the cell surface was conducted. Bivalent ligands with rigid linkers of an appropriate length showed a remarkable increase of competitive inhibition activity for the receptor, CXCR4. It has been suggested that homo- and hetero-oligomerization of CXCR4 is important in cancer metastasis and the significance of oligomeric forms of GPCR has been gaining acceptance. Thus, the evaluation of the state of dimerized receptors could lead to the functional disclosure of complex GPCR

5 Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University Chiyoda-ku, Tokyo 101-0062, Japan E-mail: tamamura.mr@tdm.ac.jp
6 Electronic Supplementary Information (ESI) available: [details of ligand synthesis, HPLC charts, ESI-MS, and results of binding analyses are described in Supporting Information]. See DOI: 10.1039/x0xx00000x

This journal is © The Royal Society of Chemistry 20xx
J. Name., 2013, 00, 1-3 | 1
copper(I)-catalyzed Huisgen-Meldal-Sharpless cycloaddition click chemistry.\textsuperscript{12}

![Scheme 1](image1)

Scheme 1. Synthesis of trivalent cFC131 ligands with polyproline linkers. Reagents and conditions: a) Fmoc-based solid-phase peptide synthesis (SPPS), rt; b) 6-azidohexanoic acid, DIPC, HOBT•H₂O, DMF, rt; c) 95% TFA aq., rt; d) chloroacetyl chloride, DIPEA, DMF, rt; e) tripropargylamine, CuSO₄, sodium ascorbate, 0.2 M phosphate buffer (pH 7.8), rt; f) cFC131, HCl, 0.1 M phosphate buffer (pH 7.8), 37 °C.

The lengths of the polyprolines were allowed to vary from 3 to 18 proline units. Azidoheptanoic acid was condensed to the N-terminus of the polyproline, to produce a reaction moiety with the amino group of lysine was modified with a chloroacetyl group for the conjugation with cFC131 (Scheme 1, Fig. 1). cFC131 is a derivative of FC131 produced by substitution of Gly by d-cysteine. The synthesis of cFC131 is described in Supporting Information.

![Fig. 1](image2)

Fig. 1 The structure of the designed trivalent ligand with rigid polyproline helices. Blue spheres indicate cFC131, which has ε-Cys in place of Gly in FC131 for conjugation with the chloroacetyl moiety at the termini of the linkers.

Fluorophore-labeled TZ14011\textsuperscript{13} was utilized in the evaluation of the IC₅₀ for CXCR4. In the traditional method, isotopically-labeled SDF-1α was utilized and in our previous report, fluorophore-labeled TZ14011 showed parallel values in binding affinities. Thus, this fluorescence method was applied in this study and the measured IC₅₀ values are shown on Table 1. Among the trivalent ligands with variable linkers, 3a and 3c showed the comparable highest IC₅₀ values. In addition to the trivalent ligands with variable polyproline linkers, control peptides, the monomer (7) and the dimer (8) were synthesized (Fig. 2, S7, and S8). The free amino termini of linkers lacking FC131 conjugation were capped by acetyl groups to prevent further reaction. The linker length for the monomer and dimer controls was fixed to 9 prolines in order to address the binding property of trivalent ligand with 9-proline linkers, 3c.

![Fig. 2](image3)

Fig. 2 Summary for Structures of synthesized ligands with or without TAMRA labeling, 7, 8, 11a, 11b, 15, 18. The blue and red spheres represent cFC131 and TAMRA-GABA, respectively.

The monomer cFC131 with 9-proline linkers (7) showed a remarkable decrease in the competitive inhibition activity (IC₅₀ = 2822 nM) but the dimeric cFC131 with 9-proline linkers (8) showed only a three-fold decrease in the competitive inhibition activity (IC₅₀ = 167 nM) compared with the corresponding trivalent ligand (3c). In the binding of FC131 alone, the competitive inhibition activity was shown to be much higher than the current monomer ligand (7). The polyproline template for trivalency would be bulky and the associated steric effect could prevent interaction of cFC131 with CXCR4. In comparison with the monomer ligand, the dimer and the trimer ligands showed 17- and 47-fold increases in the competitive inhibition activity, respectively. Such increase in the competitive inhibition activity suggests synergistic effect in binding of the ligand units. However, the IC₅₀ of the trimer is approximately 3-fold higher than that of the dimer. This could be rationalized in that there are three patterns for dimer recognition in the trivalent ligand while the divalent ligand has only single pattern. If the third unit in the trimer binds to a multimeric form of CXCR4, the competitive inhibition activity of the trimer should show a further increase. The ligands with longer linkers, in excess of 9 prolines, showed decreased competitive inhibition activity as the linkers are longer. IC₅₀ values in the range of 56-161 nM were observed for the ligands with 3-15 proline units. These values are all in the same range and do not differ significantly. Only the value for 3f and 7 do not differ significantly. Only the value for 3f and 7
differ from those of ligands 3a-e. It would be possible the flexibility at the azidohexanoic acid moieties allow the branched linker structures to adjust to the binding to multimerized CXCR4. Thus, the differences in the length of polyproline structure did not affect greatly to the IC_{50} values. The results could suggest that the trivalent ligands bind to CXCR4 in dimer form, and as shown in the case of the monomer, the bulk of the template can cause increased hindrance in binding when the linker is longer. The structure of the trimer ligand suggests that there might be more potential in the design of the trivalent template to involve three of the cFC131 units for CXCR4 recognition.

Table 1. The competitive inhibition activity of trivalent ligands for CXCR4 against TAMRA-T214011

<table>
<thead>
<tr>
<th>Compounds</th>
<th>IC_{50} (nM)</th>
<th>Length of linker units (nm)[[i]] / proline numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>56.6</td>
<td>2.5 / 3 prolines</td>
</tr>
<tr>
<td>3b</td>
<td>98.0</td>
<td>3.4 / 6 prolines</td>
</tr>
<tr>
<td>3c</td>
<td>59.8</td>
<td>4.3 / 9 prolines</td>
</tr>
<tr>
<td>3d</td>
<td>152</td>
<td>5.2 / 12 prolines</td>
</tr>
<tr>
<td>3e</td>
<td>161</td>
<td>6.1 / 15 prolines</td>
</tr>
<tr>
<td>3f</td>
<td>695</td>
<td>7.0 / 18 prolines</td>
</tr>
<tr>
<td>7</td>
<td>2922</td>
<td>4.3 / 9 prolines (monomer)</td>
</tr>
<tr>
<td>8</td>
<td>167</td>
<td>4.3 / 9 prolines (dimer)</td>
</tr>
<tr>
<td>18</td>
<td>151</td>
<td>4.2 / NA[5]</td>
</tr>
<tr>
<td>FC131</td>
<td>32.7</td>
<td>NA / NA</td>
</tr>
</tbody>
</table>

[a] Ligands with or without polyproline helices. [b] Standard errors are shown in Supporting information in logIC_{50} values. [c] The length of linker structure except oligoproline was estimated by modelling with Spartan. The length of the oligoproline was calculated based on the previous report. [d] NA, not applicable. [e] From the previous report, the IC_{50} value of acetamide-cFC131 is 1.7-fold weaker than FC131.

To investigate whether the rigidity of the polyproline structure is important to maintain the effective ligand binding in the dimer form, a trivalent template with a penta-gamma-aminobutyric acid (GABA) linker was synthesized (17) and a trivalent ligand with flexible linkers (18) was obtained by conjugation with cFC131 (Fig. 2). The length of the penta-GABA linker was estimated to be similar to that with the 9-proline rigid linker. The IC_{50} of 18 was 2.5-fold weaker (151 nM) than that of 3c with 9-proline linkers (59.8 nM) (Table 1). The difference between the IC_{50} values of 3c (59.8 nM) and the “flexible trimer (18)” (151 nM) is not large, which corroborates that the oligoproline based tripodal ligands are rather flexible.

Among the synthesized trivalent CXCR4 ligands, 3a and 3c were found to bind effectively to CXCR4. Specificity to the target receptors by the multivalent-type ligands should enable fluorescent imaging of the receptor on the cellular surface upon addition of a fluorescent dye as in the previous report. For CXCR4, a target receptor in this study, it has been reported that the degree of receptor expression is correlated with cellular malignancy. A further question arises from the previous result: how does 3a recognize CXCR4 in spite of the fact that the linker is too short to bridge the ligand binding sites between the dimeric CXCR4. A fluorescent dye, 5(6)-tetramethylrhodamine (TAMRA) was conjugated to 3a, 3c, and monomeric cFC131 with trivalent linkers (7) via GABA linker and fluorescent imaging analysis by confocal laser-scanning microscope (FV10i, Olympus) using living cells revealed that trivalent 3c labeled with TAMRA (11b) was clearly merged into EGFP fused CXCR4 at the cellular surface at 100 nM (Fig. 3G). The image of 11a showed non-specific internalization of the ligands to cytosol at 100 nM (Fig. 3B). In contrast, the monomer ligand with TAMRA (15, Fig. 2) showed very low fluorescent intensity on the cellular surface even at 750 nM (Fig. 3J). In addition, the specificity was not as high as that of the trivalent ligands.

In addition to fluorescent imaging analysis by confocal laser-scanning microscope, FACS analysis was used for detection of receptor expressions by trivalent CXCR4 ligands. If the ligand can detect the expression amount of CXCR4 on the cellular surface, flow cytometry should be a facile system for its detection. In the previous study, bivalent ligands with an appropriate length of polyproline helix could recognize the expression amount of CXCR4 on the cellular surface. Following the previous results, Jurkat, HeLa, and K562 cells were used in the analysis. The previous report noted that, these cells show different expression levels of CXCR4 (Jurkat > HeLa > K562). As the ligand concentration was increased, the amount of binding by the trimer ligand (11b) was increased in each cell line. In addition, the amount of bound ligands was increased proportionally in accordance with expression amount of CXCR4 in each cell line. In contrast, binding by the monomer ligand (15) did not show any increase or difference between cell lines (Fig. 4). The trivalent ligand with 3-proline linkers (11a) also showed an increase in the binding amount as the concentration was increased (Fig. S20). Considering the results of the fluorescent imaging study, the values for 11a
could be the sum of the fluorescence from the ligands that bind to CXCR4 and are internalized. Although the detailed mechanism of the increased amount of internalization without specificity to CXCR4-expressing cells is unclear, it suggests that the binding properties of trivalent ligands with 3 and 9 prolines are different. In view of the length of the proline linkers, the 9-proline linker is sufficiently long for dimer recognition. Although apparent IC_{50} of trivalent ligand with 3 proline linkers is comparable with that of the ligand with 9-proline linkers, the amount of ligand showing the specific binding could be lower than the ligand with 9-proline linkers.

![Fig. 4](image)

**Fig. 4** The results of flow cytometry analysis for binding of TAMRA-labeled trivalent (A, 11b) or monomer (B, 15) ligands. The bars represent the fold increase of fluorescence at the indicated ligand concentration (nM). Each bar shows the results of Jurkat (white), HeLa (gray), and K562 (black). The error bars indicate S.E.M of values from three independent experiments.

**Conclusions**

In summary, a newly synthesized trivalent CXCR4 ligand with polyproline linkers shows dimer recognition specific to CXCR4. The ligand was designed for detection of CXCR4, in particular, three receptors in the proximal distance. The recognition for the dimeric CXCR4 was successfully observed, but that for the third receptor was not observed in spite of the flexibility by triazidohexanionic acid moiety. This result may indicate that CXCR4 does not exist within the distance which the polyproline linker could reach. In addition, the trivalent ligand with flexible linker showed similar IC_{50} with the ligands with polyproline linkers. When the third CXCR4 receptor exists within the range of the linker length, the increase of inhibition rate would be observed by these ligands. For the recognition of the multiple receptors more than three, the different design of the template may be required. It is of special interest that ligands with short linkers show increased affinity and non-specific internalization at relatively lower concentration (100 nM). The mechanism of the internalization is not clear, but the cationic moiety of cFC131 could be concentrated within the molecule because of the short polyproline linker, and the accumulated cationic moiety could interact with cellular membrane and/or membrane-associated factors, such as heparan sulfate. There are many reports of the design of bivalent ligands, but there are fewer reports about the study of the design of multimeric ligands recognizing membrane receptors. The use of oligoproline for targeting GPCRs to maintain the recognition moieties in defined distance is one of the strategic approaches. The crystal structure of an oligoproline PPII helix would provide useful information for detailed design of linkers. Our results suggest requirements for the strategic design of the multimeric ligands such as the combination of rigid and flexible linkers, to expand the possibility of receptor recognition by a single molecule.

**Experimental**

**General methods**

For analytical HPLC, a Cosmosil 5C_{18}-ARII column (4.6 × 250 mm, Nacalai Tesque, Inc., Kyoto, Japan) was employed with a linear gradient of CH_{3}CN containing 0.1% (v/v) TFA at a flow rate of 1 cm/min$^{-1}$ on a LaChrom Elite HTA system (Hitachi High-Technologies Corporation, Ltd., Tokyo, Japan). Eluting products were detected by UV-absorption at 220 nm. Preparative HPLC was performed using a Cosmosil 5C_{18}-ARII column (20 × 250 mm, Nacalai Tesque, Inc.) on a JASCO PU-2089 plus (JASCO Corporation, Ltd., Tokyo, Japan) in a suitable gradient mode of CH_{3}CN solution containing 0.1% (v/v) TFA at a flow rate of 7 cm/min$^{-1}$. Eluting products were detected by UV-absorption at 220 nm. UV spectra were recorded using a JASCO V650 UV-vis spectrophotometer. ESI-TOF-MS was recorded on a micrOTOF-2 focus (Bruker Daltonics) mass spectrometer.

**General procedure for Fmoc-solid phase peptide synthesis**

Peptide synthesis was performed through the Fmoc solid-phase peptide synthesis. Fmoc-amino acids were coupled by 1,3-diisopropylcarbodiimide (DIPCI, 5.0 eq.) and N-hydroxybenzotriazole hydrate (HOBt·H_{2}O, 5.0 eq.). For the coupling of-(nitrobenzo-2-oxa-1,3-diazole) L-α,2,3-diaminopropionic acid, that is, Dap(NBD) to a peptide, 2-(7-Azabenzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HATU, 2.9 eq.),1-hydroxy-7-azabenzotriazole (HOAt, 3.0 eq.) and N,N-diisopropylethylamine (DIPEA, 10 eq.) were employed in place of DIPCI/HOBt. The Fmoc group was deprotected by treatment with 20% (v/v) piperidine/DMF for 20 min.

**Competitive binding analysis of CXCR4 ligand candidates**

Jurkat cells cultured in RPMI1640/10% FBS were utilized in the experiment. TAMRA-Ac-TZ14011 in PBS was diluted to 100 nM with the binding buffer (RPMI1640/20 mM HEPES buffer). The test compounds were added at x4 concentration to the TAMRA-Ac-TZ14011 solution in the same volume. Jurkat cells were added to the binding solution and incubated for 30 min at 4 ºC. Cells were collected and washed twice with PBS. Cells were transferred to 96-well plates and fluorescent intensity was evaluated by GENios Pro (TECAN, Ex/Em = 535/590 nm).

**Calculation of IC_{50} values in competitive binding of CXCR4 ligands**

Inhibition (%) = (Et – Ea)/(Et – Ec)×100

Et = Fluorescent intensity in the absence of test compounds (TAMRA-Ac-TZ14011 only)

Ec = Fluorescent intensity in the presence of excess non-labeled Ax-TZ14011
Organic & Biomolecular Chemistry Accepted Manuscript

CXCR4 ligands were added to each well, and the mixture was incubated for 15 min at room temperature. After washing and resuspension with PBS, cells were analyzed on FACSCalibur HG (BD Bioscience).

FACS analysis
For each measurement, 1×10^6 cells were washed once with PBS, then cells were suspended in PBS (900 µL). Fluorescently-labeled CXCR4 ligands were added to each well, and the mixture was incubated for 15 min at room temperature. After washing and resuspension with PBS, cells were analyzed on FACSCalibur HG (BD Bioscience).

Acknowledgements
This work was supported in part by KAKENHI, Grant-in-Aid for Scientific Research (B) (24390024 to H. T.); JSPS Core-to-Core Program, A. Advanced Research Networks; and the Platform for Drug Discovery, Informatics, and Structural Life Science of MEXT, Japan. The authors thank to M. Tanabe for assistance in binding analysis of the ligands.

References
A graphical and textual abstract for the Table of contents entry

The trivalent ligand with rigid linkers designed for exploration of GPCR multimerization shows specific recognition for overexpressed CXCR4.