

Submitted date 2022-09-06

Application R400-2022-509

Name Walden Bjørn-Yoshimoto
Position Assistant Professor

Institution University of Copenhagen
Institute Health and Medical Sciences
Department Department of Biomedical Sciences

Budget subtotals & total

Budget line	Year 1	Year 2	Total
Salary - postdoc			
	602.869,00 kr.	615.650,00 kr.	<u>1.218.519,00 kr.</u>
Consumables			
	230.000,00 kr.	100.000,00 kr.	<u>330.000,00 kr.</u>
Other			
	250.000,00 kr.	0,00 kr.	<u>250.000,00 kr.</u>
Project-related indirect cost			
	120.319,00 kr.	79.517,00 kr.	<u>199.836,00 kr.</u>
Total	1.203.188,00 kr.	795.167,00 kr.	1.998.355,00 kr.

Project title

Elucidating the cone snail toxinome

Summary of application

This project proposes will use new approaches to map interactions between novel peptide sequences and G protein-coupled receptors (GPCRs). The peptide sequences are derived from cone snail venom, a rich source of bioactive peptides that are indispensable in neuroscience research, as well as clinical applications. Here, I will map these interactions at scale in the entire chemical diversity of cone snail toxins, the 'toxinome'.

I have initiated two new collaborations for this (see attached letters). For work package (WP) 2, I will collaborate with Assoc Prof Alexander Hauser and Dr. Kasper Harpsøe at the Department of Drug Design and Pharmacology, University of Copenhagen. They are experts in computational biology, specifically of GPCRs. We will use the core computational facility run by Kasper Harpsøe for peptide docking and simulations. For WP3, I will collaborate with

Assoc Prof Markus Muttenthaler at the Institute for Molecular Bioscience, University of Queensland, Australia. His lab has extensive expertise in the synthesis of the type of disulfide-rich peptides used in this project. I will be heavily involved in WP2, while WP3 will be carried out primarily by Dr. Muttenthaler.

With this project, we will provide unique tools for the scientific community, as well as illuminate the evolution of neuronal signaling systems. As has been the case for many previous conotoxins, the peptide identified here are likely to fuel neuroscience research for decades to come.

Project start date	2023-06-01
Project end date	2025-05-31

Personal information

First name	Walden
Last name	Bjørn-Yoshimoto
Email	waldenb@gmail.com
Date of birth	1983-06-27
Address	Rødvig Alle 27
Postal code	2650
City	Hvidovre
Country	Denmark
Phone number	+45 24494408
Academic degree	PhD
Position	Assistant Professor
Institution	University of Copenhagen

The project

Project title

Elucidating the cone snail toxinome

Project

Project title

Elucidating the cone snail toxinome

In a nutshell

From the complex mixture of cone snail venom, we will identify and map novel peptide toxins, called ‘conotoxins’, that modulate G protein-coupled receptors (GPCRs). Fewer than 1 % of the more than 100,000 existing conotoxin sequences have been assigned a molecular target, largely due to technical limitations. Despite this, numerous conotoxins have already revolutionized neuroscientific research and have become useful drugs and valuable diagnostic tools. We will exploit new features of AlphaFold2 (AF2), the groundbreaking protein structure prediction tool, to increase this number substantially.

Key question: Which of the many uncharacterized conotoxins target which human GPCRs?

Key result: Dozens of new selective and potent modulators of human GPCRs.

Key impact: Toxin tools that will revolutionize the study of GPCR biology, and potential drug leads for associated diseases.

Research idea and context

The ~1000 living species of cone snails are venomous marine predators that utilize a complex mix of conotoxins to incapacitate their prey. Many conotoxins modulate human targets and have proven indispensable in both basic and applied neuroscience^{1,2}. More than 95 % of conotoxins characterized to date modulate ion channels, such as the paralytic α -conotoxins or the calcium channel blocking ω -conotoxins, e.g., GVIA, which has been used in thousands of research articles to date^{2,3}. While ion channel modulators have traditionally been much easier to discover, a small group of conotoxins with other targets

have been identified, such as the venom insulins or the neurotensin receptor modulator, contulakin G^{4,5}. However, these were discovered largely by happenstance^{2,4-7}. For example, the GPCR modulating conotoxin TIA, which targets the $\alpha 1$ -adrenoceptor, was coincidentally discovered during a search for α -conotoxins⁸. Likewise, conotoxins targeting different subtypes of the human somatostatin receptors were discovered by *in vivo* screening of venom fractions in mice, followed by a large *in vitro* screen⁷. This study relied on extracting enormous amounts of venom, which is unfeasible for most species. These examples show that conotoxins targeting GPCRs do indeed exist, and we hypothesize that they are more common than what is currently known. However, they also show that radically new strategies are needed to unravel the full diversity of GPCR-targeting conotoxins at scale. With GPCRs being such an important group of receptors in human physiology and disease, potent and selective modulators would be a boon to both the scientific community and direct drug development efforts, as has already been the case for several ion-channel-targeting conotoxins^{2,9}.

Proposed method of solution or concept

We aim to employ novel computational methods that will allow us to interrogate the cone snail peptide sequence databases at scale, something never before done. The project will be divided into four work packages (WP1–4, see Fig. 1).

WP1: Preparing mature peptide sequences from a toxin precursor database of 26,392 assembled sequences from large public repositories, and grouping sequences based on gene superfamily annotation, cysteine framework, and shared evolution for creating custom conotoxin multiple sequence alignment (MSA) for AlphaFold2 (AF2)¹⁰.

WP2: Predicting interactions between conotoxins and GPCR targets. The emergence of AF2 in 2020 began a transformation in structural biology, continuing in 2022 when 200 million protein structure predictions were released^{11,12}. Recently, Alphafold2 has shown significant promise in *de novo* prediction of peptide-protein docking^{13,14}. Our preliminary results suggest that this can revolutionize conotoxin-receptor complex prediction (Fig. 2). We will optimize this method with conotoxin-specific parameters such as custom conotoxin MSAs, and develop benchmarking and scoring techniques, including deep learning methods previously shown to be successful for AF2 multimeric predictions in other fields¹⁵.

WP3: Synthesizing the conotoxins identified in WP2 by standard solid phase peptide synthesis techniques. Approximately 150 – 300 peptides will be synthesized in total.

WP4: *In vitro* characterization of the peptides. Each peptide will be tested at the *in silico* predicted “hit” receptor(s), as well as closely related receptors utilizing a range of pharmacological GPCR assays that I have already established.

Major gain and obstacles

Outcome:

New potent and selective modulators of several human GPCRs, including those for which selective ligands are currently missing. These will serve as valuable research tools for neuroscience, as drug leads, or as diagnostic tools. The elucidation of new conotoxin targets would also advance the understanding of cone snail ecology and evolution, and the complex chemical and pharmacological interactions between predators and their prey.

Obstacles:

(1) There could be no additional GPCR modulators in the conotoxin sequences we have obtained. With the vast majority of conotoxins not yet characterized (>99 %), and existing examples of GPCR-modulating conotoxins found even without concentrated efforts, this appears highly unlikely.

(2) The methodology could be insufficient to identify the conotoxins of interest. Our initial results suggest otherwise. We might have false negatives, but this will largely be circumvented by some of the optimizations in WP1-2.

(3) We do not know the sequence of the mature peptide, only the precursors, and conotoxin processing is not easy to predict. However, we have observed that the shortest version of mature peptides that we predict typically has some effect at the target, and would hence be identified, something also shown for other peptide-GPCR interactions¹⁶.

Success criteria:

Novel and selective modulators of GPCRs, either with a known impact on human health or orphan GPCRs, providing the much-needed tools to study them.

Next step:

Potent and selective modulators should be used to elucidate novel biology and disease etiology by allowing differentiation of receptors and subtypes with an unprecedented accuracy, exactly as many other conotoxins have already done. While initial work on this can be done in-house, we will focus on continuing existing, and establishing new collaborations with experts in the respective target receptors.

Appropriateness

This application takes advantage of very recent advances in structural predictions and

modeling. While these methodologies have begun to revolutionize other fields of research there is no precedent that this will be the case for the project proposed here. However, cone snail venom is, due to the “evolutionary pre-screen”, a concentrated source of extremely high-value compounds, as proven repeatedly^{2,10}. Thus, this high-risk, high-gain project makes this proposal ideally suited for the strategy of the Lundbeck Experiment grant, and unlikely to attract funding from other, traditional funding schemes.

Probable objections

1. “There could be no new conotoxins modulating GPCR targets in the database. This is essentially a ‘fishing expedition’ which could yield nothing.”

Correct, this is the main “high risk” part and the reason to apply specifically for a Lundbeck Experiment grant. However, cone snail venom is already evolutionarily pre-screened, and in the <1 % of conotoxins characterized to date, examples of GPCR modulating conotoxins are already known. The fact that these have been found with the rudimentary techniques employed is nothing short of astonishing and implies that they are indeed quite plentiful. As such, what is needed is new tools, not a new source of compounds. Yes, this is a “fishing expedition”, but a fishing expedition in a lake that is overflowing with fish, where no one else seems to be using the right equipment yet.

2. “The methodology is not novel. *In silico* docking of peptides has been done.”

While some attempts have been made to use either traditional docking, or AF2 multimer to predict peptide-protein interactions, both are still in early stages, especially for disulfide-rich peptides, such as conotoxins. Despite very promising pilot studies the methods can, with our specific expertise, be improved to allow an unprecedented accuracy in determining conotoxin-receptor interactions. Additionally, these methods have never been applied at the scale this project proposes, or to such a rich source of proven bioactive peptides as conotoxins.

3. “Someone else will do this, regardless of whether the Lundbeck Foundation supports this.”

Eventually, yes, I should hope so. But we are in a unique position due to our experience and expertise. For conotoxins, it is not trivial to predict the mature toxin from the precursor databases that are available. However, the experience within our group gives us a unique advantage over efforts that focus on the computational aspects without experience in the specific biology.

References

1. Jin, A.-H., et al., *Chemical Reviews* **2019**, *119* (21), 11510-11549.
2. Bjørn-Yoshimoto, W. E., et al., *Biomedicines* **2020**, *8* (8), 235.
3. McIntosh, J. M., et al., *J. Biol. Chem.* **1994**, *269* (24), 16733-16739.
4. Craig, A. G., et al., *J. Biol. Chem.* **1999**, *274* (20), 13752-13759.
5. Safavi-Hemami, H., et al., *Proc. Natl. Acad. Sci.* **2015**, *112* (6), 1743-1748.
6. Kaas, Q., et al., *Nucleic Acids Res.* **2012**, *40* (D1), D325-D330.
7. Ramiro, I. B. L., et al., *Science Advances* **2022**, *8* (12), eabk1410.
8. Sharpe, I. A., et al., *Nat. Neurosci.* **2001**, *4* (9), 902-907.
9. Safavi-Hemami, H., et al., *J Proteomics* **2019**, *190*, 12-20.
10. Halai, R., et al., *Nat. Prod. Rep.* **2009**, *26* (4), 526-536.
11. Jumper, J., et al., *Nature* **2021**, *596* (7873), 583-589.
12. Callaway, E. 'The entire protein universe': AI predicts shape of nearly every known protein. <https://doi.org/10.1038/d41586-022-02083-2> (accessed 2022-08-16).
13. Mirdita, M., et al., *Nat. Methods* **2022**, *19* (6), 679-682.
14. Tsaban, T., et al., *Nat. Commun.* **2022**, *13* (1), 176.
15. Gao, M., et al., *Nat. Commun.* **2022**, *13* (1), 1744.
16. Foster, S. R., et al., *Cell* **2019**, *179* (4), 895-908.e21.

Potential benefit for neuroscience of the proposed project

The expected benefit is novel tool compounds for basic and translational neuroscience. These venom mixtures include numerous peptides that have high selectivity and potency at a range of targets expressed in both the CNS and PNS of humans, and many are already essential for current neuroscience research. This includes ω -conotoxins that are subtype selective blockers of voltage-gated calcium channels (e.g., ω -GVIA which has been used in thousands of publications), α -conotoxins that are subtype-selective inhibitors of nicotinic acetylcholine receptors, and Con-ikot-ikot that has been used to crystalize the AMPA receptor. Many of the receptors in this proposal are expressed in both the CNS and PNS and are current or potential drug targets. Thus, hits at these would benefit basic neuroscience and drug development efforts alike. We will likely discover surprises, conotoxins targeting utterly unexpected receptors that, nonetheless, will be valuable tools. No one expected that neurotensin, insulin, or somatostatin would be chemical weapons used by cone snails until toxins mimicking these GPCR ligands were found in venom. Further, the insight into snail and prey neurophysiology gained by elucidating novel conotoxin targets will be valuable for basic neuroscience efforts, as exemplified by the somatostatin-like conotoxins, that recently enabled tracing this neuropeptide signaling system (previously thought to exist exclusively in vertebrates) across numerous phyla.

Anticipated project impact

- Conotoxins that can be used in basic neuroscience, biomedical research, as well as for drug leads.
- Elucidation of novel cone snail and prey signaling systems and predator-prey evolutionary relationships.
- Novel in silico screening approach, that will be useful for any parts of the scientific community investigating peptides that interact with proteins.

Budget information

This project, due to the interdisciplinary nature where it will need to be able to finance both computational efforts and extensive synthetic efforts, stretches the budget to the maximum applicable. I will thus apply for alternate grants to attend conferences and cover travel costs.

I have applied for 1 FTE salary for the duration of the project for myself, consumables for peptide synthesis and pharmacological assays, as well as financial support for the computational efforts, which will be approximately equivalent to an additional FTE for 5 months.

Budget

Year	Type	Name	Amount of money	Details
1				
Applied from the Lundbeck Foundation				
		Salary - postdoc	602.869,00 kr.	I am currently a postdoc, but will be assistant professor (non tenured) from September 15th, 2022. Salary is same as postdoc.
		Consumables	230.000,00 kr.	Most of this (~175000) will be reagents for peptide synthesis. Amount estimated by collaborator 2, who will provide the labor.
		Other	250.000,00 kr.	Support for work package 2. While I will be heavily involved in this aspect of the project, additional support will be needed for the facilities of collaborator 2. This is to finance some of the expenses in work package 2.
		Project-related indirect cost	120.319,00 kr.	Indirect cost to direct project related costs to HR, Economy etc.
			<u>1.203.188,00 kr.</u>	
2				
Applied from the Lundbeck Foundation				
		Salary - postdoc	615.650,00 kr.	I am currently a postdoc, but will be assistant professor (non tenured) from September 15th, 2022. Salary is same as postdoc.
		Consumables	100.000,00 kr.	Some will be used for follow-up peptide synthesis, but most of this is for cell culture and assay consumables
		Project-related indirect cost	79.517,00 kr.	Indirect cost to direct project related costs to HR, Economy etc.
			<u>795.167,00 kr.</u>	
			<u>1.998.355,00 kr.</u>	

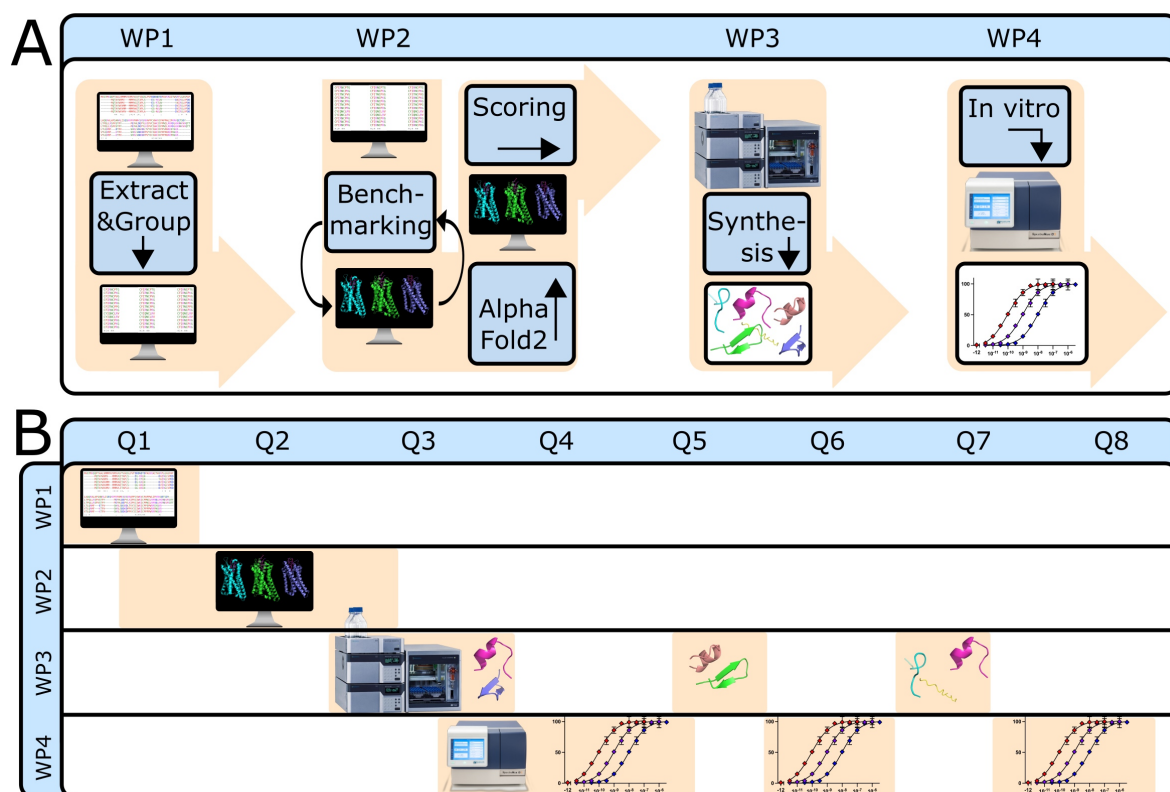


Figure 1 - Workflow and timeline:

Overview of workflow in work packages (WP) (A), and expected timeline of activities (B). WP1 will prepare the peptide database from the precursor database, process the sequences to mature peptide sequences, and create a custom multiple sequence alignment by grouping evolutionarily related conotoxins. This is expected to take ~3 months and will feed into WP2. First, we will benchmark different optimizations to AlphaFold2 Multimer and determine the best scoring method. Preliminary results will inform potential changes to WP1. Subsequently, the entire peptide database will be “docked” in the entire GPCR database, and results analyzed. This is expected to take up to 6 months, though with significant overlap with WP1 and WP3. WP3 is the peptide synthesis by standard solid-phase techniques, followed by folding (including forming the correct disulfide connectivity), purification, and quality control. This is expected to take 4-5 months initially, though results from WP4 will inform new rounds of synthesis later in the project, which will include alternate cleavage patterns and predicted amino acid modifications (e.g., D-amino acids, hydroxyprolines, etc., that are prevalent in conotoxins), as well as optimizations predicted from structural data or structural predictions from WP2. WP4 is the pharmacological characterization. The initial phase is expected to take 6 months, though subsequent rounds of synthesis from WP3 will be tested in later WP4 rounds. Late Q4 or early Q5, we will begin to know which targets are of particular interest, and this is where we will initiate collaborations with appropriate experts in our networks. Timelines include time for disseminating (articles, conferences, etc.).

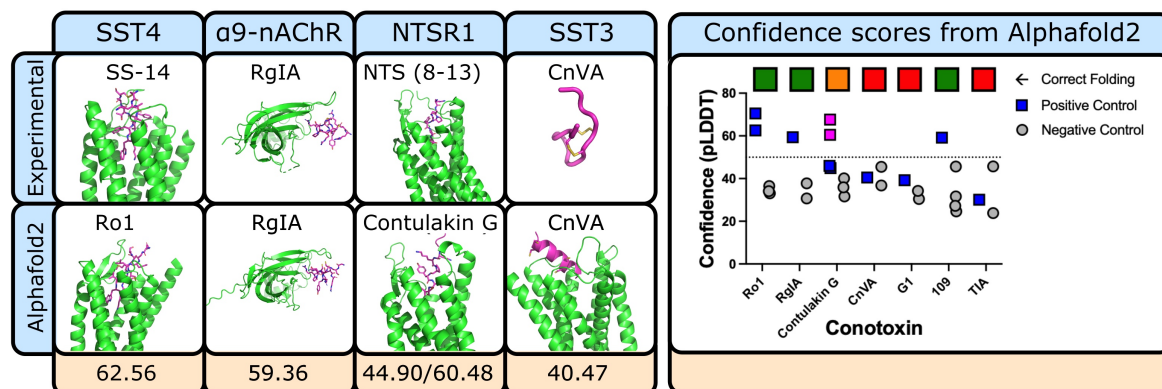


Figure 2 – Preliminary AlphaFold2 docking shows the feasibility of the approach:

Left: Structures of conotoxins or related neuropeptides (in pink) bound to the receptor they modulate (in green), or without a receptor (top right frame, no such structure exists). The top panels show experimentally determined structures, while the bottom panels show AlphaFold2 predictions. The top blue labels denote the receptor, while the labels within the frame denotes the conotoxin or neuropeptide shown. The numbers in the bottom orange is the average pLDDT scores of the peptide, a measure of the certainty of the prediction, and our rudimentary score in the pilot screen. This will be developed during early stages of WP2 but serves to demonstrate that even a very simple scoring yields separation of positive and negative controls for every conotoxin that folds to form correct disulfide bonds. For Contulakin G, the majority of the peptide is flexible, thus the average pLDDT of the entire peptide was low (44.90). However, the average of the last six amino acids (corresponding to the experimentally verified structure of the last six amino acids of neurotensin) is quite high (60.48). The AlphaFold2 structure of CnVA did not correctly predict the cysteine bonds (compare to solution NMR structure in the bottom panel), thus yielding a low score. This will likely be improved by using a custom MSA for related conotoxins. Right: Plots of the average pLDDT scores from AlphaFold predictions of conotoxins with their known interactors (blue squares) or with receptors known to not interact with them (grey circles). The colored squares above the plot indicates whether the cysteines were in positions where they could form disulfides. Green indicates yes, red indicates no, and orange indicates that no cysteines were present in the conotoxin. Note that for all the examples where the cysteines in the predicted conotoxin structure were close enough to form disulfide bonds, there is a good separation of positive and negative controls. Using a simple filter of pLDDT>50 would separate all correct interactors from the negative controls (see horizontal dashed line). For the rest, we expect that some of the planned optimizations can alleviate this. For Contulakin G, the pink squares represent the pLDDT values when only the 6 C-terminal residues are considered. SST4 = somatostatin receptor 4; α9-nAChR = α9-nicotinic acetylcholine receptor; NTSR1 = neurotensin receptor 1; SST3 = somatostatin receptor 3; SS-14 = somatostatin-14; NTS (8-13) = 6 C-terminal amino acids from neurotensin.

Appendix

Letters of support from collaborators:

letters_of_collaboration.pdf

CV and publication list:

cv_publications.pdf

Letter of Support form Host Institution:

Letter of Support.pdf