EMBnet.journal
Bioinformatics in Action

Immunoinformatics taught to middle and high school pupils in IMGT®

An in silico and in vitro pipeline for the rapid screening of helicase modulators

Dark Suite: a comprehensive toolbox for computer-aided drug design

AND MORE...

25
2019

EMBnet.journal
Bioinformatics in Action

Contents

Editorial

EMBnet journal is turning the page as it fast moves forward and fights to be fully indexed in Medline and all major indexing resources. In this direction the current volume of EMBnet journal sponsors two articles that have both been indexed in Medline. The first one is about a very fast and cost efficient assay to screen for antiviral agents that inhibit the viral helicase enzyme. In the COVID19 pandemic era quick tests for the screening and eventual discovery of antiviral drugs is of huge value to the scientific community and the society. The second article is the Dark Suite set of bioinformatics tools. This suite boasts a collection of open source bioinformatics tools for genetic and structural analyses of protein in pre-designed pipelines. This makes rather tedious and complicated bioinformatics experiments possible for the novice bioinformatician by eliminating the chances of human induced error. Through the advanced option on the menu the most experienced bioinformaticians can edit and customize all options of the experiment. Organizing everything in a seamless pipeline opens up potential for new users to perform advanced bioinformatics experiments. All in all EMBnet journal is keeping up with all the major advances in the realm of bioinformatics while aligning with the high standards required from all major scientific and indexing resources so that full dissemination and visibility for all articles published in EMBnet journal is ensured with the highest of academic and editorial standards.

Dimitrios P. Vlachakis
Deputy Editor-in-Chief
dimvl@aua.gr

http://dx.doi.org/10.14806/ej.25.0.939

EMBnet.journal
Bioinformatics in Action

D'Elia et al. (2020) EMBnet.journal 25, e931
http://dx.doi.org/10.14806/ej.25.0.931

Standards make the world go round

Domenica D'Elia[1]✉, Chris Evelo[2,3], Babette Regierer[4,5], Susanne Hollmann[4,6]

[1]Institute for Biomedical Technologies, National Research Council (CNR), Bari, Italy
[2]Department of Bioinformatics - BiGCaT, Maastricht University, Maastricht, Netherlands
[3]Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, Netherlands
[4]SB Science Management UG, Berlin, Germany
[5]Leibniz Institute for Vegetable and Ornamental Crops (IGZ), Großbeeren, Germany
[6]Potsdam University, Potsdam, Germany
Competing interests: DD none; CE none; BR none; SH none

On the 2 March 2020, COST Action - *Harmonising standardisation strategies to increase efficiency and competitiveness of European life-science research* (CHARME[1]) - held its final conference in Brussels.

After four years of successful work, the members of the COST Action CHARME met in Brussels to summarise the achievements and to discuss future perspectives and challenges for standardisation in the life sciences.

Standards represent important drivers in the life-sciences and technology transfer because they guarantee that data become accessible, shareable and comparable along the value chain.

The CHARME network, chaired by Dr Susanne Hollmann, fostered collaboration between researchers from 31 countries to increase awareness for the need for standards, enabling the reuse of research data and their interoperability within the community. CHARME provides a common ground for researchers from academia, research institutes, SMEs and multinational organisations.

Following the motto "Standards make the world go round", the outcomes of the COST Action are manifold and introduced some basic concepts and definitions that support a better understanding of the challenges and requirements.

A challenge identified is the digitalisation and interoperability of data and tools in wet- and in silico-labs, because there is an urgent need for common languages and ontologies to enable data reuse and process automation. This becomes an even bigger challenge if researchers work with data derived from different scientific fields. Relevant instruments to support the implementation of standardisation are tools for data and process documentation. Unfortunately, the interoperability between the existing instruments is limited. Hence, the development of new tools is necessary to allow the transfer of data from one system to another and thus allowing the reuse of data from databases and data repositories. Despite the fact that many researchers already make their data compliant to the FAIR principles (Findable, Accessible, Interoperable and Reusable), identifiers or terms used are not harmonised (e.g. database identifiers, ontologies and chemical [sub] structures).

The extensive involvement of the scientific community in CHARME is one of the most important achievements of this COST Action. CHARME participants have attracted, interacted and cooperated intensively with international organisations such as the ISO TC 276[2], CEN/CENELEC[3], with initiatives and scientific organisations like ELIXIR[4], ORPHANET[5], COMBINE[6], FAIRDOM[7], GA4GH[8].

To harmonise activities with the efforts done outside Europe, CHARME was supported by the COST Association[9] to run a joint workshop[10] with the Massive Analysis and QC (MAQC[11]) Society from the US where both communities discussed fundamental themes of research and machine learning reproducibility in the context of standard's needs.

During European and international conferences and workshops[12], CHARME has presented possible solutions in the development and implementation of a uniform

[1]https://www.cost.eu/actions/CA15110/#tabs|Name:overview

[2]https://www.iso.org/committee/4514241.html
[3]https://www.cencenelec.eu/Pages/default.aspx
[4]https://elixir-europe.org/
[5]https://www.orpha.net/consor/cgi-bin/index.php
[6]http://co.mbine.org/
[7]https://fair-dom.org/
[8]https://www.ga4gh.org/
[9]https://www.cost.eu/
[10]https://maqc2019.fbk.eu/
[11]https://www.pmgenomics.ca/maqcsociety/
[12]https://www.cost-charme.eu/news

Article history
Received: 20 March 2020
Published: 26 March 2020

D'Elia et al. (2020) EMBnet.journal 25, e931
http://dx.doi.org/10.14806/ej.25.0.931

Figure 1. Group picture of attendants at the CHARME Final Conference, 2-3 March 2020, Brussels.

European Education & Training programme in existing curricula.

An excellent training programme was realised through the organisation of Think Tank events, training schools and an effective set of STSM actions. Young researchers participating in the first training school, organised in 2017, were so much motivated by the standardisation topic that they initiated follow-up schools and activities. Noteworthy, 15 Early Stage Researchers coming from 12 member countries and hosted in another 7 member countries benefited from STSMs. Furthermore, as part of this success story, the outcome of these STSMs inspired new collaborations, master theses, new project ideas and are also reflected in publications. All beneficiaries of the STSM networking tool agreed that the STSM has been of high value for both training and career.

The results of this COST Action's network will be subject of a White Paper addressing the needs of standardisation, including a catalogue of requirements and recommendations to be disseminated to decision-makers at all levels to enable the implementation of standards in the daily workflow of research in academia and industry.

The Action, ending this month, has given the opportunity to its members to tentatively presenting a set of preliminary requirements to develop further the harmonisation of standards. We hope this will inspire other Actions for the future.

Contacts:
Domenica D'Elia - Science Communication Manager
Susanne Hollmann - Action Chair
Erik Bongcam-Rudloff - Action Vice Chair

Action website: https://www.cost-charme.eu/

Acknowledgements

The authors thank all the members of the CHARME network (COST Action CA15110) for their contribution.

The CHARME COST Action (CA15110) has been funded by COST - The European Cooperation in Science and Technology, MoU - 042/15.

EMBnet.journal
Bioinformatics in Action

Ngoune et al. (2020) EMBnet.journal 25, e926
http://dx.doi.org/10.14806/ej.25.0.926

Immunoinformatics taught to middle and high school pupils in IMGT®

Viviane Nguefack Ngoune, Perrine Pegorier, Morgane Bertignac, Imène Chentli, Veronique Giudicelli, Saida Hadi-Saljoqi, Joumana Jabado-Michaloud, Geraldine Folch✉

IMGT®, the international ImMunoGeneTics information system®, Institut de Génétique Humaine, UMR 9002 CNRS-UM, Université de Montpellier, Montpellier, France

Competing interests: VNN none; PP none; MB none; IC none; VG none; SHS none; JJM none; GF none

Abstract

IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) (Lefranc *et al.*, 2015), is the global reference in immunogenetics and immunoinformatics (Lefranc 2014). In 2019, for the first time, two pupils were welcomed within the team and introduced to immunoinformatics thank to the "Apprentis Chercheurs" program founded by the association l'Arbre des Connaissances. This article describes the method, and pedagogical approaches used, and demonstrates that immunoinformatics can be successfully taught to pupils of middle and high school.

Introduction

The association l'Arbre des Connaissances[1] founded by researchers to promote dialogue between scientists and the general public, set up since 2004 a program named "Apprentis Chercheurs" in Paris. This program allows middle and high school pupils to explore the different facets of research throughout the school year by immersion in research laboratories. The program was quickly replicated in other regions of the country. The local project coordinator is Genopolys[2]. Once per month, the pupils work on their scientific project according to the thematic area of the laboratory. At the end of the school year, Genopolys organises a conference, where the pupils present their work in public (school director, rector of the academy, Arbre des connaissances, researchers, supervisors, teachers, family, *etc.*) and receive a certificate of research initiation.

For the 2019 program, IMGT®, for the first time, received Tessa (middle school pupil) and Jean-Baptiste (high school pupil) in the laboratory to introduce them to a relatively recent scientific field resulting from the fusion of two domains, namely informatics and immunology. To accomplish this goal, the IMGT® biocuration team established planning of the different sessions, topics and resources to be discussed (cf Table 1). This planning was divided into three parts, the first part covered the basics of biology and immunology, the second was focused on bioinformatics, and the last part was an introduction to immunoinformatics. The purpose of these sessions was to introduce the pupils to the analysis of antibodies using bioinformatics tools.

Methods and pedagogical approaches

The delivered sessions had both a theoretical and a practical component. At the beginning of each session and before carrying out the project, it was essential to discuss with the pupils to make sure that they had understood what they were about to do, how they were supposed to do it and the importance of the task. Therefore, each topic started with a lecture or a video projection, which was a reportage or a documentary. After each projection, the content was addressed with many more details and the different questions of the pupils were discussed. Concerning the practical part of the sessions, the pupils' project differed based on the topic. For example on the topic of animal and plant cells, the IMGT® curators had drawn down, separately, all the components of those cells and asked the pupils to put together all these pieces to build a 2D mock-up of an animal and a plant cell. For the immunology project, the pupils played an online game called Leucowar[3] to understand better the immune response.

[1]http://arbre-des-connaissances-apsr.org/
[2]http://www.genopolys.fr
[3]http://philippe.cosentino.free.fr/productions/leucowar/

Article history
Received: 04 October 2019
Published: 09 January 2020

Ngoune et al. (2020) EMBnet.journal 25, e926
http://dx.doi.org/10.14806/ej.25.0.926

Table 1. Detailed program of the different topics discussed during the sessions as well as the projects carried out and resources.

Contents	Topics	Resources	Project applications
Biology and immunology	- Overview of the body systems - Immune system - Animal and plant cells - DNA/RNA structure - Protein synthesis - Genetics engineering	https://www.youtube.com/watch?v=vAz_9_Jesbo https://www.youtube.com/watch?v=w_MsGXYMv3c https://www.youtube.com/watch?v=jSRMqO0e908 https://www.youtube.com/watch?v=to_62MXI8-o https://www.youtube.com/watch?v=tf3_k0-MQzc https://www.youtube.com/watch?v=iJE_FlrA0io	- Build a 2D mock-up of an animal and plant cells - Play an online game Leucowar - Quizzes on DNA/RNA, protein synthesis and genetic engineering
Bioinformatics	- Introduction to Bioinformatics - Function of Bioinformatics - Presentation of generalists databases - Common sequence analysis	https://www.youtube.com/watch?v=CLV1nrzGyBA https://www.youtube.com/watch?v=xK9vOjvYoUM https://www.youtube.com/watch?v=0N3ize9CDwY	- Exploration of the databases NCBI and EBI - Blast analysis to compare sequences (Blast) - Multiple sequence alignment (ClustalOmega) - Phylogenetic tree (Phylogeny)
Immunoinformatics	- Introduction to Immunoinformatics - Presentation of specialised databases for Immunogenetics - Analyse of antibody genes	http://www.imgt.org/IMGTindex/factsbook.php http://www.imgt.org	- Exploration of IMGT/GENE-DB and IMGT/LIGM-DB specialised databases - Annotation of antibody genes (internal tool) - Analysis of rearranged antibody genes (IMGT/V-QUEST)

Quizzes to test the knowledge acquired by pupils were carried out, which helped us to know what topics were well understood by the pupils and which subjects needed to be revisited. At the end of each session, there was some time dedicated to collect and organise the information the pupils wanted for their end of year presentation. The last session was devoted to a rehearsal of the pupils' presentation in the IMGT® team.

Topics discussed

The first two sessions were used as a reminder of biology and immunology topics. We started with an overview of the different body systems, their functions, and how they interact with each other. We paid particular attention to the immune system and introduced some basics of immunology such as the different types of immune response, tolerance and system disorder. The different cells types, their structure, their component as well as their function were discussed in a session entitled cellular biology. Genetic engineering and what can be done with it as well as protein synthesis were addressed afterwards. The different types of antibody genes, their rearrangement process with an example, at the molecular level, of the synthesis of an antibody in humans, were presented in detail.

The huge amount of data accumulated from human and others species and the necessity to analyse them was an excellent topic to introduce the field of Bioinformatics, its use and the type of questions that can be answered with it. The pupils explored different publicly available databases and servers such as the NCBI[4] and EBI[5] resources.

They extracted a gene sequence, which gave them a lot of satisfaction, and in parallel demystified the procedure of getting hold of gene sequences. Different available output files format were introduced such as fasta, genbank and embl, along with their features. Additionally, the pupils performed various analyses such as the pairwise sequence comparison and the extraction of similar sequences from these databases for functional assignment (Johnson *et al.*, 2008). The pupils carried out multiple sequence alignment using different online tools and reconstructed a phylogenetic tree to visualise the molecular relationships of species through their genes.

Immunoinformatics is a discipline which uses informatics to the study of molecules of the immune system (Tomar and De, 2014). Therefore, it is crucial to have various information on those molecules, which is the primary responsibility of the curators of IMGT® who analyse, interpret and integrate immunological information into databases and data repositories. For this part, as for the bioinformatics topic, the pupils explored two specialised databases, IMGT/LIGM-DB (Giudicelli *et al.*, 2006) and IMGT/GENE-DB (Giudicelli *et al.*, 2005). The pupils were shown how to identify and annotate different types of antibody genes on genomic sequences (Lane *et al.*, 2010) and analysed rearranged sequences (Giudicelli *et al.*, 2004).

Conclusion

Based on the pupils' final presentation, their involvement and engagement during the monthly sessions as well as the evaluation of the internship from both the pupils and their teachers, we feel confident that this undertaking was a successful first-time introduction of immunoinformatics to middle and high school pupils.

[4]https://www.ncbi.nlm.nih.gov/
[5]https://www.ebi.ac.uk/

We are convinced that immunoinformatics can be successfully taught to this public and we presented this article in the hope that more people in the field will be encouraged to introduce pupils to this exciting, fast-moving and promising scientific field.

Acknowledgements

We are grateful to Magali Kitzmann and Marie Pequignot, in charge of scientific culture at GENOPOLYS, to set up and manage the program "Apprentis Chercheurs" in Montpellier as well as Sofia Kossida, director of IMGT® to include IMGT® to the program and help us draft the manuscript. We are grateful to Sophie Lacas and Christian Jaubert, Life and Earth Science school teachers at high school Jean Jaurès and middle school Frédéric Bazille in Montpellier for their flawless cooperation during the process. We are also grateful to "l'Arbre des Connaissances" for providing feedback, consultation, and support at all phases of the project.

References

1. Giudicelli, Véronique, Denys Chaume, and Marie-Paule Lefranc (2004) "IMGT/V-QUEST, an Integrated Software Program for Immunoglobulin and T Cell Receptor V–J and V–D–J Rearrangement Analysis." Nucleic Acids Res. **32** (Web Server issue):W435–40. http://dx.doi.org/10.1093/nar/gkh412.
2. Giudicelli, Véronique, Denys Chaume, and Marie-Paule Lefranc (2005) "IMGT/GENE-DB: A Comprehensive Database for Human and Mouse Immunoglobulin and T Cell Receptor Genes." Nucleic Acids Res. **33** (Database issue): D256-261. http://dx.doi.org/10.1093/nar/gki010.
3. Giudicelli, Véronique, Patrice Duroux, Chantal Ginestoux, Géraldine Folch, Joumana Jabado-Michaloud, Denys Chaume, and Marie-Paule Lefranc (2006) "IMGT/LIGM-DB, the IMGT Comprehensive Database of Immunoglobulin and T Cell Receptor Nucleotide Sequences." Nucleic Acids Research **34** (Database issue): D781-784. http://dx.doi.org/10.1093/nar/gkj088.
4. Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden (2008) "NCBI BLAST: A Better Web Interface." Nucleic Acids Research **36** (Web Server): W5–9. http://dx.doi.org/10.1093/nar/gkn201.
5. Lane, Jérôme, Patrice Duroux, and Marie-Paule Lefranc (2010) "From IMGT-ONTOLOGY to IMGT/LIGMotif: The IMGT Standardized Approach for Immunoglobulin and T Cell Receptor Gene Identification and Description in Large Genomic Sequences." BMC Bioinformatics **11**: 223. http://dx.doi.org/10.1186/1471-2105-11-223.
6. Lefranc, Marie-Paule (2014) "Immunoglobulin and T Cell Receptor Genes: IMGT(®) and the Birth and Rise of Immunoinformatics." Frontiers in Immunology **5**:22. http://dx.doi.org/10.3389/fimmu.2014.00022.
7. Lefranc, Marie-Paule, Véronique Giudicelli, Patrice Duroux, Joumana Jabado-Michaloud, Géraldine Folch, Safa Aouinti, Emilie Carillon, *et al.* (2015) "IMGT®, the International ImMunoGeneTics Information System® 25 Years On." Nucleic Acids Research **43** (Database issue): D413-422. http://dx.doi.org/10.1093/nar/gku1056.
8. Tomar, Namrata and Rajat K. De (2014) "Immunoinformatics: A Brief Review." Methods in Molecular Biology (Clifton, N.J.) **1184**: 23–55. http://dx.doi.org/10.1007/978-1-4939-1115-8_3.

EMBnet.journal
Bioinformatics in Action

Papakonstantinou et al. (2020) EMBnet.journal 25, e927
http://dx.doi.org/10.14806/ej.25.0.927

Reasearch Papers

An in silico and in vitro pipeline for the rapid screening of helicase modulators

Eleni Papakonstantinou[1,2], Flora Bacopoulou[3], Vasileios Megalooikonomou[4], Aspasia Efthimiadou[5], Dimitrios Vlachakis[1,2,6]✉

[1]Laboratory of Genetics, Department of Biotechnology, School of Food, Biotechnology and Development, Agricultural University of Athens, Athens, Greece

[2]Lab of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[3]Center for Adolescent Medicine and UNESCO Chair on Adolescent Health Care, First Department of Pediatrics, Medical School, National and Kapodistrian University of Athens, Aghia Sophia Children's Hospital, Athens, Greece

[4]Computer Engineering and Informatics Department, School of Engineering, University of Patras, Patras, Greece

[5]Hellenic Agricultural Organization-Demeter, Institute of Soil and Water Resources, Department of Soil Science of Athens, Lycovrisi, Greece

[6]Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, Strand, London, Uinted Kingdom

Competing interests: EP none; FB none; VM none; AE none; DV none

Abstract

To evaluate the potency of potential helicase modulators, we developed an assay of helicase enzyme activity. Using a DNA or RNA biotin labelled oligonucleotide and after the addition of a recombinant helicase, the nucleic acid unwinds, causing the emission of luminescence, which is quantified with a particular antibody. In our assay, one of the DNA oligos was biotinylated, while the other was labelled with digoxygenin (DIG), both in their 5' termini. The biotin molecule immobilises the DNA duplex on a neutravidin-coated plate and the helicase activity is measured through the unwinding of DNA, due to ATP activation. The subsequent release of DIG-labelled oligos results in a luminescence signal measured with a chemiluminescence antibody. Our goal was to provide a high throughput screening method for potential helicase inhibitors. The method described in this paper has been demonstrated to be fast, easy and reproducible and doesn't use radiochemicals.

Introduction

Helicase activity assays include analysis of ATPase activity, but it was shown that measuring the helicase unwinding activity is the best method for evaluating modulators of this class of enzyme (Borowski *et al.*, 2002). This type of assay depends on the ability of the enzyme to separate the release strand of DNA or RNA from the template strand. Other methods measure the deposition of radio-labelled release strands after gel electrophoresis, thin-layered chromatography or scintillation counting (Borowski *et al.*, 2001; Bartelma and Padmanabhan 2002; Alaoui-Ismaili *et al.*, 2000). These methods could be enhanced via high-throughput screening, although the radioactive materials would be a problem. Another method detects DIG-labelled release strands by ELISA (Hsu *et al.*, 1998). We propose a combination of the methods as mentioned above without radio-labelled molecules that can detect the residual release strand with a chemiluminescent antibody, giving a robust helicase assay and a stable read-out, well suited to high-throughput screening.

Our goal was to provide a high throughput screening method for potential helicase inhibitors, and that is why we developed this fast, easy and reproducible assay, without using radiochemicals. The high reproducibility of the assay is obvious after the observation of only insignificant variations on a single 96-reaction plate. Helicase from the Hepatitis C Virus (HCV) was expressed and isolated through recombinant protein methods and later used in our assay.

Materials

All solutions were prepared using ultrapure water obtained by purifying deionised water. All reagents were stored at room temperature unless indicated otherwise.

Article history
Received: 21 December 2019
Accepted: 31 December 2019
Published: 13 February 2020

Papakonstantinou et al. (2020) EMBnet.journal 25, e927
http://dx.doi.org/10.14806/ej.25.0.927

HCV recombinant helicase preparation

1. Escherichia coli induction: Prepare 4 conical flasks with 750 mL of LB and add 34 μg/mL chloramphenicol and 25 μg/mL kanamycin to each;
2. cell lysis buffer: 20 mM sodium phosphate pH 7.5, 300 mM NaCl;
3. buffer S: 20 mM sodium phosphate pH 7.4, 500 mM NaCl;
4. exchange buffer: 25 mM Tris–HCl pH 7.5, 0.05% CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulphonate), 20% glycerol, 5 mM DTT (dithiothreitol).

HCV helicase assay procedure

1. Oligonucleotide mix:
   - oligonucleotides (5'-biotin-GCTGACCCTGCTCCCAATCGTAATCTATAGTGTCACCTA-3',
   - 5'-DIG-CGATTGGGAGCAGGGTCAGC-3') 1:1 molar, HEPES 2mM, NaCl 0.05M, EDTA 0.1 mM, SDS 0.01% w/v;
2. neutravidin stock solution: add neutravidin in final concentration 1 mg/ml in phosphate buffered saline (1 M PBS - pH 7.0);
3. BSA solution: 0.1% w/v BSA;
4. substrate solution: mix 2.5 ng partially annealed DNA duplex and 75μL 1 M PBS containing 1M NaCl, for each well;
5. substrate wash solution: 50 mM Tris HCl pH 7.5, 50 mM NaCl;
6. helicase reaction mix: 11 nM purified full-length HCV NS3 protein, 25 mM 4-morpholine-propanesulphonic acid (MOPS) pH 7.0, 2 mM DTT, 3mM MnCl2, 100 μg/ml of BSA and 5 mM ATP. The reaction mix for the negative control lacks ATP;
7. reaction wash solution: 150 mM NaCl;
8. detection washing buffer: 0.1 M maleic acid, 0.15 M NaCl, 0.3%, Tween20, pH 7.5;
9. blocking solution: 10% BSA w/v, 0.1 M maleic acid, 0.15 M NaCl, pH 7.5;
10. antibody solution: 1:10.000 solution of the anti-Dig antibody (75 mU/mL) in Blocking solution;
11. detection buffer: 0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5;
12. chemiluminescence substrate solution: CSPD – 0.25 mM.

Methods

All procedures have to be carried out at room temperature unless specified otherwise.

HCV recombinant helicase NS3 preparation

1. Insert the full-length HCV helicase coding region in a pET28a vector, with a N-terminal 6xHis-Tag region;
2. verify the intactness of the gene before inducing the protein production;
3. transform Escherichia coli cells (strain C41, DE3) with the helicase plasmid and inoculate the prepared LB flasks with them;
4. induce the recombinant helicase production, adding 0.5 mM IPTG to each flask and then allow the cultures to grow for 3 hours at 18°C;
5. resuspend the cell pellet from the 4 cultures in 30 mL Lysis buffer and then add lysozyme 100μg/mL and Triton X-100 0.1%;
6. incubate on ice for 30 min and then sonicate four times for 20 sec with 15 sec intervals;
7. centrifuge the suspension at 15.000 x g for 20 min;
8. adjust the clarified homogenates to 10 mM imidazole and filter them through a 0.45 μm membrane;
9. load the homogenates twice on nickel affinity columns (Ni-NTA);
10. wash each column with five times the column volume of buffer S, containing 10 mM imidazole;
11. elute the NS3 helicase with buffer S containing 300 mM imidazole;
12. immediately after the elution, exchange the buffer in the helicase-containing fractions for the Exchange buffer, via dialysis; this step is critical in order to avoid precipitation;
13. evaluate the protein concentration using the Bradford assay with BSA as standard;
14. create aliquots of the NS3 helicase and store at -80°C.

This recombinant protein preparation is estimated to be more than 85% pure by SDS gel electrophoresis and Coomassie blue staining, yielding almost 1.6 mg of HCV NS3 per 3 litres of E. coli cultures (Figure 1).

Chemiluminesence readings were taken using all different control combinations (presence/absence of helicase, DNA substrate and ATP) of the experiment to ensure the reliability of the measurements (Table 1).

We demonstrated that the NS3-mediated unwinding is proportional to the amount of DNA substrate in the well, but also to the HCV helicase concentration in the reaction. The reactions were ATP-dependent (Table 2 and Figure 2).

Annealing reaction

1. Prepare for annealing by heating the oligonucleotide mix at 100°C for 5 min;
2. incubate at 65°C for 30 min and then at 22° for 4 h, to allow gradual annealing;
3. store the annealed NS3 helicase substrate at -20°C (See Note 1).

Figure 1. The SDS gel (left) and western blot (right, anti-His-tag antibody, penta-His conjugated) for the HCV helicase protein.

Table 1. Each enzymatic activity assay was performed in triplicate and the results were averaged. All reactions were allowed to proceed for 60 minutes (concentrations as described in the methods section).

	CHEMILUMINESENCE
+ DNA substrate - helicase - ATP	0.532 (± 2%)
+ DNA substrate - helicase + ATP	0.529 (± 4%)
+ DNA substrate + helicase - ATP	0.525 (± 5%)
- DNA substrate + helicase + ATP	0.040 (± 3%)

Neutravidin coating of the 96-well plates

1. Coat each of the 96 wells overnight at 4°C with 100 µl/well of a 5 µg/ml neutravidin solution in 0.5 M sodium carbonate buffer pH 9.3;
2. wash the plates three times with 100µl/well of PBS and air-dry at room temperature.

Blocking with BSA

1. Add 100 µL of the 0.1% w/v BSA solution;
2. incubate at 22°C for 2 h;
3. wash the plate three times with PBS, 200 µl/well and air-dry at room temperature;
4. store the plate at 4°C with desiccant (See Note 2).

Table 2. Different DNA substrate concentrations (60-minute run).

DNA (ng) immobilised per well	CHEMILUMINESENCE without ATP	CHEMILUMINESENCE with ATP
0	0.039	0.040
0.5	0.128	0.044
1	0.255	0.045
2	0.495	0.045
2.5	0.525	0.048
5	0.902	0.065

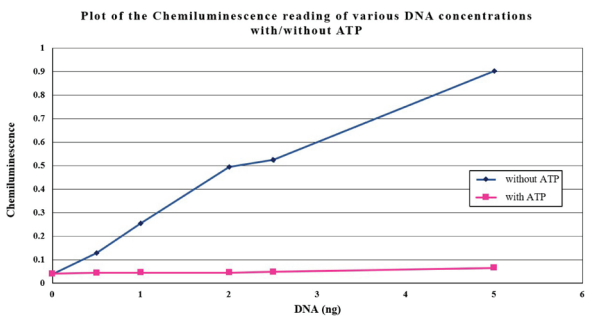


Figure 2. Graphical representation of Table 2 data.

Substrate application in the 96-well plate

1. Pre-warm all solutions to 37°C (See Note 3);
2. mix 75 µl of the Substrate solution with 2.5 ng of the partially annealed DNA duplex for each well;
3. incubate at 22°C for 4 h;
4. wash each well twice with 200 µl PBS per well and once with Substrate Wash solution.

Helicase Reaction

1. Add 90 µL of the Reaction mix per well;
2. incubate the reactions at 37°C for 1 hour;
3. wash the plate twice with 200 µL Reaction Washing Buffer per well and let dry for 15 minutes at room temperature.

Activity determination – chemiluminescence preparation

1. Wash all wells for 5 min with the Detection washing buffer;
2. then fill up each well with Blocking solution for 30 min and then incubate for 30 min in 20 µl Antibody solution;
3. wash twice the plate with 100 µl of Detection buffer;
4. apply 20 µL of Detection buffer for equilibration and 1 µL of chemiluminescence substrate working solution per well and incubate for 5 min at 17oC;
5. drain the wells and incubate the plate at 37°C for 30 min to allow any remaining solution to evaporate;
6. the luminescence has a constant intensity for about 24 hours and continues for approximately 48 hours. The remaining DIG in each well is counted for 10 min against controls (one of which lacks protein and the other lacks ATP) in a luminescence plate reader (See Note 4).

Notes

Note 1: The oligonucleotides used in this protocol are the ones described by Alaoui-Ismaili et al. (Alaoui-Ismaili *et al.*, 2000), modified by DIG labelling of the release strand. Other sequences could work, but a 3' single stranded region in the substrate is necessary to initiate the strand displacement (Tai *et al.*, 1996).
Note 2: As far as the blocking procedure is concerned, it is important to ensure that all potential binding sites are occupied, to prevent direct binding of the detection antibody to the well in a later step. The wells should be filled with blocking solution, fully coating the plate.
Note 3: Pre-warming the solutions allows reactions to proceed at their optimum temperatures and avoids rate changes due to temperature equilibration.
Note 4: All the reaction wells should be filled up with blocking solution, to ensure that the whole well has been blocked, preventing non-specific binding of any components of the detection system.

Key Points

- Testing new antiviral drugs.
- Stopping viral enzymes as therapy.
- Using supercomputers to design new drugs.
- Fast, cheap and reliable test for antiviral agents.

Acknowledgements

DV would like to acknowledge funding from: i. Microsoft Azure for Genomics Research Grant (CRM:0740983) ii. FrailSafe Project (H2020-PHC-21-2015 - 690140) "Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions", co-funded by the European Commission under the Horizon 2020 research and innovation program. iii. Amazon Web Services Cloud for Genomics Research Grant (309211522729). iv. AdjustEBOVGP-Dx (RIA2018EF-2081): Biochemical Adjustments of native EBOV Glycoprotein in Patient Sample to Unmask target Epitopes for Rapid Diagnostic Testing. A European & Developing Countries Clinical Trials Partnership (EDCTP2) under the Horizon 2020 "Research and Innovation Actions" DESCA.

References

1. Alaoui-Ismaili MH, Gervais C, Brunette S, Gouin G, Hamel M *et al.* (2000) A novel high throughput screening assay for HCV NS3 helicase activity. Antiviral Res. **46**, 181-193. http://dx.doi.org/10.1016/s0166-3542(00)00085-1
2. Bartelma G & Padmanabhan R (2002) Expression, purification, and characterization of the RNA 5'-triphosphatase activity of dengue virus type 2 nonstructural protein 3. Virology **299**, 122-132. http://dx.doi.org/10.1006/viro.2002.1504
3. Borowski P, Niebuhr A, Mueller O, Bretner M, Felczak K *et al.* (2001) Purification and characterization of West Nile virus nucleoside triphosphatase (NTPase)/helicase: evidence for dissociation of the NTPase and helicase activities of the enzyme. J Virol **75**, 3220-3229. http://dx.doi.org/10.1128/JVI.75.7.3220-3229.2001
4. Borowski P, Niebuhr A, Schmitz H, Hosmane RS, Bretner M *et al.* (2002) NTPase/helicase of Flaviviridae: inhibitors and inhibition of the enzyme. Acta Biochim Pol **49**, 597-614. http://dx.doi.org/024903597
5. Hsu CC, Hwang LH, Huang YW, Chi WK, Chu YD *et al.* (1998) An ELISA for RNA helicase activity: application as an assay of the NS3 helicase of hepatitis C virus. Biochem Biophys Res Commun **253**, 594-599. http://dx.doi.org/10.1006/bbrc.1998.9813
6. Tai CL, Chi WK, Chen DS & Hwang LH (1996) The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). J Virol **70**, 8477-8484. http://dx.doi.org/10.1128/JVI.70.12.8477-8484.1996

EMBnet.journal
Bioinformatics in Action

Papakonstantinou et al. (2020) EMBnet.journal 25, e928
http://dx.doi.org/10.14806/ej.25.0.928

Reasearch Papers

Dark Suite: a comprehensive toolbox for computer-aided drug design

Eleni Papakonstantinou[1,2], Vasileios Megalooikonomou[3], Dimitrios Vlachakis[1,2,4]✉

[1]Laboratory of Genetics, Department of Biotechnology, School of Food, Biotechnology and Development, Agricultural University of Athens, Athens, Greece

[2]Lab of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[3]Computer Engineering and Informatics Department, School of Engineering, University of Patras, Patras, Greece

[4]Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, Strand, London, Uinted Kingdom

Competing interests: EP none; VM none; DV none

Abstract

Dark Suite is a complete interactive software pipeline aiming to exploit the advantages of free software and modern programming. Besides two commands (installation and opening) on the command line, the handling and the program operation is done through a user's friendly interface. This platform has a central graphical interface which allows the user to choose in what computational tool to work. Each computational tool has an interface. Dark Suite combines the functions of other programmes to create a pipeline for high-quality secondary effects through a friendly user interface. It is made to run on GNU / Linux distributions and its interface was built using JAVA to seamlessly integrate scientific tools written in Perl, Java, R and Python.

Introduction

The discovery and development of a new drug is still a time and cost consuming process, estimating for approximately 10-15 years for a new drug to enter the market (Paul *et al.*, 2010; Pan *et al.*, 2013). In pharmaceutical research, computational methods are being employed to reduce the cost and time of drug development, while enabling the synthesis of large component drug libraries (Clark *et al.*, 2010; Szymański *et al.*, 2012). Computer-aided drug design (CADD) is based on the combination of computational techniques that can enable the simulation of molecular interactions between proteins and target molecules and predict the effectiveness of a new lead molecule (Veselovsky *et al.*, 2014). Structure-based and ligand-based drug design are the primary methodologies for drug discovery and lead optimisation; depending on the available biological information. Ligand-based approaches generate SAR models based on known active and/or inactive molecules, whereas structure-based approaches use the structural information of the protein target to discover lead molecules as potent inhibitors. Molecular Dynamics, Quantum Mechanics and Linear Interaction Energy (LIE) are included in the drug discovery process optimisation to evaluate in silico the effectiveness of the lead molecule.

CADD approaches have led to many successful applications, through lead discovery or drug repurposing. However, conventional CADD has its limitations and their results have to be validated in real biological systems, as numerous molecules that have been recognised in silico, do not eventually exhibit the predicted activity and efficiency. Apart from the complexity of the biological systems and the difficulty in interpreting and simulating them in silico, an additional limitation of CADD applications is that the tools used to discover and design a new drug are based on specific algorithms and each tool has its limitations. For that purpose, it is considered necessary to continually update the tools and algorithms to improve the accuracy and the provision of new drugs. The employment of CADD applications is an integral part of pharmaceutical research and novel approaches need to be implemented (Baig *et al.*, 2018).

Herein, we describe a novel solution for computer-aided drug design through the Dark Suite application.

Article history
Received: 28 December 2019
Accepted: 31 December 2019
Published: 13 March 2020

Papakonstantinou et al. (2020) EMBnet.journal 25, e928
http://dx.doi.org/10.14806/ej.25.0.928

The suite is designed to facilitate the drug discovery process by introducing freely available tools that have been shown to outperform conventional approaches. In the first step of the structure-based drug design and the discovery of homologous proteins, an enhanced sequence similarity search can be performed by exploiting secondary structural elements and the hydropathy profiles of the proteins. Homology modelling can then be achieved using additional restraints based on the shape and size similarity of the homologous proteins, leading to a refined 3D model structure. In the next step, the lead discovery is approached by a fully integrated platform set to automate the drug design through efficient algorithms for protein preparation, energy minimisation, pharmacophore elucidation, building and growing new drug-like moieties, and molecular dynamics simulations. Additionally, the suite includes tools for protein clustering based on gene ontology annotations, and for enhanced prediction of protein interactions based on protein-protein interaction networks.

Description of the program

The Dark Suite is a complete interactive platform that integrates free software and modern programming solutions for a high-quality research pipeline in computational drug design in the drug discovery research. It is a UNIX based software, compatible with all major GNU/LINUX distributions and its interface was built using Java language programming. The integrated tools are written in Perl, Java and bash. It is easily installed through the command line and can be downloaded freely at www.darkdna.gr.

The platform includes 2D, 3D and 4D protein analysis tools, an automated drug design platform and a 3D structure viewer (Figure 1).

Figure 1. The front-end window of Dark Suite.

Namely, the Dark Suite integrates the PSSP tool (Figure 2), TAGGO, Space, GIBA, Drugster and Jmol that are described below. Each of the tools can be assessed through its autonomous interface after installing and running the suite. Noteworthy, the suite enables the user to run all biological tools in parallel. In such a manner, the drug design pipeline can be adjusted to the user's needs.

Integrated tools

1. 2D analysis

The protein secondary structure profile (**PSSP**) tool is developed to perform fast, efficient similarity searches based not only on the protein sequences but also on the secondary structure profile of the proteins (Vlachakis et al., 2017). PSSP first performs a conventional blast search for the query protein sequence and then uses a custom made hydropathy substitution matrix by which the entries are re-scored and re-ranked. The hydropathy profile is constructed for each protein using the hydropathy index from IMGT (Lefranc *et al.*, 1999). Moreover, PSSP exploits the secondary structural information by searching against the RCSB PDB secondary elements database and is also able to predict secondary elements if the query protein is not crystallographically determined.

2. 3D analysis

TAGGO module is designed for an automated clustering of a set of proteins based on Gene Ontology (GO) resources (Roubelakis *et al.*, 2009). The module has a 5-step interface, in which the user has to select: the path for the input file that includes a list of proteins to be annotated, the appropriate GO file and its respective format, the organism for which the clustering is performed, the Evidence Codes to be considered, and the path to the output directory. TAGGO then evaluates the Information Content (IC) of each ontology term assigned to each protein of the input set for all three GO aspects; Molecular Function (MF), Cellular Component (CC) and Biological Process (BP). The most general terms (with lower information content) are considered and the proteins in the dataset are assigned to categories. The overall output of the process is the percentage of the annotated GO categories of the protein dataset, for each GO aspect. A directory called "Results" is created that includes Venn lists for each aspect, a visualisation of the results in pie and bar charts, and text files describing the parameters and annotation steps throughout the process.

Space represents a novel, efficient methodology for homology modelling that improves the quality and credibility of the resulting model compared to conventional approaches (Vlachakis *et al.*, 2013). Space can be implemented in cases of low sequence identity by considering the proteins' shape and size similarity. Based on the fact that structure is more conserved than the sequence in nature, Space performs 3D modelling

Figure 2. The interface of the PSSP toll which can be run autonomously.

by setting additional constraints determined by the conformational shape of the template protein.

3. 4D analysis

GIBA is an effective and user-friendly tool for the identification of accurate protein complexes through clustering of protein-protein interaction (PPI) networks (Moschopoulos *et al.*, 2009). GIBA encompasses two different clustering algorithms, the MCL (Enright *et al.*, 2002) and the RNSC algorithm (King *et al.*, 2004) and a set of different user-defined parameters for clustering and filtering. Once the workflow is executed, the output is a set of the final clusters with the interacting proteins predicted. GIBA surpasses other methods in quality approximations of protein complexes.

4. Drug Design

Drugster is a freeware platform aimed to assist scientists in the field of computer-aided drug design (Vlachakis *et al.*, 2013). Drugster integrates the algorithms of PDB2PQR v.1.8 (Dolinsky *et al.*, 2004; Dolinsky *et al.*, 2007), Ligbuilder v.1.2 and v.2.0 (Yuan *et al.*, 2011), Gromacs v.4.5.5 (Hess *et al.*, 2008) and Dock v.6.5 (Lang *et al.*, 2009). It is designed to automate the process of structure-based drug design and lead optimization through an easy to use interface. The complete workflow consists of five steps:

1. Input preparation; the 3D structure of a PDB file is refined and problems are automatically fixed;
2. energy minimization; the receptor is optimized using the Gromacs suite;
3. *de novo* structure-based drug design; the Ligbuilder module is used for ligand building after a pharmacophore preparation and the determination of a 3D scaffold to generate novel moieties;
4. ligand optimization; all the candidates are docked to the receptor using the Dock module and ranked after energy minimization, and
5. complex optimization; the ligand-receptor complex is energetically minimized and the system undergoes molecular dynamics simulations.

5. PDB viewer

The structures can be visualised through the PDB viewer that uses Jmol[1] (Jmol: an open-source Java viewer for chemical structures in 3D), a free, open-source molecule viewer.

Platform compatibility:

Dark Suite has been tested on the following GNU / Linux distributions:

- Ubuntu 14.04.4 LTS (32&64 bit);
- Ubuntu 15.04 LTS (32&64 bit);
- Ubuntu 16.04 LTS (32&64 bit);
- Kubuntu 14.04.4 LTS (32&64 bit);
- Debian 8.0 (32&64 bit).

[1]http://www.jmol.org/

Dark Suite Installation steps:

Open a terminal and initially go in the Dark Suite directory and then in the medical directory located inside the Dark Suite that contains all files, *e.g.* `cd DarkSuite/medical/`. Within the medical directory, you can locate the installation file. Run the command: `<./install.sh>`.

Installation process:

Run `./install.sh` and your password will be prompted. Press "enter" to begin the installation. You will be asked again to press "enter" after a bit to continue. Installation continues normally for some time and all necessary tools and libraries are being installed. Drugster application will be installed through a dedicated interface, after following instructions. You will be asked to choose the Linux distribution you are using. Choose and press "enter", and you will be prompted to the graphical installation environment of Drugster. Follow the instructions until completion.

Dark Suite Operation steps:

1. Run the command: `<java -jar darksuite.jar>` and Dark Suite will load;
2. press the button TAGGO to execute TAGGO;
3. press the button SPACE and choose the pdb file you want; a separate window is used to display information about file which was selected;
4. press the button PSSP to execute PSSP;
5. press the button Drugster to execute Drugster;
6. press the button GIBBA to execute GIBBA;
7. press the button Run Viewer to open a window to insert the pdb file you want to view; Then after selecting and pressing "ok", you will see the 3D protein structure hat selected;
8. press the button Cloud to create a new directory in the directory with the executable file on a cloud file location.

Key Points

- An innovative solution for computer-aided drug design to facilitate and refine the drug discovery process.
- Quickly installed, fully interactive platform integrating state of the art protein analysis tools.
- Efficient pipeline and novel approaches for structure-based drug design.
- Dark Suite is compatible with all major GNU / Linux distributions and is freely downloadable at darkdna.gr

Conclusions

Dark Suite is a stand-alone application for computer-aided drug design performing all fundamental steps for the process of lead discovery and optimisation, in a user-friendly environment. The Dark Suite pipeline introduces novel and efficient methods for surpassing the limitations on homologous protein discovery based on the primary structural information, traditional homology modelling and drug design approaches. Moreover, it encompasses tools for refined gene ontology annotations and predicted protein-protein interactions. Dark Suite is freely available for the scientific community and allows for a user-defined workflow by selecting the appropriate tools for use.

Acknowledgements

DV would like to acknowledge funding from: i. Microsoft Azure for Genomics Research Grant (CRM:0740983) ii. FrailSafe Project (H2020-PHC-21-2015 - 690140) "Sensing and predictive treatment of frailty and associated co-morbidities using advanced personalized models and advanced interventions", co-funded by the European Commission under the Horizon 2020 research and innovation program. iii. Amazon Web Services Cloud for Genomics Research Grant (309211522729). iv. AdjustEBOVGP-Dx (RIA2018EF-2081): Biochemical Adjustments of native EBOV Glycoprotein in Patient Sample to Unmask target Epitopes for Rapid Diagnostic Testing. A European & Developing Countries Clinical Trials Partnership (EDCTP2) under the Horizon 2020 "Research and Innovation Actions" DESCA. EP would like to acknowledge funding by the State Scholarships Foundation (IKY) - European Union (European Social Fund-ESF) and Greek national funds through the action entitled "Strengthening Human Resources Research Potential via Doctorate Research" in the framework of the Operational Program "Human Resources Development Program, Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) 2014 – 2020.

References

1. Baig, M. H., K. Ahmad, G. Rabbani, M. Danishuddin and I. Choi (2018). "Computer Aided Drug Design and its Application to the Development of Potential Drugs for Neurodegenerative Disorders." Curr Neuropharmacol **16**(6),740-748. http://dx.doi.org/10.2174/1570159X15666171016163510
2. Clark, R. L., B. F. Johnston, S. P. Mackay, C. J. Breslin, M. N. Robertson and A. L. Harvey (2010). "The Drug Discovery Portal: a resource to enhance drug discovery from academia." Drug Discov Today **15**(15-16), 679-683. http://dx.doi.org/10.1016/j.drudis.2010.06.003
3. Dolinsky, T. J., P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker (2007). "PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations." Nucleic Acids Res **35**(Web Server issue): W522-525. http://dx.doi.org/10.1093/nar/gkm276

4. Dolinsky, T. J., J. E. Nielsen, J. A. McCammon and N. A. Baker (2004). "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations." Nucleic Acids Res **32**(Web Server issue): W665-667. doi: http://dx.doi.org/10.1093/nar/gkh381
5. Enright, A. J., S. Van Dongen and C. A. Ouzounis (2002). "An efficient algorithm for large-scale detection of protein families." Nucleic Acids Res **30**(7), 1575-1584. http://dx.doi.org/10.1093/nar/30.7.1575
6. Hess, B., C. Kutzner, D. van der Spoel and E. Lindahl (2008). "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." J Chem Theory Comput **4**(3): 435-447. http://dx.doi.org/10.1021/ct700301q
7. King, A. D., N. Przulj and I. Jurisica (2004). "Protein complex prediction via cost-based clustering." Bioinformatics **20**(17), 3013-3020. http://dx.doi.org/10.1093/bioinformatics/bth351
8. Lang, P. T., S. R. Brozell, S. Mukherjee, E. F. Pettersen, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James and I. D. Kuntz (2009). "DOCK 6: combining techniques to model RNA-small molecule complexes." RNA **15**(6), 1219-1230. http://dx.doi.org/10.1261/rna.1563609
9. Lefranc, M. P., V. Giudicelli, C. Ginestoux, J. Bodmer, W. Muller, R. Bontrop, M. Lemaitre, A. Malik, V. Barbie and D. Chaume (1999). "IMGT, the international ImMunoGeneTics database." Nucleic Acids Res **27**(1), 209-212. http://dx.doi.org/10.1093/nar/27.1.209
10. Moschopoulos, C. N., G. A. Pavlopoulos, R. Schneider, S. D. Likothanassis and S. Kossida (2009). "GIBA: a clustering tool for detecting protein complexes." BMC Bioinformatics **10** Suppl 6: S11. http://dx.doi.org/10.1186/1471-2105-10-S6-S11
11. Pan, S. Y., S. F. Zhou, S. H. Gao, Z. L. Yu, S. F. Zhang, M. K. Tang, J. N. Sun, D. L. Ma, Y. F. Han, W. F. Fong and K. M. Ko (2013). "New Perspectives on How to Discover Drugs from Herbal Medicines: CAM's Outstanding Contribution to Modern Therapeutics." Evid Based Complement Alternat Med 2013: 627375. http://dx.doi.org/10.1155/2013/627375
12. Paul, S. M., D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg and A. L. Schacht (2010). "How to improve R&D productivity: the pharmaceutical industry's grand challenge." Nat Rev Drug Discov **9**(3), 203-214. http://dx.doi.org/10.1038/nrd3078
13. Roubelakis, M. G., P. Zotos, G. Papachristoudis, I. Michalopoulos, K. I. Pappa, N. P. Anagnou and S. Kossida (2009). "Human microRNA target analysis and gene ontology clustering by GOmir, a novel stand-alone application." BMC Bioinformatics **10** Suppl 6: S20. http://dx.doi.org/10.1186/1471-2105-10-S6-S20
14. Szymanski, P., M. Markowicz and E. Mikiciuk-Olasik (2012). "Adaptation of high-throughput screening in drug discovery-toxicological screening tests." Int J Mol Sci **13**(1), 427-452. http://dx.doi.org/10.3390/ijms13010427
15. Veselovsky, A. V., M. S. Zharkova, V. V. Poroikov and M. C. Nicklaus (2014). "Computer-aided design and discovery of protein-protein interaction inhibitors as agents for anti-HIV therapy." SAR QSAR Environ Res **25**(6), 457-471. http://dx.doi.org/10.1080/1062936X.2014.898689
16. Vlachakis, D., A. Armaos and S. Kossida (2017). "Advanced Protein Alignments Based on Sequence, Structure and Hydropathy Profiles; The Paradigm of the Viral Polymerase Enzyme." Mathematics in Computer Science **11**(2), 197-208. http://dx.doi.org/10.1007/s11786-016-0287-8
17. Vlachakis, D., D. G. Kontopoulos and S. Kossida (2013). "Space constrained homology modelling: the paradigm of the RNA-dependent RNA polymerase of dengue (type II) virus." Comput Math Methods Med 2013: 108910. http://dx.doi.org/10.1155/2013/108910
18. Vlachakis, D., D. Tsagrasoulis, V. Megalooikonomou and S. Kossida (2013). "Introducing Drugster: a comprehensive and fully integrated drug design, lead and structure optimization toolkit." Bioinformatics **29**(1), 126-128. http://dx.doi.org/10.4018/ijsbbt.2013070105
19. Yuan, Y., J. Pei and L. Lai (2011). "LigBuilder 2: a practical de novo drug design approach." J Chem Inf Model **51**(5), 1083-1091. http://dx.doi.org/10.1021/ci100350u

silent walls

Vivienne Baillie Gerritsen

Though it may seem a paradox, life is riddled with barriers. This is because it is sometimes necessary to create dead ends to keep things at a healthy distance. Obstructions of this kind exist at all levels of living matter. Specialized pores are found in membranes surrounding cells but also within cells, to ensure that only specific molecules are able to cross while the transit of others is barred. Aquaporin and sodium channels are two examples through which only water molecules or sodium ions fit, respectively. Another vital barrier is the one that keeps spermatozoa that belongs to one species from fertilizing eggs that belong to another – which would only bring about chaos. Though the mingling of germinal fluids and how life ensues have been discussed since the days of Aristotle, on the molecular level very little is known, still, on how species keep to themselves. A recent find in zebrafish has lifted a veil: scientists discovered a protein on the membranes of zebrafish eggs, which only allows access to zebrafish sperm. They called it Bouncer.

by Fanny Vaucher

Fertilisation is complex. Biologists have been acquainted with advanced technology for well over 20 years now, which has given them the means to become quite intimate with DNA and proteins. However, up until the late 19th century, how living entities emerged from a parent and carrying very similar features, had remained obscure for centuries. In the Middle Ages explanations flirted with metaphysics. Where is what? What penetrates what? What triggers what? When the optical microscope was perfected in 1677, it became possible to observe spermatozoa, and scientists began to focus their attention on germ cells. There were the spermists who believed that one minute living being pre-existed in each spermatozoon waiting for its cue to grow, while the ovists believed they lay waiting in the ovum. By the 1880s, much had been understood about cells, their nucleus and even chromosomes – and the genetic continuity of a species had been clarified. 100 years later, how a spermatozoon meets an ovum, where the genetic heritage lies and how it is handed down to its descendants had almost become common knowledge. What remained vague, however, was how it all happens at the molecular level.

For progeny to be viable, one partner has to mate with a second partner from the same species. In this way, the species itself will survive. How, though, do male and female germ cells recognise each other? How do they make contact, and how do sperm cells enter an egg? Furthermore, an egg must not be fertilised by more than one sperm cell – so what is it that stops other sperm cells from penetrating the egg? There is little doubt that the process of fertilisation must summon hordes of molecules to get things right. In humans, sperm are thought to undergo some sort of preselection in the uterus just to make sure they belong to the same species. Once they reach the egg, they have to forage their way through its coat. The first one that manages to do this triggers off a chemical reaction that hardens the coat, thus making it impossible for other sperm cells to penetrate. The laureate sperm cell then continues to pierce a hole through the egg's membrane to release its nucleus into the cytoplasm. Things are a little different in zebrafish. For one, fertilisation occurs in the external medium. Second, while human sperm can fertilise an egg anywhere on its membrane, zebrafish sperm has to find a specific opening on the egg membrane, known as the micropyle. It is the zebrafish egg that selects a

compatible sperm cell thanks to a species-specific receptor on its surface: Bouncer.

Bouncer belongs to the Ly6/uPAR (Ly6/urikinase plasminogen activator receptor) family – so called because Ly6 and uPAR are two of its representatives. This family of proteins features domains whose stamp is a beta-structural core stabilised by disulphide bonds that create three characteristic extended loops, or fingers. Ly6/uPAR proteins are wide-spread and found in mammals, birds, reptiles, amphibians, fish and insects where they are part of the nervous, immune and reproductive systems. The very diverse functions that Ly6/UPAR proteins carry out are probably due to structural differences in the three-fingered domains. In zebrafish, Bouncer is bound to the egg's membrane by way of a GPI anchor.

What exactly does Bouncer do? Zebrafish eggs that have been deprived of Bouncer have no effect on sperm behaviour prior to contact – which means that the protein doesn't act as a cue to attract sperm. Neither does it seem to have a role in maturation once the egg is released into the spawning medium. However, when there is no Bouncer, sperm are neither able to locate the micropyle – where the protein is mainly expressed – nor interact with the egg. This means that Bouncer is no doubt necessary for sperm-egg interaction. Now, what happens to zebrafish eggs whose Bouncer protein is replaced by one from a fish distantly related to them? Say, from the medaka fish which diverged from zebrafish about 200 million years ago. The answer? The medaka Bouncer did not restore fertility to the zebrafish eggs. However, the medaka sperm were able to interact with the zebrafish eggs. The resulting embryos were hybrids with a maternal zebrafish genome and a paternal medaka genome, but they didn't survive early embryogenesis and gastrulation. This implies that Bouncer is both sufficient and necessary for species-specific egg-sperm recognition.

So, Bouncer only lets in sperm that will not disturb the internal harmony – hence the use of "bouncer" to echo the name given to security guards at the entrance of nightclubs and bars, whose job is to deny access to troublemakers. This is the first time scientists have discovered a protein that is actually part of the species-specific barrier – but it does not act on its own. There must be a ligand on the sperm cell that binds specifically to Bouncer, and numerous other proteins will be involved in attracting the sperm to the micropyle and then barring the entrance to those who have arrived too late. Though there is still a lot to understand, species-specific studies like these are essential to grasp how species keep to themselves and how evolution is driven. They are, of course, also crucial in fertility studies and could inspire therapies to help people who seem unable to conceive. If Bouncer is necessary for sperm-egg binding, there is probably a human equivalent which, when deficient, may be one of the causes of human infertility. The hunt is surely underway.

Cross-references to UniProt

Protein Bouncer, *Danio rerio* (Zebrafish): P0DPQ9
Protein Bouncer, *Oryzias latipes* (Japanese rice fish): H2LID1

References

1. Herberg S., Gert K.R., Schleiffer A., Pauli A.
The Ly6/uPAR protein Bouncer is necessary and sufficient for species-specific fertilization
Science 361:1029-1033(2018)
PMID: 30190407

2. Vasilyeva N.A., Loktyushov E.V., Bychkov M.L., Shenkarev Z.O., Lyukmanova E.N.
Three-finger proteins from the LY6/uPAR family : Functional diversity within one structural motif
Biochemistry (Moscow) 82:1702-1715(2017)
PMID: 29523067

proteinspotlight

Swiss Institute of Bioinformatics

Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

another kind of harmony

Vivienne Baillie Gerritsen

Would Nature not tend instinctively towards symmetry? In our eyes, symmetry often spells equilibrium, a source of beauty. Consider the work of architects, or engineers. Houses, skyscrapers, bridges and dams are usually symmetric which is not only, and in a mysterious way, emotionally reassuring but also in keeping with the laws of physics. In the same vein, a face that strikes us as being attractive is a face whose sides echo one another. Anything that drifts from these unsaid boundaries strikes us as being odd, if not ugly: think of the Elephant Man. Despite this, life is defined by an underlying lack of symmetry. In fact, Nature frequently seeks a way to break symmetry. Take humans: our heart is not symmetric, neither is the arrangement of our organs inside us, and our brain hemispheres are involved in very different aspects of intelligence. Why has Nature chosen asymmetry? And how does it occur in the first place? The field of research is relatively recent and the answers to these questions are still far from satisfactory. However, we do know that a certain form of myosin, known as myosin 1D, is directly involved in paving the paths of asymmetry in zebrafish.

The Great Wave

by Katsushika Hokusai (1760 -1849)

Many forms of asymmetry are immediately apparent: take the claws of lobsters and crabs, or the coiling of snail shells or pigs' tails. Now slice a human in two, from the head down. At a first glance, it all seems quite symmetric: two eyes, two ears, two arms, two legs. But peel away the skin and asymmetry becomes obvious: our heart is placed a little to the left, our spleen seems to have been squeezed into a dark corner and our liver shifted to one side. This is exactly what may have happened: our guts are rather long and awkward, and to fit them inside us in the most efficient way would have required the relocation of a few organs from our middle. This, however, only explains organisational asymmetry but it does not explain functional asymmetry – such as the brain – or morphological asymmetry, such as the heart, and even behavioural asymmetry that defines left- or right-handedness. There must be a deeper-lying reason.

There are various theories. One is molecular chirality. Chirality is a particular form of asymmetry. Your left hand and right hand are chiral – that is to say, though they are mirror images of one another, you cannot superimpose them. It so happens that molecules at the very basis of life are also chiral, like all but one amino acid and the double helix. Could it be that chirality, already so deeply rooted in an organism's biology, forms a sort of building block to greater asymmetry? Perhaps so. A cell's cytoskeleton can also be chiral and thus act as a scaffolding for asymmetry as it places some parts of a cell on one side rather than on the other, creating what is known as lateral left-right (LR) asymmetry. In fact, scientists are beginning to realise, to their surprise, that LR asymmetry seems to occur differently in different organisms. In the very early stages of zebrafish development, for instance, a unique fluid flow is generated, literally twisting original symmetry into LR asymmetry. This singular flow arises within an organ known as Kupffer's vesicle (KV) in the zebrafish embryo.

The German anatomist Karl Wilhelm von Kupffer was the first to describe the vesicle in the 1800s. This proto-organ is spherical in shape and formed by about 24 dorsal forerunner cells. The inside

lumen is filled with fluid into which protrude cilia. The cilia are oriented in such a way in the lumen that their beat forms a sort of liquid cue that brings about downstream lateral LR asymmetry. So, in this case, cilia orientation seems to be at the heart of LR asymmetry. Cilia are ubiquitous in organisms and have many diverse roles: they line the Fallopian tube to carry ovules, some plant gametes use them for mobility and they even amplify sound in the inner ear. What is the particular property of KV cilia which interferes with an embryo's symmetry?

Myosin. Myosins are actin-binding proteins that hydrolyse ATP to produce power – such as in muscle tissue contraction or the beat of cilia for example. Tom Pollard and Ed Korn were the first to describe a myosin protein extracted from an amoeba in the 1970s – although it had already been discovered in muscle in the 1900s, hence its name. Since then, over 30 different myosin isoforms have been identified whose structure and function have been well-conserved over time and across species. Their structure? They look like golf clubs, with a large globular head on one end from which protrudes a long chain. Typically, one myosin molecule is composed of two "golf clubs", so that two globular heads appear side by side at one extremity and their chains intertwine to form a tail. It is the globular heads of myosin which bind to actin and hydrolyse ATP to produce the power stroke.

A myosin named myo1D is directly involved in shaping KVs as well as organising left-right morphogenesis and laterality in zebrafish. How? Firstly, it ensures the correct trafficking and delivery of vacuoles into the KV. Once there, the vacuoles evacuate fluid into the lumen as they pump up the KV so it expands like a balloon. Secondly, the singular spherical shape of Kupffer's vesicle lends a unique beat to the cilia creating a counter-clockwise fluid flow. This particular flow then mediates asymmetric gene expression, which finally brings about LR asymmetry in the growing embryo. It is an astonishing biological process and the reason KVs have also been coined LR organizers.

Why, you wonder, would Nature consider asymmetry in the first place? Is it all just a question of survival and adaptation? Breaking symmetry seems to have increased the evolutionary fitness of species over time. Certainly, left-right orientations are important for many organs – you cannot replace one side of our brain, or one side of our heart, with the other. There are, however, rare cases of *situs inversus* where people have their heart on their right side and the location of their abdominal organs is reversed. However, because the relationship between each organ is untouched, people suffering from this condition are otherwise perfectly healthy individuals. The fact that chirality already exists at the molecular level is also a sign that asymmetry is not to be taken lightly. This said, to date, still very little is known about this fascinating field of research. Though, in mammals, myo1D seems to have little to do with LR asymmetry, getting to know the protein more intimately, as well as the proteins it interacts with, should give biologists and medical researchers valuable insight. Certainly, Nature continues to surprise us by choosing – and very early on – to shed a certain kind of balance, for another.

Cross-references to UniProt

Unconventional myosin-1D, *Danio rerio* (Zebrafish): E7F9L8

References

1. Lebreton G., Géminard C., Lapraz F. et al.
Molecular to organismal chirality is induced by the conserved myosin 1D
Science 362:949-952(2018)
PMID: 30467170

2. Juan T., Géminard C., Coutelis J.-B. et al.
Myosin1D is an evolutionarily conserved regulator of animal left-right asymmetry
Nature Communications 9:1942(2018)
PMID: 29769531

3. Blum M., Ott T.
Animal left-right asymmetry
Current Biology 28 :R293-R305(2018)
https://doi.org/10.1016/j.cub.2018.02.073

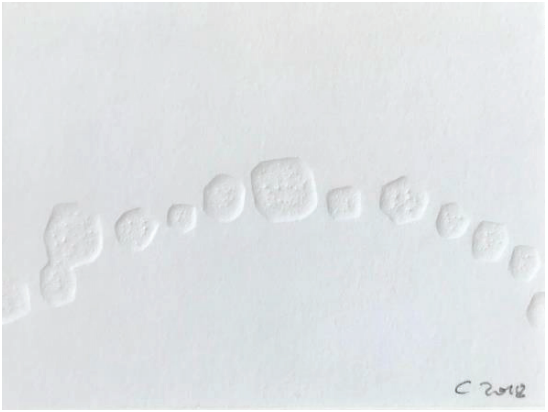
proteinspotlight

Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

a sense of direction

Vivienne Baillie Gerritsen

Survival depends on cues, mobility and a medium to evolve in. Cues – such as scents, sounds or colours for example – will attract organisms towards food, mating grounds and an environment in which they feel protected and are happy to stay. Thanks to them, organisms usually head off in a direction they expect will be to their advantage, using the means of locomotion they have, to cross all sorts of media. Some organisms make use of additional systems to reach a given destination. An example? Magnetotactic bacteria have learned to use the Earth's magnetic field as a speedy highway to travel to nutrients of interest. They do this by way of minute iron-rich pouches – or magnetosomes – that are aligned along their middle and act much like a compass would. Many macromolecules are required to model this fascinating system. One of particular interest is a protein known as MamB which is at the heart of magnetosome initiation. Magnetosomes have also long intrigued those behind the microbiology blog *Small Things Considered*, and this article echoes a lovely piece on magnetotactic bacteria and their navigation skills written by Christoph Weigel earlier this week, and whose artwork is shown below.



frottage on paper by Christoph Weigel

Courtesy of the microbiologist/artist

Magnetotactic bacteria proliferate in freshwater sediments but also in brackish, marine and hypersaline environments all over the world. In the late 1950s, an Italian doctor, Salvatore Bellini, observed for the first time a group of bacteria which all seemed to be heading in the same direction: northwards. He termed them magnetosensitive bacteria and described what he had seen in an article published in 1963. Though a short summary was translated into several languages, including English, his findings were given little to no attention at all. Seemingly unaware of Bellini's findings, in the mid-1970s Richard Blakemore, then a graduate student in microbiology, also observed bacteria that were using the Earth's magnetic field to navigate. Thanks to electron microscopy, Blackmore was able to go a step further and distinguish the chain of magnetosomes within the bacteria. Blakemore coined the bacteria's mode of transit 'magnetotaxis' – which is the term used ever since.

Why use the Earth's magnetic field to direct movement? Speed and efficiency may well be the answer. If you use our planet's magnetic field as a multi-lane highway, you are travelling in one plane only. Instead of dashing off in random directions – which costs energy – you follow a single direction. Fair enough, but why would this lead you to where you need to go, you may ask. How do bacteria know that what they are looking for lies along the Earth's magnetic field? It is a cunning system. Magnetotactic bacteria such as the spiral-shaped *Magnetospirillum gryphiswaldense* propel themselves forward by way of two flagella – one at each of their ends. They use very little oxygen and "know" that the best place for them to find nutrients is in a zone known as the oxic-anoxic transition zone (OATZ) of water. This particular zone is sandwiched between two layers of water: an upper oxygen-rich (oxic) layer supplied by the atmosphere and aqueous photosynthesis, and a lower oxygen-free (anoxic) layer. The anoxic layer releases all sorts of fermentation products that drift into OATZ while oxygen is fed into it thanks to the oxic layer. As a result, OATZ is an ideal feeding place for *Magnetospirillum*, and one of the fastest and most convenient ways of getting there is by hitching a ride on the geomagnetic field.

Magnetotactic bacteria use their magnetosomes to do just this. Scientists have been looking into these singular vesicles since the 1970s and, though knowledge is still scant, they are beginning to understand magnetosome biogenesis at the molecular scale. Several magnetosomes – up to 60 in *Magnetospirillum* – form simultaneously, and independently, by invagination of the bacterial membrane. As invagination occurs, a host of magnetosome-specific proteins, known as Mam proteins, rush to the site. Once formed, the vesicles align along the bacteria's middle held stable by an underlying actin-like cytoskeleton. Meanwhile, membrane-bound magnetite (Fe_3O_4) crystals nucleate and mature inside each magnetosome – typically one perfect cubo-octahedral crystal per magnetosome in *Magnetospirillum*. From the bending of the bacterial membrane and its invagination to magnetosome formation, iron transport, the shape, size and number of magnetite crystals and the final alignment of the magnetosomes, every stage is controlled by different Mam proteins, of which there are many.

One essential protein for magnetosome initiation in particular is MamB. MamB seems to play an important part in two stages of magnetosome formation: membrane invagination and magnetite nucleation. In the process of invagination, MamB may act as a landmark protein by forming a complex with other Mam proteins which, together, bend the membrane by sheer physical force to form a vesicle. Concomitantly, though iron ions may float passively from the cytoplasm into the forming magnetosomes, there is reason to believe that MamB is actively involved in their transport and triggers off magnetite nucleation. It is difficult, however, to talk about mamB without mentioning MamM, which has a stabilizing effect on MamB and is necessary for its correct function.

MamB and MamM belong to the cation diffusion facilitator (CDF) family. CDFs are found throughout all domains of life where they transport divalent metal cations from the cytoplasm into the extracellular space or indeed into intracellular compartments like magnetosomes. MamB and MamM are transmembrane proteins and share very similar structures; V-shaped, they probably form homodimers and interact with each other via their cytoplasmic regions. Both are needed for magnetosome invagination and magnetite nucleation although it seems that MamB has a more central role as MamM keeps it stable.

Magnetite biocrystals are very pure as opposed to their geocrystal equivalents which can harbour foreign inclusions. As a consequence, applications of magnetotactic bacteria and magnetosomes are currently being considered to resolve medical and environmental issues. Magnetosomes would be more effective contrast agents than their current chemical counterparts for magnetic resonance imaging (MRI) for instance. They could also be tailored to target malignant cells to which a magnetic field is applied thus inducing heat induction and cell disruption. On the environmental front, besides iron, magnetotactic bacteria are also capable of absorbing other metals such as cadmium, selenium, cobalt, manganese and copper. Engineered, magnetotactic bacteria could then be used to clean up toxic waste. Certainly, the applications seem promising and endless, but it is this faculty nature has of devising such sophisticated and unlikely ways of surviving which is the most intriguing. Magnetotactic bacteria have been around since the Cretaceous at least and are thought by a few to even exist on Mars but it is this capacity some living beings have of knowing how to sculpt inorganic crystals which is thought-provoking.

Cross-references to UniProt

Magnetosome protein MamB, *Magnetospirillum gryphiswaldense*: V6F510
Magnetosome protein MamM, *Magnetospirillum gryphiswaldense*: V6F235

References

1. Uebe R., Keren-Khadmy N., Zeytuni N. *et al*.
The dual role of MamB in magnetosome membrane assembly and magnetite biomineralization
Molecular Microbiology 107:542-557(2018)
PMID: 29243866

2. Uebe R., Schüler D.
Magnetosome biogenesis in magnetotactic bacteria
Nature Reviews Microbiology 14:621-637(2016)
PMID: 27620945

3. Dieudonné A., Pignol D., Prévéral S.
Magnetosomes : biogenic iron nanpoarticles produced by environmental bacteria
Applied Microbiology and Biotechnology 103:3637-3649(2019)
PMID: 30903215

proteinspotlight

Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

sting

Vivienne Baillie Gerritsen

Venom has a language of its own. The recurring message is not a nice one, and usually expresses one thing: back off. Certain animals use venom – a cocktail of molecules – to ward off predators or, at the very least, to divert oncoming danger. We all know what a wasp's sting is like and many of us may have felt the sting of a jellyfish, or perhaps even the bite of a snake. It is a painful experience. To what end? The reason is twofold: one, we at once recoil from the animal that has just caused pain and two, our body is instantly told where it hurts. Concomitantly, the animal takes flight while our body attends to our wound. The feeling of pain itself is caused by the opening and closing of minute channels that riddle the membranes of our nerve cells just under our skin. This gives rise to pain signals that originate at the location of the sting, or bite, and are relayed to our brain. Understanding how pain occurs on the molecular plane helps scientists find ways of designing pain relievers. However, more often than not, pain is usually accompanied by swelling which has a protective role. So we face a conundrum: how do you relieve pain while preserving inflammation? One particular scorpion toxin, the Black Rock scorpion toxin known as the wasabi receptor toxin or WaTx, may well provide an answer.

An artwork by Asuka Hishiki

Courtesy of the artist
(drawing of a wasabi plant)

Why call a scorpion toxin a wasabi receptor toxin? One belongs to an animal, the other to a plant, neither of which bears any obvious resemblance to the other. It is because WaTx binds to an animal receptor known as Transient Receptor Potential A1, or TRPA1[1]. TRPA1s are membrane channels known to be stimulated by molecules – allyl isothiocyanates (AITCs) – that are typically found in wasabi but also in mustard, onions, and ginger for instance and bring about that disagreeable and often painful sensation inside our nose, or sometimes make us cry. This is why TRPA1s are also known as wasabi receptors.

Scorpion toxin causes pain. Fair enough. But why do natural products such as mustard and wasabi become a painful – albeit controlled – experience on our palate? The thing is, although many of us enjoy the savoury kick lent by these plants on our sensory system, AITCs probably originally evolved as a means of defence against herbivores. Besides humans, few animals would take an interest in a slice of ginger or wasabi. WaTx binds to the same receptors as AITCs – could it be that WaTx is used to ward off predators in a similar way that plant defensins are? And if we go one step further: could scorpion venom have evolved by the diversion of scorpion defensins, themselves part of a primitive immune system?

TRPA1s, or wasabi receptors, are embedded in the membranes of sensory nerve endings throughout our body and literally act as alarms against chemical irritants that can cause considerable damage to cells – like cigarette smoke and noxious gases, or irritants that are far

less harmful such as those found in wasabi, mustard or ginger. The scorpion wasabi toxin binds to wasabi receptors on their intracellular region and at a particular location that has been called the "allosteric nexus". Thanks to an unusual stretch of amino acids, WaTx can simply drift across the nerve cell membrane to reach its destination – no need for endocytosis or any other sophisticated means of transport. Such passive diffusion is particularly attractive for designing therapeutic drugs.

Structurally, WaTx looks like a rigid and compact helical hairpin which is kept stable by way of two disulphide bonds, and thus resembles many other scorpion toxins. However, its mode of action is very different: as already mentioned, unlike other toxins, WaTx is able to diffuse passively across cell membranes, and then binds to the allosteric nexus of the wasabi receptor. This is a surprising finding that prompts scientists to believe that wasabi toxins and scorpion defensins may share a common ancestor. In fact, when comparing their sequences, scientists found that only one amino-acid change was needed to go from one to the other. Such a mutation would represent a rather convenient way for scorpions to widen their hunting ground and colonise new niches as they crawled out of the water about 430 million years ago using venom to ward off predators, or indeed kill prey.

When WaTx wedges itself into TRPA1, the channel undergoes a conformational change causing it to open up for an uncharacteristically long period. Calcium levels in particular are modified; the levels inside the cells are sufficient to send pain signals to the brain, while being too low, however, to cause any swelling of the tissue. Inflammation is believed to be a response to tissue damage and used by organisms to isolate tissue which needs to be repaired. In short, WaTx hurts but does not signal to our body that part of us needs to be protected and tended to. This is precisely one of the key processes that becomes dysregulated in chronic pain.

WaTx receptors only exist in mammals and WaTx itself, in all likelihood, is purely defensive. WaTx seems an ideal candidate for developing novel therapeutic drugs such as analgesics. Not only is it able to drift through cell membranes unhindered but – like all AITCs – it binds to a very specific cytoplasmic region on wasabi receptors and, most importantly: it causes no swelling. This means that WaTx is an ideal candidate on which drug designers could build to develop pain relievers while not abolishing the essential protective role played by inflammation. Why WaTx causes no swelling of tissue is not understood. It may be strategic. Perhaps it demands less energy than the dual signal of pain and inflammation signal does. Who knows. For millennia, in human culture, scorpions have either been the embodiment of evil or indeed the incarnation of protection against evil so, in a certain way, they continue to give rise to opposing signals.

[1] Read *Protein Spotlight issue 82: The power behind pain*

Cross-references to UniProt

Wasabi receptor toxin, *Urodacus manicatus* (Black rock scorpion): C0HLG4

References

1. Lin King J.V., Emrick J.J., Kelly M.J.S., Herzig V., King G.F., Medzihradszky K.F., Julius D.
A cell-penetrating scorpion toxin enables mode-specific modulation of TRPA1 and pain
Cell 178: 1362-1374(2019)
PMID: 31447178

2. Green D., Dong X.
A pungent and painful toxin
Cell 178:1279-1281(2019)
PMID: 31474364

proteinspotlight

Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

EMBnet.journal
Bioinformatics in Action

ISSN 1023-4144

DEAR READER,

If you have any comments or suggestions regarding this journal we would be very glad to hear from you. If you have a tip you feel we can publish then please let us know. Before submitting your contribution read the "Authors guidelines"[1] and send your manuscript and supplementary files using our on-line submission system[2].

Past issues are available as PDF files from the web archive[3].

Visit EMBnet website for more information: www.journal.embnet.org

EMBNET.JOURNAL EXECUTIVE EDITORIAL BOARD

[1]http://journal.embnet.org/index.php/embnetjournal/about/submissions#authorGuidelines
[2]http://journal.embnet.org/index.php/embnetjournal/author/submit
[3]http://journal.embnet.org/index.php/embnetjournal/issue/archive

PUBLISHER

EMBnet Stichting p/a
CMBI Radboud University
Nijmegen Medical Centre
6581 GB Nijmegen
Th e Netherlands

Email: erik.bongcam@slu.se
Tel: +46-18-67 21 21