Deliverable D7.2

Project Title:	World-wide E-infrastructure for structural biology	
Project Acronym:	West-Life	
Grant agreement no.:	**675858**	
Deliverable title:	ProteinCCD with new analysis options	
WP No.	7	
Lead Beneficiary:	1: NKI AVL	
WP Title	Joint research	
Contractual delivery date:	Month 11	
Actual delivery date:	Month 11	
WP leader:	A. PERRAKIS	NKI AVL
Contributing partners:	N/A	

Deliverable written by Anastassis Perrakis

Contents

1 Executive summary

A common bottleneck and a frequent task in structural biology and structural genomics is the production of soluble protein in amounts suitable for structural studies. The Protein Crystallographic Construct Design [1] ProteinCCD (https://xtal.nki.nl/ccd/) software aims to increase the efficiency of researchers performing that task, by facilitating the design of several truncation constructs of a protein under investigation. ProteinCCD functions as a meta server that collects information from (web-based) external software that predicts from sequence secondary structure, disorder, coiled coils, transmembrane segments, domains and domain linkers. A clear and concise view of the protein sequence annotated with the prediction results allows users to interactively choose possible starts and ends for suitable protein constructs. ProteinCCD can help designing the primers needed for PCR amplification of all constructs, as the required user input is the DNA and not the protein sequence. ProteinCCD outputs a comprehensive view of all constructs and all primers needed for bookkeeping and/or ordering of the designed primers.

The functionality of ProteinCCD is now extended: we implemented it in a new computational platform allowing a more interactive and efficient interface to the user, and we are providing new analysis options, including: parallel processing of server requests, more efficient interface for construct design, more cloning methods, an extended collection of existing vectors, local execution of some algorithms for improving response time, new servers for meta-analysis, easy bookkeeping, and better data security.

2 Project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives:

No.	Objective	Yes	No
1	**Provide analysis solutions for the different Structural Biology approaches**	X	
2	**Provide automated pipelines to handle multi-technique datasets in an integrative manner**		X
3	**Provide integrated data management for single and multi-technique projects, based on existing e-infrastructure**	X	
4	**Foster best practices, collaboration and training of end users**	X	

3 Detailed report on the deliverable

3.1 Background

The production of soluble proteins in amounts suitable for structural studies has been a common bottleneck in structural biology and structural genomics alike. Cloning techniques are high-throughput, inexpensive and compatible with robotic implementations, allowing parallel construction of tens of expression constructs for each protein under study: that is a standard practice in many labs. Expression constructs can be designed based on experimental information, but also by computationally. Despite significant progress in sequence analysis, there is no definitive method of choice and submitting many queries to different servers, and subsequent collection of the results of different methods, is the normal laboratory practice. The researcher then typically decides what are promising domain boundaries for the protein in hand, and designs oligonucleotides to be used for PCR-based amplification of all these fragments. At this stage a trivial but time consuming additional bottleneck is encountered: the protein-based analysis has to be transformed back to the DNA sequence. Although the task is by all means trivial, it is time consuming and error prone, since the direct mapping between protein and DNA sequence is lost in the analysis step.

The ProteinCCD (https://xtal.nki.nl/ccd/), introduced for the first time in 2008, is a meta-server to cater for the needs of the above tasks. It automates sequence analysis, and provides interaction with the user for the optimal design of protein constructs that are good candidates for structural analysis. The collection of sequence analysis tools, includes servers for the prediction of secondary structure, disorder, coiled coils, transmembrane segments, domains and domain linkers. A clear and concise view of the protein sequence annotated with the prediction results allows users to interactively choose possible starts and ends for suitable protein constructs. ProteinCCD can help designing the primers needed for PCR amplification of all constructs, as the required user input is the DNA and not the protein sequence. ProteinCCD outputs a comprehensive view of all constructs and all primers needed for bookkeeping and/or ordering of the designed primers.

3.2 Current work

3.2.1 Redesigning software to allow new analysis methods

The application was originally delivered as a Java applet to the client, and half a decade after its conception remains a popular tool for many scientists, with 250-350 users per month making use of the online service. The Java approach however, has security issues and Java applets cannot be run on all devices. To overcome these limitations, we re-implemented ProteinCCD as a web application. The new requirements were based on user feedback. The Python Flask backend uses Biopython. The frontend relies on Bootstrap and Javascript. The new implementation eliminates security concerns, makes the application available to any device able to run a web browser, and can be extended with new functionality easily. Other West-Life work packages are making similar technology choices.

3.2.1 A new interface

The user interface has been given a modern and more functional look, emphasising on the most commonly used features based on user feedback. Elimination of all Java dependencies eliminates loading times and the needs for certificates validation, improving significantly the user experience. Screenshots of the design are shown in the Poster presentation in Appendix 1, and can also be viewed interactively in the server at http://www/nki.nl/ccd.

3.2.1 New analysis

The new design allowed several analysis options to be implemented as planned. Notably this also allowed progress towards Milestone 24. New analysis options are as follows:

- We have enabled parallel processing of server requests. While data are being collected, the software allows the user to use the current information and interact with the software, without waiting time.
- A more efficient interface for construct design is available. Specifically an improved primer design interface is implemented. Previously, the software was geared towards the NKI Ligation Independent Cloning (LIC) system [2]. The new options allow the user to extend the plasmid vectors to other methods (for a recent review see [3]), such as restriction based cloning, and implements custom laboratory collections for target plasmids.

- The internal database is checked on the fly for vectors with compatible ligation sequences that are often used in other plasmids in the database, increasing user awareness of the experimental possibilities.
- ProteinCCD now provides better support for restriction based cloning, by including a small user database of most common restriction enzymes.
- Local execution of some algorithms improves response time.
- New servers for the meta-analysis have been added; meta-servers now include HNN, MLR, DPM, PREDATOR, IUPRED, GLOBPLOT and COILS (https://xtal.nki.nl/ccd/credits).
- Easy saving of predictions, primers and resulting peptides has been implemented to enable efficient book-keeping.

References cited

[1] Mooij WT, Mitsiki E, Perrakis A. ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. Nucleic Acids Res. 2009 Jul;37:W402-5. doi: 10.1093/nar/gkp256.

[2] Luna-Vargas MP, Christodoulou E, Alfieri A, van Dijk WJ, Stadnik M, Hibbert RG, Sahtoe DD, Clerici M, Marco VD, Littler D, Celie PH, Sixma TK, Perrakis A. J Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases. Struct Biol. 2011 Aug;175(2):113-9. doi: 10.1016/j.jsb.2011.03.017.

[3] Recombinant cloning strategies for protein expression. Celie PH, Parret AH, Perrakis A. Curr Opin Struct Biol. 2016 Jun;38:145-54. doi: 10.1016/j.sbi.2016.06.010. Epub 2016 Jul 5. Review.

Appendix 1: A poster for the presentation of ProteinCCD[2]

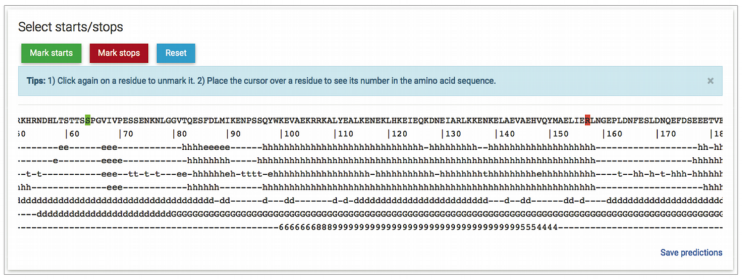
proteinCCD²: Constructs for the People

Giorgos Damaskos, Andrea Murachelli, Wouter G. Touw, Anastassis Perrakis

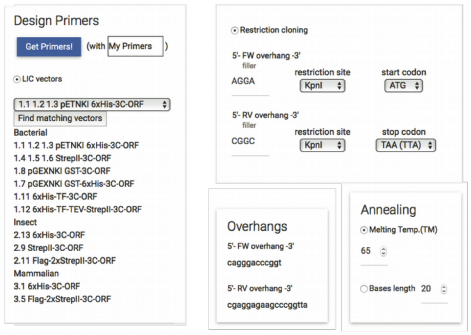
The Protein Crystallographic Construct Design (ProteinCCD) software aims to facilitate a common practice in structural biology, namely the design of several truncation constructs of a protein under investigation. ProteinCCD is a meta server that collects information from (web-based) external software that predicts from sequence secondary structure, disorder, coiled coils, transmembrane segments, domains and domain linkers. A clear and concise view of the protein sequence annotated with the prediction results allows users to interactively choose possible starts and ends for suitable protein constructs. ProteinCCD can help designing the primers needed for PCR amplification of all constructs, as the required user input is the DNA and not the protein sequence. ProteinCCD outputs a comprehensive view of all constructs and all primers needed for bookkeeping and/or ordering of the designed primers. The server is available at htttp://xtal.nki.nl/ccd.

Input

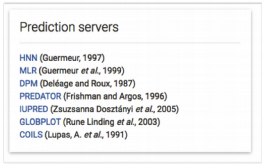
Meta-server output and interaction with user



Options for primer design



Servers



Visit us!

Primers for PCR and resulting polypeptides

Technical data

Half a decade after its conception CCD [1] remains a popular tool for many scientists, with 250-350 users per month. The application was originally delivered as a Java applet to the client. The Java approach however, has security issues and Java applets cannot be run on all devices. To overcome these limitations, we re-implemented ProteinCCD as a web application. The new requirements were based on user feedback. The Python Flask backend uses Biopython. The frontend relies on Bootstrap and Javascript. The new implementation eliminates security concerns, makes the application available to any device able to run a web browser, and can be extended with new functionality easily.

New features

An example of the new functionality is the improved primer design interface.
Previously, the software was geared towards the NKI Ligation Independent Cloning system [2]. The new implementation allows the user to extend the plasmid vectors to other methods (for a recent review see [3]), and implement custom laboratory collections for target plasmids. The reimplementation also automatically checks the internal database for vectors with compatible ligation sequences that are often used in other plasmids in the database, increasing user awareness of the experimental possibilities. Finally, ProteinCCD now provides better support for restriction based cloning, by including a small user database of most common restriction enzymes.
In addition some of the analysis software now runs locally to increase response speed, and the requests to external servers is handled in parallel, also increasing efficiency.

Future plans

An additional feature under implementation is the calculation or prediction of physico-chemical properties of the designed protein constructs (e.g. molecular weight, isoelectric point, absorbance coefficient, and in the future solubility and crystallisation prediction scores).
We will also implement the generation of plasmid maps of the new constructs. The format of the plasmid maps will be compatible with all common cloning software.
Finally, the system might provide a valuable data source for automating construct design when users make available both the designed constructs and the validated results.

References

1. *Mooij WT, Mitsiki E, Perrakis A.*
ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments.
Nucleic Acids Research
2009 Jul;37:W402-5. doi: 10.1093/nar/gkp256.

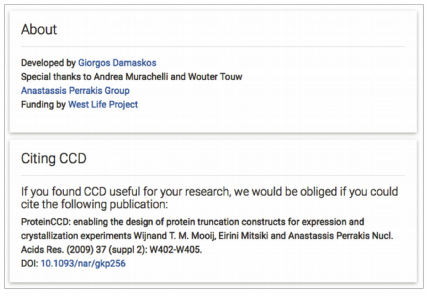
2. *Luna-Vargas MP, Christodoulou E, Alfieri A, van Dijk WJ, Stadnik M, Hibbert RG, Sahtoe DD, Clerici M, Marco VD, Littler D, Celie PH, Sixma TK, Perrakis A. J*
Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases.
Journal Structural Biology
2011 Aug;175(2):113-9. doi: 10.1016/j.jsb.2011.03.017.

3. *Celie PH, Parret AH, Perrakis A.*
Recombinant cloning strategies for protein expression.
Current Opinion Structural Biology
2016 Jun;38:145-54. doi: 10.1016/j.sbi.2016.06.010. Epub 2016 Jul 5. Review.

Funding

West-Life

ANTONI VAN LEEUWENHOEK
NEDERLANDS KANKER INSTITUUT



Background information

This deliverable relates to WP7; background information on this WP as originally indicated in the description of work (DOW) is included below.

WP7 Title:
Lead:
Participants:

Work package number	7	**Start date or starting event:**	
Work package title	Joint research		
Activity Type	COORD		
Participant number	2		
Person-months per participant:	3		
Objectives			
Description of work and role of participants			

Deliverables

No.	Name	Due month
		2
		6
		9
		36
		48