brought to you by



VALUE ADDED WHEAT CRC PROJECT REPORT

Workshop:

Protein Structural Prediction for Mutagenesis

4th October 2002

Presented by: Dr Warren Kaplan

Date: October 2002

VAWCRC Report No: 26 Copy No: 1

(Not to be copied)

Value Added Wheat CRC has taken all reasonable care in preparing this publication. Value Added Wheat CRC expressly disclaims all and any liability to any person for any damage, loss or injury (including economic loss) arising from their use of, or reliance on, the contents of this publication.

Protein Structural Prediction for Mutagenesis



Presented by Dr. Warren Kaplan

Peter Wills Bioinformatics Centre

Garvan Institute of Medical'Research

A workshop sponsored by Value Added Wheat CRC Ltd. ABN 65 070 001 839 www.wheat-research.com.au

41



The Presenter:

Warren Kaplan, PhD

University of the Witwatersrand (Wits), Johannesburg, South Africa, 1998.

PhD thesis:

The conformational stability of a detoxification enzyme widely used as a fusion-protein affinity tag.

Research and Work Experience

Dr Kaplan has a strong background in bioinformatics development and training. His PhD involved the heterologous expression, purification and characterisation of a glutathione S-transferase. Solvent- and thermal-denaturation studies were used to monitor the unfolding/refolding of the enzyme via the following physico-chemical techniques: steady-state kinetics, spectroscopy (UV/visible and fluorescence, electrophoresis (urea- and thermal-gradient gel) and differential scanning calorimetry (DSC).

From 1998-1999 he was Unix system administrator and molecular modeller, Department of Biochemistry, Wits University. His work at Entigen (1999-2001, http://www.entigen.com) included the design and integration of the Molecular Modelling programs for BioNavigator (www.bionavigator.com). He also wrote the relevant help documentation for the Molecular Modelling programs, and designed the Molecular Modelling protocols in BioNavigator to serve as an introduction to users new to Molecular Modelling. He has also written a training manual for users to integrate programs into BioNavigator using the Ajax Command Definition Language. His programming experience included writing parsers in the Python programming language, and he has experience in UNIX shell programming. He was involved in Molecular Modelling related customer support and conducted very extensive testing of the latest BioNavigator system.

He is familiar with most of the protein sequence related bioinformatics software and am very familiar with the molecular modelling packages WHAT IF, ICM, MolMol, and other related software packages such as Swiss PDB Viewer (Deep View), Molscript, GRASP, RasMol and Ligplot. He has experience with setting up and installing software on SGI, SUN, and Linux machines.

Teaching Experience

He designed and ran the MSc biotechnology course offered in the department of biochemistry at Wits over a 4 year period, lectured to undergraduate students and ran workshops introducing users to the computer programs What If, GRASP and ICM.

While at Entigen he was involved in staff training of the Molecular Modelling programs in BioNavigator, and lectured to customers on using BioNavigator.



A workshop sponsored by Value Added Wheat CRC Ltd. ABN 65 070 001 839 www.wheat-research.com.au

A word of welcome!



Dear Student

Welcome to today's workshop on Protein Structural Prediction for Mutagenesis. At the beginning of this year Clare Johnson, from Value Added Wheat CRC Ltd., approached me with the idea of running today's series of lectures. From the outset our intention for this course has been to show you how enormously valuable three-dimensional structures of protein are. In so doing we hope to add value, and introduce a new way in which you go about designing experiments and interpreting results.

In order to do this, I have designed the course around two research papers: the first paper describes the structures of two bound proteins that were solved using X-ray crystallography. The second paper uses the structures in the first paper to build models of homologous proteins. Once the models are built the structures are examined and a decision is made to mutate 4 amino acids in the models. The paper then goes on to describe the observable effects of these mutations. My aim for today is to familiarise you with the structures of the first paper and then emulate the *'in silico'* experiments of the second paper. By doing this I hope that you will be able to apply these similar steps to the protein that you may be interested in.

The notes in this booklet include all the slides that I will be using. Browsing through them you will notice that they are brief with only keywords, and the occasional diagram. I hope that you will not be disappointed by their brevity, but we're in the business of learning about the three-dimensional structures of proteins and for most of my talks I will be using a protein structure viewer which will substitute for cumbersome, and non-interactive slides. The viewer we will be using is the Swiss PDB Viewer, also know as Deep View. The reasons for choosing Deep View are that it is fairly easy to use, it runs on most operating systems, it's free, and offers the most functionality of all the free modelling programs that I know. I do need to emphasise that I don't get paid to say nice things about it too, but what I want to stress is that this course is about proteins structures, not Deep View. What I hope you will learn from it is the kind of things a modelling program should be able to do, the detail on how to do it will be available from the manual of the modelling program that you eventually use.

~· ;

الوجيد

في د

I thought I should tell you a little about myself too – I am a protein chemist that spent most of my research involved in protein folding and protein structurefunction relationships. I spent some time as a molecular modeller before taking a job with the startup Bioinformatics company, Entigen that produced the BioNavigator system. Since March 2002 I have held the position of Bioinformatics Specialist in the newly created Peter Wills Bioinformatics Centre at the Garvan Institute, where I spend most of my time meeting very interesting people and doing exciting things, like what we will be doing today.

I hope you will find the day enjoyable and that it will assist you in achieving greater success in your research.

Warren Kaplan (PhD) w.kaplan@garvan.org.au Program:

Protein Structural Prediction for Mutagenesis

Friday 4 October 2002

9.30 - 10.00

Registration and coffee

10.00-11.00 Introduction to protein structure

Beginning with a primary sequence of a protein we will build up the secondary, tertiary and quaternary structures, and also describe the bonds that are used to build these structures. We will then introduce the repository of macromolecular biological structures, the Protein Data Bank (PDB).

11.00-11.20 Morning tea

11.20-12.20 Protein viewers and analysis tools

Using the SwissPDBViewer, Deep View, we will discuss the various methods used for displaying proteins and then describe fundamental tools for analysing them. These will include the measuring of distances, or angles, between atoms and also the use of a Ramachandran plot.

12.20-1.00 Lunch

1.00-2.00 In Silico Site-Directed Mutagenesis

Site-directed mutagenesis is a very powerful technique used in wet-lab biology, but blindly mutating amino acids is not only expensive, it can also be very disappointing if the wrong residues are mutated. In this section we will discuss considerations that should be made when using this technique, by emulating these experiments on a computer.

2.00-3.00 Homology Modelling and Threading

Very often no known structure exists for the protein sequence that you are interested in. Both homology modelling and threading are very powerful techniques that are used to build theoretical models that may provide us with the structural information we require. In homology modeling, the alignment of the sequence of interest with the sequence of a homologous protein, with a known structure, is crucial for correct model building. Using an example, we will do a homology modelling experiment, and will also discuss threading.

3.00-3.30

Afternoon tea and close



Recommended Reading

Misura, Kira MS, Scheller, Richard H & Weis William I (2000), Three-dimensional structure of the neuronal-Sec1-syntaxin 1a complex, **Nature** <u>404</u>: 355-361.

Kaupp, Maria; Wohlfart, Gerd; & Olkkonen, Vesa M (2002), Analysis of the Munc 18b-syntaxin binding interface: Use of a mutant Munc 18b to dissect the functions of syntaxins 2 and 3, **JBC Papers in Press, published August 26, 2002 as Manuscript M208315200**.















	Render image: O Let	fteye 🌘) Standard Solid 3D st	Right eye	
	Banda:	siowery	JOIN 3D 30	ays in Motion	
	bonas: Line Width	2.000	pixels		
ŕ	Radius	0.200	Å		
	H-bond radius	0.075	Å ⊡ Dott	ed H-bonds	
	Smoothness (113)	7			
	Atoms:				
	Radius			0.300 Å	
	Alpha Carbons size (100 %			
	Atoms Smoothness (7			
	Spacefilled atoms smoothness (113) 7				
	🗹 Show Atoms 🔲 same color as	bonds for	Catoms 🗹	and others	
	Keep atoms prop by 1 1 for H 1 5 for	ortions by N. 1.7 for C	multiplying at 1.4 for 0, 1.85	tom radius Sfor Sam	







Primary Structure

>1DN1:A SYNTAXIN BINDING PROTEIN 1

MAPIGLKAVVGEKIMHDVIKKVKKGEWKVLVVDQLSMRMLSSCCKMTDIMTEGITIVED INKRREPLPSLEAVYLITPSEKSVHSLISDFKDPPTAKYRAAHVFFTDSCPDALFNELVK SRAAKVIKTLTEINIAFLPYESQVYSLDSADSFQSFYSPHKAQMKNPILERLAEQIATLC ATLKEYPAVRYRGEYKDNALLAQLIQDKLDAYKADDPTMGEGPDKARSQLLILDRGFDPS SPVLHELTFQAMSYDLLPIENDVYKYETSGIGEARVKEVLLDEDDDLWIALRHKHIAEVS QEVTRSLKDFSSSKRMNTGEKTTMRDLSQMLKKMPQYQKELSKYSTHLHLAEDCMKHYQG TVDKLCRVEQDLAMGTDAEGEKIKDPMRAIVPILLDANVSTYDKIRIILLYIFLKNGITE ENLNKLIQHAQIPPEDSEIITNMAHLGVPIVTDSTLRRRSKPERKERISEQTYQLSRWTP IKDIMEDTIEDKLDTKHYPYISTRSSASFSTTAVSARYGHWHKNKAPGEYRSGPRLIIF ILGGVSLNEMRCAYEVTQANGKWEVLIGSTHILTPQKLLDTLKKLNKTDEEISS

>1DN1:B SYNTAXIN 1A

CRC

MKDRTQELRTAKDSDDDDDVTVTVDRDRFMDEFFEQVEEIRGFIDKIAENVEEVKRKHSA ILASPNPDEKTKEELEELMSDIKKTANKVRSKLKSIEQSIEQEEGLNRSSADLRIRKTQH STLSRKFVEVMSEYNATQSDYRERCKGRIQRQLEITGRTTTSEELEDMLESGNPAIFASG IIMDSSISKQALSEIETRHSEIIKLENSIRELHDMFMDMAMLVESQGEMIDRIEYNVEHA VDYVERAVSDTKKAVKYQSKARRKKIM











VUHEAT CRC	Anatomy of a PDB file
	Header
	PDB code
	Remark
	Resolution
	Remark 350
	Sequence
	SHEET
	SSBOND
	CRYST1
	ATOM
	TER
	HETATM
	CONNECT
	END



















While the high precision structures required for detailed studies of protein-ligand interaction can only be obtained experimentally, theoretical protein modelling provides the molecular biologists with "low-resolution" models which hold enough essential information about the spatial arrangement of important residues to guide the design of experiments. The rational design of many sitedirected mutagenesis experiments could therefore be improved if more of these "low-resolution" theoretical model structures were available.

N. Guex and M. Peitsch





CRC

Aligning the model sequence with the template sequence

Align core regions and ignore the non-conserved loops

Building the Model

Framework Construction Loop insertion using 'spare parts' (only Ca) Backbone Completion (C=O and N) -library of pentapeptides < 2.0 Angstroms Adding side Chains -table of rotamers -check for bumps Model Refinement -energy minimization -molecular dynamic simulation (http://www.chemaxon.com/marvin/doc/example-view3.2.html)



Assessing the quality of the Model

General Considerations Mistake if not within 0.5 Angstroms RMSD of the control structure (<u>http://www.chemaxon.com/marvin/doc/example-</u> view3.2.html) Deviation of stereochemical values (bond lengths and angles) WhatCheck

Causes of Errors Wrong Sequence Rubbish Template Bad Alignment





Display Model

-confidence factor Secondary Structure Succession -trace the molecule

Ramachandran Plot

Only GLY and ALA in the forbidden regions

Select

-amino acids making clashes

- -amino acids making clashes with the backbone
- -sidechains lacking proper Hydrogen Bonds

Display

- -sidechains even when backbone is hidden
- -colour: type
- -select group kind: Nonpolar see 1PRC
- -select group kind: Acidic
- -select group kind: Basic











BioLateral Bioinformatics Courses:

book online at http://www.biolateral.com.au

Protein Structure Bioinformatics

This course will provide training in molecular modelling and is intended for biologists who would like to build three-dimensional protein structures from a protein sequence of interest. By using the Swiss Modelling server and its dedicated molecular modelling program, Swiss PDB Viewer (Deep View), the course will cover a basic introduction to protein structure and tools used for the analysis of macromolecular structures. Databases related to protein structure will be used do comparative homology modelling and threading. Assessing the model's quality will also be addressed. There are no pre-requisites for the course.

Microarray Bioinformatics

This course will cover subjects including Microarray Technologies, Design of Microarray Experiments, Image Analysis, Normalisation, and Interpretation of Results. A range of bioinformatics tools will be used in the course. The course will review some of the more popular programs in the field and introduce the statistical programming language "R" as a flexible tool for all stages in the analysis of array data. There are no pre-requisites for the course.

Bioinformatics Computer Systems

This course will cover the technical aspects of running of UNIX® like workstations and servers for biologists and bioinformaticians, providing an understanding of the power and wealth of bioinformatics applications available under UNIX® both for the desktop and as a shared server. The course will cover from a practical viewpoint how to setup and maintain a UNIX® like systems and provide basics services such as shared network file systems, printers, web servers and the installation of bioinformatics tools. Participants are encouraged to bring their own laptops or desktop machines to install a UNIX® like operating system or use the computers provided by BioLateral. There are no pre-requisites for the course.

Bioinformatics Programming

The popular and easy programming language "Python" will be used to introduce the basics of programming for bioinformatics applications. There will be an emphasis on biological applications, including wrapping of unix programs and tools found on web sites. The course will introduce participants to the technical aspects of bioinformatics software development and will "demystify" the bioinformatics software development process. Suitable for people who have little or no programming experience. Participants are encouraged to bring their own problems and projects to work on throughout the course.

Bioinformatics Web Services

This course is designed for bioinformatics researchers who need to publish databases or make software available on the VWW or who repetitively use other WWW sites and want to learn techniques for "wrapping" these sites so that they can be accessed automatically. The course covers how to master technologies of web serving, HTML authoring, web forms (POST and GET methods), and running applications on WWW sites using CGI. The course will reviews the use of the Unix Shell and Python scripts to integrate remote web sites. Participants will learn how to set up local web services and write CGI scripts to serve web contents. The course will focus on on biological data and problems. Pre-requisites for this course is Python programming.

Bioinformatics Primer

A brief introduction to the many aspects of bioinformatics required by researchers and / or educators. Topics touched on will include Bioinformatics software and databanks (Biological Databanks, Molecular Sequence Annotation, Using the Complete Human Genome Sequence, Protein Structure Prediction and Modelling, Phylogenetic analysis) and Bioinformatics development (Programming, Integrating Remote Bioinformatics Services, Relational Database Systems for bioinformatics, Developing Web Services, Choosing and Upgrading Computer Hardware, Administration of bioinformatics Workstations and Installing Bioinformatics Tools). There are no pre-requisites for the course.

Bioinformatics Databases

5

This course will explain the problems involved with the use of various databases systems and why Relational Database Management Systems can help in data storage and querying. The example RDBM's used is the freely available and powerful MySQL system. The aim is to show basic operation of a RDBMS including installation, design and creation of databases, and learning the standard Structured Query Language (SQL) in order to query the database. There are no pre-requisites for the course. Participants are encouraged to bring their own laptops- otherwise BioLateral will provide desktop computers for the duration of the course. Participants with PC laptops who would like to install UNIX on their laptops during the course are required to backup all their data prior to attending the course.

n at

.

,

		Early Bird Prices (payment 1 week prior to course)			Prices			
	Date	Student	Academic	Commercial	Student	Academic	Commercial	
Protein	Wed 6 - Fri							
Structure	8 Nov	\$1,000	\$1,250	\$1,500	\$2,000	\$2,500	\$3,000	
Bioinformatics	2002							
Microarray Bioinformatics	Mon 25 - Fri 29 Nov 2002	\$1,500	\$2,000	\$2,500	\$3,000	\$4,000	\$5,000	
BioInformatics Databases	Mon 21 Wed 23 Oct 2002	\$1,000	\$1,250	\$1,500	\$2,000	\$2,500	\$3,000	
Bioinformatics Programming	Mon 2 - Fri 6 Dec2002	\$1,500	\$2,000	\$2,500	. \$3,000	\$4,000	\$5,000	
Bioinformatics Hands-on Workshop	To be announced	\$150	\$250	\$350	\$150	\$250	\$350	

CURRENT TIMETABLE AND PRICING

ASSISTANCE FOR NSW BIOTECHNOLOGY FIRMS AND RESEARCH GROUPS

The Department of State and Regional Development has recognised the BioLateral Bioinformatics Courses as an approved element of its BioBusiness Strategy. Biotechnology companies and academic research groups who can clearly demonstrate that they are actively involved in commercial projects may be able to access financial assistance of \$2,500 towards the cost of the BioLateral Bioinformatics Courses.

For information on applying for assistance to attend the course please contact BioLateral Education at education@biolateral.com.au.

ASSISTANCE FOR OTHER GROUPS

For all groups other than NSW based biotechnology firms, please contact your state or territory office or organisation to investigate the possibility of assistance to attend this course.

FOR INFORMATION

BioLateral Pty Ltd, P.O. Box A51 Enfield Sth, 2133, NSW Australia Ph. 02 9036 3007 Toll Free: 1300 131 821 Fax 02 9036 3001 Email education@biolateral.com.au www.bidateral.com 2002 - 111 - 110-

BIOLATERAL BIOINFORMATICS COURSES



SIGLATERAL BIOINFORMATICS COURSES 2002

The BioLateral BioInformatics Courses will be run around Australia at the locations shown. To ensure your place fill and return this form or register online at www.biolateral.com.au.

2002 TIMETABLE & REGISTRATION FORM INSW based biotechnology companies may be eligible for assistance - see section below 1

To register, please write the number of places you wish to reserve in the "Please Reserve" column next to the applicable course, location and pricing option. Fill out your contact details and submit this form to BioLateral by post or fax. An invoice will be promptly dispatched and if your payment is received at least one week prior to course commencement the **EarlyBird rates** will apply.

Course	Max.Places	Location	Oates	Regular Pric	U	EarlyBird Rate	Please Reve	V LL
Bloinformatics for Macintosh® OS X	20	Sydney	Nov 1				15401	
		Melbourne	Nov 4	\$400 Includes copy of Bioinformac CD-ROM / Book		\$300	(MELB.)	
		Adetaide	Nov 5			Includes copy of BioInformac CD-ROM / Book	IADEL.I	
		Perth	Nov 6				IPERTHI	
		Brisbane	Nov 7				IBRIS.I	
		Other	Enquire]				coli
Protein Structure Bioinformatics	10	Sydney	Nov 6-8	Student	\$2,000	\$1,000		
				Academic	\$2,500	\$1,250		
				Industry	\$3,000	\$1,500		
Microarray Bioinformatics - Introduction	20	Melbourne	Nov 25-27	Student	\$1,500	\$750		
				Academic	\$2,000	\$1,000		
				Industry	\$2,500	\$1,250		
Microarray Bioinformatics -	4	Melbourne	Nov 28-29	Student	\$3,000	\$1,500		
				Academic	\$4,000	\$2,000		
Analyse your data				Industry	\$5,000	\$2,500		
Bioinformatics Programming	10	Sydney	Dec 2-6	Student	\$3,000	\$1,500		
				Academic	\$4,000	\$2,000		
				Industry	\$5,000	\$2,500		
Bioinformatics Computer Systems	2	Sydney	On Demand / Enquire	Student	n/a	\$3,000		-sait
				Academic	n/a	\$4,000		cat
				Industry	rv/a	\$5,000	Ne	tall

Contact Name :		
Organistation :		
Street Address :		
Suburb :	Postcode :	
State / Country :	Email :	
Telephone :	Fax :	

LOCATIONS

Sydney : BloLateral, 92-94 Parramatta Rd, Camperdown Melbourne : VPAC, 110 Victoria Street, Carlton South Adelaide : BioInnovation SA, level 13, 33 King William Street, Adelaide. Perth/Brisbane : TBA

AVAILABLE ASSISTANCE AND FURTHER INFORMATION

Assistance for NSW Biotechnology Firms

The Department of State and Regional Development has recognised the BioLateral Bioinformatics Courses as an element of its BioFriet Strategy. Assistance is available in NSW biotechnology companies and NSW academic participants who can clearly demonstrate that they are actively involved in commercial projects. Assistance for eligible applicants is up to a further 50% of the course fees. For more information about this support please context :



Assistance for Other Groups

For all groups other than NSW histochiology firms, contact your state/territory office or organisation, or Biol aterat, to unvestigate possible assistance in attending this course.

For More Information / Online Registration :

For more information or to register online, please access the BinLateral web site at www.binlateral.com.au. Alternatively, contact BioLateral by phene or email (details below)

FORM RETURN / CONTACTS



BioLateral Pty Limited. P.O. Box A51, Enfield South, N.S.W 2133 AUSTRALIA. tel : 1300 131 821 (Aust) or +61 2 9036 3007 (International) fax : +61 2 9642 5011 web : www.biolateral.com.au email : education@biolateral.com.au