

2010

CHARACTERIZATION OF THE R&R CONSENSUS REGION USING IN SILICO MOLECULAR MODELING

Nethaniah Dorh
Northern Michigan University

Follow this and additional works at: <https://commons.nmu.edu/theses>

Recommended Citation

Dorh, Nethaniah, "CHARACTERIZATION OF THE R&R CONSENSUS REGION USING IN SILICO MOLECULAR MODELING" (2010). *All NMU Master's Theses*. 384.
<https://commons.nmu.edu/theses/384>

This Open Access is brought to you for free and open access by the Student Works at NMU Commons. It has been accepted for inclusion in All NMU Master's Theses by an authorized administrator of NMU Commons. For more information, please contact kmcdonou@nmu.edu, bsarjean@nmu.edu.

CHARACTERIZATION OF THE R&R CONSENSUS REGION USING IN SILICO
MOLECULAR MODELING

By

NETHANIAH DORH

THESIS

Submitted to
Northern Michigan University
In partial fulfillment of the requirements
For the degree of

MASTER OF SCIENCE

Graduate Studies Office

2010

SIGNATURE APPROVAL FORM

This thesis by Nethaniah Dorh is recommended for approval by the student's thesis committee in the department of Biology and by the Dean of Graduate Studies.

Committee Chair: Dr John Rebers

Date:

First Reader: Dr Mark Paulsen

Date:

Second Reader: Dr Robert Winn

Date:

Department Head: Dr Patrick Brown

Date:

Dean of Graduate Studies: Dr Cynthia Prosen

Date:

**OLSEN LIBRARY
NORTHERN MICHIGAN UNIVERSITY**

THESIS DATA FORM

In order to catalog your thesis properly and enter a record in the OCLC international bibliographic data base, Olson Library must have the following requested information to distinguish you from others with the same or similar names and to provide appropriate subject access for other researchers.

NAME: Dorh, Nethaniah

DATE OF BIRTH: December 12, 1982

ABSTRACT

CHARACTERIZATION OF THE R&R CONSENSUS REGION USING IN SILICO MOLECULAR MODELING

By

Nethaniah Dorh

The R&R consensus sequence is a sequence of amino acids that have been found through sequence alignment studies to be relevant in chitin binding thus forming a chitin binding domain. Previous studies indicated that mutations of Y128 and F136 (tyrosine and phenylalanine) or mutations of T95 and D97 (threonine and aspartic acid) eliminated chitin binding (Rebers and Willis 2001). This suggested that these particular amino acids were crucial in the protein structure thus allowing chitin to bind. Threonine at position 95 was later hypothesized by Hamodrakas et al. (2005) to play a crucial role in the binding cleft of the RR-1 cuticular protein HCCP12 which was known to bind chitin. In order to achieve this goal, molecular models of the native protein and other models of the mutated proteins were created. A series of programs including AutoDockTools were used for the docking and interpretation of binding interactions between a six-unit macromolecule of chitin and the proposed models of the AGCP2B chitin binding protein from *Anopheles gambiae*. The results of the study showed that the intermolecular energies were all relatively close for the mutations made. Analysis of an original mutation attempted by Rebers and Willis (2001), (GST+65YF) yielded similar results suggesting that the docking analysis was ineffective in determining the dynamics of the chitin binding between the *Anopheles gambiae* protein and chitin.

Copyright by

Nethaniah Dorh

2010

DEDICATION

This Thesis is dedicated to my family, friends, mentors and most importantly my Lord and Savior Jesus Christ. None of this would have been possible without them.

ACKNOWLEDGMENTS

I would first like to acknowledge my advisor Dr John Rebers for his tireless efforts in helping me complete a suitable project for my master's thesis. I hope that this work is but a just measure of the patience and dedication exhibited by Dr Rebers towards myself in the completion of the final product.

Also of worthy mention is Dr Mark Paulsen who also offered unconditional advice and mentoring where it was warranted and as a result allowed me to produce the current work which was merely an extension of a greater vision. His expertise was truly priceless.

It would not be complete if I also did not offer my gratitude to Dr Suzanne Williams and Dr Robert Winn for first being a part of the thesis committee as well as for offering very good critique of procedures, drafts and the project in general.

In closing I would also like to spare some time to acknowledge my wife Jeanette and my daughter Kaelan who went many hours without me allowing me to progress and complete this project. It goes without saying that my mother, father, sister and brothers are also worthy of mention. I truly appreciate all of the effort and support from all of you. I am at this point today because of you.

This thesis follows the format prescribed by the Journal of cell biology (<http://jcb.rupress.org/site/misc/ifora.xhtml>) and the NMU Department of Biology.

TABLE OF CONTENTS

List of tables.....	(vii)
List of figures.....	(viii)
List of Abbreviations.....	(xi)
Chapter One: Introduction.....	1
1. R&R consensus	
1.1 R&R consensus region Overview	
1.2 Experimental background on R&R Consensus region	
1.3 Why is characterizing this region important?	
1.4 How will this region be further characterized?	
2. Molecular modeling using AutoDockTools	
2.1 In silico Molecular Modeling overview	
2.2 Using Autogrid	
2.3 Using Autodock	
2.4 Analyzing data recovered	
3. Hypothesis	
Chapter Two: Methods.....	24
Chapter Three: Results.....	34
Chapter Four: Discussion.....	53
References.....	56
Appendix A: Quickphyre results.....	61

Appendix B: PSI BLAST results.....	65
Appendix C: Output information from AutoGrid for various mutations attempted.....	72
Appendix D: Clustering Histogram data for AutoDock runs including the mutated AGCP2B..	75

List of Tables

Table 1. Abbreviations used for defining results from ProteinPredict	39
Table 2. Docking run results from Clustering histogram.....	41
Table 3. Results from most favorable dockings for the mutations made.....	44
Table 4. Final intermolecular energies of AutoDock runs for the mutations made...	45
Table 5. Clustering histogram for AGCP2B_M1 Run	76
Table 6. Clustering histogram for AGCP2B_M2 Run.....	77
Table 7. Clustering histogram for AGCP2B_M3 Run.....	78
Table 8. Clustering histogram for AGCP2B_M4 Run.....	79
Table 9. Clustering histogram for AGCP2B_M5 Run.....	80
Table 10. Clustering histogram for AGCP2B_M6 Run.....	81
Table 11. Clustering histogram for AGCP2B_M7 Run.....	82

List of Figures

Fig. 1. Chitin binding protein sequence of <i>Anopheles gambiae</i>	5
Fig. 2. Sequences used in Rebers and Willis, 2001 for chitin binding assays.....	6
Fig. 3. Amino acid sequence of GST+65 and the proposed constructs... ..	10
Fig. 4. AutoDock Force field.....	15
Fig. 5. Different states accounted for in the AutoDock calculations	16
Fig. 6. Grid box preparation using grid toolbar in AutoDockTools.....	19
Fig. 7. Ribbon view of PHM using RasWin.....	28
Fig. 8. 3D Model of AGCP2B sequence	29
Fig. 9. NAG base unit of the polymer Chitin	30
Fig. 10. Model of Chitin 6Mer built in Spartan.. ..	31
Fig.11. Secondary Structure Prediction from QuickPhyre	34
Fig. 12. PSI-Blast results for AGCP2B and PHM.....	35

Fig. 13. Blastp results for retinol binding protein and HCCP12 protein sequence.....	36
Fig. 14. ClustalW Report of sequence similarity between AGCP2B and PHM.....	37
Fig. 15. ProteinPredict secondary structure predictions for AGCP2B	38
Fig. 16. Preliminary energy calculations as made by AutoGrid for original AGCP2B.....	40
Fig. 17. Energy report for docked conformation.....	42
Fig. 18. Tif Image of AGCP2B original sequence and Chitin_6Mer interaction.....	47
Fig. 19. Solvent excluded surface view for AGCP2B-Chitin interaction.....	48
Fig. 20. Secondary structure of AGCP2B original sequence in area of interaction.....	48
Fig. 21. AGCP2B protein sequence interacting with chitin at its N-terminus position.....	49
Fig. 22. Interaction view of the AGCP2B_M1 model with chitin.....	50
Fig. 23. Interaction view for the AGCP2B_M3 model with chitin.....	50
Fig. 24. Interaction view for the AGCP2B_M7 model with chitin.....	51
Fig. 25. Interaction view of the GST+65 YF and the chitin oligomer	52
Fig. 26. Preliminary energy calculations as made by AutoGrid for AGCP2B_M1.....	72

Fig. 27. Preliminary energy calculations as made by AutoGrid for AGCP2B_M2.....72

Fig. 28. Preliminary energy calculations as made by AutoGrid for AGCP2B_M3.....73

Fig. 29. Preliminary energy calculations as made by AutoGrid for AGCP2B_M4.....73

Fig. 30. Preliminary energy calculations as made by AutoGrid for AGCP2B_M5.....73

Fig. 31. Preliminary energy calculations as made by AutoGrid for AGCP2B_M6.....74

Fig. 32. Preliminary energy calculations as made by AutoGrid for AGCP2B_M7.....74

LIST OF ABBREVIATIONS

Assisted Model Building and Energy Refinement.....	AMBER
Basic Load Alignment Search Tool.....	BLAST
Conformation player.....	CP
CPK.....	Corey, Pauling, and Koltun
Dexoyribonucleic acid.....	DNA
Genetic Algorithm.....	GA
Genetic Algorithm Local search.....	GALS
Glutathione-S-transferase.....	GST
Larmarckian Genetic Algorithm.....	LGA
Local Search.....	LS
National Center for Biotechnology Information.....	NCBI
Peptidylglycine alpha-hydroxylating monooxygenase.....	PHM

Position specific iterative BLAST.....	PSI-BLAST
Protein DataBank.....	PDB
Protein BLAST.....	BLASTp
Protein Homology/analogy Recognition Engine.....	PHYRE
Retinol binding Protein.....	RBP
Root Mean Square Deviation.....	RMSD
Simulated annealing.....	SA

INTRODUCTION

Chitin is a polysaccharide made from a polymer of β -1-4 linked N-acetyl-D-glucosamine and is one of the most abundant biological polymers worldwide (Gooday, 1990). Chitin is a major component in the exoskeletons of arthropods such as crustaceans and insects like beetles and mosquitoes, and it is also found in the cell walls of fungi. Chitin fibers help to make up the cuticles of arthropods which are essentially composed of microfibrils of chitin embedded in a matrix of cuticular proteins (Anderson, 1998).

A specific domain called the chitin binding domain was identified in cuticular proteins later found to bind chitin. Sequence alignments of six insect cuticular proteins revealed a conserved sequence that was suggested to play an important role in cuticular protein structure (Rebers and Riddiford, 1988). As additional cuticular protein sequences became available from a variety of arthropods, this consensus was recognized in a large number of proteins (Andersen et al., 1995; Willis et al., 2005; Willis, 2010), and came to be called the R&R consensus. Rebers and Willis (2001) showed that GST fusion proteins with 65 amino acids from the R&R consensus bound to chitin, showing that this region functioned as a chitin-binding domain. A few point mutations were attempted on this cuticular protein sequence from *Anopheles gambiae* (Rebers and Willis, 2001) which resulted in the loss of chitin binding. The current study focused on the particular reason these amino acids resulted in the loss of chitin binding. Further detail on the previous studies conducted and the information revealed will be emphasized later.

This study investigated the hypothesis that conservative amino acid changes in the chitin binding domain result in binding comparable to the original protein sequence. In order to validate the hypothesis, the intermolecular energies between chitin and the chitin binding protein of *Anopheles gambiae* was used for comparison.

1. R&R Consensus

1.1 R&R Consensus region Overview

The class Insecta is the one of the most diverse group of animals on the Earth representing a large number of the life forms on the planet (Erwin, 1982). It is because of their diversity and prominence that they have been studied so extensively. This group of animals is also known for their life cycles usually involving molting of their exoskeleton forming a new external body. In arthropods, this exoskeleton can function as protection, support and even as food (Bengston, 2004). Chitin can be mixed with silica or calcium carbonate (Campbell, 1996) giving it the much harder form that most people are familiar with.

The first chitin binding domain characterized in arthropods was found in chitinases and peritrophic membrane proteins (Shen and Jacobs-Lorena, 1999). The peritrophic membrane is a semi-permeable chitinous matrix lining in the gut of most insects and is thought to have important roles in the maintenance of the insect's gut structure, facilitation of digestion, and protection from invasion by microorganisms and parasites (Elvin et al., 1996). Chitinases are enzymes involved in the break down of glycosidic bonds in chitin (Campbell, 1996). The chitin-binding domain in these proteins has a conserved sequence with cysteine residues that are important in chitin-protein interactions.

More recently, the R&R consensus domain found in cuticular proteins (Rebers and Riddiford, 1988) was shown to bind chitin (Rebers and Willis, 2001). Cuticular proteins do not include cysteine (Willis, 1999), so Rebers and Willis (2001) proposed that the R&R domain represented a new type of chitin-binding domain in arthropods, which they designated the "non-Cys CBD".

The R&R consensus was originally subdivided into one of two possible groups, the RR-1 sequence in proteins from soft cuticles and the RR-2 sequence in proteins from hard cuticles (Andersen, 1998). A third group of proteins called the RR-3 group had since been identified, but had not been clearly classified (Togawa et al., 2004) due to the fact that it had been identified in postecdysial cuticle of insects as well as in preecdysial cuticle (Iconomidou et al., 2005) which would put it in both RR-1 and RR-2 classification (Karouzou et al., 2007). The RR-1, RR-2 and RR-3 sequences have since been hypothesized to be defined by the region of the cuticle in which they are found (Karouzou et al., 2007).

The R&R-2 consensus sequence from *Anopheles gambiae* is shown below in caps (Red text, underlined and bold font), outlined from the known chitin binding protein sequence (65 amino acid sequence) of *Anopheles gambiae* (Rebers and Willis, 2001). Figure 1 shows the consensus sequence for *Anopheles gambiae*, *Drosophila melanogaster* and *Hyalphora cecropia* and identifies this region in both RR-1 and RR-2 cuticular proteins.

⁸⁰apan**Ye**Fs**Ys**Vh**Deh**TG**Di**Ks**Qh**ET**Rh**GD**eV**hG**YS**LL**Ds**DG**hq**Ri**VD**Yh**AD**hh**tG**FN**AV**VR**Re**P**¹⁴⁴******************** (1)**************************************

¹⁸shpq**Ys**Fn**Yd**Vq**Dpe**TG**Dv**Ks**Qs**ES**Rd**GD**VV**hG**YS**Vn**Da**DG**yr**RT**VD**Yt**AD**dv**rG**FN**AV**VR**Re**P********************** (2)**************************************

⁴²**Yg**ye**Ts**nG**iq**hq**es**Gq**ln**nv**gt**e--ne**gi**eV**rG**qF**sY**vg**pD**Gv**ty**sV**tY**tA**g-Qe**GF**kP**vG**ah**IP********************** (3)**************************************

xxPxYxYxYxVxDxHTGDxKSQxExRDGDVVxGxYSLxExDGxxRTVxYTADxxNGFNAAVxxEx RR2

G Y DG Y AD GF P RR

Fig. 1. Amino acids in bold, uppercase, red color and underlined represent the R&R-2 consensus. (1) is the chitin binding sequence found in *Anopheles gambiae* (AGCP2B with genbank accession number [AAC05657](#)). (2) is the chitin binding sequence found in *Drosophila melanogaster* (EDG84 with genpept accession number [NP_524247](#)). (3) is the chitin binding sequence found in *Hyalophora cecropia* (HCCP12 with Swiss-Prot accession number [P45589](#)). Superscript numbers indicate starting and end of sequence where applicable (Iconomidou et al. (1999). Differences in conserved residues, such as between (1) and (3), are allocated to differences in hard and soft cuticular proteins (Iconomidou et al., 1999). (1) and (2) are both RR-2 cuticular protein sequences while (3) is a RR-1 cuticular protein sequence. RR-2, and RR consensus are all indicated below aligned sequences. RR-2 is in blue font color; RR is in orange font color. It is clear that the RR-2 is a version of the RR consensus such that the RR consensus is a subset of the RR-2. Please note that xEYDxx is not included at the N-terminus of the RR-2.

1.2 Experimental background on R&R Consensus region

Rebers and Willis (2001), conducted experiments with GST (glutathione-S-transferase; an enzyme used to catalyze Glutathione reactions) fusion proteins. Using the AGCP2b cuticular protein sequence from the mosquito *Anopheles gambiae* (Fig. 1) as the test sequence, two different fusion clones were constructed using either 40 residues ((2) in Fig. 2) to GST or 65 residues ((1) in Fig. 2) to GST, consequently called the GST+40 and GST+65 respectively (Fig. 2).

<p>(1) apanYeFsYsVhDehTGDiKsQhETRhGDeVhGqYSLLDsDGhqRiVDYhADhhtGFNAVVRReP</p> <p>(GST+65)</p> <p>(2) TRhGDeVhGqYSLLDsDGhqRiVDYhADhhtGFNAVVRReP</p> <p>(GST+40)</p>

Fig. 2. Sequence (1) identifies the original full length transcript thought to have chitin binding abilities. The sequence (2) represents the second clone used which contained only 40 of the amino acids from the full length transcript.

Rebers and Willis (2001), noted that the GST+65 clones bound chitin while the GST+40 clones did not bind chitin. The experiment also used site directed mutagenesis, making amino acid substitutions, in order to investigate whether chitin binding would be affected. This was of particular interest due to the proposed role of aromatic amino acids in conferring chitin binding to the protein sequence (Quiocho, 1989; Rebers and Willis, 2001). Threonine and aspartic acid residues were also found to be highly conserved, especially in the RR-2 cuticular protein sequences, in a multiple sequence alignment (Iconomidou et al., 1999) and as a result, Rebers and Willis (2001) sought to investigate the effects of mutations in the locations of these particular residues.

The results indicated that mutating Y128 (tyrosine at position 128) and F136 (phenylalanine at position 136) to produce the new construct (fusion protein GST+65.YF) resulted in a loss of chitin binding. Also, mutating T95 (threonine at position 95) and D97 (aspartic acid at position 97) to produce the construct (fusion protein GST+65.TD) eliminated chitin binding. These findings helped to further characterize the consensus sequence indicating that a loss of tyrosine (136), phenylalanine (128), threonine (95) and aspartic acid (97) results in a loss of chitin binding ability.

In a subsequent study conducted by Hamodrakas et al. (2002), a structural model for the protein HCCP12 (RR-1 type cuticular protein) from *Hyalophora cecropia* was proposed. This model contained several attractive features which made it popular as a basic structural model for interactions with the chitin chains in the cuticle. The structure was made using the retinol binding protein as a template. The template shared similar secondary structure with the HCCP12 protein. The retinol binding protein is a mammalian protein (Rask et al., 1987) which is not closely related to the protein HCCP12 but due to similarities in structure, the retinol binding protein served as a good template. The authors used molecular modeling as a means to investigate the binding cleft of the HCCP12 protein as it related to binding with chitin. The results indicated that the structure was consistent with evidence shown in Rebers and Willis (2001). Rebers and Willis (2001), showed that chitin binding had been affected with mutations of key amino acids further supporting the claim for the structure's accuracy (Hamodrakas et al., 2002).

1.3 Why is characterizing this region important?

Essentially the binding cleft responsible for chitin binding in the HCCP12 protein was identified which raised new questions about the binding cleft used in hard cuticle proteins such as that of *Anopheles gambiae*. Understanding the kind of cleft used, particularly the functional groups located in that region would allow for improved visualization of the chitin-protein interaction.

The R&R consensus region defines the interaction between cuticular proteins and chitin. Since its discovery, several hypotheses have been circulating as to the reason this sequence is necessary for chitin binding. An understanding of the specifics of the

sequence provides a key that opens the door to a plethora of applications for chitin and protein interactions.

Interactions of chitin with cuticular proteins is of great importance because it will allow the scientific community to understand the mechanics involved with the chitin binding sequences. A secondary effect or application of this kind of study and other future advanced studies that will help to fully characterize the chitin binding domain could make it possible to control vectors like mosquitoes that transfer viruses and parasites efficiently. Control at this level could help to curtail the spread of these dangerous diseases that reduce the quality of human life worldwide.

1.4 How will this region be further characterized?

Keeping all of the aforementioned in mind, the specific objectives were to further characterize the R&R consensus sequence by making even more amino acid substitutions to analyze the effect on chitin binding. The amino acids chosen to be changed (tested) were selected on the knowledge that the aromatic amino acids are essential in chitin binding as well as the fact that threonine 95 was believed to participate in the binding cleft as well as in the maintenance of the structure of a sharp turn in the beta strand (Hamodrakas et al., 2002). An analogous substitution with serine should maintain the binding capacity and would provide support for threonine's proposed role in the chitin binding domain. Similar amino acids in charge and size will be selected as replacements for the test amino acids within the consensus sequence in an attempt to fully characterize the chitin binding.

Initially, these substitutions were attempted using *in vitro* site directed mutagenesis. However; due to several time consuming challenges, the method was

switched to a computer modeling approach. As a result, substitutions were achieved by simply changing the test amino acid in the GST+65 sequence which was then transformed into an electronic format (pdb – protein databank file). This file was then used to model the protein *in silico* and allowed for further analysis. The sequences would be used to model chitin binding and observe the theoretical binding energy achieved by the interaction of each amino acid substitution with the chitin sequence.

Figure 3 indicates the GST+65 amino acid sequence showing substitutions already made by Rebers and Willis (2001) along with the proposed constructs indicating the substitutions to be made in order to allow the comparative analysis of chitin binding. Changes to be made include conservative substitutions at seven different positions using the GST+65 clone. The amino acids to be switched are tyrosine at 3 different locations, aspartic acid, phenylalanine at 2 different locations and threonine. Any information retrieved from these experiments will help to define the R&R consensus region even further.

		* *		* *
GST+65:		APANYEFSYSVHDEHT <u>T</u> <u>G</u> DIKSQHETRHGDEVHGQYSLLDSDGHQRIVD <u>Y</u> HADHHT <u>G</u> NAVVRREP		
Proposed models:				
#1		APANYEFSYSVHDEHTGDIKSQHETRHGDEVHGQYSLLDSE <u>E</u> GHQRIVDYHADHHTGNAVVRREP		
#2		APANYEFSYSVHDEHTGDIKSQHETRHGDEVHGQYSLLDSDGHQRIVD <u>F</u> HADHHTGNAVVRREP		
#3		APANYEFSYSVHDEHTGDIKSQHETRHGDEVHGQYSLLDSDGHQRIVDYHADHHT <u>G</u> NAVVRREP		
#4		APANYEFSYSVHDEHTGDIKSQHETRHGDEVHGQ <u>F</u> SLLDSDGHQRIVDYHADHHTGNAVVRREP		
#5		APANYEFSYSVHDEHTGDIKSQHETRHGDEVHGQYSLLDSDGHQRIVDYHADHHT <u>G</u> NAVVRREP		
#6		APANYEFSYSVHDEHTGDIKSQHETRHGDEVHGQYSLLDSDGHQRIVD <u>W</u> HADHHTGNAVVRREP		
#7		APANYEFSYSVHDEH <u>S</u> GDIKSQHETRHGDEVHGQYSLLDSEGHQRIVDYHADHHTGNAVVRREP		

Fig. 3. GST+65 amino acid sequence of *Anopheles gambiae* with alanine substitutions already made (Rebers and Willis, 2001) indicated with an overhead asterisk, a bold underline () and bold text. The proposed models (constructs) with substitutions to be made are underlined and in red text. In models shown, the original amino acid is replaced with the similar amino acid in red font.

Models one through six indicate the substitutions for the aromatic amino acids with similar side chains while model seven is a serine substitution for threonine. Serine is the closest substitution in side chain and charge which can be made for threonine.

Another method, mentioned earlier, by which the R&R consensus region could be further characterized is through *in vitro* site directed mutagenesis. Using this method, clones of the mutations mentioned above are made targeting the amino acid of interest and changing it to the desired mutation. Specifics of the mutation process depend on the method being used. Analysis of binding would be done by attempting a chitin binding assay which allows for the protein to either bind to chitin beads or not bind to chitin beads. The amount of binding and comparison of binding between mutants would then have to be analyzed by western blotting paying attention to the amount of luminescence which would be recorded in order to create a clear understanding of the difference in

binding between clones. However; as was mentioned earlier, this technique suffered significant setbacks and as a result was not used.

2. Molecular Modeling using AutoDockTools

2.1 In Silico Molecular Modeling Overview

Molecular modeling uses a macromolecule (protein in this case) along with the use of a ligand (ion, molecule, protein) to mimic, analyze and predict interactions. It is a multidisciplinary field that requires the use of theoretical and computational techniques to make such predictions. Fields known to benefit from this unique tool include chemistry (organic and physical), biology and materials science (Leach, 2001). This tool relies heavily on computers to perform computational analyses on anything from a small scale interaction to large biomolecular interactions by taking an atom first type approach. The entire interaction can be defined in whole, but must first be defined in terms of the different atoms available in the system (Frenkel and Smit, 1996).

Coupling molecular modeling with molecular dynamics allows for even greater applications. Molecular dynamics is essentially the link between the structure and function of a protein. It uses classical approximations of known physics concepts to estimate equilibrium constants, structure, dynamics and the thermodynamics of complex systems (Rapaport, 2004). This is done by first computing the vectors for every atom in the system for both ligand and macromolecule, using the force field to predict the different conformations. The force field typically takes into consideration the system internal energy (U) which is the sum of the kinetic and potential energies of a system. The force field used in molecular modeling includes energies describing bonding, torsions (rotations), angles, van der Waals interactions, non-bonding and electrostatic energies (see equations below). The total energy of a system is referenced as the sum of all of these individual energies.

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{non-bonded}}$$

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}$$

$$E_{\text{non-bonded}} = E_{\text{electrostatic}} + E_{\text{vanderWaals}}$$

Collectively, these energies or force field define the potential function of a system and depending on the implementation of the system can vary in terms included in the calculation. For molecular dynamics, time is important as this process generates a computational simulation of the motion of the particles during interactions (Rapaport, 1996)

The essential steps used in molecular modeling require the description of each atom in vector form inclusive of time, followed by the predictions of the different conformations given the specified force fields, and finally the analysis of the results to verify that the conformations are all plausible. Specifically for AutoDockTools, the procedure is further simplified to the creation of the coordinate files which provide information about the atoms and the location; setting of the grid points so as to specify the coordinates of localized minimum binding and allow for precalculation of the atomic affinities; docking of the ligand to the macromolecule; and finally the analysis of the results obtained. The force field employed in AutoDock uses all of the mentioned energies in addition to other parameters such as desolvation and directionality of hydrogen bonds (Huey et al., 2007).

This force field is the AMBER (Assisted Model Building and Energy Refinement) force field originally defined in Pearlman et al. (1995) which is mostly used for proteins and DNA. According to the UserGuide for AutoDock4.2, the force field specifically includes six pair wise evaluations (labeled V) as well as an estimate of the

conformational entropy lost through binding (see equation below) (Goodsell and Olson, 1990).

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf})$$

The L represents evaluations done for the ligand while the P represents evaluations done for the protein. The ΔS represents the entropy of the system. The ΔG refers to Gibbs free energy which in this case is a measure of the work exchanged by the system with its surroundings, minus the work of the pressure forces observed in a reversible transformation of the system from the same initial state to the same final state (Perrot, 1998). It is further specified that each energy evaluation is calculated using the following equation relating the dispersion/repulsion, hydrogen bonding, electrostatics and desolvation. The W indicates weighting for the particular component, i.e., van der Waals, etc.

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{i,j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2 / 2\sigma^2)}$$

The AutoDock4.2 UserGuide (Morris et al., 2009) describes the equation as follows:

“The first term is a typical 6/12 potential for dispersion/repulsion interactions. The parameters are based on the AMBER force field. The second term is a directional H-bond term based on a 10/12 potential. The parameters C and D are assigned to give a maximal well depth (resolution) of 5 kcal/mol at 1.9Å for hydrogen bonds with oxygen and nitrogen, and a well depth of 1 kcal/mol at 2.5Å for hydrogen bonds with sulfur. The function E(t) provides directionality based on the angle t from ideal h-bonding geometry. The third term is a screened Coulomb potential for electrostatics. The final term is a

desolvation potential based on the volume of atoms (V) that surround a given atom and shelter it from solvent, weighted by a solvation parameter (S) and exponential term with distance-weighting factor $\sigma=3.5\text{\AA}$ ” (Goodsell and Olson, 1990). The four main contributions to the AutoDock force field are outlined in figure 4.

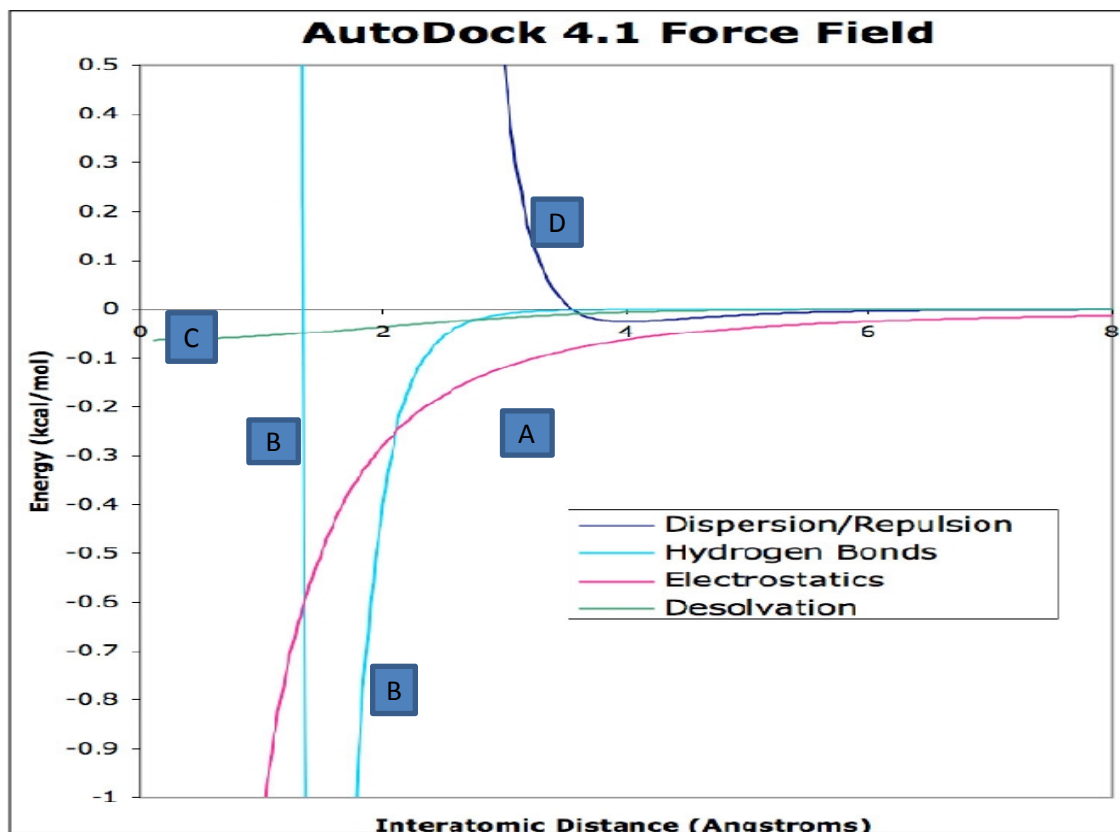


Fig. 4. The four main components of the AutoDock force field. “The desolvation potential is shown for a carbon atom, with approximately 10 atoms displacing water at each distance. The hydrogen bond potential, which extends down to a minimum of about -2 kcal/mol, is shown for an oxygen-hydrogen interaction. The dispersion/repulsion potential is for interaction between two carbon atoms and the electrostatic potential is shown for interaction of two oppositely charged atoms with a full atomic charge” as described by the AutoDock UserGuide 4.2 (Morris et al., 2009). A = Electrostatics curve. B = Hydrogen Bonds curve. C = Desolvation curve. D = Dispersion/Repulsion curve.

Binding energy in AutoDock is based on a semiempirical free energy force field (Huey et al., 2007), such that the intramolecular energy for the unbound state of ligand and protein are first evaluated followed by the estimation of the intermolecular energy

required to transition ligand and protein into their bound states (Goodsell and Olson, 1990). A simplified version of this is shown in figure 5 exaggerating the bound and unbound states of ligand and macromolecule.

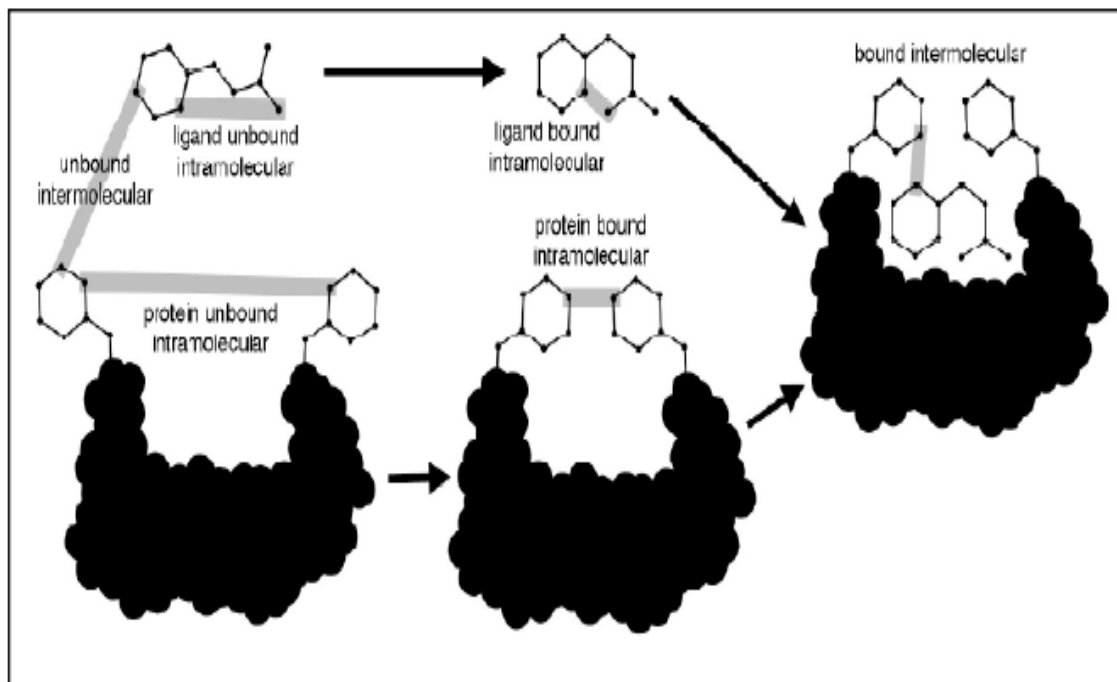


Fig.5. Bound and unbound states of ligand and macromolecule as they contribute to the overall prediction of the interaction between ligand and macromolecule (Morris et al., 2009).

2.2 Using Autogrid

Autogrid is the software package included in AutoDockTools (Huey et al., 2007) which serves a number of preparatory roles before docking can be attempted. Autogrid uses a grid map (potential energies) to represent the receptor and thus define the area of interaction. One grid map is precalculated for each atom type present in the ligand. A grid map consists of a three dimensional lattice of regularly spaced points, surrounding the receptor (protein in this case) and is centered on the region of interest of the macromolecule under study (Huey et al., 2004).

Each point within the grid map is determined by the sum of the pairwise potential interaction energy of a predefined probe atom of a particular type with each of the atoms in the macromolecule. The resulting 3-dimensional volume allocated by the grid maps together with the 'n' active torsions in the ligand defines the 6 + 'n' dimensioned search space (Huey et al., 2004).

When using the program the sequence of events required to create a grid plot is much more simplistic. The first step involves the preparation of the macromolecule file to be run in Autogrid. The automated steps first include the reading of the .pdb file (Protein Databank file). This file is then initialized allowing for computing of Kollman or Gasteiger charges depending on whether the macromolecule is a peptide or not respectively. Gasteiger charges are an iterative partial equalization of the orbital electronegativity for all atoms in the molecules (Gasteiger and Marsili, 1980). Kollman charges are derived from the electrostatic potential of point charges through semiempirical means and have been referenced to produce close to empirical expectations (Besler et al., 1990). The use of either Kollman or Gasteiger charge calculations is set as default in Autodock, however; the option can be changed to user preference. The default represents the current use of the charge calculations in experimental literature. The next automated steps include checking and merging of non-polar hydrogens and the determination of the atom types present in the macromolecule (default of 7 atom types can be tolerated). The macromolecule can also be modeled as flexible but the flexible portion would need to be specified and labeled accordingly (Goodsell et al., 1996).

In like manner, the ligand file is initialized as well from its pdb format involving;

checking and merging of non-polar hydrogens; checking of charges and computing and/or assignment of Kollman or Gasteiger charges depending on whether the ligand is a peptide or not respectively; and finally the renaming of planar carbons. In addition to this initial step, the number of torsions (rotatable bonds) can then be set if required. Rotatable bonds can be set manually or automatically depending on user preference. These changes made to macromolecule and ligand files are then saved in the applicable file format for use in Autogrid.

The grid coordinates for the grid box can then be set manually by the user or can be set to center on the macromolecule, ligand or any atom. An example of a grid box setup is indicated below (Fig. 6)

On completion of the gridbox setup, the grid parameter file is created retaining all of the specified information for running AutoGrid. A standard set of parameters and weights are used for the force field by AutoGrid. Once all the preliminary information is obtained, the AutoGrid program is used to perform all precalculations and returns useful information such as the minimum and maximum energy for each atom type as well as the electrostatic potential and desolvation potential. This information is all maintained in the output log file at the end of an AutoGrid run.

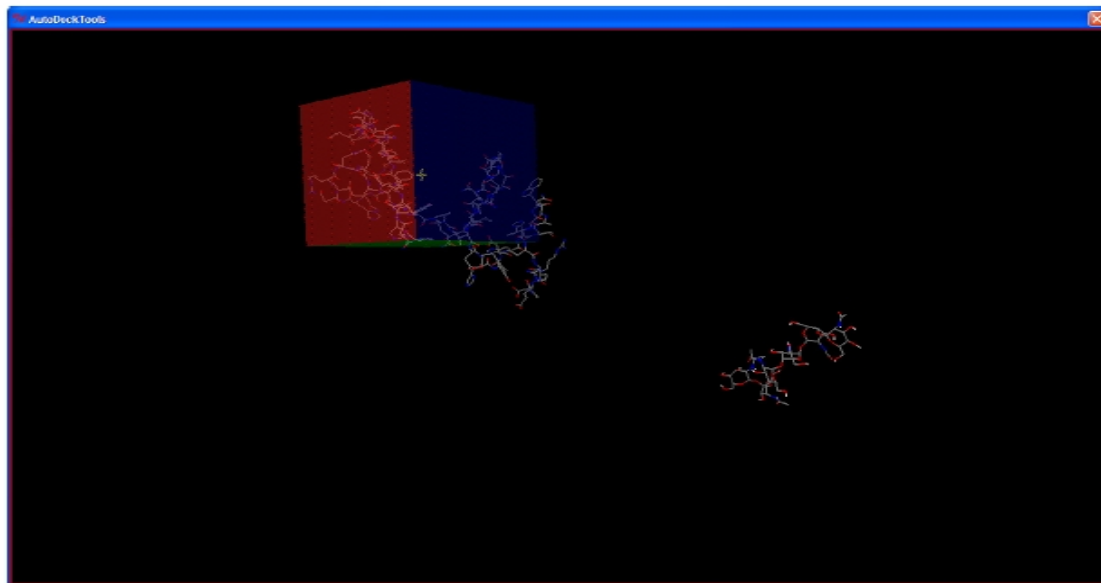


Fig.6. Grid box setup of the protein and ligand in AutoDockTools. The molecule to the right is the ligand and the molecule to the left with the multicolored box is the protein of interest. The blue areas indicate favorable carbon atom binding. The red areas indicate favorable oxygen atom binding.

2.3 Using AutoDock

The next step in molecular modeling using AutoDockTools requires the use of AutoDock4 (Huey et al., 2007). In order to use AutoDock, a docking parameter file is created. This file indicates to AutoDock which map files are to be used, the ligand molecule that needs to be moved, what the center and number of torsions are for this ligand are, where to start the ligand, which docking algorithm to use and how many runs that need to be done (Morris et al., 1998). AutoDock currently uses four different docking algorithms: SA (simulated annealing), the original *Monte Carlo* simulated annealing; GA (Genetic Algorithm), a traditional Darwinian genetic algorithm; LS, local search; and finally, the GALS (Genetic algorithm local search also known as LGA or Lamarckian Genetic Algorithm), which is a hybrid genetic algorithm inclusive of a local search (Morris et al., 1998).

The Monte Carlo simulated annealing is essentially an extensive global search which gives a good approximation for the global minimum of a given function in a relatively large search area (Metropolis et al., 1953). It relies on repetitive random sampling in order to generate the approximations and as a result can easily be used on a wide range of scenarios (Granville et al., 1994). Genetic algorithms are different such that they rely on a different subset of parameters to generate the solutions. As such they are classified as evolutionary algorithms (Wang et al., 2007). This kind of algorithm takes into consideration parameters such as mutation, crossover, inheritance and even selection which can all be accounted for via various means such as successive generations. In a molecular modeling point of view, the best conformations are the “best fit” which get to supply information for the next generation. A local search algorithm is essentially a metaheuristic defined by the restrictive search area where an optimal match is retrieved. The search pattern is such that it moves from one possible solution to the next until the best approximation is made (Hoos and Stützle, 2005). The Lamarckian genetic algorithm is similar to the traditional genetic algorithm with the exception that adaptive traits of a phenotype become inheritable and as a result improve fitness (Morris et al., 1998). This allows for increased sensitivity and since it also combines a local search improves the focus of the problem.

The docking algorithm used in this experiment was the Lamarckian Genetic Algorithm (LGA) which is basically an adaptive genetic algorithm global-local search. Huey et al. (2007), referenced the LGA as the most effective genetic algorithm for ligands with rotatable bonds of 10 and more. AutoDock essentially attempts to dock the ligand and protein using multiple conformations of the ligand and uses the grid maps

produced by AutoGrid to evaluate the intermolecular interaction for each docked simulation. In order to attempt the simulations, the program requires that a docking parameter file be written to specify parameters to be used in the simulation.

The docking parameter file specifies the number of runs to be attempted, the random number generator seeds, the energy outside the grid, the maximum allowable initial energy, the maximum number of retries, the step size parameters, the output format parameters and finally, whether or not to do a cluster analysis. In this case, the results are clustered to identify similar conformations using the root mean square deviation method. There are a few other parameters which can be changed but they are specialized parameters and are not required for this experiment.

After loading the ligand file, the ligand parameters can be specified such as the initial position and dihedral offset. After the macromolecule is loaded, then the searching parameter can be specified. In this case, the Lamarckian Genetic Algorithm is specified. After the search parameters are chosen, then the docking parameters files are written with all of the specified information.

A docking log file is created on completion of the docking run which produces valuable information such as the different components of the force field, the state variables for each conformation, the clustering histogram, the RMSD (Root Mean Square Deviation) table and the docked conformations for the lowest energy docked conformation from each cluster.

This log file signifies the end of the docking simulations and allows for analysis to determine the best docked conformation (hopefully a binding cleft) based on location of minimization and the type of atoms within that cleft.

2.4 Analyzing data recovered

In order to analyze the results of docking simulations AutoDockTools includes several tools that allow for clustering results by conformational similarity, visualizing conformations, visualizing interactions between ligands and proteins, and finally, visualizing the affinity potentials (Goodsell et al., 1996). The first step involved in analysis of the docking from AutoDock requires a review of the docking log file. The clustering histogram is the best place to start as it gives an indication of the range of binding energies retrieved from the simulations. After reviewing this information, the actual output from each docking simulation is the next step. A favorable binding energy is negative and helps to provide some validation for the docked conformation. The other energies can then be analyzed to verify that they are in accordance with the binding energy provided. A typical docked result provides the atom information, Cartesian coordinates and the associated state variables that help define the docked structure.

Using this docking log file, all of the docked conformations can be visualized using the Conformation Player (CP) supplied with AutoDockTools. Loading the docking log file into the CP allows visualization of all of the different conformations retrieved by AutoDock. At this point, it is possible to analyze the different conformations for location of docking as well as for the type of atoms located at this point. A favorable docked conformation which is not located in a binding cleft is an unlikely candidate for a real approximation of the binding between ligand and protein.

On critical analysis of the results, a clear indication of the best docked conformation as well as the thermodynamics of the system should be evident.

3. Hypothesis

The previous experiments on the R&R consensus region created some questions as to the reason the chitin binding was lost when alanine was substituted for other amino acids.. Two reasons were suggested for the loss of chitin binding: either the specific amino acid eliminated was required due to its R group which allowed further interactions with the chitin or that the substitution to alanine changed the overall protein structure causing it to interact differently with chitin. Using molecular modeling and attempting seven similar substitutions using the original AGCP2B sequence from *Anopheles gambiae* will allow for some insight into these specific questions. The amino acids switched were three tyrosine residues at different locations, two aspartic acid residues at different location, two phenylalanine residues at 2 different locations and one threonine residue. Any information retrieved from these experiments will help to define the R&R consensus region even further.

Due to the fact that these are all similar amino acids, chitin binding observed through similar binding energies should be maintained with all of the different models created. This was further evaluated by paying attention to the intermolecular energies and binding energies of the constructed models as they compare to the original model. In keeping with the expected binding, all of the sequences should have close to and similar intermolecular and binding energies. A control model using the original mutation made by Rebers and Willis (2001) (GST+65 YF), was also attempted in order to validate the findings that suggest that similar intermolecular energies are due to the amino acids chosen for these models.

Methods

Site directed mutagenesis

This method was attempted first but because of some unexpected difficulty and time restraints, molecular modeling was conducted. This method involved the creation of mutant clones (Fig. 3) using a commercially available site directed mutagenesis system. The first kit employed was the GeneEditor by Promega. Seemingly successful transformants were created using this kit, however; neither of the transformants contained any of the attempted mutant sequences and thus no successful clones. Subsequent trials to create mutant clones were also unsuccessful. As a result, a second kit was chosen to attempt the mutations, the GeneTailor kit by Invitrogen. Due to technical issues with the positive control and the proposed mutants, successful mutants for study were not obtained. As a result of this scenario, the *in silico* method was chosen for characterization of the R&R consensus region.

Search for template model for AGCP2B

The first step to creating a good model of any sequence which lacks crystallographic data is to find a suitable template. A template is important because it provides the atomic bonding information, bond length and positions that allow modeling to be done. The ideal model uses crystallographic data of the protein of interest. However; when getting this information is difficult, creation of a comparable model is acceptable. A template which already has crystallographic data can be used to produce a 3-D model of the protein using only the protein's sequence. In order to do this, there are various tools available to retrieve such information.

A free program provided by NCBI (National Center for Biotechnology

Information) called Blastp (Altschul et al., 1990) was used to do an initial search for prospective modeling candidates. Blastp is a program which can be used to identify the protein sequence of interest or find protein sequences that are similar. The only parameter which was limited in order to conduct the initial Blastp search was that the databases option was set to “Protein Databank Proteins” which limited the similarity search to proteins with X-ray or NMR structures which total approximately 60,000 proteins..

A PSI-Blast (Altschul et al., 1990) search was also conducted to increase the chances of finding a meaningful match in the search. PSI-Blast can essentially be used to identify proteins in a particular family as well as to build a position-specific score matrix which shows much greater resolution than the Blastp for purposes of this experiment. The position-specific score matrix is an often used representation of motifs in biological sequences and assumes independence between positions in the pattern as it calculates scores at each position independently from the symbols at other position (Ben-Gal et al., 2005).

A third search was conducted using QuickPhyre (Kelley and Sternberg, 2009), which essentially searches the Protein DataBank database as well as other protein databases for proteins that are similar in structure. QuickPhyre also produced secondary structure predictions.

Sequence alignment analysis

After obtaining suitable modeling template candidates for the AGCP2B sequence, ClustalW (Larkin et al., 2007) was used to run sequence alignments to identify similarities in primary structure. The sequence for the oxidized form of the M314i mutant of peptidylglycine alpha-hydroxylating monooxygenase (PHM, Protein Data

Bank Accession code 1YI9; Siebert et al., 2005) protein (discussed later) showed the greatest sequence similarity and as a result proposed the best template to model the AGCP2B sequence.

The sequence similarity was comparable to that of the retinol binding protein (RBP) template used to model the HCCP12 sequence in Hamodrakas et al., (2002). The sequence analysis with the retinol binding protein using Blastp (Altschul et al., 1990) showed a sequence identity of 25% (10/40), an E value of 6.1 and an alignment score of 17.7. For the sequence analysis of the PHM sequence with the AGCP2B protein sequence, the identity was 40% (9/22) with an E value of 7.9 and an alignment score of 24. An E value closer to zero is preferred for identifying a significant match. Essentially, the E value determines the likelihood of finding another match like the one found given the database size. Although, the E values for either sequence alignment were not favorable, the model protein chosen (PHM) is still a reasonable choice when compared to sequence alignment data for the model (RBP) used by Hamodrakas et al. (2002). The AGCP2B sequence was also aligned to the RBP (data not shown), with sequence similarity almost identical to that of the chosen template (PHM).

In addition to sequence alignment analysis with the original AGCP2B protein, alignments were also conducted using the seven specified models mentioned in Fig. 3. This required the simple editing of the sequence which was then saved and stored using a format such as: AGCP2B_MX.txt where X is the number of the mutation as it corresponds to Fig. 3. All of the new sequences were aligned to the PHM protein sequence and recorded. This alignment provided sequence information regarding similarities in the sequences. This provided information necessary for creating a protein

data bank file (pdb) of all of the mutated sequences.

Determination of the secondary structure of AGCP2B sequence.

In order to generate a model of the AGCP2B sequence, it is useful to predict the secondary structure of the sequence. This is useful because it can be used for comparison purposes when choosing suitable crystallographic data upon which to model the sequence.

Using the program ProteinPredict (Rost et al., 2004), the secondary structure of the AGCP2B sequence was predicted using several prediction programs including PROF and PHDsec included in the ProteinPredict analysis. The results reported were obtained from PROF and are presented in later figures documenting only the first few rows specifying secondary structure. This secondary structure was compared to the output retrieved from QuickPhyre. Since all seven mutated sequence alignments were almost identical (with the exception of one conservative amino acid substitution), secondary structure predictions were not attempted for these sequences as drastic changes in secondary structure were not expected.

Modeling the AGCP2B sequence

Modeling the AGCP2B sequence required the use of the crystallographic data for the template model PHM (the Oxidized Form Of The M314i Mutant Of Peptidylglycine Alpha-Hydroxylating Monooxygenase (Siebert et al., 2005); PDB ID: 1YI9: A) (Fig. 7) which was obtained from the PDB database for the organism *Rattus norvegicus*.

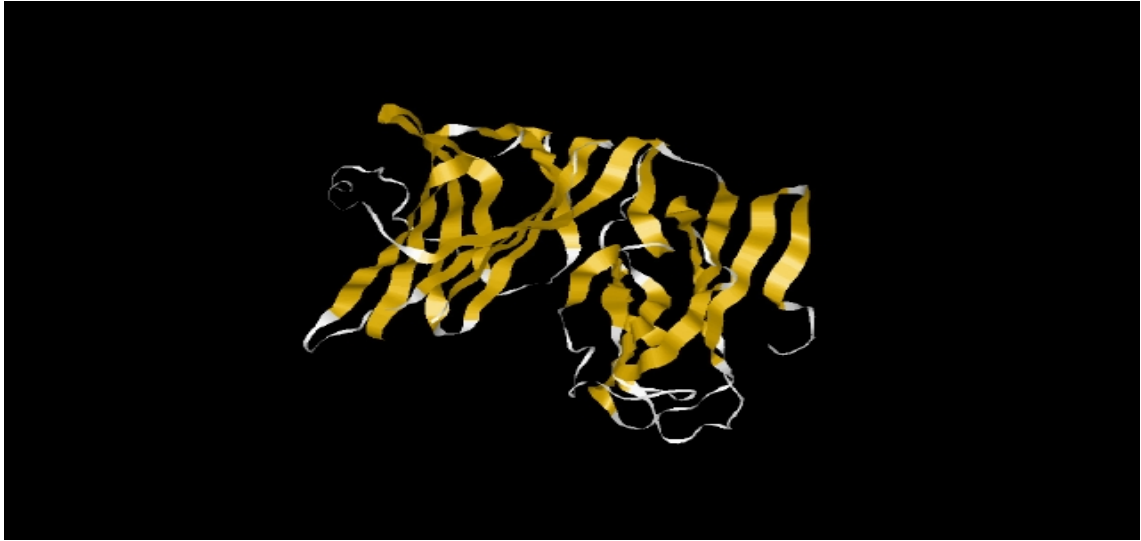


Fig. 7. RasWin (Sayle and Milner-White, 2005) was used to view and rotate PHM into current position exposing predominantly beta sheet and loop structure.

After the acquisition of the correct pdb file, the program Modeller9v6 (Sali and Blundell, 1993) was used to create the AGCP2B model in pdb format. The main use of Modeller9v6 is in homology modeling and three dimensional protein structures (Eswar et al., 2006; Sali and Blundell, 1993). Modeller works by using a sequence alignment of the model sequence to the sequence to be modeled. Python scripts are written which direct the program to the appropriate files (pdb file of template model; parameter files; sequence alignment file). On completion of modeling, a pdb file of the protein sequence was retrieved (Fig. 8). A review of the new model was then conducted using RasWin to verify that the produced structure was similar to that predicted by the structure predicting programs used previously.

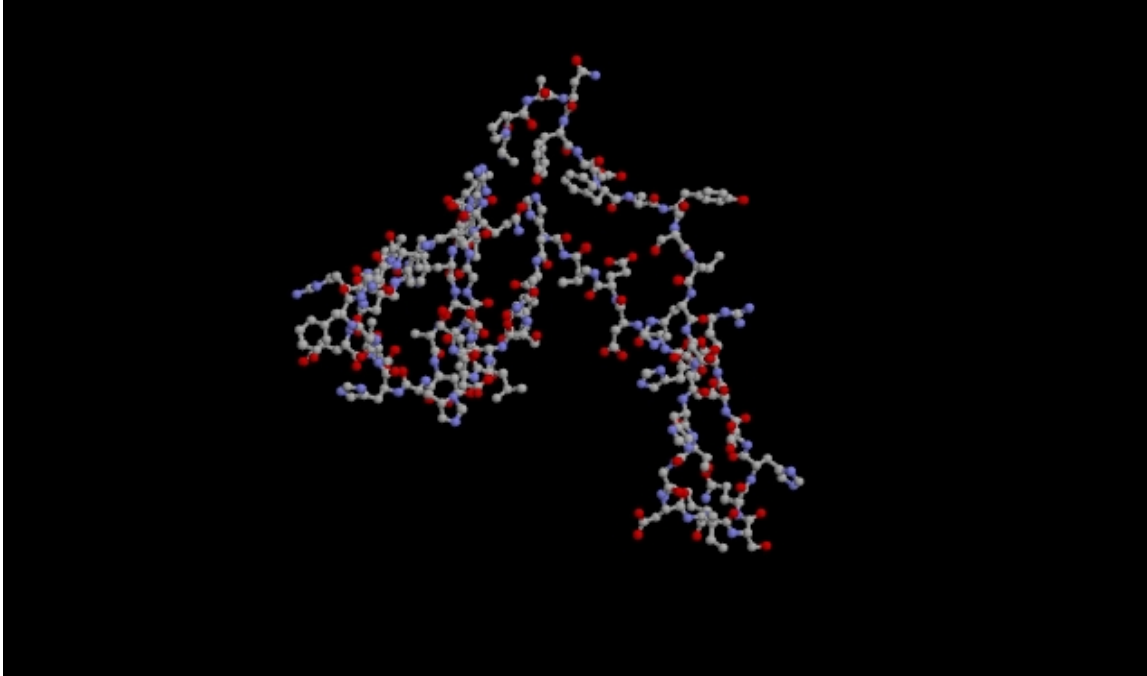


Fig.8. Ball and stick model of AGCP2B sequence. Sequence starts from amino acid 80 on the bottom right to amino acid 144 on the left as referenced in figure 1. Protein sequence is modeled using RasWin 2.7.5 (Sayle and Milner-White, 2005). This sequence contains the R&R consensus region described by Rebers and Willis (2001) and was the sequence used in the experiment.

Using the seven sequences in text file (.txt) format from the sequence alignments, the modeling for each of these point substitutions was attempted using Modeller9v6. The resulting pdb file was saved in the format: AGCP2B_MX.pdb where X represented the corresponding substitution number as referenced in Fig. 3.

Preparing Chitin pdb file

Since chitin is believed to interact with the AGCP2B protein sequence in the form of a four - six macromolecular unit (4Mer or 6Mer) (Togawa et al., 2004), a suitable model was built in order to facilitate this form. This was accomplished using the program Spartan (Kong et al., 2000). Each atom was connected using the following figure of a monomer as a guideline for a 6Mer unit (Fig. 9). The original file was stored in Spartan format which would be converted to the required pdb format.

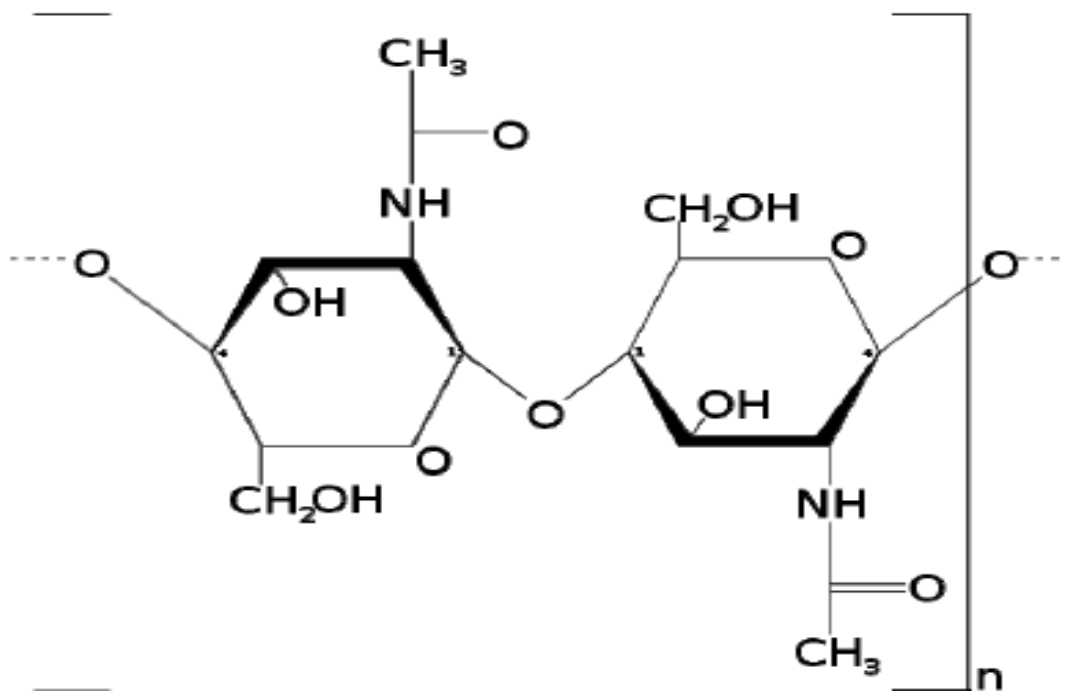


Fig. 9. Figure of chitin dimer from Cambell (1996), showing two of the N-acetylglucosamine units that repeat to form long chains in the beta-1,4 linkage of Chitin.

On completion of the 6Mer unit, the entire structure was minimized to reduce constraints on bonds. The completed file was converted and exported as a pdb file as is shown modeled in RasWin (Fig. 10). The newly acquired pdb file, labeled as Chitin_6Mer.pdb for the ligand, was then used in the molecular mechanics interaction to characterize the chitin-protein interaction.

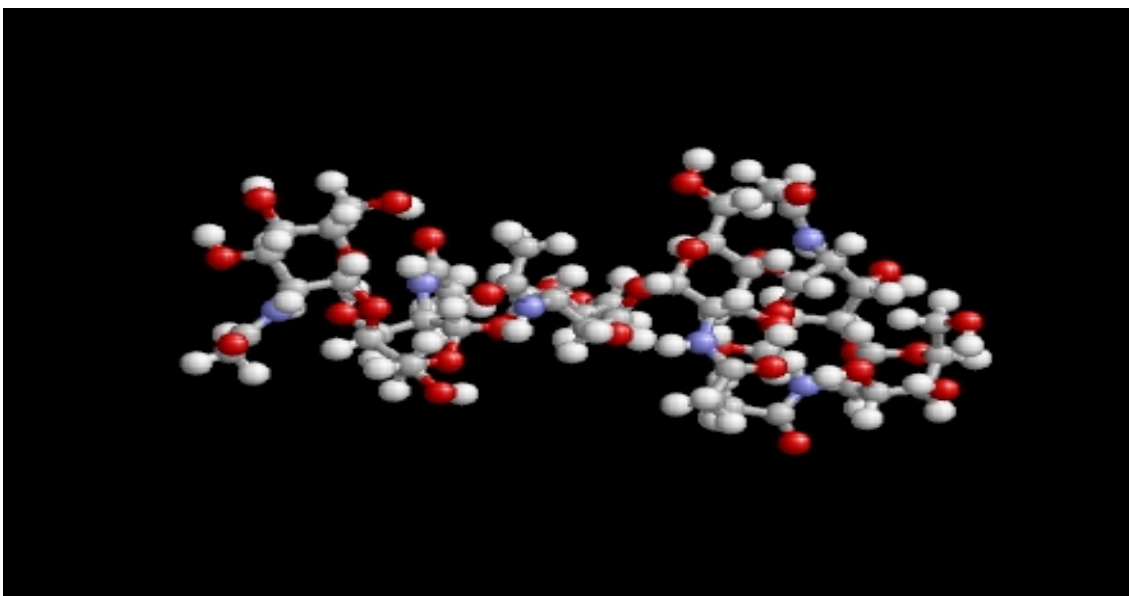


Fig.10. RasWin visualization of the chitin 6mer built in Spartan. Six NAG units are coupled by O-glycosidic bonds.

Molecular docking of the AGCP2B protein with chitin

After the required files were prepared for AutoDockTools, intermolecular interactions between the AGCP2B sequence and the chitin 6Mer unit were investigated. As previously described, the pdb files for the macromolecule and ligand were loaded and initialized to account for parameters such as charge and the merging of polar hydrogens. After this initial stage, the respective files for ligand and protein required by the program for running AutoGrid were created. The total number of available torsions for the chitin 6Mer was 42, however; only 32 were made active. The AutoDock program allows a total of 32 torsion and a maximum number of atoms set to 2000. In order to remedy this situation, the number of torsions for use was set to 32. The grid box parameters were then specified. Using information about nature of the binding pocket from Hamodrakas et al. (2002), the initial grid was set to coordinates containing a beta loop beta structure which contained aromatic amino acids that mimicked a binding cleft.

Subsequent analysis revealed the minimized energy area to be located at grid coordinates of 37, 26, and 44 (x, y, and z respectively). The grid box was centered along the macromolecule and then set to these coordinates manually. The box size was set to a volume of 58 X 58 X 58 with a spacing of 0.375 angstroms. The parameters for the AutoGrid were then saved using the file format AGCP2B_MX_Chitin_6Mer_AX.gpf with the exception of the original AGCP2B sequence which had the format AGCP2B_Y_Chitin_6Mer.gpf. In this case X represented the corresponding point substitution and Y represented the trial number for that run. This creation of the grid parameter files was repeated for the other AGCP2B sequences files containing the single point substitution.

At this point, the AutoGrid program was initialized and set to run using the applicable grid parameter file. The output of the program was recorded in several files containing the individual atom map files, the log file indicating specifics of the AutoGrid run, the atomic affinity and electrostatics map and finally the minimum and maximum coordinates map file for the gridbox. The output provided all of the preliminary information required by the AutoDock program to evaluate and predict the possible conformations.

Having created the files containing the area of interest in the macromolecule, the AutoDock docking parameter file was then created. This file was used to specify the docking parameters and the algorithms used to perform such a task. The algorithm of choice, as described earlier, was the Lamarckian Genetic Algorithm (LGA) and the parameters specified were the number of runs to be attempted, the number of evaluations and the maximum number of generations. The number of runs attempted was set to 50

while the maximum number of evaluations was set to 2,500,000 and finally the number of generations set to 25,000. These specific parameters were guidelines for macromolecule of the particular size being used. The docking parameters were then saved in the output format specified as Lamarckian GA (LGA) with a filename set as AGCP2B_MX_Chitin_6Mer_AX.dpf where X specifies the number corresponding to the attempted amino acid substitution referenced in Fig. 3. The completed docking performed by AutoDock produces a log file which contains the information pertaining to all of the conformations examined in the docking run. The docking energy is a function of the sum of the intermolecular energies (Van der Waals, hydrogen bond, desolvation energy and electrostatic energy) and the internal energy of the system. The binding energy is a function of the sum of the intermolecular energies and the torsional free energy. For the purposes of this study, only the intermolecular energies were considered for comparative purposes. This was due to an issue with the calculation of the torsional free energy which was an overestimation and as result created an offset in the calculation of the binding energy. The docking software is usually used with relatively small ligands with very few torsions and as a result, the torsional free energy calculation would need correction to account for the differences. This resulted in binding energies which were not favorable for protein-chitin interactions.

Results

The QuickPhyre search revealed less than favorable candidates for a model template listing E values as high as 31 with the lowest at 16 and sequence identities of less than 20% (See Appendix A). An E value closer to zero is considered a significant match. As a result, a suitable candidate for modeling was not chosen from this set of possibilities. Along with the listing of sequences, QuickPhyre also provided secondary structure information which served as insight into what should be expected when modeling the AGCP2B protein sequence. The secondary structure information reported is shown in Fig. 11.

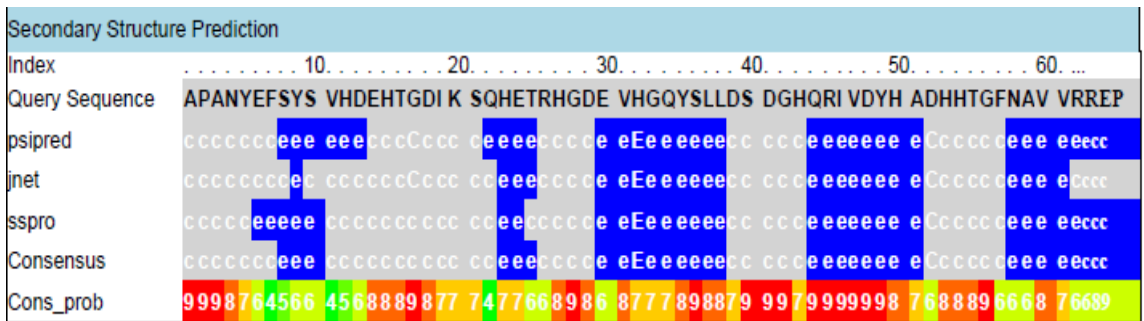


Fig.11. Secondary structure prediction for AGCP2B using QuickPhyre. The e represents the areas of interest where a beta sheet is expected to form. The C represents amino acids which are part of the coil and thus do not have a defined secondary structure type. Capital fonts show strong likelihood for the defined secondary structure. Despite the low resolution, it is evident that the structure is primarily beta sheets.

The general consensus is that the AGCP2B protein sequence is primarily beta sheets interconnected with loops (Fig. 11; consensus line). This is in line with what was reported for the HCCP12 protein sequence by Hamodrakas et al. (2002) as well as in many carbohydrate binding domains (Simpson and Barras, 1999).

The successful candidate for a model template was retrieved from a PSI-Blast search (Appendix B) which uses continuous iterations to produce greater resolution in the possible matches. The data retrieved suggested that the best option was the crystal

structure of the PHM (M314i Mutant of Peptidylglycine Alpha-Hydroxylating Monooxygenase) (Fig. 12) (Siebert et al., 2005). This is apparent in the data obtained from the search which lists a reasonable identity between sequences (9/22) as well as the crystal structure. The E-value was 7.6 which essentially meant that for a database of that size, one could expect to get a match like PHM 7.6 times. The lower or closer the E value to zero, the more significant the match. When templates are chosen, crystal structures are generally preferred over NMR structures due to the greater resolution observed in crystal structures.

```

pdb|1YI9|A Chain A, Crystal Structure Analysis Of The Oxidized Form Of The
M314i Mutant Of Peptidylglycine Alpha-Hydroxylating Monooxygenase
Length=309
Score = 24.6 bits (52), Expect = 7.9, Method: Composition-based stats.
Identities = 9/22 (40%), Positives = 13/22 (59%), Gaps = 0/22 (0%)
Query 7   FSYSVHDEHTGDIKSQHEIRHG 28
          F+Y VH  H G + S +  R+G
Sbjct 191  FAIRVHTHHLGKVVSGYRVRNG 212

```

Fig.12. The data returned from PSI-Blast after searching using the AGCP2B protein sequence. The referenced E value is 7.9. Three iterations were attempted in total yielding similar results and no changes in identities or significant changes in E values.

Since the alignment between PHM and the cuticular protein AGCP2B was not very significant, RBP and HCCP12 were aligned using Blastp (Altschul et al., 1990) for comparison. These proteins also gave a high E value (6.1) with 16/40 residues identical or similar. However, proteins that are not closely related at the primary sequence level can still provide useful templates to predict tertiary structures (Baker and Sali, 2001). This information was very necessary to validate the use of PHM as a suitable model for the AGCP2B template. The specific data is shown in Fig. 13.

```

>sp|P45589.1|CU12_HYACE RecName: Full=Flexible cuticle protein 12; Flags: Precursor gbl
AAA85640.1 cuticle protein 12 [Hyalophora cecropia]
Length=105 Score = 17.7 bits (34), Expect = 6.1, Method: Compositional matrix adjust.
Identities = 10/40 (25%), Positives = 16/40 (40%), Gaps = 3/40 (7%)
Query 107 IDTDYETFAVQYSCRLLNLDGTCADSYSFVFARDPSGFSP 146
      + T+ E   V+   + DG   +YS +   GF P
Sbjct 61 VGTENEGIEVRGQFSYVGPDPGV---TYSVTYTAGQEGFKP 97

```

Fig.13. Blastp (Altschul et al., 1990) results for retinol binding protein and HCCP12 protein sequence alignment. Note the E value was stated as 6.1 with 16/40 residues being identical or similar.

Peptidylglycine alpha-hydroxylating monooxygenase is the enzyme that catalyzes the first of the two steps of the amidation reaction of many bioactive peptides at their carboxy terminus allowing them to exhibit full biological activity (Kolhekar et al., 1997). In the mutant form, PHM has the methionine (Met) at position 314 switched to isoleucine (Ile) in order to investigate structure and function differences between native and mutated protein. Kolhekar et al. (1997) revealed that the PHM mutant was inactivated with the change from methionine to isoleucine. The crystal structure of this mutant protein provided the best candidate for molecular modeling of the AGCP2B protein.

Having selected the best template for modeling, the AGCP2B sequence and the PHM sequence were then reanalyzed by ClustalW (Larkin et al., 2007) to produce a sequence alignment file (Fig.14). This alignment file was the basis on which the modeled protein was erected.

```

CLUSTAL 2.0.10 multiple sequence alignment

M314i      CLGTIGPVTPLDASDFALDIRMPGVTPKESDITYFCMSMRLPVDEEAFVIDFKPRASMDTV 60
AGCP2b     -----

M314i      HHMLLFGCNMPSSSTGSYWFCDEGTCTDKANILYAWARNAPPTRLPKGVGFRVGGGETGSKY 120
AGCP2b     -----APANYEFS-----Y 9
                        * . * *

M314i      FVLQVHYGDISAFRDNHKDCSGVSVHLTRVPQPLIAGMYLMMSVDTVIPPGEKVVNADIS 180
AGCP2b     SVHDEHTGDI---KSQHETRHGDEVH-----GQYSLDSD-----G----- 42
      * : * ***  :.:*:  * .**          * * :.. * *

M314i      CQYKMPMHVFAYRVHTHHLGKVVSGYRVRNGQWTLIGRQNPQLPQAFYFVEHPVDVTFG 240
AGCP2b     -----HQRIVDYHAD-HHTG-----FNAVVRREP----- 65
              :.. *:. ** *          :. : *:*

M314i      DILAARCVFTGEGRTEATHIGGTSSDEICNLYIMYYMEAKYALSFMTCTKNVAPDMFRTI 300
AGCP2b     -----

M314i      PAEANIPIP 309
AGCP2b     -----

```

Fig. 14. ClustalW sequence alignment to be used in the modeling of the AGCP2B protein sequence. The output also shows consensus symbols denoting the degree of conservation observed in each column. See text for description of symbols. Alignment of PHM and AGCP2B is shown.

The consensus symbols employed by ClustalW can be described as follows: "*" indicates that the residues or nucleotides in that column are identical in all sequences in the alignment. ":" indicates that conserved substitutions have been observed, and "."

means that semi-conserved substitutions are observed (Larkin et al., 2007). Due to the short length of the AGCP2B protein sequence, only the region between amino acid 99 to about amino acid 230 of the PHM protein was used in the modeling.

Before modeling was attempted, the secondary structure was predicted to give a glimpse as to what to expect in the complete model. ProteinPredict, which uses several prediction programs, yielded results proposing that the AGCP2B protein sequence was mostly beta sheets with interconnected loops in its structure (Fig. 15; Table 1). This was in agreement with the data collected by QuickPhyre described earlier.

PROF results (normal)	
10.....20.....30.....40.....50.....60.....,
AA	APANYEFSYSVHDEHTGDIKSQHETRHGDEVHGQYSLLDSDGHQRIVDYHADHHTGFNAVVRREP
OBS_sec	
PROF_sec	EEEEEE E EEEEEEEEE EEEEEEEE EEEEEEE
Rel_sec	98750132000268878755530203677034467878078715888888457774488986348
SUB_sec	LLLL.....LLLLLLLLL.....LLL...EEEE.LLL.EEEEE.LLLL..EEEE..L

Fig. 15. ProteinPredict Secondary structure approximations for AGCP2B. E represents the beta sheets and L represents the interconnecting loops. The structure predicted is approximately 4 beta strands forming a sheet.

Table 1. Abbreviations used for defining results from ProteinPredict

AA :	amino acid sequence
OBS_sec:	observed secondary structure: H=helix, E=extended (sheet), blank=other (loop)
PROF_sec:	PROF predicted secondary structure: H=helix, E=extended (sheet), blank=other (loop) PROF = PROF: Profile network prediction HeiDelberg
Rel_sec:	reliability index for PROFsec prediction (0=low to 9=high) Note: for the brief presentation strong predictions marked by '*'
SUB_sec	subset of the PROFsec prediction, for all residues with an expected average accuracy > 82% (tables in header) NOTE: for this subset the following symbols are used: L: is loop (for which above ' ' is used) . : means that no prediction is made for this residue, as the reliability is: Rel < 5

Table 1. Description of the symbols used in the PROF report box from Fig.14.

The PHM sequence was successfully used to generate a plausible molecular structure for the AGCP2B protein sequence (see Fig. 8). This allowed for use of AutoDockTools to generate feasible conformations that represent binding between the chitin 6Mer and the AGCP2B protein.

After running AutoGrid, the results indicated that the energy calculation was within the range of what was expected (Fig. 16).

Grid	Atom	Minimum	Maximum	
Map	Type	Energy	Energy	
		(kcal/mol)	kcal/mol)	
1	C	-0.72	2.02e+05	
2	HD	-0.68	1.09e+05	
3	OA	-1.40	2.00e+05	
4	N	-0.73	2.00e+05	
5	e	-31.43	2.88e+01	Electrostatic Potential
6	d	0.00	1.02e+00	Desolvation Potential

Fig. 16. Typical results from AutoGrid energy calculations. Minimum and maximum energy estimations are reported in kcal/mol. Minimum van der Waals' energies and hydrogen bonding energies are typically -10 to -1 kcal/mol, while maximum van der Waals' energies are clamped at +105 kcal/mol. Electrostatic potentials tend to range from around -103 to +103 kcal/mol/e. C represents carbon; HD represents hydrogen; OA represents oxygen; N represents a nitrogen that cannot accept hydrogen bonds.

Grid calculations for all of the 7 point mutations to be used in the AutoDock were also successful yielding an acceptable range of energy results. The specific values are referenced in Appendix C.

The completion of a successful AutoGrid run allows the next step to occur which produces the docked conformations of the ligand and the energies associated with such conformations. After AutoDock completed its run, a docking parameter file was produced. The clustering histogram provides a synopsis of all the attempted runs and gives a summary in a clear and concise manner. The results obtained for the docking run are shown (Table 2).

Table 2. Docking run results from Clustering histogram

Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram
					5 10 15 20 25 30
					: : : : :
1	+1.05	29	+2.40	2	##
2	+1.57	12	+3.22	8	#####
3	+2.73	34	+3.52	2	##
4	+3.11	19	+3.11	1	#
5	+3.68	6	+3.68	1	#
6	+3.84	20	+3.84	1	#
7	+4.15	27	+4.15	1	#
8	+4.23	41	+4.23	1	#
9	+4.34	5	+4.34	1	#
10	+4.39	47	+4.39	1	#
11	+4.64	28	+4.64	1	#
12	+4.94	21	+4.94	1	#
13	+5.08	37	+5.08	1	#
14	+5.20	42	+5.20	1	#
15	+5.23	9	+5.23	1	#
16	+5.32	30	+5.32	1	#
17	+5.34	40	+5.34	1	#
18	+5.37	38	+5.37	1	#
19	+5.63	3	+5.63	1	#
20	+5.67	22	+5.67	1	#
21	+5.69	48	+5.69	1	#
22	+5.70	39	+5.70	1	#
23	+5.74	10	+6.00	2	##
24	+5.75	50	+5.75	1	#
25	+5.84	18	+5.84	1	#
26	+5.89	31	+5.89	1	#
27	+5.90	24	+5.90	1	#
28	+6.06	1	+6.06	1	#
29	+6.09	15	+6.09	1	#
30	+6.15	35	+6.15	1	#
31	+6.51	2	+6.51	1	#
32	+6.58	44	+6.58	1	#
33	+6.59	14	+6.59	1	#
34	+6.63	26	+6.63	1	#
35	+6.66	25	+6.66	1	#
36	+6.75	23	+6.75	1	#
37	+6.92	8	+6.92	1	#
38	+6.95	13	+6.95	1	#
39	+7.27	16	+7.27	1	#
40	+7.87	17	+7.87	1	#

Table 2. Clustering results for AutoDock run using the original AGCP2B protein and the Chitin 6Mer. The histogram reveals the minimum binding energy in Kcal/mol as well as the mean binding energy in kcal/mol. Also note that binding energies are all positive which suggest unfavorable interactions.

Using the root mean square deviation (RMSD) method, 4 total multi-member

conformation clusters were defined out of 50 runs. In order to continue the analysis of the results, the run with the lowest binding energy was chosen for further review. The lowest binding energy indicates favorable binding for the conformations of ligand and protein at the specified coordinates and parameters set. The data for this run specified the following (Fig. 17):

```
Run = 29
Cluster Rank = 1
Number of conformations in this cluster = 2

RMSD from reference structure      = 54.571 A

Estimated Free Energy of Binding   = +1.05 kcal/mol  [(1)+(2)+(3)-(4)]

(1) Final Intermolecular Energy    = -9.68 kcal/mol
    vdW + Hbond + desolv Energy    = -9.27 kcal/mol
    Electrostatic Energy           = -0.42 kcal/mol
(2) Final Total Internal Energy     = -9.81 kcal/mol
(3) Torsional Free Energy           = +10.74 kcal/mol
(4) Unbound System's Energy        = -9.81 kcal/mol
```

Fig.17. Energy report for run 29 of 50 showing the lowest binding energy. The different energies included in the AutoDock force field are all shown.

Due to the significant contribution of the torsional free energy, the estimated free energy of binding was proposed to be +1.05 kcal/mol. The binding energy calculation as mentioned earlier is the sum of the final intermolecular energy and the torsional free energy. The torsional free energy is usually a product of a predefined constant (0.3113;

which is the AutoDock 4 force field torsional free energy parameter) and the total number of rotatable bonds. In this case, the chitin 6Mer had the maximum number of torsions allowed also demonstrated a relatively large torsional energy. This indicates that AutoDock is including unfavorable conformations of the ligand which would not be allowed in an *in vivo* or *in vitro* experiment and marking it as favorable. As a result, for comparison purposes, the favorable conformations were first identified by the lowest binding energies, but then compared using the final intermolecular energy. In this case, the final intermolecular energy for the untouched AGCP2B (original sequence without substitutions) and the chitin 6Mer was -9.68 kcal/mol (Fig. 17). This served as the reference to which all other conformations obtained from the mutations were compared given the same general position and parameters.

The data recovered from all of the AutoDock runs for the amino acid substitutions was compiled and recorded in Table 3.

Table 3. Results from most favorable dockings for the point mutations

Mutation number	Lowest binding energy (kcal/mol)	Mean Binding energy (kcal/mol)	Run number
1	0.70	0.70	7
2	1.91	1.91	21
3	1.21	1.21	10
4	0.96	1.49	39
5	1.82	1.82	38
6	1.97	1.97	42
7	0.04	0.04	40
GST+65 YF	1.91	1.91	26

Table 3. Table showing the lowest binding energy dockings for each of the AutoDock runs. The mutation number corresponds to the construct number shown in Fig. 3. The lowest binding energy (kcal/mol) as well as the mean binding energy (kcal/mol) and the run number are also included for convenience. GST+65 YF is an original mutation attempted by Rebers and Willis (2001) known to eliminate chitin binding.

The data indicates that the serine substitution for threonine produced a much lower binding energy than the original binding energy. The standard deviation for the list of binding energies, not counting the GST+65 YF, was calculated as 0.543 kcal/mol. For comparison purposes, however; the final intermolecular energies are taken into consideration instead of the binding energy. Table 4 lists similar information but lists the final intermolecular energy and torsional free energy as well as the run and mutation number.

Table 4 Final intermolecular energies of AutoDock runs for the mutations made

Mutation number	Final intermolecular energy (kcal/mol)	Torsional free energy (kcal/mol)	Run number
1	-10.04	10.74	7
2	-8.82	10.74	21
3	-9.52	10.74	10
4	-9.78	10.74	39
5	-8.92	10.74	38
6	-8.77	10.74	42
7	-10.70	10.74	40
GST+65YF	-8.83	10.74	26

Table 4. Range of final intermolecular energies as supplied by the AutoDock docking log file. All intermolecular energies are negative which suggest favorableness. The energies range from -8.77 kcal/mol to -10.70 kcal/mol. GST+65 YF is an original mutation by Rebers and Willis (2001) known to eliminate chitin binding.

This data suggests that using a torsional free energy, which is a closer representation of the real life scenario, a more plausible binding energy could be calculated. Intermolecular energies are all negative indicating favorableness of the specified conformation. As a result, one can conclude that making the specified mutations (Fig. 3) would result in approximately the same level of binding to chitin as the original AGCP2B protein (see Figure 17) which yields a final intermolecular energy of -9.68 kcal/mol. Specific clustering histograms for each docking run are included and referenced as Appendix D. Mutations made by substitution in models one, four and seven all show intermolecular energies that are relatively better than that of the native protein while the other constructs either show relatively close or lower intermolecular energies (Table 4).

When compared to a model of a protein which is known to eliminate binding, the results are less convincing. The GST+65 YF model yielded an intermolecular energy of

-8.83 kcal/mol which is within the range of all of the attempted models reviewed. Standard error for the Autodock4 forcefield with reference to intermolecular energy could be as much as about 2.5 kcal/mol (Huey et al., 2007). Since all energies are within this range, then they are not considered significantly different.

The final portion of the AutoDock analysis required the visual screening of the docking conformation of interest for anomalies that would render the conformation invalid or unrealistic. This was conducted using the CP tool provided by AutoDockTools. On review of the favorable docking conformation for the original AGCP2B sequence with the chitin 6Mer, the conformation revealed that at its lowest binding energy conformation, the chitin 6Mer was located in the N-terminus region of the protein. Figure 18 shows a ball and stick picture of the AGCP2B protein with the chitin 6mer at its lowest energy conformation for cluster one.

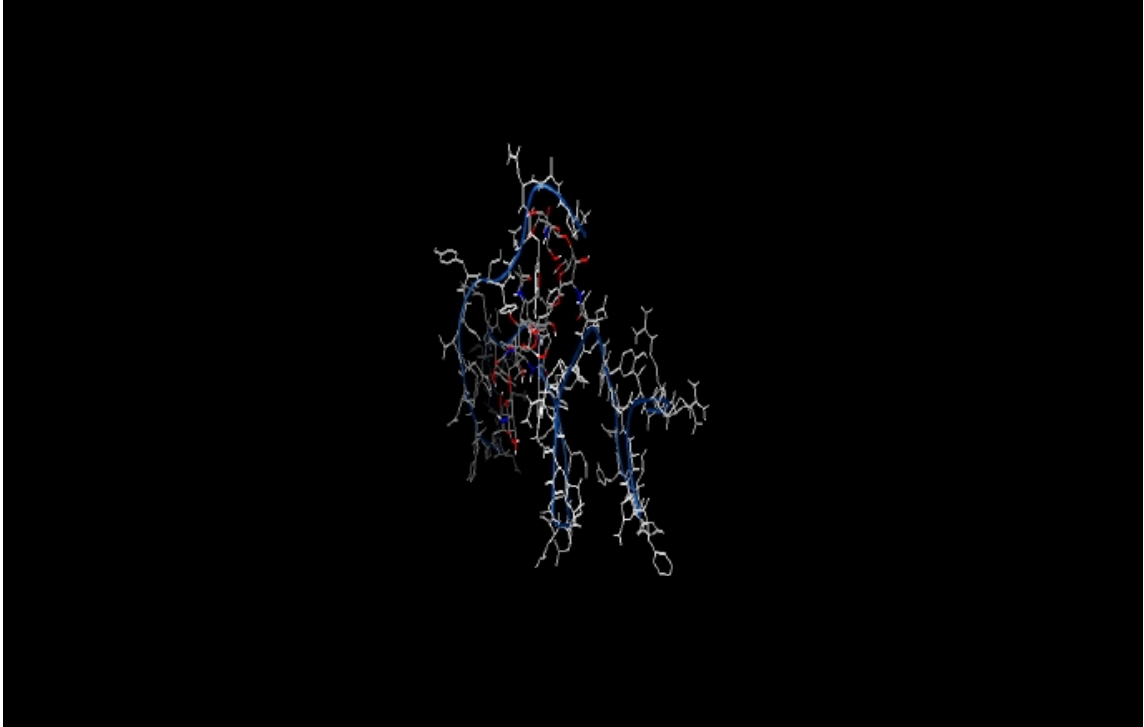


Fig.18 A rendered picture of the AGCP2B original sequence and the chitin 6Mer in the lowest energy conformation. The solid blue line indicates the secondary structure of the AGCP2B sequence. The grey molecules are those of the AGCP2B sequence while the colored molecules (red, etc) are those of the chitin 6mer.

The following image shows the interaction between molecules as a surface view (Fig. 19). Using this visualization tool, it appears that the ligand (grey) is sitting in an area which appears to look like a binding pocket. This new view shows support for the chemical reasonableness of the docking conformation given the constraints and parameters previously specified in the methods. The surprising part is that this pocket appears to include the N-terminus amino acids where most of the extended loop structure is located.

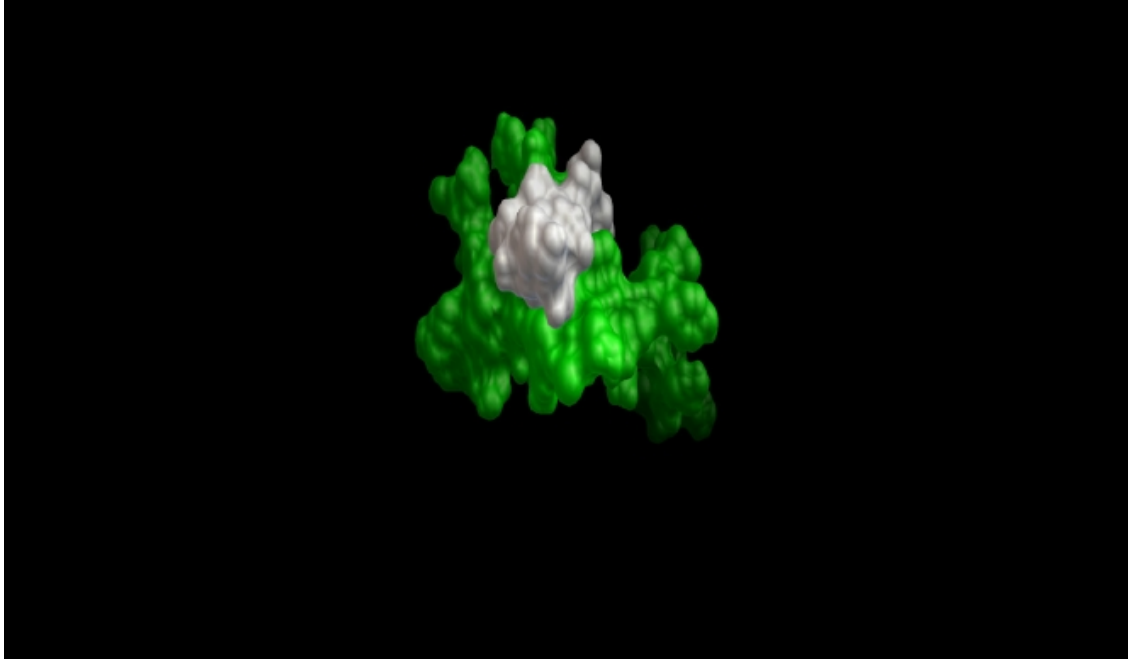


Fig. 19. Surface view of the chitin and original AGCP2B protein interaction. The green molecules represent the AGCP2B protein and the grey represents the chitin 6Mer. The picture suggests that chitin is interacting with the protein using a binding pocket formed at the start of the sequence.

This view is even more evident when visualizing the secondary structure of AGCP2B interacting with the chitin 6Mer (Fig. 20).

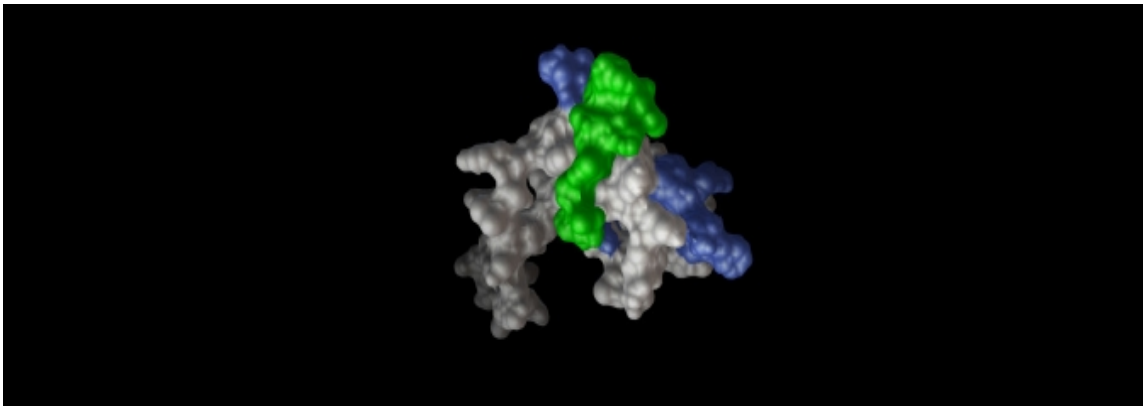


Fig.20. Interaction between chitin 6mer (green) and AGCP2B (blue and grey) original sequence showing secondary structure position relative to that of the area of interaction. This suggests that the area is indeed a binding cleft. Blue regions indicate area of beta sheets.

This information clearly suggests that a binding cleft is located in this area outlined by the beta sheets (blue) in which the aromatic amino acids are known to reside (Fig. 20).

As a result, the information retrieved supports the hypothesis suggesting that the binding cleft is an area of beta sheets in which the aromatic amino acids are housed and also the area of interaction for the chitin 6mer ligand.

In order to identify the amino acids located in this binding pocket, a new view showing interaction areas of the specific amino acids and chitin as spheres was acquired and is shown below (Fig. 21).

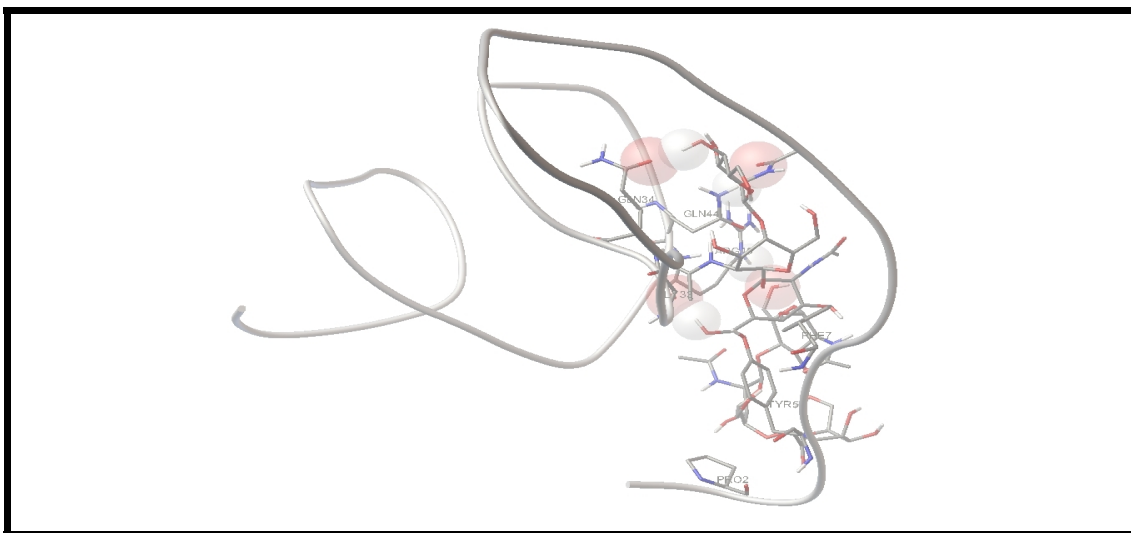


Fig. 21. AGCP2B protein sequence interacting with chitin at its N-terminus position. The color scheme used is CPK (where red and blue spheres represent oxygen and nitrogen respectively, while white/gray spheres represent hydrogen) indicating hydrogen bonding for GLN 34, GLN 44, ARG 45 and GLY 33. PRO 2 indicates the approximate start of the protein sequence. This supports that the binding pocket is located relatively towards the N-terminus.

In comparison to the other models, the position of the binding pocket was in the same region using some of the same amino acids, but not all. The overall binding concept was that the binding pocket's location was maintained regardless of changes in amino acids. This was a concept observed through even the modeled original mutation of Rebers and Willis (2001) (GST+65 YF). This can be seen in the following figures (Fig. 22, 23, and 24) containing interaction views of select models with chitin. Only select views are visualized due to the fact that the figures all show the same region.

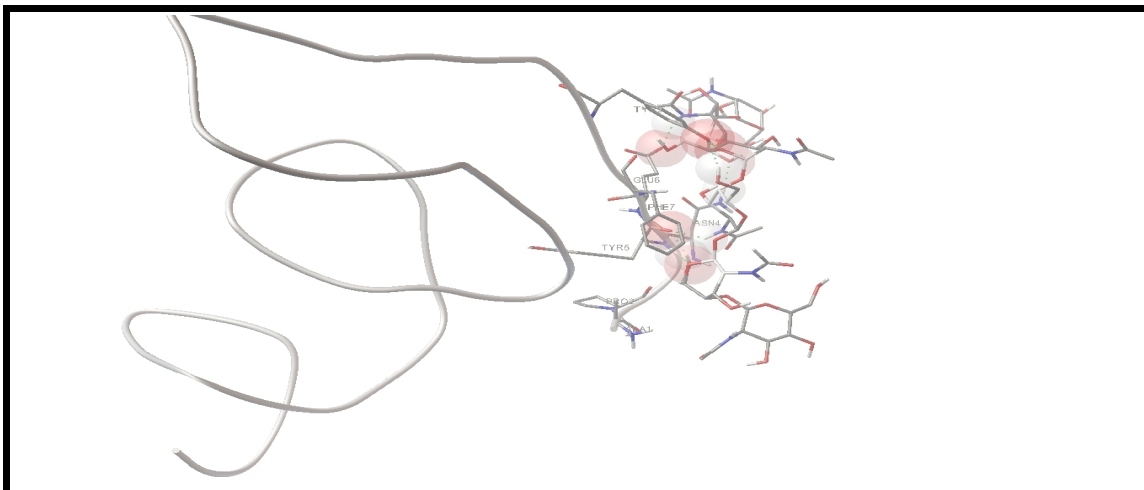


Fig. 22. Interaction view of the AGCP2B_M1 model with chitin. This interaction view is based on CPK colors where red and blue spheres represent oxygen and nitrogen respectively, while white/gray spheres represent hydrogen. Hydrogen bonding is observed for TYR 9, TYR 5, ASN 4 and PHE 7. In this case, only N-terminus amino acids are used in binding as opposed to the original sequence which had chitin in close proximity to the N-terminus as well as the middle amino acids where most of the bonding took place.

The next interaction view for the AGCP2B_M3 model also shows the same general region with hydrogen bonding including one of the same amino acids as the original AGCP2B sequence.

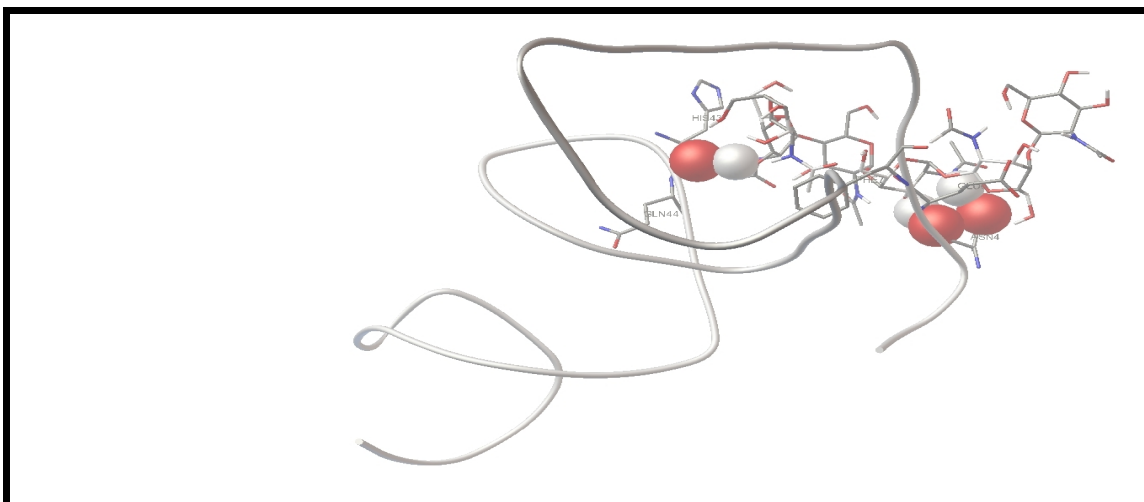


Fig. 23. Interaction view for the AGCP2B_M3 model with chitin. Amino acids noted to conduct hydrogen bonding with chitin are ASN4 and GLN 44. Color scheme is the same CPK mentioned previously. Once again the area identified in binding to chitin is the N-terminus as well as a the middle amino acids (GLN 44 in this case).

The AGCP2B_M7 model showed the least intermolecular energy as well as the largest binding energy, however; in contrast to the other models, the binding region as well as some of the amino acids used in binding were no different. This is clearly seen in Figure 24.

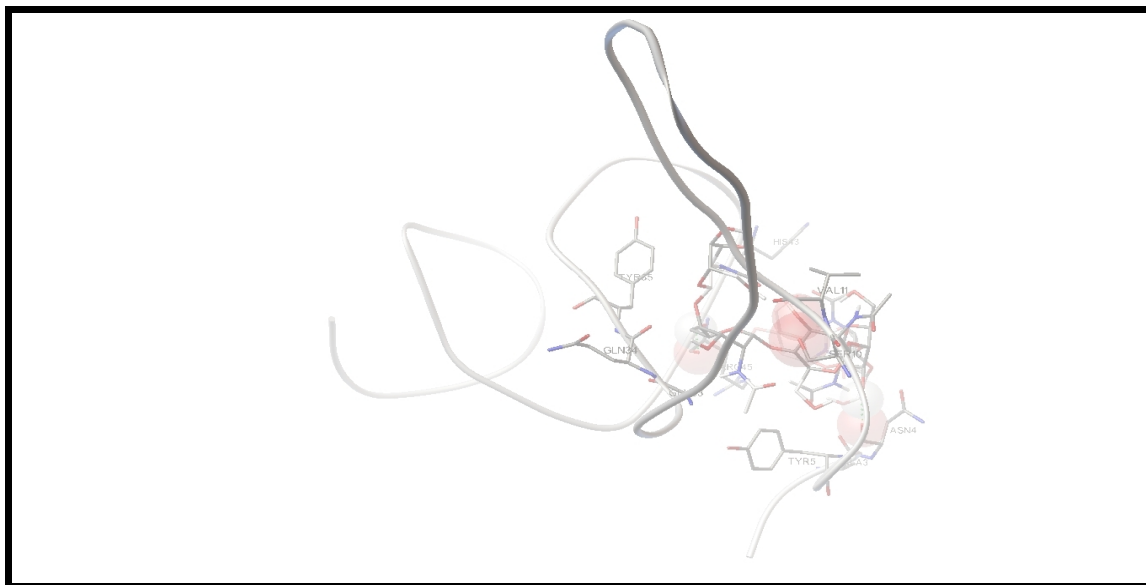


Fig. 24. Interaction view for the AGCP2B_M7 model with chitin. Color scheme used here is also CPK already described earlier. Amino acids identified as participating in hydrogen bonding with chitin are ASN 4 and ARG 45. ARG 45 was also seen in the interaction view of the original AGCP2B sequence with chitin.

Visual analysis of the docking conformations for the original Rebers and Willis (2001) mutation model (GST+65 YF) was also conducted. The interaction view shows two very important details. First, the overall protein folding is clearly different to that of the original sequence as it is much more condensed. The second detail is that the interaction is conducted in the N-terminus region with at least one of the amino acids seen in previous interactions with a few of the models (TYR 5). This can be seen in the figure 25 outlining the details of the interaction between the AGCP2B mutant by Rebers and Willis (2001), the GST+65 YF and the chitin oligomer.

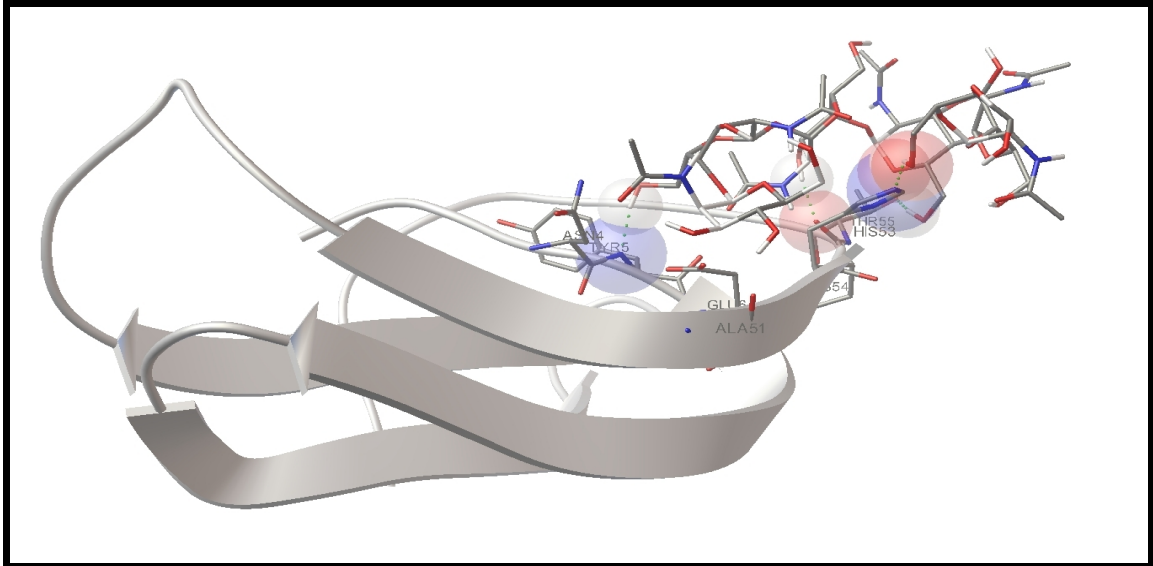


Fig. 25. Interaction view of the GST+65 YF and the chitin oligomer. The color scheme used was CPK. The amino acids indicated to be involved in hydrogen bonding with chitin were TYR 5, THR 55 and HIS 53. The chitin oligomer appears to be interacting with amino acids from both ends of the sequence which supports the alternative folding scenario.

This data thus suggests that the main area of interaction for chitin and the AGCP2B protein is at the N-terminus position with varied amino acids interacting with the chitin oligomer. Note that the bulk of the molecular surface of chitin is away from the GST+65_YF suggesting that interaction is occurring with surface amino acids due to convenience and availability. Greater resolution of the current torsional free energy calculations should allow better comparison of the models based on energies as well as docked conformations.

Discussion

A model for the AGCP2B protein was successfully built using the PHM protein as a template. This model showed greater resolution than a model built using the 1FEN template (Hamodrakas et al., 2002) due to better alignment scores (data not included). This suggests that the PSI-Blast search was indeed successful in rating the sequence similarity and as a result choosing a potential template. This allowed for increased sensitivity in the procedures that were used to compare binding of the original AGCP2B sequence to that of the mutated sequences.

The analysis of the results proposed that the interaction between chitin and the AGCP2B original sequence is composed primarily of oxygen and hydrogen interactions as shown in the 3D isocontour views of the maps constructed by AutoGrid (data not shown). This is what would be expected given the particular molecules in the binding cleft and particularly the secondary structure of the protein at that point.

Given all of the information collected, it is plausible that the reason the original mutations made by Rebers and Willis (2001) eliminated binding was due to the change in the overall conformation of the binding cleft. This could be attributed to a lack of the substituted aromatic amino acid resulting in incomplete folding which rendered the binding cleft unfavorable for chitin binding and interaction. This claim is based solely on the difference in overall structure of the GST+65 YF (Fig. 25) as compared to the original AGCP2B sequence seen in previous figures.

It is interesting to note that despite the 3D model for the GST+65 YF (data not shown) which predicted a mixture of alpha helices and beta strands which was different to the other constructed models which contained primarily beta strands, that the

intermolecular energies of all of the models were still rather close. It is thus clear that in addition to binding energy, that the intermolecular energy may not have been the best guide for determining the worthiness of any of the models in question. Due to the fact that the program makes its calculations based on the differences in unbound and bound states, perhaps the unbound system's energy would have been a good start for investigating differences between models as compared to the original AGCP2B model. Although, this would not be ideal since the point of the experiment was to determine differences in models when interacting with chitin, it would still serve as reasonable starting point for troubleshooting as the GST+65 YF had an unbound energy of -7.81 kcal/mol while the original AGCP2B and all of the other models created had an unbound energy of -9.81 kcal/mol or higher (data not shown). Values ranged from -9.01 kcal/mol to -13.22 kcal/mol with the exception of the M2 model (tyrosine to phenylalanine change) which had an energy of -8.17 kcal/mol. One could, however; argue that the differences between models is based primarily on the fact that each model is different in the number of atoms present. As a result, certain models with more or less atoms will show higher or lower unbound energies compared to the original as was indicated above.

On reviewing the binding energies produced by AutoDock, the positive energy seemed to be the entropy penalty (Torsions Free Energy) associated with the unusual number of torsions. The original AutoDock scoring function provided an additive type term for each torsion which was basically tuned to work with relatively small ligands (Morris et al., 2009). The specific calculation mentioned above can be tuned to more accurately address the torsional energy associated with the chitin ligand. A more ideal representation of the torsion will then allow for explicit comparison of the binding

energies and docking conformation for the original AGCP2B sequence and the mutated sequences. It would be interesting to further investigate the serine mutation due to the higher intermolecular energy compared to that reported for the native AGCP2B protein.

A second option would be to have simpler formulations of the system. This can be accomplished by breaking up the ligand into two pieces and docking them separately, or by freezing some rotatable bonds. The latter was already attempted (data not included) yielding results which were no better than when the maximum number of torsions were specified. As a result, the only reasonable options would be to change the way the torsional free energy is calculated or to break up the ligand into a dimer unit and a tetramer unit. In general, an increase in the number of rotatable bonds in the ligand decreases the likelihood that a good docking conformation will be retrieved even in repeated docking experiments. As a future experiment, it would be plausible to adjust the chitin unit to a tetramer (Togawa et al., 2004), and observe how the calculation for torsional free energy affected the overall binding energy of the system. This would be a reasonable experiment as Togawa et al. (2004) showed that a chitin binding protein bound itself to a chitin tetramer. This would fix the issue with the number of rotatable bonds and thus yield better results as far the binding energy is concerned. Despite the fact that the kinetics of the chitin binding was not resolved, suitable models were built successfully which can be utilized in future experiments to further characterize the chitin binding domain.

AutoDock4.2 was still successfully used for its initial proposed use which was in virtual screening. By virtually screening sequences, one can make inferences as to what to expect in an *in vitro* approach. As a result, the two techniques can be used together to

formulate a strong hypotheses, but, neither gives the entire picture if used on its own.

REFERENCES

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-10.
- Andersen, S.O. 1998. Amino acid sequence studies on endocuticular proteins from the desert locust, *Schistocerca gregaria*. *Insect Biochem. Molec. Biol.* 28:421–434.
- Baker, D., and A. Sali. 2001. Protein structure prediction and structural genomics. *Science*. 294:93-96.
- Bengston, S. 2004. Early skeletal fossils. *Paleontological Society Papers*. 10:67-78.
- Ben-Gal, I., A. Shani, A. Gohr, J. Grau, S. Arviv, A. Shmilovici, S. Posch, and I. Grosse. 2005. Identification of Transcription Factor Binding Sites with Variable-order Bayesian Networks. *Bioinformatics*. 21:2657–2666.
- Besler, B.H., K.M. Merz Jr., and P.A. Kollman. 1990. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* 11:431-439.
- Campbell, N. A. 1996. *Biology* (4th edition). Benjamin Cummings, New York. 1206 pp.
- Elvin, C.M., T. Vuocolo, R.D. Pearson, I.J. East, G.A. Riding, C.H. Eisemann, and R.L. Tellam. 1996. Characterization of a Major Peritrophic Membrane Protein, Peritrophin-44, from the Larvae of *Lucilia cuprina*. *J. Biol. Chem.* 271:8925–8935.
- Erwin, T.L. 1982. Tropical forests: their richness in Coleoptera and other arthropod species. *Coleopt. Bull.* 36:74-5.
- Eswar, N., M.A. Marti-Renom, B. Webb, M.S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, and A. Sali. 2006. Comparative Protein Structure Modeling With MODELLER. *Current Protocols in Bioinformatics, Supplement*. 15:5.6.1-5.6.30.
- Frenkel, D., and B. Smit. 1996. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, London. 9-13.
- Gasteiger, J., and M. Marsili. 1980. Iterative partial equalization of orbital electronegativity A rapid access to atomic charges. *Tetrahedron*. 36:3219–3228.

- Gooday, G.W. 1990. The ecology of chitin degradation. In: Marshall, K.C. (Ed.) *Adv. in Microb. Ecol.* 11:387–430.
- Goodsell, D. S., and A.J. Olson. 1990. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins: Str. Func. and Genet.* 8:195-202.
- Goodsell, D. S., G.M. Morris, A.J. Olson. 1996. Docking of Flexible Ligands: Applications of AutoDock. *J. Mol. Recogn.* 9:1-5.
- Granville, V., M. Krivanek, and J.P. Rasson. 1994. Simulated annealing: a proof of convergence. *IEEE Trans. Pattern Anal. Mach. Intell.* 16, 652 – 656.
- Hamodrakas, S. J., J.H. Willis, and V.A. Iconomidou. 2002. A structural model of the chitin-binding domain of cuticle proteins. *Insect Biochem. and Mol. Biol.* 32:1577–1583.
- Hoos, H.H., and T. Stützle. 2005. Stochastic Local Search: Foundations and Applications, Morgan Kaufmann, San Francisco. 37-51.
- Huey, R., D.S. Goodsell, G.M. Morris, and A.J. Olson. 2004. Grid-based hydrogen bond potentials with improved directionality. *Lett. Drug Des. Discov.* 1:178-183.
- Huey, R., G.M. Morris, A.J. Olson, and D.S. Goodsell. 2007. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* 28:1145-1152.
- Iconomidou, V.A., J.H. Willis, and S.J. Hamodrakas. 1999. Is beta-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle? *Insect Biochem. and Mol. Biol.* 29:285–292.
- Iconomidou, V.A., J.H. Willis, and S.J. Hamodrakas. 2005. Unique features of the structural model of ‘hard’ cuticle proteins: implications for chitin–protein interactions and cross-linking in cuticle. *Insect Biochem. and Mol. Biol.* 35:553-560.
- Karouzou, M.V., Y. Spyropoulos, V.A. Iconomidou, R.S. Cornman, S.J. Hamodrakas, and J.H. Willis. 2007. Drosophila cuticular proteins with the R&R Consensus: Annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences. *Insect Biochem. and Mol. Biol.* 37:754-760.
- Kelley, L.A., and M.J.E. Sternberg. 2009. Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols.* 4:363–371.
- Kolhekar A. S., H.T. Keutmann, R.E. Mains, A.S. Quon, and B.A. Eipper. 1997. Peptidylglycine alpha -hydroxylating monooxygenase: active site residues, disulfide linkages, and a two-domain model of the catalytic core. *Biochemistry.* 36:10901–10909.

Kong, J., C.A. White, A.I. Krylov, D. Sherrill, R.D. Adamson, T.R. Furlani, M.S. Lee, A.M. Lee, S.R. Gwaltney, T.R. Adams, C. Ochsenfeld, A.T.B. Gilbert, G.S. Kedziora, V.A. Rassolov, D.R. Maurice, N. Nair, Y. Shao, N.A. Besley, P.E. Maslen, J.P. Dombroski, H. Daschel, W. Zhang, P.P. Korambath, J. Baker, E.F.C. Byrd, T.V. Voorhis, M. Oumi, S. Hirata, C.P. Hsu, N. Ishikawa, J. Florian, A. Warshel, B.G. Johnson, P.M.W. Gill, M. Head-Gordon, and J.A. Pople. 2000. Q-Chem 2.0: a high-performance ab initio electronic structure program package. *J. Comput. Chem.* 21:1532-1548.

Larkin M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. ClustalW and ClustalX version 2. *Bioinformatics.* 23:2947-2948.

Leach, A.R. 2001. *Molecular Modelling: Principles and Applications* (2nd Ed). Prentice Hall, Harlow. 768 pp.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. 1953. Equations of State Calculations by Fast Computing Machines. *J. Chem. Phys.* 21:1087-1092.

Morris, G.M., D.S. Goodsell, M.E. Pique, W. Lindstrom, R. Huey, S. Forli, W.E. Hart, S. Halliday, R. Belew, and A.J. Olson. 2009. Autodock4.2 Userguide. The Scripps Research Institute. Available at: [HTTP://autodock.scripps.edu/faqshelp/manual/autodock-4-2-user-guide](http://autodock.scripps.edu/faqshelp/manual/autodock-4-2-user-guide)

Morris, G. M., D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, and A.J. Olson. 1998. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.* 19:1639-1662.

Pearlman, D. A., D.A. Case, J.W. Caldwell, W.S. Ross, I.T.E Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P.A. Kollman. 1995. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comp. Phys. Commun.* 91:1-41.

Perrot, P. 1998. *A to Z of Thermodynamics*. Oxford University Press, New York. 336 pp.

Quioco, F.A. 1989. Protein-carbohydrate interactions: basic molecular features. *Pure and Appl. Chem.* 61:1293-1306.

Rapaport, D.C. 2004. *The Art of Molecular Dynamics Simulation* (2nd Ed). Cambridge University Press, Cambridge. 1-44.


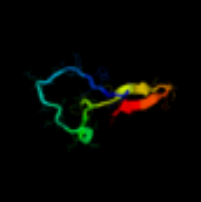


Rask L, H. Anundi, J. Fohlman, and P.A. Peterson. 1987. The complete amino acid sequence of human serum retinol-binding protein. *Ups. J. Med. Sci.* 92:115-46.


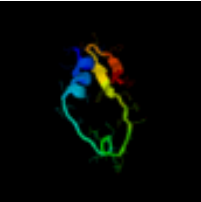


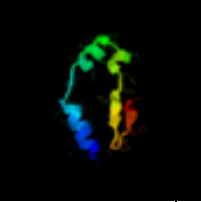


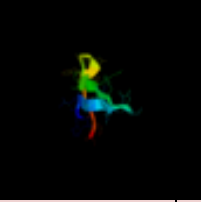

- Rebers, J.E., and L.M. Riddiford. 1988. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. *J. Mol. Biol.* 203:411–423.
- Rebers, J.E., and J.H. Willis. 2001. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem. and Mol. Biol.* 31:1083-1093.
- Rost, B., G. Yachdav, and J. Liu. 2004. The PredictProtein Server. *Nucleic Acids Res.* 32 (Web Server issue):W321-W326.
- Sali, A., and T.L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779-815.
- Sayle, R., and E.J. Milner-White. 1995. RasMol: Biomolecular graphics for all. *Trends Biochem. Sci.* 20:374.
- Shen, Z., and M. Jacobs-Lorena. 1999. Evolution of chitin-binding proteins in invertebrates. *J. Mol. Evol.* 48:341–347.
- Siebert, X., B.A. Eipper, R.E. Mains, S.T. Prigge, N.J. Blackburn, and L.M. Amzel. 2005. The Catalytic Copper of Peptidylglycine α -Hydroxylating Monooxygenase also plays a Critical Structural Role. *Biophys J.* 89:3312–3319.
- Simpson, H.D., and F. Barras. 1999. Functional Analysis of the Carbohydrate-Binding Domains of *Erwinia chrysanthemi* Cel5 (Endoglucanase Z) and an *Escherichia coli* Putative Chitinase. *J. Bacteriol.* 181:4611-4616.
- Togawa T., H. Nakato, and S. Izumi. 2004. Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, *Bombyx mori*. *Insect Biochem. and Mol. Biol.* 34:1059-1067.
- Willis, J.H. 1999. Cuticular proteins in insects and crustaceans. *Am. Zool.* 39:600–609.
- Willis, J.H. 2010. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem. and Mol. Biol.* (Epub ahead of print).
- Willis, J.H., V.A. Iconomidou, R.F. Smith, and S.J. Hamodrakas. 2005. Cuticular Proteins. In: Gilbert, L.I., Iatrou, K., Gill, S.S., editors. *Comprehensive Molecular Insect Science*. Elsevier, Oxford. 79–110.
- Wang, S., Y. Wang, W. Du, F. Sun, X. Wang, C. Zhou, and Y. Liang. 2007. A multi-approaches-guided genetic algorithm with application to operon prediction. *Artif. Intell. Med.* 41:151–159.


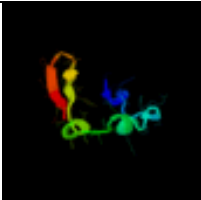


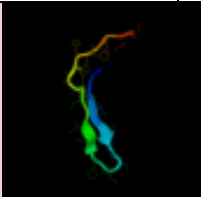


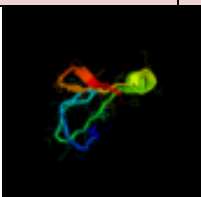

Appendix A

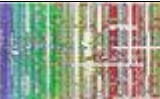
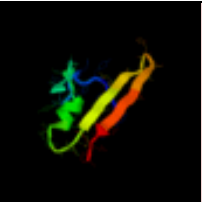


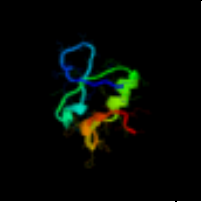

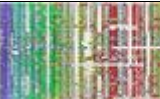
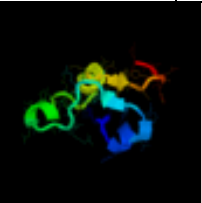

QuickPhyre results

Sequence alignment search result for QuickPhyre

View Alignments	SCOP Code	View Model	E-value	Estimated Precision	BioText	Fold/PDB descriptor	Superfamily	Family	(beta-test)
	d1tza_a (length:132) 11% i.d.	  	16	10 %	n/a	Immunoglobulin-like beta-sandwich	ApaG-like	ApaG-like	n/a

 c2f1eA (length:127) 14% i.d.	 	18	10 %	n/a	PDB header: structural genomics, unknown function	Chain: A: PDB Molecule: protein apag;	PDBTitle: solution structure of apag protein	n/a
 d1xq4a (length:123) 14% i.d.	 	19	10 %	n/a	Immunoglobulin-like beta-sandwich	ApaG-like	ApaG-like	n/a
 c2pnwA (length:380) 12% i.d.	 	21	10 %	n/a	PDB header: hydrolase	Chain: A: PDB Molecule: membrane-bound lytic murein transglycosylase;	PDBTitle: crystal structure of membrane-bound lytic murein2 transglycosylase from agrobacterium tumefaciens	n/a


	d1xvs_a (length:124) 14% i.d.		23	5 %	n/a	Immunoglobulin-like beta-sandwich	ApaG-like	ApaG-like	n/a
									
	c2g5d_A (length:422) 6% i.d.		24	5 %	n/a	PDB header: hydrolase	Chain: A: PDB Molecule: gna33;	PDBTitle: crystal structure of mlta from neisseria gonorrhoeae2 monoclinic form	n/a
									
	d1h8l_a1 (length:79) 18% i.d.		28	5 %	n/a	Prealbumin-like	Carboxypeptidase regulatory domain-like	Carboxypeptidase regulatory domain	n/a
									

 c2apnA (length:114) 20% i.d.	 	31	5 %	n/a	PDB header: structural genomics, unknown function	Chain: A: PDB Molecule: protein hi1723;	PDBTitle: hi1723 solution structure	n/a
 d1s98a (length:97) 20% i.d.	 	31	5 %	n/a	HesB-like domain	HesB-like domain	HesB-like domain	n/a
 c1hfiA (length:62) 15% i.d.	 	31	5 %	n/a	PDB header: glycoprotein	Chain: A: PDB Molecule: factor h, 15th c-module pair;	PDBTitle: solution structure of a pair of complement modules by2 nuclear magnetic resonance	n/a

Appendix B


PSI-BLAST Results

PSI-Blast results from search using AGCP2B sequence

[pdb|1Q5F|A](#)  Chain A, Nmr Structure Of Type Ivb Pilin (Pils) From Salmonella Typhi
Length=156

Score = 26.6 bits (57), Expect = 2.6, Method: Compositional matrix adjust.
Identities = 13/38 (34%), Positives = 18/38 (47%), Gaps = 0/38 (0%)

```
Query 21 SQHETRHGDEVHVGQYSLLDSDGHQRIVDYHADHHTGFN 58
      ++ T GD G +L +S G Q +V TGFN
Sbjct 49 AKGMTVSGDPASGSATLWNSWGGQIVVAPDTAGGTGFN 86
```

[pdb|1YI9|A](#)  Chain A, Crystal Structure Analysis Of The Oxidized Form Of The M314i Mutant Of Peptidylglycine Alpha-Hydroxylating Monooxygenase
Length=309

Score = 24.6 bits (52), Expect = 7.9, Method: Composition-based stats.
Identities = 9/22 (40%), Positives = 13/22 (59%), Gaps = 0/22 (0%)

```
Query 7 FSYSVHDEHTGDIKSQHETRHG 28
      F+Y VH H G + S + R+G
Sbjct 191 FAYRVHTHHLGKVVSGYRVRNG 212
```

▣ [pdb|1SDW|A](#) **S** Chain A, Reduced (Cu+) Peptidylglycine Alpha-Hydroxylating
Monooxygenase
With Bound Peptide And Dioxygen
Length=314

Score = 24.6 bits (52), Expect = 7.9, Method: Composition-based stats.
Identities = 9/22 (40%), Positives = 13/22 (59%), Gaps = 0/22 (0%)

```
Query   7      FSYSVHDEHTGDIKSQHETRHG   28
          F+Y VH  H G + S +  R+G
Sbjct  195    FAYRVHTHHLGKVVSGYRVRNG   216
```

▣ [pdb|2VDU|B](#) **S** Chain B, Structure Of Trm8-Trm82, The Yeast Trna M7g Methylation
Complex
[pdb|2VDU|D](#) **S** Chain D, Structure Of Trm8-Trm82, The Yeast Trna M7g Methylation
Complex
Length=450

Score = 24.6 bits (52), Expect = 8.3, Method: Composition-based stats.
Identities = 18/50 (36%), Positives = 24/50 (48%), Gaps = 8/50 (16%)

```
Query   13      DEHTGDIKSQHETRHGDE-VHGQYSLL-----DSDGHQRIVDYHADHH   54
          D ++ DI S  E +   E + G  S+L           DSDGHQ I+   D H
Sbjct  172    DVYSIDINSIPEEKFTQEPILGHVSMMLTDVHHLIKDSDGHQFIITSDRDEH   221
```

☐ [pdb|1YIP|A](#) **S** Chain A, Oxidized Peptidylglycine Alpha-Hydroxylating Monooxygenase (Phm) In A New Crystal Form
Length=311

Score = 24.6 bits (52), Expect = 8.3, Method: Composition-based stats.
Identities = 9/22 (40%), Positives = 13/22 (59%), Gaps = 0/22 (0%)

```
Query 7 FSYSVHDEHTGDIKSQHETRHG 28
      F+Y VH H G + S + R+G
Sbjct 193 FAYRVHTHHLGKVVSGYRVRNG 214
```


☐ [pdb|1OPM|A](#) **S** Chain A, Oxidized (Cu²⁺) Peptidylglycine Alpha-Hydroxylating Monooxygenase (Phm) With Bound Substrate

[pdb|3PHM|A](#) **S** Chain A, Reduced (Cu⁺) Peptidylglycine Alpha-Hydroxylating Monooxygenase (Phm)

[pdb|1PHM|A](#) **S** Chain A, Peptidylglycine Alpha-Hydroxylating Monooxygenase (Phm)
From Rat
Length=310


Score = 24.6 bits (52), Expect = 8.4, Method: Composition-based stats.
Identities = 9/22 (40%), Positives = 13/22 (59%), Gaps = 0/22 (0%)

```
Query 7 FSYSVHDEHTGDIKSQHETRHG 28
      F+Y VH H G + S + R+G
Sbjct 193 FAYRVHTHHLGKVVSGYRVRNG 214
```

☐ [pdb|1YJK|A](#)  Chain A, Reduced Peptidylglycine Alpha-Hydroxylating Monooxygenase (Phm) In A New Crystal Form
Length=306

Score = 24.6 bits (52), Expect = 8.8, Method: Composition-based stats.
Identities = 9/22 (40%), Positives = 13/22 (59%), Gaps = 0/22 (0%)

```
Query 7 FSYSVHDEHTGDIKSQHETRHG 28
      F+Y VH H G + S + R+G
Sbjct 188 FAYRVHTHHLGKVVSGYRVRNG 209
```

☐ [pdb|1YJL|A](#)  Chain A, Reduced Peptidylglycine Alpha-Hydroxylating Monooxygenase In A New Crystal Form
Length=306

Score = 24.6 bits (52), Expect = 8.8, Method: Composition-based stats.
Identities = 9/22 (40%), Positives = 13/22 (59%), Gaps = 0/22 (0%)

```
Query 7 FSYSVHDEHTGDIKSQHETRHG 28
      F+Y VH H G + S + R+G
Sbjct 188 FAYRVHTHHLGKVVSGYRVRNG 209
```

[pdb|2JIX|B](#) **S** Chain B, Crystal Structure Of Abt-007 Fab Fragment With The Soluble Domain Of Epo Receptor
[pdb|2JIX|C](#) **S** Chain C, Crystal Structure Of Abt-007 Fab Fragment With The Soluble Domain Of Epo Receptor
[pdb|2JIX|E](#) **S** Chain E, Crystal Structure Of Abt-007 Fab Fragment With The Soluble Domain Of Epo Receptor
Length=225

Score = 24.6 bits (52), Expect = 9.0, Method: Composition-based stats.
Identities = 8/14 (57%), Positives = 9/14 (64%), Gaps = 0/14 (0%)

```
Query 1 APANYEFSYSVHDE 14
      P NY FSY + DE
Sbjct 49 GPGNYSFSYQLEDE 62
```

[pdb|1ERN|A](#) **S** Chain A, Native Structure Of The Extracellular Domain Of Erythropoietin (Epo) Receptor [ebp]
[pdb|1ERN|B](#) **S** Chain B, Native Structure Of The Extracellular Domain Of Erythropoietin (Epo) Receptor [ebp]
Length=213

Score = 24.6 bits (52), Expect = 9.0, Method: Composition-based stats.
Identities = 8/14 (57%), Positives = 9/14 (64%), Gaps = 0/14 (0%)

```
Query 1 APANYEFSYSVHDE 14
      P NY FSY + DE
Sbjct 40 GPGNYSFSYQLEDE 53
```

☐ [pdb|1EBP|A](#) **S** Chain A, Complex Between The Extracellular Domain Of Erythropoietin (Epo) Receptor [ebp] And An Agonist Peptide [emp1]
[pdb|1EBP|B](#) **S** Chain B, Complex Between The Extracellular Domain Of Erythropoietin (Epo) Receptor [ebp] And An Agonist Peptide [emp1]
Length=211

Score = 24.6 bits (52), Expect = 9.0, Method: Composition-based stats.
Identities = 8/14 (57%), Positives = 9/14 (64%), Gaps = 0/14 (0%)

```
Query 1 APANYEFSYSVHDE 14
      P NY FSY + DE
Sbjct 40 GPGNYSFSYQLEDE 53
```

☐ [pdb|1EBA|A](#) **S** Chain A, Complex Between The Extracellular Domain Of Erythropoietin (Epo) Receptor [ebp] And An Inactive Peptide [emp33]
Contains 3,5-Dibromotyrosine In Position 4 (Denoted Dby)
[pdb|1EBA|B](#) **S** Chain B, Complex Between The Extracellular Domain Of Erythropoietin (Epo) Receptor [ebp] And An Inactive Peptide [emp33]
Contains 3,5-Dibromotyrosine In Position 4 (Denoted Dby)
Length=215

Score = 24.6 bits (52), Expect = 9.0, Method: Composition-based stats.
Identities = 8/14 (57%), Positives = 9/14 (64%), Gaps = 0/14 (0%)

```
Query 1 APANYEFSYSVHDE 14
      P NY FSY + DE
Sbjct 40 GPGNYSFSYQLEDE 53
```

▣ [pdb|2GRY|A](#) **S** Chain A, Crystal Structure Of The Human Kif2 Motor Domain In
Complex With Adp
Length=420

Score = 24.6 bits (52), Expect = 9.6, Method: Composition-based stats.
Identities = 12/31 (38%), Positives = 19/31 (61%), Gaps = 1/31 (3%)

```
Query  26  RHGDEVHGQYSLLDSDGHQRIVDY-HADHHT  55
      R   ++HG++SL+D  G++R  D   AD  T
Sbjct  311  RRKGKLGKFLIDLGNERNAGADTSSADRQT  341
```

Appendix C

Output information from AutoGrid for various mutations attempted

Grid Map	Atom Type	Minimum Energy (kcal/mol)	Maximum Energy (kcal/mol)	
1	C	-0.70	2.01e+05	
2	HD	-0.68	1.13e+05	
3	OA	-1.32	2.00e+05	
4	N	-0.67	2.00e+05	
5	e	-34.21	2.91e+01	Electrostatic Potential
6	d	0.00	1.02e+00	Desolvation Potential

Fig. 26 Typical results from AutoGrid energy calculations for AGCP2B_M1. Minimum and maximum energy estimations are reported in kcal/mol. Minimum van der Waals' energies and hydrogen bonding energies are typically -10 to -1 kcal/mol, while maximum van der Waals' energies are clamped at +105 kcal/mol. Electrostatic potentials tend to range from around -103 to +103 kcal/mol/e. C represents carbon; HD represents hydrogen; OA represents oxygen; N represents a nitrogen that cannot accept hydrogen bonds.

Grid Map	Atom Type	Minimum Energy (kcal/mol)	Maximum Energy (kcal/mol)	
1	C	-0.75	2.02e+05	
2	HD	-0.69	1.09e+05	
3	OA	-0.83	2.00e+05	
4	N	-0.73	2.01e+05	
5	e	-40.66	3.55e+01	Electrostatic Potential
6	d	0.00	1.02e+00	Desolvation Potential

Fig. 27 Typical results from AutoGrid energy calculations for AGCP2B_M2. Minimum and maximum energy estimations are reported in kcal/mol. Minimum van der Waals' energies and hydrogen bonding energies are typically -10 to -1 kcal/mol, while maximum van der Waals' energies are clamped at +105 kcal/mol. Electrostatic potentials tend to range from around -103 to +103 kcal/mol/e. C represents carbon; HD represents hydrogen; OA represents oxygen; N represents a nitrogen that cannot accept hydrogen bonds.

Grid Map	Atom Type	Minimum Energy (kcal/mol)	Maximum Energy (kcal/mol)	
1	C	-0.77	2.03e+05	
2	HD	-0.68	1.12e+05	
3	OA	-0.85	2.00e+05	
4	N	-0.75	2.00e+05	
5	e	-34.31	4.44e+01	Electrostatic Potential
6	d	0.00	1.04e+00	Desolvation Potential

Fig. 28 Typical results from AutoGrid energy calculations for AGCP2B_M3. Minimum and maximum energy estimations are reported in kcal/mol. Minimum van der Waals' energies and hydrogen bonding energies are typically -10 to -1 kcal/mol, while maximum van der Waals' energies are clamped at +105 kcal/mol. Electrostatic potentials tend to range from around -103 to +103 kcal/mol/e. C represents carbon; HD represents hydrogen; OA represents oxygen; N represents a nitrogen that cannot accept hydrogen bonds.

Grid Map	Atom Type	Minimum Energy (kcal/mol)	Maximum Energy (kcal/mol)	
1	C	-0.76	2.01e+05	
2	HD	-0.68	1.08e+05	
3	OA	-0.84	2.00e+05	
4	N	-0.75	2.00e+05	
5	e	-36.92	3.20e+01	Electrostatic Potential
6	d	0.00	1.06e+00	Desolvation Potential

Fig. 29 Typical results from AutoGrid energy calculations for AGCP2B_M4. Minimum and maximum energy estimations are reported in kcal/mol. Minimum van der Waals' energies and hydrogen bonding energies are typically -10 to -1 kcal/mol, while maximum van der Waals' energies are clamped at +105 kcal/mol. Electrostatic potentials tend to range from around -103 to +103 kcal/mol/e. C represents carbon; HD represents hydrogen; OA represents oxygen; N represents a nitrogen that cannot accept hydrogen bonds.

Grid Map	Atom Type	Minimum Energy (kcal/mol)	Maximum Energy (kcal/mol)	
1	C	-0.67	2.01e+05	
2	HD	-0.67	1.10e+05	
3	OA	-0.74	2.00e+05	
4	N	-0.65	2.00e+05	
5	e	-37.88	4.38e+01	Electrostatic Potential
6	d	0.00	9.28e-01	Desolvation Potential

Fig. 30 Typical results from AutoGrid energy calculations for AGCP2B_M5. Minimum and maximum energy estimations are reported in kcal/mol. Minimum van der Waals' energies and hydrogen bonding energies are typically -10 to -1 kcal/mol, while maximum van der Waals' energies are clamped at +105 kcal/mol. Electrostatic potentials tend to range from around -103 to +103 kcal/mol/e. C represents carbon; HD represents hydrogen; OA represents oxygen; N represents a nitrogen that cannot accept hydrogen bonds.

Grid Map	Atom Type	Minimum Energy (kcal/mol)	Maximum Energy (kcal/mol)	
1	C	-0.81	2.01e+05	
2	HD	-0.68	1.10e+05	
3	OA	-0.87	2.00e+05	
4	N	-0.76	2.00e+05	
5	e	-30.97	4.23e+01	Electrostatic Potential
6	d	0.00	1.08e+00	Desolvation Potential

Fig. 31 Typical results from AutoGrid energy calculations for AGCP2B_M6. Minimum and maximum energy estimations are reported in kcal/mol. Minimum van der Waals' energies and hydrogen bonding energies are typically -10 to -1 kcal/mol, while maximum van der Waals' energies are clamped at +105 kcal/mol. Electrostatic potentials tend to range from around -103 to +103 kcal/mol/e. C represents carbon; HD represents hydrogen; OA represents oxygen; N represents a nitrogen that cannot accept hydrogen bonds.

Grid Map	Atom Type	Minimum Energy (kcal/mol)	Maximum Energy (kcal/mol)	
1	C	-0.71	2.03e+05	
2	HD	-0.69	1.09e+05	
3	OA	-0.89	2.00e+05	
4	N	-0.71	2.00e+05	
5	e	-36.97	4.01e+01	Electrostatic Potential
6	d	0.00	1.04e+00	Desolvation Potential

Fig. 32 Typical results from AutoGrid energy calculations for AGCP2B_M7. Minimum and maximum energy estimations are reported in kcal/mol. Minimum van der Waals' energies and hydrogen bonding energies are typically -10 to -1 kcal/mol, while maximum van der Waals' energies are clamped at +105 kcal/mol. Electrostatic potentials tend to range from around -103 to +103 kcal/mol/e. C represents carbon; HD represents hydrogen; OA represents oxygen; N represents a nitrogen that cannot accept hydrogen bonds.

Appendix D

Clustering Histogram data for AutoDock runs including the mutated AGCP2B

Table. 5 Clustering histogram for AGCP2B_M1 Run

Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram
30					5 10 15 20 25
					____:____ ____:____ ____:____
1	+0.70	7	+0.70	1	#
2	+2.42	14	+2.42	1	#
3	+3.21	43	+3.21	1	#
4	+3.31	19	+3.31	1	#
5	+3.40	37	+3.40	1	#
6	+3.41	17	+3.41	1	#
7	+3.61	2	+3.61	1	#
8	+3.62	47	+4.49	2	##
9	+3.66	41	+3.66	1	#
10	+3.69	8	+3.69	1	#
11	+4.03	20	+4.03	1	#
12	+4.14	15	+4.14	1	#
13	+4.21	32	+4.21	1	#
14	+4.27	4	+4.27	1	#
15	+4.32	38	+4.32	1	#
16	+4.40	9	+4.40	1	#
17	+4.45	11	+4.45	1	#
18	+4.73	13	+4.73	1	#
19	+4.77	26	+4.77	1	#
20	+4.98	12	+4.98	1	#
21	+5.03	42	+5.03	1	#
22	+5.14	25	+5.38	2	##
23	+5.16	18	+5.16	1	#
24	+5.31	50	+5.31	1	#
25	+5.31	48	+5.31	1	#
26	+5.44	27	+5.44	1	#
27	+5.67	10	+5.67	1	#
28	+5.73	22	+5.73	1	#
29	+5.80	1	+5.80	1	#
30	+5.87	49	+5.87	1	#
31	+6.00	46	+6.00	1	#
32	+6.05	45	+6.05	1	#
33	+6.10	3	+6.10	1	#
34	+6.15	5	+6.15	1	#
35	+6.18	21	+6.18	1	#
36	+6.19	23	+6.19	1	#
37	+6.31	39	+6.31	1	#
38	+6.68	30	+6.68	1	#
39	+6.72	33	+6.72	1	#
40	+6.79	35	+6.79	1	#
41	+6.88	40	+6.88	1	#
42	+6.89	34	+6.89	1	#
43	+6.94	6	+6.94	1	#
44	+6.95	24	+6.95	1	#
45	+7.14	36	+7.14	1	#
46	+7.15	16	+7.15	1	#
47	+7.62	44	+7.62	1	#
48	+8.75	31	+8.75	1	#

Table 5. Clustering histogram showing the results of the AutoDock run for the first mutation.

Table. 6 Clustering histogram for AGCP2B_M2 Run

Clus	Lowest	Run	Mean	Num	Histogram				
-ter	Binding		Binding	in					
Rank	Energy		Energy	Clus	5	10	15	20	25
30									
					:	:	:	:	:
1	+1.91	21	+1.91	1	#				
2	+2.13	11	+2.23	2	##				
3	+2.98	37	+2.98	1	#				
4	+3.22	35	+4.03	2	##				
5	+3.23	4	+3.23	1	#				
6	+3.23	27	+3.23	1	#				
7	+3.46	32	+4.09	2	##				
8	+3.82	50	+4.17	2	##				
9	+3.95	26	+3.95	1	#				
10	+3.99	6	+3.99	1	#				
11	+4.02	9	+4.52	2	##				
12	+4.07	44	+4.07	1	#				
13	+4.14	39	+4.14	1	#				
14	+4.14	5	+4.14	1	#				
15	+4.16	19	+4.16	1	#				
16	+4.29	40	+4.29	1	#				
17	+4.44	48	+4.44	1	#				
18	+4.47	2	+4.47	1	#				
19	+4.49	16	+4.49	1	#				
20	+4.51	33	+4.51	1	#				
21	+4.70	17	+4.70	1	#				
22	+4.72	7	+4.93	2	##				
23	+4.75	34	+4.75	1	#				
24	+4.91	13	+4.91	1	#				
25	+5.18	18	+5.18	1	#				
26	+5.25	1	+5.25	1	#				
27	+5.27	24	+5.65	2	##				
28	+5.30	12	+5.30	1	#				
29	+5.31	28	+5.53	2	##				
30	+5.36	25	+5.36	1	#				
31	+5.37	43	+5.37	1	#				
32	+5.49	8	+5.49	1	#				
33	+5.94	47	+5.94	1	#				
34	+6.14	46	+6.14	1	#				
35	+6.21	22	+6.21	1	#				
36	+6.29	10	+6.29	1	#				
37	+6.42	38	+6.42	1	#				
38	+6.56	15	+6.56	1	#				
39	+6.62	23	+6.62	1	#				
40	+7.28	45	+7.28	1	#				
41	+7.78	20	+7.78	1	#				
42	+8.15	41	+8.15	1	#				

Table 6. Clustering histogram showing the results of the AutoDock run for the second mutation.

Table. 7 Clustering histogram for AGCP2B_M3 Run

Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram
					5 10 15 20 25 30
					____:____ ____:____ ____:____
1	+1.21	10	+1.21	1	#
2	+1.66	27	+2.93	2	##
3	+1.98	21	+1.98	1	#
4	+1.98	31	+1.98	1	#
5	+2.60	4	+2.60	1	#
6	+2.83	41	+2.83	1	#
7	+2.86	26	+2.86	1	#
8	+2.91	47	+2.91	1	#
9	+2.94	45	+2.94	1	#
10	+3.01	24	+3.01	1	#
11	+3.04	38	+3.04	1	#
12	+3.12	14	+3.12	1	#
13	+3.45	42	+3.45	1	#
14	+3.48	36	+3.48	1	#
15	+3.60	46	+3.60	1	#
16	+3.61	37	+3.61	1	#
17	+3.70	30	+3.70	1	#
18	+3.75	11	+3.75	1	#
19	+3.77	2	+4.38	2	##
20	+3.79	25	+3.79	1	#
21	+3.90	50	+3.90	1	#
22	+3.98	28	+3.98	1	#
23	+4.04	34	+4.04	1	#
24	+4.10	9	+4.10	1	#
25	+4.16	32	+4.16	1	#
26	+4.24	23	+4.24	1	#
27	+4.50	8	+4.50	1	#
28	+4.51	15	+4.51	1	#
29	+4.58	17	+4.58	1	#
30	+4.59	49	+4.59	1	#
31	+4.65	18	+4.65	1	#
32	+4.67	12	+4.67	1	#
33	+4.69	35	+4.69	1	#
34	+4.82	29	+4.82	1	#
35	+4.88	1	+4.88	1	#
36	+4.95	6	+4.95	1	#
37	+4.96	40	+4.96	1	#
38	+4.98	44	+4.98	1	#
39	+5.32	43	+5.32	1	#
40	+5.36	20	+5.36	1	#
41	+5.47	39	+5.47	1	#
42	+5.55	48	+5.55	1	#
43	+5.57	3	+5.57	1	#
44	+5.60	33	+5.60	1	#
45	+5.61	22	+5.61	1	#
46	+6.11	7	+6.11	1	#
47	+6.69	5	+6.69	1	#
48	+6.92	13	+6.92	1	#

Table 7. Clustering histogram showing the results of the AutoDock run for the third mutation.

Table. 8 Clustering histogram for AGCP2B_M4 Run

Clus	Lowest	Run	Mean	Num	Histogram				
-ter	Binding		Binding	in					
Rank	Energy		Energy	Clus	5	10	15	20	25
30									
					:	:	:	:	:
1	+0.96	39	+1.49	2	##				
2	+0.99	3	+0.99	1	#				
3	+1.34	29	+2.91	3	###				
4	+1.52	45	+1.52	1	#				
5	+2.28	26	+3.81	2	##				
6	+2.31	41	+2.31	1	#				
7	+2.43	14	+3.98	3	###				
8	+2.73	28	+2.89	2	##				
9	+2.93	50	+2.93	1	#				
10	+3.35	23	+3.35	1	#				
11	+3.58	21	+3.58	1	#				
12	+3.58	7	+3.58	1	#				
13	+3.59	36	+3.59	1	#				
14	+3.71	11	+3.71	1	#				
15	+3.86	42	+3.86	1	#				
16	+3.96	17	+4.92	2	##				
17	+4.15	10	+4.15	1	#				
18	+4.25	18	+4.73	2	##				
19	+4.47	12	+4.47	1	#				
20	+4.64	1	+4.64	1	#				
21	+4.66	34	+4.66	1	#				
22	+4.69	40	+4.69	1	#				
23	+4.74	27	+4.74	1	#				
24	+4.87	37	+4.87	1	#				
25	+5.02	19	+5.02	1	#				
26	+5.10	31	+5.10	1	#				
27	+5.51	4	+5.51	1	#				
28	+5.55	13	+5.55	1	#				
29	+5.60	25	+5.60	1	#				
30	+5.60	43	+5.60	1	#				
31	+5.78	5	+5.78	1	#				
32	+5.99	49	+5.99	1	#				
33	+6.09	2	+6.09	1	#				
34	+6.28	6	+6.28	1	#				
35	+6.43	9	+6.43	1	#				
36	+6.46	15	+6.46	1	#				
37	+6.48	48	+6.48	1	#				
38	+6.63	22	+6.63	1	#				
39	+6.70	24	+6.70	1	#				
40	+7.65	8	+7.65	1	#				
41	+7.84	16	+7.84	1	#				

Table 8. Clustering histogram showing the results of the AutoDock run for the fourth mutation.

Table. 9 Clustering histogram for AGCP2B_M5 Run

Clus	Lowest	Run	Mean	Num	Histogram				
-ter	Binding		Binding	in					
Rank	Energy		Energy	Clus	5	10	15	20	25
30									
					_____:	_____:	_____:	_____:	_____:
1	+1.82	38	+1.82	1	#				
2	+1.86	40	+1.86	1	#				
3	+1.96	5	+1.96	1	#				
4	+2.11	11	+2.11	1	#				
5	+2.35	35	+2.35	1	#				
6	+2.66	28	+2.66	1	#				
7	+2.96	19	+3.98	2	##				
8	+3.10	16	+3.10	1	#				
9	+3.12	34	+3.65	2	##				
10	+3.23	36	+3.23	1	#				
11	+3.43	32	+3.43	1	#				
12	+3.43	17	+3.43	1	#				
13	+3.47	31	+3.47	1	#				
14	+3.49	46	+3.49	1	#				
15	+3.49	24	+3.49	1	#				
16	+3.49	21	+3.49	1	#				
17	+3.62	18	+3.62	1	#				
18	+3.64	47	+3.64	1	#				
19	+3.90	41	+3.90	1	#				
20	+3.93	9	+3.93	1	#				
21	+4.06	15	+4.06	1	#				
22	+4.06	39	+4.06	1	#				
23	+4.24	30	+4.24	1	#				
24	+4.25	25	+4.50	2	##				
25	+4.27	42	+4.27	1	#				
26	+4.29	13	+4.29	1	#				
27	+4.38	4	+4.38	1	#				
28	+4.48	45	+4.48	1	#				
29	+4.58	29	+4.58	1	#				
30	+4.78	22	+4.78	1	#				
31	+4.84	50	+4.84	1	#				
32	+4.91	26	+4.91	1	#				
33	+4.97	33	+4.97	1	#				
34	+4.99	43	+4.99	1	#				
35	+5.03	49	+5.03	1	#				
36	+5.07	6	+5.07	1	#				
37	+5.10	20	+5.10	1	#				
38	+5.14	14	+5.14	1	#				
39	+5.22	37	+5.22	1	#				
40	+5.46	27	+5.46	1	#				
41	+5.56	1	+5.56	1	#				
42	+5.64	3	+5.87	2	##				
43	+5.69	48	+5.69	1	#				
44	+6.13	7	+6.13	1	#				
45	+6.41	10	+6.41	1	#				
46	+6.86	44	+6.86	1	#				

Table 9. Clustering histogram showing the results of the AutoDock run for the fifth mutation.

Table. 10 Clustering histogram for AGCP2B_M6 Run

Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram						
					5	10	15	20	25	30	
					:	:	:	:	:	:	
1	+1.97	42	+1.97	1	#						
2	+2.69	44	+2.96	2	##						
3	+2.86	1	+2.86	1	#						
4	+3.20	35	+3.20	1	#						
5	+3.26	41	+3.26	1	#						
6	+3.38	47	+3.38	1	#						
7	+3.40	31	+3.40	1	#						
8	+3.50	2	+3.50	1	#						
9	+3.50	50	+3.50	1	#						
10	+3.51	32	+3.51	1	#						
11	+3.58	37	+3.58	1	#						
12	+3.76	5	+3.99	2	##						
13	+3.88	7	+3.88	1	#						
14	+3.93	13	+3.93	1	#						
15	+3.97	8	+3.97	1	#						
16	+4.25	27	+4.25	1	#						
17	+4.25	20	+4.25	1	#						
18	+4.57	49	+4.57	1	#						
19	+4.62	6	+4.62	1	#						
20	+4.72	14	+4.72	1	#						
21	+4.76	23	+4.76	1	#						
22	+4.98	16	+4.98	1	#						
23	+5.11	12	+5.11	1	#						
24	+5.20	25	+5.20	1	#						
25	+5.29	38	+5.29	1	#						
26	+5.39	36	+5.39	1	#						
27	+5.39	40	+5.39	1	#						
28	+5.43	3	+5.43	1	#						
29	+5.47	43	+5.47	1	#						
30	+5.47	45	+5.47	1	#						
31	+5.54	24	+5.54	1	#						
32	+5.77	4	+5.77	1	#						
33	+5.88	34	+5.88	1	#						
34	+5.93	48	+5.93	1	#						
35	+5.93	15	+5.93	1	#						
36	+6.00	33	+6.00	1	#						
37	+6.15	26	+6.15	1	#						
38	+6.15	22	+6.15	1	#						
39	+6.38	39	+6.38	1	#						
40	+6.48	18	+6.48	1	#						
41	+6.54	17	+6.54	1	#						
42	+6.66	9	+6.66	1	#						
43	+6.70	29	+6.70	1	#						
44	+6.72	28	+6.72	1	#						
45	+6.84	10	+6.84	1	#						
46	+7.04	11	+7.04	1	#						
47	+7.23	30	+7.23	1	#						
48	+7.64	19	+7.64	1	#						

Table 10. Clustering histogram showing the results of the AutoDock run for the sixth mutation.

Table. 11 Clustering histogram for AGCP2B_M7 Run

Clus-ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram
					5 10 15 20 25 30
					____:____ ____:____ ____:____
1	+0.04	40	+0.04	1	#
2	+1.03	17	+1.03	1	#
3	+1.29	13	+2.34	4	####
4	+1.50	29	+1.50	1	#
5	+1.65	1	+1.65	1	#
6	+1.66	31	+1.66	1	#
7	+1.67	41	+1.67	1	#
8	+1.93	18	+2.49	2	##
9	+2.30	27	+2.30	1	#
10	+2.50	19	+2.50	1	#
11	+2.53	36	+2.53	1	#
12	+2.64	32	+2.64	1	#
13	+2.65	48	+2.65	1	#
14	+2.78	42	+2.78	1	#
15	+2.99	34	+2.99	1	#
16	+3.07	24	+3.07	1	#
17	+3.16	21	+3.16	1	#
18	+3.27	16	+3.27	1	#
19	+3.33	37	+3.33	1	#
20	+3.37	45	+3.37	1	#
21	+3.41	35	+3.41	1	#
22	+3.50	5	+3.50	1	#
23	+3.56	14	+3.56	1	#
24	+3.58	3	+3.58	1	#
25	+3.62	46	+3.62	1	#
26	+3.72	9	+3.72	1	#
27	+3.82	30	+3.82	1	#
28	+3.94	38	+3.94	1	#
29	+4.04	22	+4.04	1	#
30	+4.11	20	+4.11	1	#
31	+4.27	2	+4.27	1	#
32	+4.36	15	+4.36	1	#
33	+4.39	50	+4.39	1	#
34	+4.41	23	+4.41	1	#
35	+4.50	7	+4.50	1	#
36	+4.51	33	+4.51	1	#
37	+4.57	6	+4.57	1	#
38	+4.69	26	+4.69	1	#
39	+4.75	49	+4.75	1	#
40	+4.92	39	+4.92	1	#
41	+4.94	47	+4.94	1	#
42	+5.11	25	+5.11	1	#
43	+5.26	43	+5.26	1	#
44	+5.36	11	+5.36	1	#
45	+6.35	44	+6.35	1	#
46	+6.77	10	+6.77	1	#

Table 11. Clustering histogram showing the results of the AutoDock run for the seventh mutation.