The use of structural modelling to infer structure and function in biocontrol agents

Colin Berry & Jason Board

Cardiff School of Biosciences, Cardiff University, Park Place, Cardiff CF10 3AT, UK

Corresponding author: Dr Colin Berry, Cardiff School of Biosciences, Cardiff University, Park Place, Cardiff CF10 3AT, UK

Email Berry@cf.ac.uk

Key words:

Insecticidal toxins; Bacillus thuringiensis; structural modelling; P20; PirB

Abstract:

Homology modelling can provide important insights into the structures of proteins when a related protein structure has already been solved. However, for many proteins, including a number of invertebrate-active toxins and accessory proteins, no such templates exist. In these cases, techniques of *ab initio*, template-independent modelling can be employed to generate models that may give insight into structure and function. In this overview, examples of both the problems and the potential benefits of *ab initio* techniques are illustrated. Consistent modelling results may indicate useful approximations to actual protein structures and can thus allow the generation of hypotheses regarding activity that can be tested experimentally.

1. Introduction

Homology modelling permits the construction of modelled structures of proteins where the actual structure of a member of the same protein family has been solved experimentally. In order to initiate the production of a homology model, a level of primary sequence identity between the solved template structure and the candidate protein of approximately 20% or higher is required and modelling accuracy declines below 50% and drops rapidly below 30% identity (Baker and Sali, 2001). Using this approach, approximations of the molecular structures of insecticidal toxins can be developed, as shown for the Cry48/Cry49 toxin pair that is active against *Culex* mosquito larvae (Jones et al., 2008; Kelker et al., 2014). However, if no suitable templates exist in the protein structure database, emerging computational solutions may allow us to develop structural models that are template independent. Such *ab initio* models must be treated with some caution since there is no direct experimental basis for their predictions, however, they can provide hypotheses regarding toxin structure. In addition, in cases where predicted structures resemble known toxins, hypotheses regarding mechanism of action can also be developed and these can subsequently be tested.

Ab initio modelling has been applied to a range of insecticidal toxins and accessory proteins as described below. The shortcomings of this methodology will be illustrated along with examples where consistent predictions of structures have been derived, and, in some cases subsequently been shown to mirror actual structures. Modelling of the Cry6Aa structure *ab initio* and the similarity of the resulting prediction with the subsequent crystallographic structure of the protein will be described elsewhere [Dementiev et al. submitted for publication].

2. Methods

2.1. Modelling software: For the majority of cases described here the Rosetta software (Bonneau et al., 2002) was used via the Robetta website (robett.bakerlab.org) with default settings. After generating secondary structure predictions, the program continued to full structure prediction using either *de novo* or database structure comparison methods. Each method returned five best model structures, which were visualised using standard molecular graphics software such as Pymol (DeLano, 2010). In the case of P20, the I-TASSER server (Roy et al., 2010; Yang et al., 2015; Zhang, 2008) was also used (with default settings) to compare predictions. Although it does not produce protein models, the HHPRED software (Soding et al., 2005) is a powerful tool for the detection of remote homologs and can predict structural homologies that simpler primary sequence comparisons such as Blast (Altschul et al., 1990) will not identify. Thus, the use of HHPRED may also be valuable in predicting structural classes to which proteins may belong and may contribute to the building of hypotheses regarding their modes of action.

3. Results and Discussion

3.1 Modelling outcomes and interpretations:

3.1.1. P19 protein: The gene encoding the P19 accessory protein (accession number AJ010753) is found in close association with the *cry11Aa* gene as the first in a 3-gene operon (*p19-cry11Aa-p20*) on the *Bacillus thuringiensis* serovar. *israelensis* pBtoxis plasmid (Berry et al., 2002; Dervyn et al., 1995). A closely-related protein is encoded by *orf1* in the operon encoding Cry2Aa. The P19 protein does not appear to influence the crystallisation of Cry11Aa or its toxicity, although it may influence toxin yield (Manasherob et al., 2001; Shi et al., 2006).

Modelling of the P19 protein using the Rosetta program returned 5 structure predictions with no significant similarity to each other (Figure 1). The lack of consistency of the models makes it extremely difficult to resolve which, if indeed any, of the models might

approximate to the real structure of the protein. I-TASSER was also inconsistent in the models that it generated for this protein and HHPRED did not identify any clear homologs. As a result, the example of P19 serves as a warning of the real limitations of *ab initio* modelling. No hypotheses regarding the structure or function of this protein could be derived from the modelling and resolution of these issues will require the application of classical structural techniques such as X-ray crystallography.

3.1.2. P20 protein: The P20 protein is encoded by the third gene on the *cry11Aa* operon. It appears to have a role in stabilising some Cry and Cyt toxins from *B. thuringiensis.* It is able to enhance the accumulation of Cyt1Aa and Cyt2Ba toxins and protects bacterial cells from the growth inhibition induced by production of Cyt toxins (Manasherob et al., 2006; Manasherob et al., 2001; Nisnevitch et al., 2006; Wu and Federici, 1993). It is also able to interact with and synergise the toxicity of Cry11Aa against *Aedes aegypti* although P20 itself does not appear to be toxic to this insect (Xu et al., 2001).

Ab initio modelling of P20 using Rosetta, returned 5 models with remarkable similarity to each other (Figure 2) and the I-TASSER predictions also produced very similar predictions (not shown). These models show a 3-layer, ∝/ß structure with the alpha helices contributed from the N-terminal end of the protein and a central anti-parallel beta sheet structure lying between two helix-turn-helix motifs with their helices running approximately parallel to the beta strands. The overall fold predicted for the P20 protein shows a high degree of similarity to the structure of *B. thuringiensis* Cyt toxins (Figure 2) with a root mean squared deviation between 400 backbone atoms of 2.7 Å (model 1 compared to Cyt1Aa pdb accession number 3RON). The HHPRED program also identifies homologies to Cyt toxins and also to volvatoxin, another toxin previously recognised as sharing the cytolysin fold (Cohen et al., 2011). Although the mechanism by which P20 might interact with Cry and Cyt toxins is not entirely clear from the predicted structure, it is interesting to note that Cyt toxins have been shown to

form dimers and Cyt and P20 co-immunoprecipitate (Visick and Whiteley, 1991). It is possible that Cyt-P20 heterodimers may form and that this may assist in masking the detrimental effects of Cyt toxin in producing bacteria. Interactions of Cyt with Cry11Aa may occur (Couche et al., 1987; Perez et al., 2005) and Cyt1Aa synergises with Cry11Aa (Wu et al., 1994). We may speculate that the known P20/Cry11Aa interactions and synergism may mirror the Cyt1Aa/Cry11Aa interactions however, the specific residues in Cyt1Aa identified as playing a role in the interaction with Cry11Aa (Perez et al., 2005) are not present in P20. Further analysis of all of the above interactions will be necessary to resolve this possibility.

3.1.3. PirB: The PirB protein from *Photorhabdus luminescens* (Waterfield et al., 2005) and *Photorhabdus asymbiotica* (Ahantarig et al., 2009) are the larger components of two part toxins PirA/PirB. The proteins have been shown to be toxic against *Aedes* mosquitoes and *Galleria mellonella* larvae and some small regions of similarity to 3-domain Cry toxins have been noted (Ahantarig et al., 2009; Waterfield et al., 2005). This similarity, however, was insufficient for homology modelling of the proteins.

Rosetta simulations for the PirA component returned inconsistent structures (not shown), all of which were rich in beta sheets. PirB models, in contrast, showed significant similarity to each other (Figure 3) and resembled domains I and II of the typical 3-domain Cry toxin structure. This similarity is consistent with the I-TASSER model of the protein generated by Maithri *et al.* (Maithri et al., 2012) and was validated by the recent publication of the structure of the PirB ortholog from *Vibrio parahaemolyticus* (PDB 3X0U along with PirA structure PDB 3X0T)(Lee et al., 2015). Structural alignments shows a high degree of similarity with the *P. luminescens* PirB model 1 giving a root mean squared deviation between 1273 backbone atoms of 3.6 Å compared to the *V. parahaemolyticus* protein. This provides a clear example of the predictive abilities of Rosetta and I-TASSER for this class of proteins.

3.2. Summary: In conclusion, it can be seen that *ab initio* modelling is able in some instances to generate models with a good match to the actual structures of the proteins as later validated by crystallographic methods (Cry6 and PirB). In other cases, such as P20, the consistency of models both within the predictions of Rosetta and between predictive tools (Rosetta and I_TASSER) enhance confidence in the models and allow us to speculate on the roles of the proteins. Models must always be treated with caution and subjected to experimental testing to validate their predictions. However, the modelling is powerful in directing us to the best experiments to perform and it is to be hoped that they will help to accelerate our understanding of the invertebrate-active toxins and their accessory proteins.

Figure Legends:

Figure 1: P19 models

The top 5 models of P19 generated by the Rosetta software are presented.

Figure 2: P20 models

The top 5 models of P20 generated by the Rosetta software are presented along with the structure of Cyt1Aa (PBD accession 3RON).

Figure 3: PirB models

The top 5 models of *P. luminescens* PirB generated by the Rosetta software are presented along with the structure of *V. parahaemolyticus* PirB (PBD accession 3X0U).

References

Ahantarig, A., et al., 2009. PirAB toxin from *Photorhabdus asymbiotica* as a larvicide against dengue vectors. Appl Environ Microbiol. 75**,** 4627-4629.

154 Altschul, S. F., et al., 1990. Basic local alignment search tool. J. Mol. Biol. 215**,** 403-410.
155 Baker, D., Sali, A., 2001. Protein structure prediction and structural genomics. Science. 294**,**
156 93-6.
157 Berry, C., et al., 2002. Complete sequence and organization of pBtoxis, the toxin-coding
158 plasmid of *Bacillus thuringiensis* subsp. *israelensis.* Appl. Environ. Microbiol. 68**,** 5082-
159 5095.
160 Bonneau, R., et al., 2002. *De novo* prediction of three-dimensional structures for major protein
161 families. J Mol Biol. 322**,** 65-78.
162 Cohen, S., et al., 2011. Cyt1Aa Toxin: High-Resolution Structure Reveals Implications for Its
163 Membrane-Perforating Function. J. Mol. Biol.
164 Couche, G. A., et al., 1987. Structural disulfide bonds in the *Bacillus thuringiensis* subsp.
165 *israelensis* protein crystal. J Bacteriol. 169**,** 3281-3288.
166 DeLano, W. L., The PyMOL Molecular Graphics System., Vol. 2010. DeLano Scientific LLC, San
167 Carlos, CA, USA., 2010.
168 Dervyn, E., et al., 1995. Transcriptional regulation of the *cryIVD* gene operon from *Bacillus*
169 *thuringiensis* subsp. *israelensis.* J. Bacteriol. 177**,** 2283-2291.
170 Jones, G. W., et al., 2008. The Cry48Aa-Cry49Aa binary toxin from *Bacillus sphaericus* exhibits
171 highly-restricted target specificity. Environ. Microbiol. 10**,** 2418-2424.
172 Kelker, M. S., et al., 2014. Structural and Biophysical Characterization of *Bacillus thuringiensis*
173 Insecticidal Proteins Cry34Ab1 and Cry35Ab1. PLoS One. 9**,** e112555.
174 Lee, C. T., et al., 2015. The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes
175 virulent by acquiring a plasmid that expresses a deadly toxin. Proc Natl Acad Sci U S A.
176 Maithri, S. K., et al., 2012. Modeling and Docking Studies of PirB Fusion Protein
177 from *Photorhabdus luminescens*. International Research Journal of Biological Sciences.
178 1**,** 7-18.

179 Manasherob, R., et al., 2006. Cyt1Ca from *Bacillus thuringiensis* subsp. *israelensis*: production
180 in *Escherichia coli* and comparison of its biological activities with those of other Cyt-
181 like proteins. Microbiol. 152**,** 2651-2659.
182 Manasherob, R., et al., 2001. Effect of accessory proteins P19 and P20 on cytolytic activity of
183 Cyt1Aa from *Bacillus thuringiensis* subsp. *israelensis* in *Escherichia coli.* Curr. Microbiol.
184 43**,** 355-364.
185 Nisnevitch, M., et al., 2006. Cyt2Ba of *Bacillus thuringiensis israelensis*: activation by putative
186 endogenous protease. Biochemical and biophysical research communications. 344**,** 99-
187 105.
188 Perez, C., et al., 2005. *Bacillus thuringiensis* subsp. *israelensis* Cyt1Aa synergizes Cry11Aa toxin
189 by functioning as a membrane-bound receptor. Proc Natl Acad Sci U S A. 102**,** 18303-
190 18308.
191 Roy, A., et al., 2010. I-TASSER: a unified platform for automated protein structure and function
192 prediction. Nat Protoc. 5**,** 725-38.
193 Shi, Y. X., et al., 2006. [Influence of accessory protein P19 from *Bacillus thuringiensis* on
194 insecticidal crystal protein Cry11Aa]. Wei sheng wu xue bao = Acta microbiologica
195 Sinica. 46**,** 353-7.
196 Soding, J., et al., 2005. The HHpred interactive server for protein homology detection and
197 structure prediction. Nucleic Acids Res. 33**,** W244-8.
198 Visick, J. E., Whiteley, H. R., 1991. Effect of a 20-kilodalton protein from *Bacillus thuringiensis*
199 subsp. *israelensis* on production of the CytA protein by *Escherichia coli*. J Bacteriol. 173**,**
200 1748-56.
201 Waterfield, N., et al., 2005. The *Photorhabdus* Pir toxins are similar to a developmentally
202 regulated insect protein but show no juvenile hormone esterase activity. FEMS
203 Microbiol Lett. 245**,** 47-52.

Wu, D., Federici, B. A., 1993. A 20-kilodalton protein preserves cell viability and promotes CytA crystal formation during sporulation in *Bacillus thuringiensis*. J. Bacteriol. 175, 5276-5280.

Wu, D., et al., 1994. Synergism of mosquitocidal toxicity between CytA and CryIVD proteins using inclusions produced from cloned genes of *Bacillus thuringiensis*. Mol Microbiol. 13, 965-72.

Xu, Y., et al., 2001. Expression of the p20 gene from Bacillus thuringiensis H-14 increases Cry11A toxin production and enhances mosquito-larvicidal activity in recombinant gram-negative bacteria. Appl Environ Microbiol. 67, 3010-5.

Yang, J., et al., 2015. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 12, 7-8.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 9, 40.

Figure 1

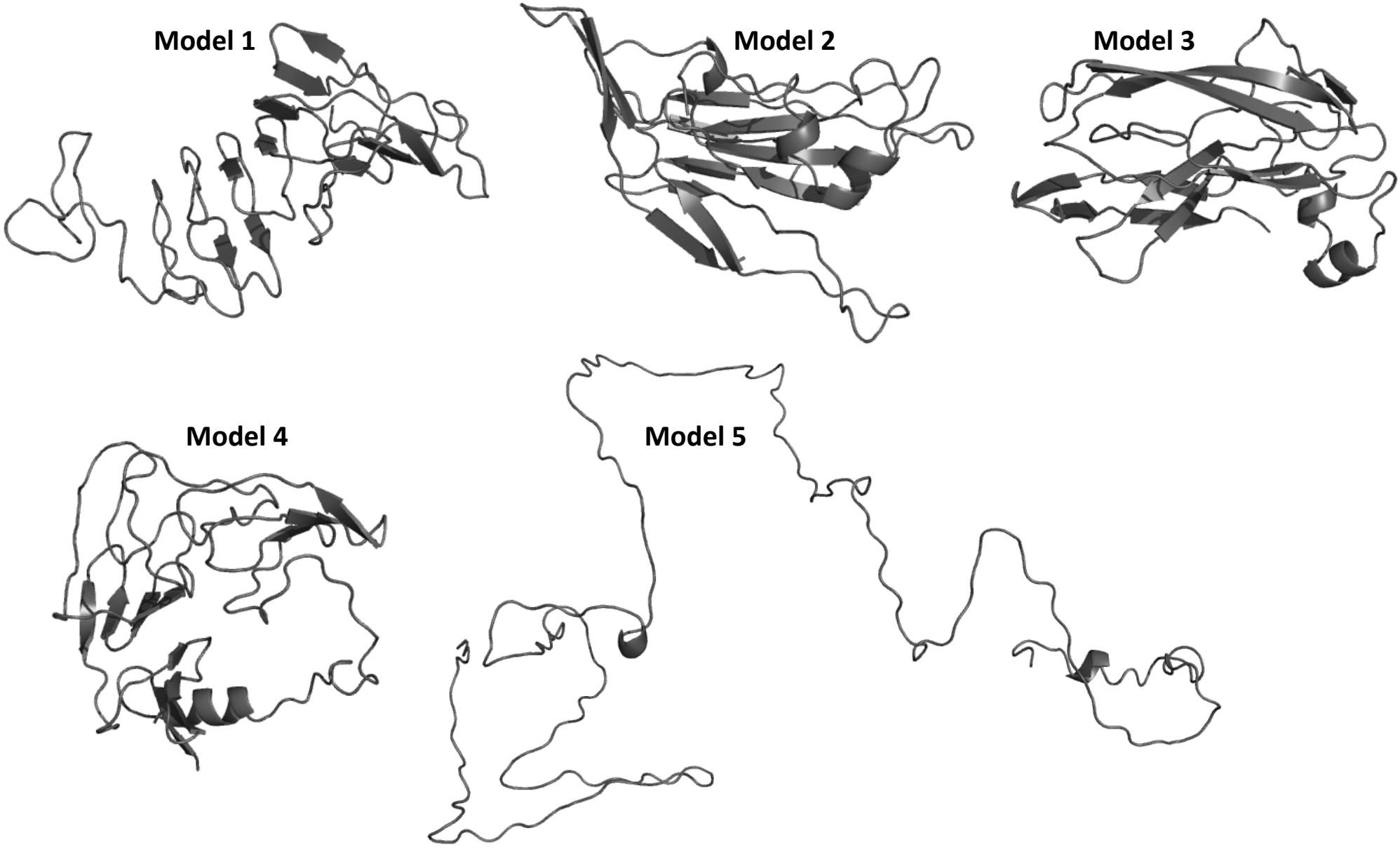
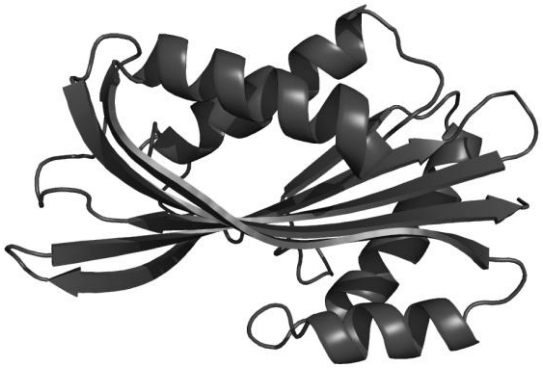


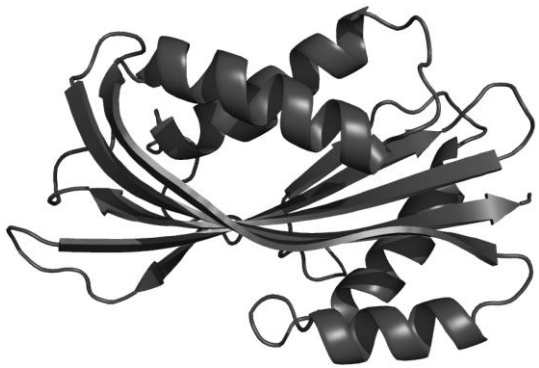
Figure 2

Figure 2

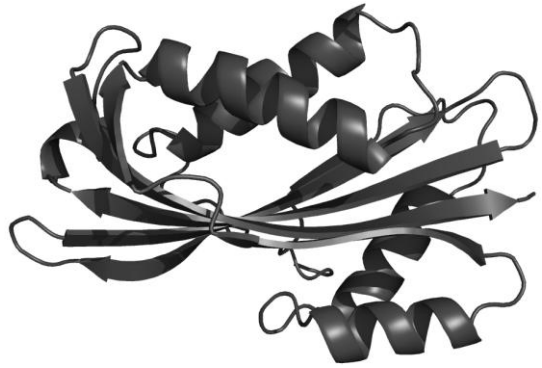
Model 1

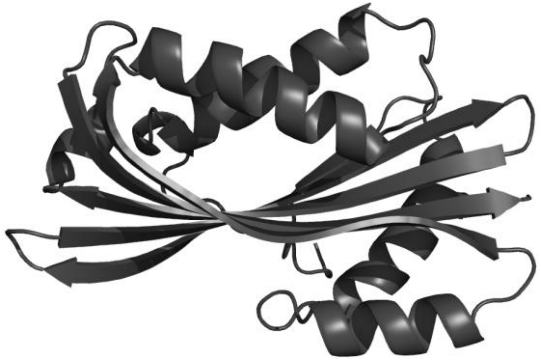


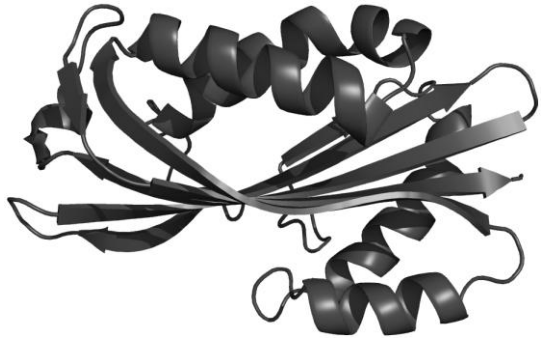
Model 2

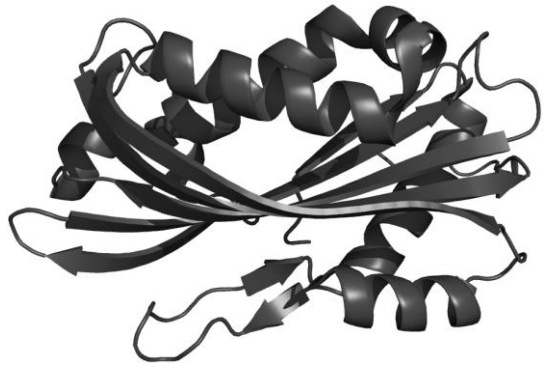


Model 3









Model 4

Model 5

Cyt1Aa

Figure 3

Figure 3