MOLECULAR MODELING PREDICTS HELIX TILTING IN HCN CHANNELS DURING cAMP BINDING

by

Adam Haesler
Bachelors of Science, University of Guelph 2005

THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
Department of Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Spring 2010

Approval

Name: **Adam Haesler**

Degree: **Master of Science**

Title of Thesis: **Molecular modeling predicts helix tilting in HCN channels during cAMP binding**

Examining Committee:

Chair: **Jonathan Choy**
Assistant Professor

_____

Edgar Young
Senior Supervisor
Assistant Professor

_____

Mark Paetzel
Supervisor
Associate Professor

_____

Glen Tibbits
Supervisor
Professor

_____

Lisa Craig
Internal Examiner
Assistant Professor

Date Defended/Approved: December 2, 2009 _____

Declaration of
Partial Copyright Licence

Abstract

Hyperpolarization-activated cyclic nucleotide-modulated (HCN) channels are believed to regulate the heartbeat. Binding of cAMP to a cytoplasmic binding domain (BD) increases probability of HCN channel opening; a C-linker region (helices A'-F' and A) links the BD to the channel pore. The C-linker was proposed to move when cAMP binds the BD; I hypothesized a movement of D'- F'. I created two models of a cytoplasmic fragment from an "apo" HCN channel (without cAMP), one by comparative modeling and one by molecular dynamics. Compared to the crystal structure of the fragment with cAMP, both apo models showed D' tilting, such that N-terminus gets farther and C-terminus closer to helix A. E' translates parallel to the long axis of A from C- to N-terminus. Helix A did not change relative to the BD. Therefore when cAMP binds the HCN channel, the N-terminus of D' is likely tilting towards the BD.

Keywords: Cyclic-nucleotide binding domain (CNBD); molecular dynamics; comparative modelling; cAMP; Hyperpolarization-activated cyclic nucleotide modulated ion channel; C-linker.

For my Father who always believed I could do anything.

Acknowledgements

I would like to thank the people who made this work possible. First off thank you to my supervisor Edgar Young for his patience throughout my Masters. It was an unforgettable journey where I learned to many things from him to mention here. Also, thank you to Mark Paetzel for the use of the processors in his laboratory to perform the molecular dynamics simulations. As well, thank you to Charles Stevens for all of his assistance in learning how to use Gromacs and understanding both the theoretical and technical aspects of performing simulations.

Also I would like to thank members of the Young laboratory. Thank you to Zarina Madden for all of her help in learning how to do soluble protein work. As well thank you to Kerry Chan for his assistance with learning how to perform frog surgery. Thank you to the rest of the Young laboratory members for their support throughout the years.

Table of Contents

List of Figures

List of Tables

Glossary

Six transmembrane superfamily	A family of ion channels that have common six transmembrane helices
Depolarization	Development of negative voltage, as a result of opening of sodium channels.
Repolarization	Development of negative voltage, as a result of opening of potassium channels.
Hyperpolarization	Development of negative voltages beyond resting voltage, as a result of continued repolarization of the membrane.
Pore	The passage through the membrane able to allow potassium ions to pass from one side of the membrane to the other. It includes the selectivity filter, the P-helix and S6 helices from each subunit.
Transmembrane domain	The six helices (S1-S6) of the potassium ion channel found within the membrane and spanning it from one side to another.
N-terminal domain	A cytoplasmic domain of the HCN channel, composed of the residues from the N-terminus to the first residue of S1 helix.
C-terminal domain	A cytoplasmic domain of the HCN channel, composed of residues C-terminal of S6. This includes the C-linker and CNBD.
Sinoatrial node	This is the part of the heart responsible for regulating the heartbeat.
Selectivity filter	A component of the the pore of the channel that selects for passage of only sodium and potassium ions.

C-linker	A structural unit of the C-terminal domain C-terminal to the S6 helix and N-terminal of the CNBD. It is composed of helices A', B', C', D', E' and F'
Transmission hinge motif	E'-F' of the C-linker (K10-F18 in HCN2)
unliganded	Describing the conformation state of a protein when no ligand is in the CNBD, as opposed to "cAMP bound".
Molecular dynamics simulation	A series of movements of atoms that are calculated based on forces due to atom to atom interactions.
Periodic box	A box within the simulation environment whose contents are repeated in three-dimensional space. The target gets placed into the box for the simulation.
PDB file	A file that contains coordinates of all atoms of the protein. Name is of the form *.pdb. It is generated from either VMD, Swiss-pdbviewer or downloaded from the RCSB protein databank (www.pdb.org).
GMX file	A combination of two files, one is *.pdb or *.gro format and the other is a *.top file. The *.pdb file contains structural coordinates where as the *.gro contains both velocities and positions.
Force field	A collection of potentials defined by a potential energy function, where the individual force on an atom is the negative of the derivative of the potential energy function for that atom.
Parameters file	A file that details types and values of parameters to be fed into an algorithm. Name is of the form *.mdp.
Topology file	A file that details the charge, type, residue origin, types of bonds and angles for each atom. Name is of the form *.top, *.tpr or itp.
Log file	A file detailing a particular run instance of an algorithm. It can include details of inputs, outputs, parameters and calculations performed during run of the algorithm. Name is of the form *.log.

Trajectory file	A file detailing the movements and forces of each atom for each step of the simulation. Name is of the form *.trr or *.xtc).
Template	The input in a comparative modeling experiment that acts as the blueprint for building the new model.
Target	The input in a modeling experiment that will be altered to create the new model.
Z-score	The quantitative measure of quality of a structural model, generated from ProSA based on atom to atom interactions.
CNBD	Cyclic nucleotide binding domain
HCN	Hyperpolarization activated cyclic nucleotide modulated ion channels.
Mlotik1	A potassium channel from *Mesorhizobium loti*.
CNG	Cyclic nucleotide gated channel
Thm	Transmission Hinge motif
RMSD	Root mean square deviation (Angstroms, Å)
H2a, H2R, and SA2	Comparative models built using various methods (Figure 8-5).
H2h, Mlh and Mla	Crystal structures as they were used in the research for modelling (Figure 8-5).
G3 and G4	Molecular dynamics models (Figure 8-5)
1q5o	The cAMP bound HCN2 cytoplasmic domain source crystal structure.
1vp6	The cAMP bound Mlotik1 cytoplasmic domain source crystal structure.
1u12	The unliganded Mlotik1 cytoplasmic domain source crystal structure.

1. Introduction

1.1 Ion channels are responsible for normal cellular function.

Ion channels are responsible for the control of many cellular functions. It would be advantageous to learn ways to alter cellular function by changing ion channel function. All ion channels of the "six transmembrane superfamily" (Figure 8-1) have three necessary components that are interconnected to allow opening of the channel (Swartz, 2004). The three components include: a voltage sensor, which is able to respond to changes in membrane voltage; an ion pore; and a gate that regulates the state of the channel as opened or closed (Swartz, 2004). Some potassium ion channels of the superfamily have a special importance to cellular functions (Swartz, 2004). One of these functions is changing the membrane voltage from more positive to more negative (Boron and Boulpaep, 2005). This process is known as "repolarization" and occurs after the completion of a nerve impulse (Boron and Boulpaep, 2005).

1.2 Hyperpolarization- activated cyclic-nucleotide modulated ion channels are unique potassium channels.

The ion channels studied in this work open optimally at very negative voltages (Biel et al., 2002). These negative voltages are called "hyperpolarized" voltages (Boron and Boulpaep, 2005). The channels are called "hyperpolarization-activated cyclic-nucleotide modulated" (HCN) channels. Like the name implies they are modulated by cAMP. They belong to the potassium channel family.

These particular potassium channels produce an inward current of ions called I_h, I_f, or I_q (DiFrancesco and Tortora, 1991; Accili et al., 2002; Biel et al., 2002; Robinson and Siegelbaum, 2003). The current is produced by conduction of sodium (inward) and potassium (outward) (Accili et al., 2002). Opening HCN channels at hyperpolarized voltages has the effect of making the membrane voltage more positive. Most other potassium channels respond to positive voltages, which is considered a "depolarized" membrane (Boron and Boulpaep, 2005). Research on HCN channels is an opportunity to investigate how the channels are able to use a similar six transmembrane protein structure to respond to hyperpolarized voltages rather than depolarized voltages (Swartz, 2004). From the structures of channels that have been solved it is obvious that opening and closing involves similar components (DiFrancesco and Tortora, 1991; Accili et al., 2002; Biel et al., 2002; Robinson and Siegelbaum, 2003). Thus the potassium ion channels studied in this work allow the opportunity to study a channel with an important function in regulation of membrane potential.

Binding of cAMP occurs in a “cyclic nucleotide binding domain” (CNBD) found within the cytoplasmic domain (HCN*) (Figure 8-2). This allows response of the channel at less hyperpolarized membrane voltages (Biel et al., 2002; Robinson and Siegelbaum, 2003). Since the channel is responding at less hyperpolarized voltages it is able to respond faster to hyperpolarizing membrane voltage, returning the membrane to resting voltage faster (Biel et al., 2002; Robinson and Siegelbaum, 2003).

1.3 HCN channels are proposed to allow proper regulation of the heartbeat.

The first (HCN1), second (HCN2) and fourth (HCN4) members of the HCN channel family are proposed to play a role in pacemaking in the heartbeat (DiFrancesco and Tortora, 1991; Biel et al., 2002; Robinson and Siegelbaum, 2003). HCN1, HCN2 and HCN4 are expressed in cells of the sinoatrial node (Noma and Irisawa, 1976; Biel et al., 2002) (Figure 8-3), which is a part of the heart (Robinson and Siegelbaum, 2003). The sinoatrial node is responsible for regulation of the heartbeat and as seen in Figure 8-3 HCN channels cause depolarization of the sinoatrial node from repolarized voltages. Assuming no other channels are able to activate at hyperpolarized voltages, without HCN1, HCN2, and HCN4 multiple action potentials could not occur (Biel et al., 2002; Robinson and Siegelbaum, 2003). The action of norepinenephrine is to increase intracellular levels of cAMP, initiating current at higher voltages to open the channel faster than simply with just voltage activation (Robinson and Siegelbaum, 2003). Therefore HCN1, HCN2, and HCN4 are believed to be vital to proper heartbeat regulation (Biel et al., 2002; Robinson and Siegelbaum, 2003).

1.4 HCN channels have three major domains.

HCN channels have four subunits and each subunit has three major domains. (Biel et al., 2002). The subunits can be identical or different (Accili et al., 2002). The "N-terminal domain" is responsible for channel expression of HCN1 and HCN2 family members (Proenza et al., 2002). The N-terminal domain is found in the cytoplasm and is connected to the first segment of the "transmembrane domain" (Proenza et al., 2002). The transmembrane domain is composed of six transmembrane segments (called S1

through S6) including a pore with a GYG protein sequence and a positively charged fourth transmembrane helix "S4" (Accili et al., 2002), Figure 8-1). The pore itself is formed by a selectivity filter composed of the sequence GYG, the P-helix, and S6 helices in the transmembrane domain (Biel et al., 2002; Roncaglia et al., 2002) (Johnson and Zagotta, 2005). The GYG sequence acts as a selectivity filter for allowing potassium ions to be conducted due to selectivity filters of all contributing subunits (Heginbotham et al., 1994; Doyle et al., 1998). C-terminal of the transmembrane domain is the HCN* region that is of interest in this thesis.

The structure of interest (HCN*) is found in the cytoplasm and consists of a "C-linker" and the CNBD. The purpose of this research was to study how HCN* is different comparing cAMP bound versus unliganded states. The C-linker comprises six helices (called A'-F') and an extra helix called "A helix" (Figure 8-2). The C-terminus of the sixth transmembrane helix is linked to the N-terminus of the A' helix (Figure 8-2) (Zagotta et al., 2003; Flynn et al., 2007). A "transmission hinge motif" (Thm) was the name assigned to E' and F'. The CNBD follows the C-terminus of the A helix (Figure 8-2). This domain consists of eight beta strands, followed by a "B helix", and a "C helix" (Zagotta et al., 2003; Flynn et al., 2007) (Figure 8-2). In some literature the A helix is considered to be part of the CNBD. However, for the purpose of analysis in this thesis (see superimpositions, Chapter 3, 4, and 6) I considered the CNBD not to include the A helix.

A large part of the C-linker is believed to form a ring structure in the cAMP bound conformation of HCN* (Figure 8-2; Zagotta et al., 2003; Flynn et al, 2007). The ring structure, known as the gating ring, is composed of helices A'-D' of each subunit

interacting (Figure 8-2). There are competing views of whether the ring structure would exist in the intact cAMP bound channel given the conformation of gating ring helices in the current HCN* crystal structure (Craven and Zagotta, 2004; Hua and Gordon, 2005).

1.5 Key CNBD residues have been identified

A cAMP binding event is the result of the interaction of cAMP with a few key residues within the beta-roll of the CNBD. Within the CNBD is a phosphate binding cassette (PBC), which is a set of residues that includes beta strand six, the P-helix, and strand seven (Berman et al., 2005). The PBC residues responsible for the binding of cAMP are believed to be an arginine (R591) interacting with the phosphate of the cAMP and a glutamate (E582) with the 2'OH of the ribose(Berman et al., 2005). It has been proposed that R591 (PBC), K638 (C-helix) and I636 (C-helix) are important for cAMP binding with no effect on channel opening (Zhou and Seigelbaum, 2007). The only residue of the C-helix that has been found to cause a change in channel opening is R632 that is proposed to make an important salt bridge with E582 to stabilize the CNBD with bound cAMP (Zhou and Seigelbaum, 2007). As originally proposed by Varnum et al., 1993 and further supported by results of Matulef et al., 1999, cAMP binding in the core of the beta-roll is a first step, followed by movement of the C-helix towards the beta-roll to allow interaction of R632 with cAMP. Lastly, there is a conserved set of six hydrophobic residues that hide the cyclic phosphate bond and rings of cAMP from the environment (Berman et al., 2005). This is a conserved mechanism among all cAMP binding proteins (Kornev et al., 2008).

1.6 Mlotik1 has a CNBD and a shorter C-linker

A potassium ion channel called Mlotik1 is found in the bacterium *Mesorhizobium loti* (Nimigean et al., 2004; Clayton et al., 2005). The channel is known to open with the binding of cAMP (Nimigean et al., 2004; Clayton et al., 2005; Cukkemane et al., 2007). Similar to HCN it has the following: an N-terminal domain; six transmembrane helices with a voltage sensor; a pore composed of the pore-helix, selectivity filter and S6 helices; and a C-terminal cytoplasmic domain with a C-linker linking the S6 to a CNBD (Clayton et al., 2005; Clayton et al., 2008). The C-linker of Mlotki1 is only 22 amino acids as opposed to HCN and CNG where the C-linker is 60-80 amino acids (Clayton et al., 2004). Thus it has been proposed that the C-linker of Mlotik1 is the minimal length necessary to confer channel opening due to binding of cAMP (Nimigean et al., 2004).

Mlotik1 defined the region of interest due to the limited structural components in its C-linker. The C-linker of Mlotik1 only spans helices D' – F' while both Mlotik1 and HCN have the A helix and CNBD in common (Figure 8-4 and Clayton et al., 2004). Due to the similarity of the CNBD for both bound and unliganded Mlotik1 structures (1vp6 and 1u12) the CNBD was not considered as a focus of this research (Clayton et al., 2004). This is further supported by other bound and unliganded crystal structures for Mlotik1 showing high similarity for the CNBD (Altieri et al, 2008); also supporting this notion is the similarity within CNBD for bound and unliganded form of another cAMP binding protein known as RIα (Kim et al., 2005).

1.7 HCN C-linker could be responsible for channel modulation.

The purpose of this research was to understand the affects of cAMP on HCN2 channel structure and function in HCN*. Does a cAMP binding message transmit

through the C-linker? Further, what are the conformational changes in the C-linker responsible for message transmission? Thus specifically it is the conformational changes responsible for message transmission from the CNBD to the pore (S6 helices in the transmembrane domain) that are the concern of this thesis. Can structural changes of HCN channels due to cAMP binding alone be proposed?

1.7.1 Hypothesis 1: The C-linker

I hypothesized that channel modulation is the result of binding cAMP as a message that is transmitted to the pore via the C-linker. This was based on connectivity of the CNBD to the C-linker and the C-linker to the pore. In addition, the hypothesis is based on results of previous studies, of the whole channel or soluble cytoplasmic domain, supporting the idea of C-linker conformation being different when the CNBD was bound with cAMP ("holo") and when the CNBD was unliganded ("apo"). Before my thesis, it was known that in the apo to holo transition, A' and D' helices changed conformation, that F' position changed and that homologous residues of Mlotik1 also changed conformation (Varnum et al., 1995; Matulef et al., 1999; Clayton et al., 2004; Nimigean et al., 2004; Hua and Gordon, 2005; Johnson and Zagotta, 2005; Berrera et al., 2006; Craven and Zagotta, 2006; Taraska and Zagotta, 2007; Zhou and Siegelbaum, 2007; Altieri et al., 2008). Also it was known that conformational change in the C-linker brought about conformational change in the pore helices (Johnson and Zagotta, 2005)

My strategy to test this hypothesis was to look for structural differences comparing holo versus apo forms of HCN* (from HCN2) for the region of interest, see next hypothesis.

1.7.2 Hypothesis 2: The region of interest

Helices D'-A were defined as the "region of interest" for this research (section 1.4). I hypothesized that D' and Thm modulate the cytoplasmic domain structure and function. This is based on D' being found at the C-terminus of the gating ring and the Thm being the last structural component before the CNBD other than A helix (Zagotta et al., 2003). The A helix has been proposed to be stationary relative to the CNBD in previous studies of CNBD containing proteins (Berrera et al., 2006; Altieri et al., 2008).

My strategy to test this hypothesis was to build molecular models of unliganded HCN* (from HCN2). The models were built first by comparative modeling (section 1.8.2) using three different methods, and next by molecular dynamics (section 1.8.3). The comparative models were based on the crystal structure of the unliganded domain from Mlotik1 (section 1.6). Following building of the molecular models, I set out to determine differences between apo models and holo crystal structures. This was done using RMSD (which is a measure of the differences between structures in a superimposition) and distance comparisons.

1.7.3 Hypothesis 3: Individual helices in the region of interest

D' helix

I hypothesized that the apo versus holo comparisons would show D' has a different conformation in the cAMP bound state than the unliganded state. This was based on previous studies supporting the proposal that the gating ring helices change conformation during the holo to apo transition (Craven and Zagotta, 2004; Hua and Gordon, 2005; Taraska and Zagotta, 2007). My strategy to test this hypothesis was to analyze the orientation and overall conformational differences in D' relative to A helix, in

a comparison of the unliganded model and cAMP bound crystal structure. I planned to do this by measuring RMSD comparing D' of the two structures. Also I planned to measure the angle of D' relative to a fixed vertical for each structure projected onto the plane of the page.

A helix

I hypothesized that A helix was stationary during the holo to apo transition. This was based on previous unliganded and cAMP bound crystal structures of RIα and Mlotik1 (Kim et al., 2005; Altieri et al., 2008). My strategy to test this hypothesis was to evaluate the RMSD of residues of A helix comparing the apo model and holo crystal structure.

Transmission hinge motif

I hypothesized that E' and F' either moved closer together or further apart during the transition from cAMP bound to unliganded. This is based on previous studies (Berrera et al., 2006; Taraska and Zagotta, 2007; Zhou and Siegelbaum, 2007; Altieri et al., 2008; Kornev et al., 2008) where F' moves in the transition from apo to holo. My strategy to test this hypothesis was to compare the distance between residues of E' and F', in both apo models and holo crystal structure, and also to measure the RMSD for E', F', and the Thm.

1.8 Computational strategies - Model building and analysis

I wished to use comparative modeling and molecular dynamics modeling to build model predictions of the unliganded state of HCN*. Quality control tests were applied to these models. The analysis then compared the C-linkers of cAMP bound and unliganded

conformations of HCN*. The structures used for the analysis were the bound crystal structure, the comparative models of unliganded HCN*, and simulation models of HCN* in the bound and unliganded states. This section introduces strategies for comparative modeling, quality control testing, and model analysis.

1.8.1 Alignments for model building

An amino acid sequence alignment is useful to determine the amount of similarity between two sequences using a pairwise alignment. The best alignment is determined by using a matrix for comparing amino acid pairs, using weightings for identity, mismatches, gaps, and insertions to determine the best residues to be aligned (Thompson et al., 1994).

1.8.2 Comparative modeling

Comparative modeling is a tool for building models based on a template. It is used to predict a protein conformation. The template is a known structure from the RCSB protein databank (Berman et al., 2000): www.pdb.org. This is a database of all known molecular structures. Important to this research it includes the crystal structures coordinate files (pdb files) used in this research. Comparative modeling is a technique of interest to create a prediction of the conformation of a protein state that is unknown (Guex and Peitsch, 1997; Schwede et al., 2003). It uses a known structure (template) for conformation of the resulting model. The target is the input protein that gets modified and it can be a pre-existing structure or a raw sequence. The building of the comparative model requires alignment of the sequences to know which residues of the template to use to model the target. Subsequently a build algorithm is run that sets the conformation of

the target to make it look similar to the template. Due to use of a separate structure for the template compared to the target input this is called a "template dependent method".

1.8.3 Molecular dynamics modeling (Simulation)

A molecular dynamics simulation is the act of allowing each atom to atom interaction total to be calculated and represented as the force acting on that atom over time to move the atom to a new location. This process is repeated until a specified total time has been reached. The process of running an MD simulation requires the use of many different applications by the program (www.gromacs.org).

Initially a structure file is created using a force field applied to the coordinates of the structure to be simulated. All hydrogen atoms of the original structure are stripped away and new ones added. The new molecule is energy minimized to make all atom to atom interactions at the lowest total energy for the whole protein as possible. An energy minimization is a movement of atoms in the molecule to get the lowest energy possible for the molecule based on existing interactions. The energy of each atom to atom interaction is re-evaluated to determine if the molecule has reached the predefined minimum energy. The minimization is stopped when the energy of the molecule has reached the minimum or the total number of steps specified to be performed has been completed.

The resulting molecule is placed into a periodic box, solvated and neutralized by adding ions of sodium or chloride, and the simulation is run. The total force between each pair of atoms is calculated for each step of the simulation and then the force simulated

resulting in atom positions. New calculations are made for all atom to atom interactions and the process is repeated to give a trajectory.

For some of the simulations the protein moved outside of the periodic box. This resulted in a picture of the protein as if it had extremely long bonds. Thus the periodic boundary conditions were removed after the trajectory was created.

1.8.4 RMSD

RMSD stands for the root mean square deviation (Guex and Peitsch, 1997). RMSD is useful for determining similarity between two structures. RMSD is used to determine a quantitative measure for amount of conformational differences between two models/structures. It is a measure of distances between atom positions. RMSD is calculated based on a superimposition of two structures/models for selected residues.

1.8.5 Comparative model quality control - template / model similarity

This is used to test if the comparative model was built correctly. A quality control I came up with was to use the superimposition of the model and the template crystal structure. The method is template dependent so the model should be similar to the template. If it was not similar then I would question if the model was built correctly. Also this should be the case for the whole model and not just the CNBDs. This could be tested both qualitatively using the superimposition and quantitatively using the RMSD.

1.8.6 Molecular dynamics quality control- RMSD vs. Time

This is used to determine how much the model is changing and after how long does it take the model to stop moving. The RMSD is calculated comparing the first frame, or input structure, to each frame of the trajectory. This would allow evaluation of

the model stability based on whether the RMSD had stopped changing for the frame of the trajectory used (www.gromacs.org).

1.8.7 Atom to atom distance

Measuring distance between atoms in both holo and apo structures/models is used to determine a predicted distance change in the real protein during the holo to apo transition. The Euclidean distance is between two points in any dimension. To be able to calculate the distance, the coordinates for the points chosen for distance calculation are required as the inputs. In the case of protein structure coordinates are in three-dimensions, so there are three ordinates to each atom defining its position in space relative to the origin, for example (x, y, z) = (8, 9, 12). The algorithm uses the equation

$$\text{Distance} = [(x_1-x_2)^2 + (y_1-y_2)^2+(z_1-z_2)^2]^{1/2} ,$$

where the ones and twos represent the first or second atom. Thus the equation uses the differences in the ordinates to determine the difference in the coordinates.

1.8.8 ProSA is used as a model and structure quality control

ProSA is used to evaluate the quality of a model, compared to previously determined crystal structures or NMR structures of similar size (Hefti et al., 2004; Wiederstein and Sippl, 2007). This program evaluates the total energy of interactions between all possible individual atom pairs on neighboring residues (Sippl, 1993).

2: Methods Part I: Explanation of Algorithms

2.1 Comparative model building

Almost all algorithms in this section are found and used within pdbviewer (Guex and Peitsch, 1997). For the purpose of this research the suite of algorithms within pdbviewer served as a protein structure viewer, environment for structural analysis and molecular modeling. The one algorithm not found in pdbviewer that was used for comparative modeling was Swiss-Model (Geux and Peitsch, 1997).

All structures used in this study were either models that I built or pre-existing structures solved by X-ray crystallography (Berman et al., 2000).

2.1.1 Fit Molecules (from selection)

An algorithm called Fit Molecules (from selection) was used to perform a superimposition of two protein structures. This algorithm was the most commonly used superimposition algorithm (section 2.1.2, Fit Raw Sequence for the alternative). The algorithm Fit Molecules (from selection) was the preferred algorithm due to the objectivity in its use and number of applications. This algorithm could be used in building a structure based model or superimpositions to determine RMSD. The inputs are two structures or models first loaded into pdbviewer that are pdb files. The residues to be superimposed from each structure/model are selected. An equal number of residues must be selected in both structures/models to run the algorithm. In all cases for this research only the homologous residues were used for superimposition based on an

alignment previously performed using ClustalW (section 2.1.7). The selected residues are superimposed based on an alignment done by pdbviewer of the selected residues for the superimposition. To perform the superimposition the user is asked to select a “reference” structure, which is the template structure you want to compare to the other (“query”), so this reference structure/model acts as an anchor. The result of the superimposition of query on anchor is a screen shot of the superimposition and RMSD. The superimposition can be saved as a "project" pdb file.

2.1.2 Fit Raw Sequence

This algorithm fits a raw sequence (meaning the file has no structural information) on to a structure. Similar to Fit Molecules (from selection) it also works by use of an alignment but offers less control and objectivity. This algorithm must be used to create the input superimposition for a “project mode” modeling project in Swiss Model (see below). The input is a raw sequence FASTA file and a pdb structure file loaded into pdbviewer. Then the user is asked to designate the reference and non reference structure without selecting the appropriate residues. The result is a new superimposition/model of the raw sequence residues fit on to the structure template. This superimposition or just the new model can be saved as a pdb file.

2.1.3 RMSD

RMSD was calculated using a superimpositon by selecting the residues in both structures to be compared. Then the algorithm “Calculate RMS...” was run for comparison of the Calpha atoms.

2.1.4 Swiss-Model

The purpose of using the Swiss-Model server (Schwede et al., 2003, Guex and Peitsch, 1997, Project Mode http://swissmodel.expasy.org/) is to build a comparative model to make predictions about protein structure currently unknown. The input is a superimposition using Fit Raw Sequence (section 2.1.2 Fit Raw Sequence for details) in pdbviewer that is uploaded to the server in the project mode as a pdb file. The pairwise alignment is improved if at all possible using the target and template superimposition. The Cα atoms of the template are matched to the Cα atoms of aligned residues of the target. For the target residues that are not aligned, in the case of insertions, these parts of the target sequence are not modeled at this stage. A loop between the ends of residues that were not aligned is formed based on the lowest amount of steric hinderance, force field energy, and optimizing ideal interactions between residues like hydrogen bonding. Alternatively for loops of 10 residues or more then a loop library of known structures is used to build the loop.

Side chain orientations are chosen based on similarity to those of the template. The same orientation of the side chain in the template (rotamer) residues are first attempted. If the rotamer is not a possible option due to steric clashes then a new rotamer is chosen from a library based on known structures to build side chain conformations for all residues. The rotamer is also chosen based on producing favourable interactions, such as hydrogen bonds, and reducing unfavourable interactions.

The algorithm runs a steepest descent energy minimization (section 2.2.3) with the force field called Gromos96 (van Gunsteren et al., 1996). This is used to correct any

problems with the alignment done in pdbviewer. The final output is a new model of the target sequence as a new pdb file.

2.1.5 Build Preliminary Model for Selected Layers…

Currently there is no documentation for this algorithm outside the pdbviewer software itself. The purpose of this application is to produce a new model created from two structures, pdb files. One of the inputs is a template (reference type structure) and the other is the target (to be modeled). A superimposition is performed using Fit Molecules (from selection) of the target on the template. Using this superimposition as the input the algorithm Build Preliminary Model for Selected Layers is run, molding the target to the template (similar to Swiss Model), and the output is a new model, in the form of a new pdb file for the target, superimposed on the template. This new model and/or the superimposition can be saved as a pdb file for further analysis.

2.1.6 ProSA-web

ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php; Wiederstein & Sippl, 2007; Sippl, 1993) is an online server used to evaluate the quality of a structural model, whether the model is derived from crystallographic / NMR experiments or from predictions. The ProSA algorithm was used in this research to evaluate the quality of all models (section 1.8.8).

2.1.7 ClustalW

This algorithm (Larkin et al., 2007) was performed using the ClustalW server (http://www.ebi.ac.uk/Tools/clustalw2/index.html). The purpose of the algorithm is to align primary sequences to determine the amount of identity between the sequences.

Inputs are copies of a sequence FASTA file, copy and paste into field provided, for all sequences you wish to align. They were performed using the matrix gonnet, gap open value of 100, no end gap value of yes, and gap extension of 10.0. Gaps were created by the algorithm itself and were left alone. The gaps represented the absence of homologous residues. Output was a text file of the sequence alignment with stars underneath identical residues. Amount of identity was determined by counting the number of stars out of the total number of residues in the alignment. Based on alignment of the full length sequences in the crystal structures, the pairs of protein sequences were later trimmed to be the same length before building models and analyzing structures.

2.2 Molecular dynamics simulations:

All these algorithms are found in GROMACS (www.gromacs.org). Gromacs is a suite of applications, which is used for molecular modeling and structure analysis. This includes the capability to run simulations of protein structures.

2.2.1 Pdb2gmx

pdb2gmx is an application that converts a pdb file to a gmx format. The input is a pdb file for the structure of interest. The input has all HETATOMs, waters, and hydrogen atom removed. This file is obtained from pdbviewer or the RCSB protein databank (Berman et al., 2000). The parameter for calculating the gmx file is a force field that determines the charge on the side chains by protonating them, which in this thesis was G43a1. The output is a structure file (*.gro), a topology file (*.top), and a restratints file (posre.itp). The structure file describes the coordinates and velocities for each atom of the structure.and the topology file describes the characteristics of the structure.

Characteristics described in the topology file include dihedral angles of the backbone and sidechains, the charge and mass of each atom, the bonding of one atom to another and the way it interacts with other atoms, parameters for interactions and linkages and classification of atom type, and position restraints on any waters found in the structure. The restraints file describes the atom to atom bonds and linkages to restraint and the type of restraint. The file posre.itp details a set of restraints associated with an "equilibration process" used in later stages of simulation flow. The restraints to place on the protein are listed in the topology file posre.itp atom by atom.

2.2.2 Grompp

Grompp is used to prepare a single input to other programs by combining the information of multiple other files. The input is a structure file (*.gro) and a topology file (*.top), and a parameters file (*.mdp) that contains the specifics of how the output file will be used. The preprocessor uses these files to generate a new single topology file (*.tpr) that can be used in subsequent programs.

In some runs of grompp, constraints are indicated in the associated parameters file (*.mdp) for that step. This makes grompp use an algorithm called LINCS. LINCS takes unconstrained position coordinates and applies bond length constraints to the bonds between all the atoms (Hess et al., 2007). The input is the updated positions of a simulation that are no longer constrained due to being allowed to move during the simulation. The output is new positions of the atoms with the bond constraints applied (Hess et al., 2007).

2.2.3 Mdrun

Mdrun is the application that allows for the simulations and energy minimizations to actually be executed. Mdrun takes as input the topology file (.tpr) from grompp, and uses it as the starting topology and coordinates for the molecule. In addition, mdrun uses a parameter file (*.mdp) that contains the information for either performing an energy minimization or a molecular dynamics simulation. The outputs are log (*.log), trajectory (*.trr or *.xtc), coordinates (*.gro), and energy files (*.edr). The first of two possible trajectory outputs (*.trr) is a full version of the trajectory and the other is a compressed version (*.xtc). The compressed version does not contain as many time points as the *.trr file. I always used *.trr for analysis of the trajectory in VMD (section 2.2.8). The trajectory and coordinates files have coordinates and velocities for individual frames of the simulation. The trajectory file contains this information for the whole simulation whereas the coordinates file only contains information for the last frame or in other words the final model.

If mdrun is used for a minimization then the parameter file includes information about how to link the algorithm to the source of the files needed. Also the parameter file includes parameters for the minimization and name of the algorithm for performing the minimization. The algorithm options are either “steep” for steepest descent or “cg” for congruent gradient (see further explanation below). Other information included in the parameter files are specific restrictions on electrostatic and van der Waals interactions. This includes cut off distances for distance to an interacting atom and if it will be in range to produce such an interaction.

An algorithm called "steep" allows for steepest descent minimization to minimize the total energy of the system. This involves the calculation of the distance and direction of the atom to move based on the force calculated to act on it. The potential energy function describes all of the atom to atom interactions. The derivative of this function with respect to the position of each atom gives the original force on the atom. The new direction of descent along a "potential energy gradient" is based on the current position described by a potential energy function for the molecule given the specific coordinates. Added to this initial position is the position change due to the force acting on the atom. This change in distance is calculated by taking the force acting on the atom divided by the maximum force in the x, y, or z direction. This ratio is then multiplied by the maximum possible distance allowed for an atom to move (0.01nm) in a single step of the minimization. This calculation results in a new vector for movement of the atom (see equation below). This algorithm works well when not close to a minimum in the potential energy.

$\mathbf{r}_n$: The position of the atom before the current step of the minimization. This is a position vector for space coordinates (x_n, y_n, z_n)

V_n: Potential energy function of the atom for the current step - a function of $\mathbf{r}_n$

$\mathbf{F}_n$: Force applied on the atom for the current step : a vector function of $\mathbf{r}_n$ where:

$$\mathbf{F} = -\text{grad } V = -(\delta V / \delta x)\,\mathbf{i} - (\delta V / \delta y)\,\mathbf{j} - (\delta V / \delta z)\,\mathbf{k}$$

where **i**, **j**, and **k** are the unit vectors for (x, y, z).

max($|\mathbf{F}_n|$): Maximum magnitude of force allowed in the routine for F_n.

h_n: the maximum distance allowed in the routine for moving the atom during the current step

$\mathbf{r}_{n+1}$: the position of the atom after the current step

$\mathbf{r}_{n+1} = \mathbf{r}_n + h_n \mathbf{F}_n / \max(|\mathbf{F}_n|)$

Conjugate gradient minimization performs a similar operation to steepest descent with the same parameters. The difference is that it determines the new direction along the potential energy function by evaluating all the previous steps. This works well when close to a minimum in the potential energy function and not when far away.

2.2.4 Editconf

Editconf is used to put a minimized protein modeled in a vacuum into a "periodic" box of specific shape and dimensions so that it can be simulated. This allows a more efficient simulation due to only having to evaluate atoms in a restricted space. To be periodic means that the box with the protein inside is repeated. The input is a structure coordinates file (*.gro). It uses parameters which are the shape and size of the box. The output is a new structure file (*.gro).

2.2.5 Genbox

This application is used to solvate the protein within the box created by Editconf to create a more natural like environment for the simulation. The input is a structure file of the protein within a box from editconf (*.gro). To perform solvation, genbox uses a water model input, which is a box filled with water molecules as a coordinate file, to place water molecules around the protein found in the box (Berensden et al., 1981). The

output is a new structure file (*.gro) and a topology (*.top) file detailing the additional water molecules,which are now a part of the protein system.

2.2.6 Genion

This application is used to neutralize the solvated sytem within the box to create a more natural like environment. The input is a structure file (*.gro) and the associated topology file (.top) of the protein solvated within the box from Genbox. In addition, it uses a parameter of concentration of sodium chloride to be used. Ions are added to random locations within the box. Output is the new topology file (*.top) and a new structure file (*.gro) now including ions**.**

2.2.7 Pbc whole

This application is used to correct the problem of a protein with extremely long bonds as seen using the visualization program VMD. Therefore the program makes the molecule whole again. The long bonds are from atoms moving outside the periodic box they began in and into an adjacent one. The action of pbc whole is to make molecules moving from one periodic box to the next a single whole molecule inside a single box. This is done by the application removing the boundaries of the periodic box.

The input is a trajectory file (*.trr) and a topology file (*.tpr) created by mdrun. The output is a new trajectory file (*.trr) where all bonds are visualized as the lengths assigned in the topology file (*.top), assuming they were all calculated correctly during the simulation.

2.2.8 VMD

VMD (Humphrey et al., 1996) is a molecular viewing program for structures and molecular dynamics simulations. It is similar to pdbviewer except that the structure can be saved in one of many formats, including but not limited to *.gro and *.pdb. To view a molecular dynamics simulation a structure representing the final structure of the simulation is loaded in the VMD program and the trajectory attached to the structure. The attachment of the trajectory informs VMD of the multiple different conformations the structure can have and the time points this occurs. Thus the structure can be viewed as a movie of the structural movements from the first to the last that occurred during the simulation.

2.3 Simulations flow

Global view of simulations: A protein structure file (*.pdb) from the RCSB (Berman, 2000) is converted to a protein in a solvated and neutralized box. The simulation of the protein movements based on atom to atom interaction forces are then performed on the protein in this box.

Step 1. Conversion of pdb file format to gmx format:

Algorithm: pdb2gmx

Input: pdb file for the structure of interest.

Parameters: force field G43a1 that determines the charge on the side chains by protonating them.

Output: Structure file (*.gro) and a topology file (*.top), goes to step 2.1 into Grompp. The result is the same structure in two new files, each containing unique

information. The structure file describes the coordinates and the topology file describes the characteristics of the structure. Characteristics described in the topology file include angles of the backbone and sidechains, the charge and mass of each atom, linkage of atoms, and position restraints on any waters found in the structure. At this point no water is found in the topology file as it has yet to be added (step 3.2). The other output is a topology file called "posre.itp" that is used in step 7.2 . This file details all required restraints for the input structure to be applied later.

2. Minimize protein in vacuum:

The purpose is to create an energy minimized version of the protein in a vacuum so that it will be ready to be placed into the periodic box.

Step 2.1 Create file to be minimized:

Algorithm: Grompp

Input: Structure file (*.gro) and topology file (*.top) from step 1.

Parameters: The file is called minim.mdp. The file includes the information that I want to do a steepest descent minimization in 2000 steps. The cut-off value for electrostatic interactions that could be possible for two atoms was set at 1.0 nm between two atoms as was done in Zhou and Siegelbaum, 2007 and Wang et al., 2008. Also the minimum force that the minimization would ideally reach was set at 1.0 kJ mol^{-1} nm^{-1}, which was set very small to ensure that only minor changes during minimization took place; it is assumed that the structure already is nearly at a minimum. Lastly the periodic boundary conditions for this minimization were not included. Also set by default was the maximum allowable distance for a single atom to move during a given step at 0.01nm.

Output: The topology file (*.tpr) represents the input file for mdrun, step 2.2, which performs the energy minimization.

Step 2.2 To run a steepest descent minimization (section 2.2.3):

Algorithm: mdrun

Input: topology file (*.tpr) from Grompp step 2.1 that includes all parameter information from grompp.

Parameters: “nice” to determine the allocation of processor resources, this was set to use all available processors. This minimization is done for 2000 steps.

Output: trajectory files (*.trr and *.xtc), structure coordinates file (*.gro) and details of the energy minimization (*.edr). Structure file (*.gro) is used by editconf in step 3.1. This is the minimized protein now ready to be put into a box.

Step 3. Create a box around the protein and fill spaces with water molecules:

The idea is to place a single molecule of the protein into repeating boxes all of the same shape and size, then fill in the spaces of the box found around the protein with water molecules.

Step 3.1 Create a box with protein:

Algorithm: editconf

Input: structure file (*.gro) from 2.2

Parameters: Shape of box (octahedron); longest distance from the outermost protein atom to the edge of the box is 0.9 nm, as was used in Zhou and Siegelbaum, 2007 and Wang et al., 2008.

Output: Structure file (*.gro) of the protein inside the new box. This file goes into genbox in 3.2.

Step 3.2 Solvate the protein within the box:

Algorithm: Genbox

Input: Structure file (*.gro) from step 3.1.

Parameters: Water model file (spc216.gro) to input water molecules into the box where the protein does not exist (Berensden et al., 1981).

Output: Structure file (*.gro) after having solvated the box, and a new topology file (*.top). This is now the protein plus water within a new periodic box.

Step 4. Create file to be used for minimization of box with ions:

Algorithm: Grompp

Input: Structure file (*.gro) and topology file (*.top) from step 3.2.

Parameters: A parameters file to determine how to create a single topology from coordinates and angles. The parameters file is called minim2.mdp and differs from minim.mdp in two ways. First, it defines a grid for the atoms that restrains their movements. This restraint is required now and not before step 2 as the atoms are now solvated and no longer in the vacuum. The atoms will move more within a solvated box than in a vacuum (step 2). Thus movement of atoms are now more of an issue for counting the neighboring atoms. Second, the minimum force, which the minimization would ideally reach between an atom and interacting partners, was set to 1000 kJ/mol^{-1} nm^{-1}. This allows the minimization to always be going down the steepest descent of the

potential energy function describing the protein prior to each step. This minimization is done for 2000 steps.

Output: topology file (*.tpr) goes to genion in step 5.

The output is the system of the protein with the waters inside the box all minimized. Also this step will output the charge on the system, which is used to determine how many sodium or chloride ions to add to the box by Genion.

Step 5. Add ions to the box:

Algorithm: Genion

Input: Topology file created by Grompp in step 4 (*.tpr)

Parameters: Parameters used are concentration of the NaCl solution (0.01M) and the assumption that the protein is already neutral.

Output: As a result of genion sodium ions and chloride ions were added to the box at random positions. The output is a structure file (*.gro), a log file (*.log) for the results of the addition of ions to the box and a topology file with the ions included (*.top). The structure and topology file is used in step 6 by grompp. The two files represent the protein in the box, solvated and neutral in charge due to the added sodium and chloride ions.

Step 6. Minimize box solvated including a neutralized protein:

The whole box, ions, water and protein must now be minimized. This is performed using the structure file (*.gro) obtained from adding ions to the box in step 5.

Step 6.1 Create a topology file for minimization:

Algorithm: Grompp

Input: Structure file (*.gro) and topology file (*.top) from step 5

Parameters: Uses the same parameters as in step 4. Thus this is exactly the same method as step 4.

Output: topology file (*.tpr) goes into mdrun in step 6.2. The output is the neutral system of a protein with the waters inside the box ready to be minimized.

Step 6.2 Minimization of the box, ions, and protein:

Algorithm: mdrun

Input: topology file (*.tpr) from Grompp step 6.1 that includes all parameter information from grompp.

Parameters: Same as step 2.2.

Output: The output is two trajectory files describing the movement of the atoms during a steepest descent minimization in full (*.trr) and compressed forms (*.xtc). Also output is a structure file (*.gro). This structure file represents a minimized neutral system of a protein, solvated within a box ("The system") and is used for further minimization in step 7.1.

Step 7. Slow minimization (conjugate gradient):

The whole system (box, ions, water and protein) must now be minimized further to get the most ideal structure prior to simulation. (section 2.2.3)

Step 7.1 Create a topology file for the conjugate gradient minimization:

Algorithm: Grompp

Input: Structure file (*.gro) from step 6.2 and topology file (*.top) from step 3.2.

Parameters: Uses a new parameter file called minim3.mdp. This file is different from the others in step 4 and 6 as it uses a different algorithm for minimization. This minimization algorithm is called "cg" and performs a conjugate gradient minimization. Also different is the minimum force to which the minimization would ideally reach, 50.0 kJ mol^{-1} nm^{-1}, which is smaller than steepest descent minimization. The steepest descent minimization should have brought the protein close to the global minimum for the potential energy function describing the protein. Also the number of steps is 20000 steps.

Output: topology file (*.tpr) goes to mdrun in step 7.2.

The output is the neutral system of the protein with the waters and ions inside the box ready to be minimized.

Step 7.2 Minimization of the box, ions, and protein:

Algorithm: mdrun

Input: topology file (*.tpr) from Grompp step 7.1 that includes all parameter information from grompp.

Parameters: Uses the nice parameter as in step 4. This is a conjugate gradient minimization rather than steepest descent. This is done for 20000 steps.

Output: The output is a two trajectory files (*.trr), a structure file (*.gro), and an energy file (*.edr) and a topology file (*.top). This structure file is now used to feed into grompp for the preparation for the simulation.

Step 8. Position restrained refinement of the protein model:

The purpose is to minimize the protein with position restraints on the atoms within the solvated box to prepare it for the 20ns simulation (step 9).

Step 8.1 Prepare structure for restraint refinement:

Algorithm: grompp

Input: The minimized structure file from the mdrun of step 7.2 (*.gro) and the topology file (*.top).

Parameters: A parameters file called pr.mdp was used. The file includes statements to tell grompp where to find appropriate files (/lib/cpp), that it must use the posre.itp file for position restraints (-DPOSRES), to include the topology file for reference (-I../top), name of integrator (md), total simulation time (1000 ps), time step in femtoseconds (0.002), number of steps (50000) and frequencies to write to the output files (5000). Also the neighboring was defined with ns_type (grid) and to produce the list neighbour list (rlist = 0.8), coulomb type was particle mesh ewald (coulombtype = PME, rcoulomb = 0.8) and fourier spacing (fourierspacing = 0.15), treatment of van der Waal's (rvdw 1.4) and periodic bound conditions (pbc = xyz). Also the file includes the constraints statement "All-bonds" that constrains all atom to atom bonds to the default distance. Also in this file is the temperature (V-rescale) and pressure coupling (Parrinello-Rahman coupling) coupling were used to maintain the set temperature (300

K) and pressure (1 bar). The temperature coupling groups were protein and non-protein and the pressure couple type was isotropic. Lastly the file includes velocity generation using default values.

Output: A new topology file (.tpr) for the system ready to be position restrained.

8.2 Perform restraint refinement:

Algorithm: mdrun

Input: topology file (*.tpr) from Grompp step 8.1 that includes all parameter information from grompp.

Parameters: Same as step 2.2.

Output: A new topology file for the position restrained protein (*.tpr), two trajectory (*.trr) files, a structure file (*.gro) and an energy file (*.edr).

Step 9. MD of the whole protein:

Step 9.1 Prepare structure for simulation:

Algorithm: Grompp

Input: Structure file (*.gro) and topology file (*.top) from step 6.2

Parameters: Grompp uses the parameter file fullmd_sol.mdp. This is the same as pr.mdp, except for setting a simulation length of 5 ns. This is performed in 250000 steps at steps of 0.002 femtoseconds. The coulomb distance (rcoulomb = 0.9) and for this parameter file rlist = 0.9. The particle mesh ewald order was set to 4 (pme_order = 4 and optimize_fft = yes).

Output: topology file (*.tpr) goes to mdrun in step 7.2.

The output is the neutral minimized system, of the position restrained protein, with the waters and ions, inside the box ready to be simulated.

Step 9.2 Simulation of the protein system:

Algorithm: mdrun

Input: topology file (*.tpr) from Grompp step 9.1 that includes all parameter information from grompp.

Parameters: Uses the nice parameter as seen in step 2.2.

Output: The trajectory files (*.xtc and *.trr) for 5 ns simulation. Also output is a structure file of the last frame of the 5 ns simulation (*.gro) and an energy file for the simulation (*.edr).

9.3 Extension of the simulation

This step involved preparing the topology file for extension of the simulation to 20 ns. This was performed by simulating for an additional 15 ns.

Algorithm: grompp

Input: structure (*.gro) file from the mdrun for 5 ns and topology file (*.top).

Parameters: Extension of simulation by 15 ns, thus this is the fullmd_sol.mdp file with 5ns changed to 15ns within the file. Step size was 0.002 and 250000.

Output: topology file (*.tpr) goes to mdrun in step 9.4.

9.4 Simulation of the extra 15 ns

Algorithm: mdrun

Input: Topology file from 9.3 (*.tpr), meaning that the topology file represents the topology after 5 ns of simulation.

Parameters: Uses the nice parameter as seen in step 2.2.

Output: The output is a trajectory file (*.trr) for full 20 ns, a structure file (*.gro), and an energy file (*.edr) for the full 20 ns.

3: Methods Part II: Model Building

Molecular modeling was performed in this research to predict the unliganded conformation for HCN* (section 1.4). To perform molecular modeling required use of previously solved X-ray crystal structures from the RCSB Protein DataBank (Berman, 2000).

Using comparative modeling and molecular dynamics (sections 1.8.2 and 2.1, 1.8.3 and 2.2) a total of five new models of HCN* were built including four in the unliganded state (three from comparative modeling and one from molecular dynamics), and one in the cAMP bound state (from molecular dynamics). Each experimental model and predicted model is described here to outline how it was used in this research and specifics for building if it was a predicted model.

3.1 Mlh – Crystal structure of the holo Mlotik1 cytoplasmic domain

At the time I began the project, the Protein Data Bank (Berman et al., 2000) contained only two source crystal structures for the Mlotik1 cytoplasmic domain (Clayton et al., 2007), one in apo form (1u12) and the other in cAMP bound (holo) form (1vp6). I wanted to confirm that the Mlotik1 cytoplasmic domain was an appropriate template for comparative modeling of the apo HCN* domain. Therefore I compared the holo Mlotik1 source crystal structure (1vp6) with the HCN* source crystal structure with cAMP bound, 1q5o (Zagotta et al., 2003). The source crystal structures 1vp6 and 1q5o I required to be similar within the CNBD because then any additional differences between

CNBDs of 1u12 (apo Mlotik1) and 1q5o (holo HCN*) would most likely be due only to the difference in presence of cAMP. Then structures of apo Mlotik1 regions outside the CNBD would be reasonable for HCN2 to adopt in the cAMP bound form as an alternative conformation.

Ideally based on the amino acid sequence alignment of HCN* and Mlotik1 (Figure2.2) it would be desired that residues be chosen from 1vp6 (holo Mlotik1) to span as much homologous HCN2 sequence as possible. This new structure file (Mlh) was made in pdbviewer by using the save algorithm "Selected Residues of Current Layer…". The structure file 1vp6 was downloaded from the RCSB PDB and opened in pdbviewer. The residues R219-R349 of chain C, and the cAMP molecule d360 were selected, and "Save / Selected Residues of Current Layer…" was run saving the file (Table Table 8-1: Crystal structures and modified crystal structures used.). Chain C was chosen as it had higher similarity to 1q5o than chain A, judging from a superimposition (CNBD = 1.17 and Region of interest = 6.01 for chain C and CNBD = 1.17 and region of interest = 19.83, Table 8-2). Residues V218 and G350 were not selected from the 1vp6 structure as they are not found in the apo structure 1u12. This way all structure files could be generated in the same way to eliminate any subjective content selection in the pdb text file. Mlh is found in Figure 8-5 coloured orange and Table 8-1.

3.2 Mla – Crystal structure of apo Mlotik1 cytoplasmic domain

A modified structure file of unliganded Mlotik1 cytoplasmic domain was required that was a single chain of just residues to be the template for comparative modeling. This Mla crystal structure is found in Figure 8-5 coloured brown and Table 8-1.

This crystal structure file was made in pdbviewer by using the same method as for Mlh. The source crystal structure file 1u12 was at the time the only source crystal structure representation of apo Mlotik1 in the RCSB Protein Data bank (Berman et al., 2000; Clayton et al., 2007), www.pdb.org. The residues R219-R349 were selected from chain A and saved as Mla. Chain A was used as the RMSD (Å) for chain A for a superimposition with 1q5o was lower than using chain B of 1u12 (For chain A: CNBD =. 2.022, region of interest = 5.361 and chain B CNBD = 3.178, region of interest = 5.849, Table 8-3)

3.3 H2h – Crystal structure fragment of the cAMP bound HCN*

A modified source crystal structure file of holo HCN2 cytoplasmic domain (HCN*) was required that was the same length as Mlh and Mla for comparative modeling (Figure 8-6) including only residues and cAMP, d201. It was required for analysis of template selection and unliganded models, both comparative and molecular dynamics. Also the crystal structures needed to have only amino acids native to the HCN2 amino acid sequence, thus all selenomethionines were substituted for methionines. This new crystal structure (H2h) was made in pdbviewer from 1q5o (Zagotta et al., 2003) by using the same method as used to create Mlh, selecting residues E1-L133. The structure file 1q5o (D443-L643) was downloaded from the RCSB PDB and opened in pdbviewer. All selenomethionine entries in the pdb text file were changed to methionines by making a series of edits. The edits necessary were changing all “HETATM” denoting atoms of a selenomethionine to “ATOM”; changing all “Se” to “S’ for each selenomethionine; lastly changing all “MSE” for lines of a selenomethionine to “MET”. The residues E1-

L133 were then saved as H2h. H2h is found in Figure 8-5 coloured in red and in Table 8-1.

In this thesis, all models of HCN* were built using the same primary sequence as H2h. This leaves out some residues from the holo source crystal structure of HCN* from HCN2. As seen in Table 8-1 residues for A'-C' and part of D' were not included in H2h as there were not homologous to residues of Mlh.

3.4 H2R – Comparative model of unliganded HCN* built using a superimposition of H2h amino acid sequence onto template

A preliminary comparative model was needed in order to submit a model-building request to Swiss Model to create another unliganded model described below (H2a). This preliminary comparative model was built using the amino acid sequence of H2h in FASTA format and pdb file Mla. The raw sequence was fit onto the structure of Mla using the pdbviewer algorithm "Fit raw sequence". The output was a pdb file for the preliminary comparative model of HCN*, which is H2R. It was named H2R as it was an HCN2 (H2) comparative model built with the algorithm Fit Raw sequence...(R). This model H2R is found in Figure 8-5 coloured in magenta and in Table 8-1.

The screen showed a superimposition of this model superimposed on Mla. After modeling the coordinate entries in the pdb file for residues K67 and G68 were all 9s This is because they do not have homologous residues in Mla so the model building was unable to produce actual coordinates (Figure 8-6). Doing a superimposition with these coordinates caused H2R to disappear from the screen prohibiting analysis. Also an RMSD was displayed of -.01 demonstrating that an error had occurred in trying to use this structure for the superimposition. Once the faulty coordinate entries were removed a

superimposition allowed H2R to be seen in the superimposition and a valid RMSD was displayed.

3.5 H2a – H2R after submission to Swiss-Model

Creating H2R and then submitting the superimposition to Swiss-Model in project mode built a new apo HCN* prediction comparative model. The output comparative model of Swiss-Model (section 2.1.4) was a new pdb file named H2a. This is found in Figure 8-5 coloured in blue and in Table 8-5.

H2a and H2R have high similarity as is described in more detail in section 4.3 and seen in Figure 8-12. Thus the amount of change created by Swiss-Model was minimal.

3.6 SA2- Comparative model of unliganded HCN* using H2h and Mla structures

A comparative model was built to determine if the type of target input and algorithm determined the model conformation and secondary structure. The comparative model was built in two steps; the first step superimposed the two crystal structure inputs and the second built the model. The new comparative model was named SA2. This model is found in Figure 8-5 coloured in purple and in Table 8-1: Crystal structures and modified crystal structures used. The name SA2 was born out of another stream of comparative modelling that is not described in this document.

Initially both structures H2h and Mla were loaded into the pdbviewer environment. The first step was performed by selecting all residues of H2h and Mla after loading in pdbviewer and running the algorithm Fit Molecules (from selection). The second step was performed by running an algorithm called “Build Preliminary Model for

Selected Layers..." The build algorithm to make the new comparative model mimic the template altered the conformation of H2h. The output was a superimposition of the altered structure of H2h along with Mla, that could be saved as a pdb file. After modeling the coordinate entries in the pdb file for residues K67 and G68 were removed from the pdb file of the new comparative model (section 3.4).

3.7 G3 molecular dynamics model of cAMP bound HCN*

To determine whether the structural predictions made by the comparative models could be confirmed with a template independent method, molecular dynamics simulations (sections 2.2 and 2.3) were performed to simulate H2h in a solvated octahedron box, including sodium and chloride ions to neutralize the system and cAMP in the binding domain. A total of two sodium and three chloride ions were added to the system. The simulation was performed for 20 ns. The resulting trajectory final frame was used as a new holo HCN* model called G3. The molecular dynamics model G3 is seen in Figure 8-1 coloured yellow and in Table 8-1.

3.8 G4 molecular dynamics model of apo HCN*

This was done the same way as G3 except cAMP was not included in the binding site. A total of two sodium ions and four chloride ions were added to the system. Again the final frame of the 20 ns trajectory (after running pbc whole) was used as a new apo model for HCN2 cytoplasmic domain called G4. This molecular dynamics model is found in Figure 8-1 coloured green and in Table 8-1.

4: Comparative modeling of HCN* and evaluation of models

I wanted a structural prediction for HCN2 cytoplasmic domain (HCN*) in the apo state, that is, when cAMP was not bound in the binding site of the cyclic nucleotide binding domain (CNBD). The concentration was on the differences between holo crystal structures and apo comparative models for the D'-A helices. I set out to build models of apo HCN* using comparative modeling. The template structure was the crystal structure of Mlotik1 cytoplasmic domain without cAMP bound.

4.1 Mlotik1 can be used as a template to build an unliganded HCN* model

I hypothesized that apo Mlotik1 would be a good protein to use for a template in protein's comparative modeling. This was based on the proteins ability to bind cAMP in a CNBD (Clayton et al., 2004; Altieri et al., 2008). Similar to HCN2 where binding cAMP causes an increase in conductance, the binding of cAMP in Mlotik1 causes opening of the channel (Zagotta et al., 2003; Clayton et al., 2004).

The similarity of the CNBDs in structure was also of particular importance when considering a reasonable template. I required that the template protein have the ability to bind cAMP. This way it would be likely that the similar structures would have a similar set of conformational changes for binding cAMP. Thus Mlotik1 made for a good template based on this criterion (Clayton et al., 2004). In turn, this would allow the

prediction that if the CNBDs are the same then the rest of the protein is likely undergoing similar conformational changes.

The first quality control for template selection was to ensure significant primary sequence alignment homology for the cytoplasmic domains of Mlotik1 and HCN2. Primary concern was the identity between the sequences of the CNBD (Figure 8-4).

I expected similarity in the CNBD based on the likely evolutionary conservation required for specific residues to give rise to the more important secondary and tertiary conservation. The reason tertiary conservation is more important is due to the necessity of the CNBD to bind cAMP. This is to say that it is not unreasonable that the same function could be performed by different residues so long as overall conformation is the same. However, if there is not significant sequence homology then we would not be able to use that template, as there would not be enough similar anchor residues to help decide which residue of the target (HCN*) sequence was matched with a residue of the template.

A threshold for sequence identity of 25% was used as this is a minimum used in the technique of molecular replacement. The threshold of 25% is a rough minimum for template and target structure sequences identity as demonstrated by (Schwarzenbacher et al., 2008). Molecular replacement is a technique used to solve a crystal structure using a known structure as a template for the known target structure (Rhodes, 2006). This method of template selection assumes that sequence identity implies tertiary structure similarity. Sequence alignment was performed using ClustalW and identity evaluated (section 1.8.1, 2.1.7) (Larkin et al., 2007).

4.1.1 Evaluation of similarity: for cAMP bound HCN* and MlotiK1

Trimming of HCN* sequence was performed based on the alignment (Figure 8-4) to create a new sequence and the resulting crystal structure was named H2h. The result of trimming is shown in Figure 8-4 in that residues E501-L633 now define the fragment of HCN* that will be used for comparative modeling. In this thesis, all models of H2h were built using this particular primary sequence. This structure's sequence does not include residues D443-R500 of the original cAMP bound crystal structure of HCN* from HCN2 (Figure 8-5). Those residues could not have a basis for comparative modeling in the selected template, which was unliganded Mlotik1.

The region of interest to this research (D' to A) was defined as residues E501 to T531 of HCN*. From here on residues in all HCN crystal structures and modesl will be referenced using the numbering of HCN* in the Dictionary (Table 8-5).

I compared H2h (E1-L133 of D443-L645) to Mlh (R219-R349 of 1vp6) and Mla (R219-R349 of 1u12) Sequence identity for H2h and Mlh exceeded the threshold of 25%. The value for identity was 29% (Table 8-4). The identity for the H2h and Mla was 28.2%. This value is less than the comparison with Mlh due to R348A mutation in 1vp6. The sequence identity for the region of interest (H2h: E1-T31; Mlh or Mla: R219-R249) was 22.6% (Figure 8-6). Although this is below the threshold of 25% it is still not unreasonable to use Mlotik1 as a template based on overall identity. It is also not unreasonable based on use of sequence identity in template selection for molecular replacement. It was previously demonstrated that with as low as 19% sequence identity for the whole protein target and template comparison a structure could be solved using molecular replacement (Schwarzenbacher et al., 2007). It was found that the ideal

sequence identity was above 27%, which gave a much higher success rate. Thus it is still reasonable to use Mlotik1 as a template to where I am particularly interested in modeling the region of interest.

RMSD was used for the comparison of three-dimensional "tertiary" structural similarity (section 2.1.1 and 2.1.3) and RMSD (section RMSD)). The threshold for RMSD was set at 3.70 Å. Thus similarity for superimposition of two structures was defined as <3.70 Å, based on the RMSD of Mlh compared to Mla (R219-R349 in both structures) after superimposing the CNBDs (R252-R349 in both structures).

The H2h structure was found to be similar to Mlh in both the CNBD and region of interest. I performed the superimposition (section 2.1.1) using residues of the CNBD in H2h (K34-E71 and L74-L133, not including the nonhomologous residues G67-K68) and Mlh (R252-R349). Based on the appearance of the superimposition (Figure 8-7) it is obvious that the CNBDs are similar (Table 8-7). One difference is that the C-helix (H2h: M120-L133; Mlh: E336-R349) in each protein is in a different position, where the N-terminus of Mlh is closer to the beta-roll than in H2h (Figure 8-7). However, overall the CNBDs demonstrated a high similarity based on the superimposition in not only secondary structure but also tertiary structure. The RMSD for the CNBDs (RMSD = 1.90 Å) supported the observations of the close superimposition.

Overall most of the two proteins show very similar secondary structure for D' (H2h: E1-F6, Mlh: R219-V224), E' (H2h: K10-V12; Mlh: Q228-V230), A (H2h: P23-T31; Mlh: P241-A250) helices and the CNBD. However, A and D' helices are not in the same position. A helix of Mlh is further from the beta-roll than in H2h; D' is tilted further away from the A helix in Mlh than in H2h.

There was also considerable similarity between H2h and Mlh in tertiary structure overall. RMSD of backbone atoms for the whole or parts of the superimposition was calculated (section 2.1.3) by selecting residues that were to be compared in both structures. This was confirmed by the RMSD for the whole domains (H2h: E1-L133; Mlh: R219-R349) (3.13 Å, Table 8-6) although similarity was demonstrated for the region of interest (H2h: E1-T31; Mlh: R219-A250) (5.49 Å, Table 8-6). Thus H2h and Mlh have overall structural similarity.

The secondary structure of D' , E' , and F' are different. F' in H2h is not a helix for homologous residues in Mlh. D' and E' form a single helix in Mlh, with the intervening residues (H2h: A13-M15; Mlh: A231-V233) being a loop between D' and E' in H2h.

Therefore, Mlh and H2h are highly similar in amino acid sequence similarity, secondary structure and tertiary structure for the CNBDs. Although the amino acid sequence identity, secondary structure and tertiary structure similarity are not overwhelmingly similar this might be an unreasonable expectation. The results of structural and sequence comparisons suggests it is reasonable to use Mlotik1 without cAMP as a template for comparative modeling of unliganded HCN*, assuming similarity of the CNBDs still stands. This is based on the assumption that due to the similarity that does exist between Mlh and H2h, it is reasonable to believe that apo HCN* could assume the same conformation as Mla.

4.1.2 Evaluation of similarity: cAMP bound HCN* and unliganded Mlotik1

I expected comparing cAMP bound HCN* fragment (H2h) and Mlotik1 unliganded (Mla) (Figure 8-8**)** tertiary structures to demonstrate minimal similarity based on differences between H2h and Mlh. I expected that they might still be defined as having similarity for CNBDs. This was demonstrated by the superimposition of the CNBDs (H2h: K34-E71 and L74-L133; Mla: R252-R349) with significant differences in C helix (H2h: M120-L133; Mla: E336-R349) only; the C-helix for Mla is further from the beta-roll than the C-helix of H2h. The similarity for the CNBDs is supported by the RMSD of 2.38 Å. Also seen in the superimposition is that A (H2h: P23-T31; Mlh: P241-A250) helices show a high level of structural similarity (Figure 8-8). As well the secondary structure for the region of interest and CNBD is still similar (Figure 8-8). This is supported by the RMSD for the whole domains (H2h: E501-L633; Mla: R219-R349) that was 3.52 Å. Thus looking at the whole domains they are considered similar but this is mostly due to similarity in the CNBDs.

As for the region of interest (H2h: P23-T31; Mlh: P241-A250) there is minimal similarity. Looking at the superimposition it is seen that there are considerable differences in the position of D', F' residues, and E' homologous residues. D' of Mla is tilted away from A much more than in H2h. Looking at K510 of H2h and Q228 it is obvious that E' homologous residues of Mla are translated parallel to A helix in the C-terminus to N-terminus direction. This is supported by the RMSD for the region of interest of was 5.84Å (Table x). This demonstrates that significant differences exist between H2h and Mla. This was to be expected considering this is a cAMP bound to

unliganded superimposition eluding to possible conformational changes within the region of interest.

4.1.3 Evaluation of structure quality

I predicted that H2h, Mlh, and Mla were good quality and can be used for the comparisons that have to be done to confirm Mlotik1 as a suitable template for comparative modeling. Quality of both crystal structures and models was tested using the online algorithm called ProSA (section 1.8.8). Each model was evaluated based on the interactions of pairs of all pairs atoms by submitting a pdb file of the structure to ProSA-web (section 2.1.6). The evaluation of the structures is expressed by a unit less number called a z-score. A value for the z-score of the protein crystal structure or model of <-3.57 was considered to be of good quality based on a study by Hefti et al., 2004 (Table 8-4). Also, having a score similar to other source crystal structures found in the RCSB PDB for the same length gave further support the structure was of good quality. Lastly, the energy profile assigned by ProSA if below zero demonstrated a good quality structure or model.

Two general factors were used to assess crystal structure quality in H2h. These included structure resolution and R-factor. Both are generally summary parameters from the diffraction experiments of the specific factors described above. The resolution describes how well specific atoms can be distinguished from neighboring atoms. R-factor describes the amount of variation between the electron density and the expected protein backbone and side orientation. The resolution of the H2h structure was 2.3 Å, the R-factor was 26.1% for R-free and 21.6% for R-work (Zagotta et al., 2003). According to the statistics for structures in the RCSB PDB statistics most structures (~95%) have a

resolution of < 3.0 Å and/or an R-free < 30% (as of December 2009). Therefore based on resolution and R-factor, H2h was a good quality structure typical of most others in the PDB.

I expected all three crystal structures to be of good quality. Each model or crystal structure was sent to ProSA and the outputs are seen in figures 3.3 and 3.5. The values for each z-score did in fact show that each structure was good quality (H2h: -6.99; Mlh: -6.25; Mla: -6.30; Table 8-4). Also looking at the energy profiles for each structure they all have profiles under zero for the use of a forty residue window for evaluation (Figure 8-9). Lastly the plots of the z-scores indicate that all of these structures have values similar to other structures of similar length far below the cut-off. Thus H2h, Mlh, and Mla are reasonable structures in terms of characteristics of the atoms.

Mlotik1 is a suitable template for comparative modeling based on small structural differences between H2h and Mlh and reasonable differences between H2h and Mla. The primary sequence, secondary structure, and tertiary structure for cAMP bound structures comparison meets the criteria meeting the first necessity for a good template for the region of interest. The CNBDs were found to be similar in structure for H2h and Mla so and differences in the region of interest are expected considering this is a comparison for structures bound (H2h) and unliganded (Mla) with cAMP. Thus not only is it reasonable that Mlh could be an alternative conformation to H2h for HCN* (E1-L133) but that HCN* could change conformation to be similar to Mla. Thus Mla can be used as a template for comparative modeling.

4.2 H2a model was built successfully using Mla as the template.

H2a was built in two steps. First, there was a superimposition of H2h amino acid sequence and Mla to create a preliminary model (H2R, section 3.4, Figure 8-5: Crystal structure models of the proteins HCN and Mlotik1.). Second, the superimposition was used for improvement of the H2R model with energy minimization, rotamer selection and alignment refinement to create H2a (section 3.5) (Schwede et al, 2006) (Figure 8-5).

Considering that Mla was the template for creating H2a, it was predicted that their tertiary structures would have high similarity if the algorithm worked properly. This is to say that the function of a comparative model algorithm is to produce a prediction of a structure based on using the coordinates of the template to build the model structure (Guex and Peitsch, 1997). Thus if model building is performed correctly then there should be high similarity between the model and template tertiary structure. An exact match in structure was defined as an RMSD of the two structures <1.00 Å (calculated based on a superimposition of the CNBDs).

Based on the Figure 3.4 it is obvious there is considerable similarity for both the region of interest and CNBD. There are no parts showing significant differences. This demonstrates that H2a was built properly based on the template being a blueprint for H2a conformation. The RMSD was determined to be 1.16 Å for whole structures (H2a: E1-L133; Mla: R219-R349) and 0.63 Å for the region of interest (H2a: E1-T131; Mla: R219-A250). This value does not meet the criteria for exact matching overall, suggesting that the refinements have been implemented to create a new model. Ideally this model is better than the original input model (H2R). Overall the RMSD values show the model building process to be correct for conformation of the model.

I predicted that the model would also pass the test of good quality based on Z-score <-3.57 (Hefti et al., 2006). Looking at Figure3.5 it is obvious that the energy of the model rises close to zero but never over. Thus there is a region of H2a that has some less than ideal structure. This is located near residues K67 and G68. These residues do not have homologous pairs in Mla, seen as spaces in the alignment of amino acid sequences of HCN* and Mlh, so it is reasonable that this would create problems in the new model (Figure 2.1). Overall the point on the plot does show that H2a has good quality compared to other structures of the same length. This is confirmed by the z-score (-6.37, Table 8-4), which is similar to that of the template. Thus the z-score is showing that the model is as good quality as the template.

So H2a is accepted as a comparative model that could be used for analysis. This was based on the confirmation and the z-score and energy profile.

4.3 H2R similarity to H2a demonstrates proper model building

During the process of building H2a, a preliminary model called H2R was built (section 3.4). H2R gives an additional comparative model, built using a strategy that was not exactly the same as the one used to make H2a. H2R was built the same way using the superimposition of H2a (section 3.5). The difference was that no energy minimization, alignment checking or rotamer searching was performed to create this new model. The purpose of examining the H2R model was to determine the structural differences between this model and H2a due to the refinements.

In building H2R, the target amino acid sequence was H2h sequence that would be modified in the modeling process to produce H2R. The template structure was Mla. The

model H2R was built by creating a superimposition of the H2h sequence and Mla structure. All residues in both the H2h amino acid sequence and Mla structure were superimposed. The output is a superimposition of the new model (H2R) and the template structure (Mla).

I hypothesized that H2R and H2a should be similar if there is no algorithm dependence in model building. Based on the superimposition of CNBDs (H2R and H2a: K34-T66, N69-L133) it is obvious that H2R and H2a are similar (Figure 8-12) (RMSD for CNBDs = 1.29 Å, for whole structures (E1-L133): 1.16 Å) (Table 8-7). There are two residues K67 and G68 that were removed from the H2R pdb file (section 3.4). This demonstrates that H2R and H2a are similar models.

The comparative model H2R is an acceptable model for apo HCN*. The acceptability of the model was based on model quality and model building correctness. The model quality was evaluated to be good as with H2a (z-score: -5.91, Table 8-4). This is higher than H2a and is validated by the energy profile of H2R compared to H2a. Both profiles increase for the same residues thus the nonhomologous residues (Figure 8-11) to Mla are an issue more in H2R since they were removed from the structure altogether. Thus the refinements made to H2a have created a more ideal model than the superimposition alone. This was true for the whole model based on Z-score values (H2a: -6.30 and H2R: -5.91, Table 8-4) and for structure sequentially near the residues K67 and G68 based on the energy profiles (Table 8-11). Specifically this is seen as a smaller peak and less significant jump in the energy profile for residues in sequentially near K67 and G68. Although, the z-score is higher than H2a it is still below the threshold of <-3.57 and so considered to be of good quality. Also the model building process was

evaluated by superimposition of H2R and Mla giving an RMSD for all Cα carbons of 0.00 Å (RMSD: CNBDs (H2R: K34-T66, N69-L133; Mla: A250-R349) = 0.00 Å and region of interest (H2h: E501-T531; Mla: R219-R349) = 0.00 Å). This indicates that no significant changes were introduced in H2R using Mla as a template. Thus H2R is a mimic of Mla supporting that comparative model building was performed correctly.

4.4 Target file type results in no differences in model structure.

To determine algorithm dependence and explore alternative algorithms for building unliganded HCN* fragment (E1-L133), two structures were used as inputs. The two structures were H2h and Mla. This differs from building H2a and H2R that used an amino acid sequence for the target and Mla for the template. This method is similar to using Swiss-Model as both are taking a superimposition as an input. Thus I wanted to know does the target input type cause differences in the model structure? So I assumed that the algorithms operations were the same for "Build Preliminary Model for Selected Residues" (section 2.1.5) and Swiss-Model (section 2.1.4).

The comparative model was built by creating a superimposition of H2h and Mla (H2h: E1-L133; Mla: R219-R349). This superimposition was then used as the input for the build algorithm. The build algorithm changes the conformation of H2h to become similar to Mla. The output is a superimposition of the new model and the template structure. The new model is called SA2.

I hypothesized that SA2 and H2a should be similar if there is no algorithm dependence in model building. Also, I hypothesized that building SA2 would have a

structure that was the same as H2R showing that target file type does not create differences in the model.

SA2 demonstrates similarity in tertiary and secondary structure to H2a (Figure3.7). It is obvious from the superimposition that SA2 and H2a superimpose well for all parts of the two structures. This is similar to the result for H2R and H2a. The only significant difference between SA2 and H2a is in the loop connecting strand 4 and 5. This is due to the necessary deletion of residues K67 and G68 to be able to use the protein for analysis. The reason for this was the same as that for H2R (section 3.4 and 4.3). Thus SA2 is similar to H2a demonstrating that there is no difference in the model building process. This is supported by the RMSDs comparing SA2 and H2a (RMSD: region of interest (E501-L633) = 0.01 Å; CNBDs (K534-T566, N569-L633) = 0.262 Å; whole domain (E501-L633) = 0.232 Å, Table 8-7). A superimposition of all atoms for H2a and SA2 except for K67 and G68 gives an RMSD of 0.542 Å. Thus SA2 model building with two structures does not produce many differences then building using the strategy for H2a. Thus the assumption that the algorithms Build Preliminary Model from Selected Layers... and Swiss-Model perform the same operations would be incorrect as an RMSD of 0 Å would have been the case if this assumption was true.

SA2 was predicted to have good quality and to have been built correctly. This was shown based on the energy profile being almost identical to H2a. This supports the hypothesis that comparative model building with a structure or raw sequence for the target does not make any difference (z-score: -6.32, Table 8-4). Also the correctness of the model building process was checked based on a superimposition with Mla. All

regions show exact match based on the RMSD (Region of Interest: 0.00 Å, CNBD: 0.00 Å and whole protein: 0.00 Å).

Comparing SA2 and H2R shows no significant difference. Based on a superimposition of the CNBDs of SA2 and H2R (SA2 and H2R: K534-T566, N569-L633) the RMSD = 0 for all residues (E1-T66, N69-L133). Thus this does support the hypothesis that template input file type (amino acid sequence or structure) does not make a significant difference in the model.

5: Unliganded HCN* comparative models contain differences in the tertiary structure of Thm and D' in comparison with bound HCN*.

I hypothesized that the region of interest (HCN2: E1-T31), that is believed to be responsible for transmitting the message of binding cAMP to the pore (Hypothesis #1: section 1.7.1), would demonstrate differences for an unliganded HCN* model compared to cAMP bound HCN* (Hypothesis 2: section 1.7.2). I wanted to compare the comparative models I had built (including H2a, H2R, and SA2, sections 3.4-3.6 and 4.2-4.4) to H2h (fragment of HCN*, section 1.4 and 3.3). This was done to determine if the models were predicting differences in the region of interest. The comparison was done by calculating RMSD values and comparing distances between reference residues in the region of interest, to highlight the differences between the apo models and holo structure.

5.1 An apo HCN* comparative model (H2a) demonstrates structural differences in D'-F' vs. holo HCN*.

5.1.1 E' and F' are likely translating together rather than away from each other.

I hypothesized that the Thm (transmission hinge motif, section 1.4) must have a different conformation in H2a than H2h. More specifically, I hypothesized that E' and F' move away from each other in upon unbinding cAMP (Hypothesis 3, section 1.7.3). This hypothesis was based on the proposal that F' changes position in the HCN cytoplasmic domain for unliganded to bound transition (Taraska and Zagotta, 2007; Zhou and Siegelbaum, 2007; Berrera et al., 2006).

From the superimposition it looks like E' and F' in H2a translate from their position in H2h (Figure 8-14). This translation of E' and F' is confirmed by looking at residues K510 and F518 position in H2a and H2h. It is seen that K10 of H2a, if projected onto the A helix, would be closer to the N-terminus (closer to the viewer) than in H2h. Similarly F18 of H2a is closer to the viewer than in H2h.

These results were confirmed by calculating the RMSD (section 4.1.1 using residues of the CNBDs for the superimposition H2h and H2a: K34-T66, N69-L133) for the whole Thm (H2h and H2a: K10-F18) and both E' (H2h and H2a: K10-V12) and F' (H2h and H2a: P16-F18) individually. Comparing H2a with H2h, the Thm RMSD was predicted to be >3.70 Å as this is a comparison of different cAMP bound states, bound (H2h) and unliganded (H2a). All of the RMSD values (Thm:5.43 Å, E': 4.98 Å , and F': 5.85 Å , Table 8-8) are greater than the criteria for difference in conformation of > 3.70 Å. Thus the whole Thm has a different conformation in the H2a model than the crystal structure H2h. So this supports the hypothesis that E' and F' move in the native protein during the transition from holo to apo transition.

I predicted the distance between E' and F' was greater in H2a than H2h. The distance was analyzed by measuring the distance between E'(at Cα for K510) and F'(at Cα for F518) with the distance tool in pdbviewer found on the toolbar. The distance measurement was performed for both H2a and H2h. Based on the superimposition appearance, it looks as though the distance has remained the same confirming the idea that E' and F' are translated rather than moving relative to each other. Comparing the distances demonstrates that the distance between E' and F' is the same in the apo state than the holo states (difference of 0.79 Å) (Figure 8-14). This distance is much lower

than the absolute distances and below the threshold of 10% of the average absolute value for any distance. A typical threshold value for a significant change in distance was set at 1 Å. This value was confirmed as reasonable based on distances reported in literature (Clayton et al., 2004; Altieri et al., 2008), where values less than 1 Å were not considered significant enough to report the exact value. Thus E' and F' do not demonstrate a change in distance relative to each other. This further supports the idea that the helices are translating together, relative to the CNBD and parallel with the A helix.

5.1.2 D' is tilting away from A helix that is a stationary helix during holo to apo transition.

D' (H2a and H2R: E1-F6) was predicted to have a change in conformation comparing H2a to H2h (Hypothesis 3, section 1.7.3). Based on the superimposition appearance, D' in H2a is definitely not in the same orientation as in H2h (Figure 8-14). Looking at the superimposition for H2a and H2h the long axis of D'-E' is tilted differently from A helix comparing N-terminus of D' and N-terminus of A helix.

I had hypothesized that A helix was stationary during the holo to apo transition relative to the CNBD (Hypothesis 3, section 1.7.3). So I predicted that the A helix in both H2a and H2h would be similar in conformation based on a superimposition of H2a and H2h CNBDs (H2a and H2h: K34-T66, N69-L133, section 4.1.1). This was defined by an RMSD of <2 Å. This was evaluated both with RMSD for single residues of A and RMSD for the whole A helix (P23-T31). The single residue RMSD values are actually changes in position of the single residues in H2a and H2h Calpha atoms. The A helix was demonstrated to be stationary for many of the residues, including F25, A28, M29 and T31 (range 1.387 Å to 1.932 Å Table 8-12) with an overall RMSD for the whole A helix

of 2.02 Å (Table x), demonstrating that A does not change. The RMSD was 7.72 Å for D' (H2a and H2h: E1-F6) (Table 8-8). This meets the criteria for being dissimilar and so D' has a different conformation in H2a compared to H2h.

I predicted that D' is tilted relative to A helix when comparing H2a to H2h. Based on the superimpositions appearance (Figure 8-14) it is clear that the long axis of D'-E' in H2a is angled with the N-terminal end away from A helix. This is compared to the long axis for D'-E' of H2h that is angled with the N-terminal end towards the A helix. This provides support that the conformational change made by D' is the N-terminus tilting away from the C-terminus of the A helix.

5.2 An apo HCN* model (H2R) built without Swiss-Model supports conclusions from H2a.

The comparative model H2R is also an apo model, built to determine the differences in using Swiss Model or not. The comparative model H2R shows all of the same changes in the region of interest as H2a (Figure 8-12 and Figure 8-5). D' N-terminus is tilted away from A helix, which is the same as in both H2a. Also the Thm is translated out of the page parallel to the A helix. The only difference between the two structures is in the loop between strands 4 and 5 that do not superimpose, based on the superimposition of H2R and H2a for a superimposition of the CNBDs (H2R and H2a: K34-T66 and N69-L133) (Figure 8-12). This is due to the necessary removal of residues K67 and G68 of H2R as they were not defined with coordinates in the pdb file. The similarity is supported by the RMSD values calculated as done in section Evaluation of similarity: for cAMP bound HCN* and MlotiK1 comparing H2R and H2a Region of interest (H2R and H2a: E501-T531) = 0.63 Å; CNBD (H2R and H2a: K34-T66 and N69-

L133) = 1.29 Å; whole domains (H2R and H2a: E1-T66 and N69-L133) = 1.16 Å, Table 8-8.

5.3 An apo HCN* model (SA2) built with two structures supports conclusions from H2a.

The comparative model SA2 is an apo model of HCN* that was built to answer whether using a raw sequence, as for H2R, or a crystal structure, for SA2, as the target would make a difference in the comparative model produced. From the superimposition of H2h and SA2 it is seen that all conclusions about the region of interest of H2a are confirmed by SA2 (Figure 8-16). The only differences between SA2 and H2a are the same small loop between strands 4 and 5 and two small differences in secondary structure. The two differences in secondary structure are at the C-terminus of A and C helices. The similarity was supported by the RMSD, calculated like in section 3.1, for the comparison of SA2 and H2a (Region of interest (H2a and SA2 (E1-T31): 0.01 Å; CNBD (H2a and SA2: K34-T566, N69-L133): 0.262 Å; Whole domains (H2a and SA2: E1-T66, N69-L133): 0.232 Å, Table 8-7).

6: Molecular dynamics simulations of HCN* and evaluation of models

I wanted to build models that did not rely on a template structure to see if they could confirm results of comparative modeling. This is different from comparative modeling where a template structure is necessary to create the model using the target sequence or structure. This was a necessary step to determine the observations that were more likely to be real in the comparative models, by showing that multiple methods produced the same model structure.

In this thesis, all molecular dynamics (MD) models of HCN* were built using the same primary sequence as H2h. This leaves out some residues from the original holo crystal structure of HCN* from HCN2 (section 4.1.1). However, this choice was made because I wanted to compare the MD models to the comparative models of the same length.

6.1 G4 is an unliganded MD model from simulation of H2h.

The first MD model that was constructed was a representation of unliganded HCN*. The model was made by first performing a series of energy minimizations of H2h in a solvated repeated box, and then simulating atom movements for 20 ns. The cAMP was not included in the binding domain. The resulting final frame of the simulation represented a new unliganded HCN* fragment structure. During the simulation a molecule structure was produced that spent some time outside of the box or

at the border. Thus to correct the protein so that it was contained in a single box an algorithm was run that removed the periodic boundary conditions post simulation (see pbc whole in chapter 2). The new final frame of the simulation without periodic boundary conditions was the new model called G4 (Figure 8-5).

To test the quality of the new MD model the total energy and amount of motion in the model were checked. The energy of the MD model was checked using ProSA as described in the methods and in section 3.1. H2h has already been demonstrated to be a good starting structure based on ProSA analysis (section 4.1.3). The z-score was performed on G4, the new MD model, and determined to be -5.74 (Table 8-4). Based on the energy profile compared to H2a the reason for the higher Z-score may be the reasonably even energy profile compared to H2a as K67 and G68 did not produce any extreme energies and neither did any other residues (Figure 8-11). Thus this would actually make G4 a better model overall. Based on the energy profile for H2a (Figure 8-11) some of the residues in the region of interest have a much lower energy than the identical ones in G4. Also some of the residues N-terminal to G68 have lower energies than those in G4. Thus due to the extreme energies of H2a this likely causes the Z-score for H2a to be lower than G4. Since both H2h and G4 structures have z-scores under the cut-off value that is a standard used to evaluate source crystal structure quality (<-3.57, Hefti et al., 2006), not only is the input a good starting structure but the new model is also a good quality crystal structure.

In addition the RMSD over time was used to determine if the model had stopped moving within the box. Thus I wanted to test if the frame that was chosen as the model represented the most stable form of the protein during the simulation, or if instead it was

only a temporary state in transition to a more stable one. I tested this by calculating the RMSD of each frame in the simulation trajectory compared with the first frame. The first frame is similar to H2h but has been subjected to several energy minimization steps including position restraint and equilibration, giving an "intermediate" model. As can be seen from Figure 8-17 it is obvious that significant changes stop occurring at 5 ns of simulation. This is seen in Figure 8-17 by looking at the general shape to the curve and determining by eye that at 5 ns the curve has a plateau at 0.3 nm. Thus the protein crystal structure has stopped the most significant movements. Thus choosing the 20 ns frame is valid as a MD model that has stopped all significant movements and is just vibrating.

Therefore the new MD model is a good quality model. The quality is determined based on the energy and movement of the structure within the box. This model was used for further analysis seen in subsequent chapters.

6.2 G3 is a cAMP bound model from simulation of H2h.

Similarly a holo HCN* model was created. The new MD model was created the same way as G4 but the cAMP was not removed from the binding domain of the starting H2h crystal structure. This new MD model is called G3 (Figure 8-5). This MD model like G4 was a good quality model based on both ProSA (G3 Z-score = -6.17, Figure 8-18 and Table 8-4) and RMSD over time, as was done for G3 (Figure 8-19), and showing most movements stopping after 5 ns. Thus it was valid to use the frame after 20 ns for the bound HCN* simulation model.

6.3 G4 is similar to H2a in CNBD conformation.

I hypothesized that G4 and H2a would be similar in conformation. This was based on both G4 and H2a being apo HCN* models. I predicted that the RMSD would be <3.70 Å. My strategy to test this hypothesis was to measure the RMSD of G4 and H2a in a superimposition of the CNBDs.

Looking at the superimposition of G4 and H2a CNBDs (G4 and H2a: K34-T66, N69-L133, section 4.1.1) (Figure 8-20) it is seen that there is much similarity in the CNBDs except for the C-helix for which there is a difference in position, where C-helix of G4 is closer to the viewer than in H2a. As for the region of interest F' identical residues have a difference in secondary structure where H2a has none but G4 has a helix. The long axis of F5 to K10 is similar in orientation although the position of identical residues is not similar. Positions of E' and F' residues, particularly K10, look to be translated into the page for G4 perpendicular to A helix when compared with H2a. Thus the region of interest is not similar but the CNBD is similar. These results demonstrate that there is similarity in the CNBDs and some differences in the region of interest. This is confirmed by the RMSD values (Region of interest (H2a and G4: E1-T31) = 4.89 Å; CNBD (H2a and G4: K34-T66, N69-L133) = 2.65 Å ; whole domain (H2a and G4: E1-T66, N69-L133) = 3.30 Å , Table 8-7).

Thus H2a conformation is in agreement with G4 for certain properties including the CNBD except for the C-helix position. However, there are some differences in the region of interest.

6.4 G4 supports predictions from H2a for conformation difference in E' and D' helices compared to H2h.

I hypothesized that G4 and H2h would demonstrate differences in the region of interest. This was based on G4 being an apo model and H2h being a known cAMP bound structure, where in a similar comparison using H2a there were differences in the region of interest. Thus I also hypothesized that comparing G4 to H2h after CNBD superimposition would support the predictions made by comparing H2a and H2h. The RMSD calculation to compare G4 and H2h was performed exactly as for comparing H2a and H2h (section 4.1.1). The models demonstrate an overall similarity based on the RMSD for whole domains (G4 and H2h: E1-T66, N69-L133 = 2.72 Å, Table 8-8). This similarity is primarily due to the significant amount of similarity in the CNBDs (RMSD = 1.42 Å, Table 8-8). Looking at the region of interest it is evident that the similarity should be low, and this is confirmed by the high RMSD ((G4 and H2h: E1-T31) = 4.96 Å, Table 8-8) for the region of interest. Thus in the comparison of G4 and H2h the CNBDs are similar but the regions of interest are not similar.

6.4.1 G4 vs. H2h supports predictions of E' translation towards N-terminus of A helix in apo HCN*.

It is seen in the superimposition (Figure 8-21) that H2h and G4 are not similar in the terms of the conformation of E' position. E' (H2a and G4: K10-V12) is translated towards the A helix N-terminus and away from the A helix C-terminus (using A28 and T31 as references that were found to not be different comparing G4 and H2a relative to the CNBDs) (Table 8-12). Residues T31 and A28 are remaining stationary based on the good superimposition of both pairs and RMSD (0.7 Å for both Table 8-12). Thus the existence of a change in the conformation of E' like in H2a versus H2h was supported.

The limitation of the confirmation is that E' and F' RMSD values for G4 versus H2h were smaller than H2h compared to H2a. This means the G4 model does not support as many changes as H2a predicted. As can be seen in the superimposition (Figure 8-21) residue F18 looks to be in the same position but K10 is closer to F18 in G4 than in H2h (Figure 8-21). The qualitative observations are supported by the RMSD values showing that E' (H2h and G4: K10-V12 = 3.99 Å, Table 8-10) is changing a lot but F' ((H2h and G4: P16-F18) = 2.04 Å, Table 8-8) is not changing a lot. Thus G4 agrees with the result of E' changing conformation as was seen for H2a by a translation, but G4 does not agree with the translation of F'.

6.4.2 G4 vs. H2h confirms D' tilt and stationary A helix in apo HCN*.

D' of G4 is tilted with the N-terminal end away from A helix compared to H2h (Figure 8-21). This confirms the tilt of D' seen in H2a compared with H2h (Hypothesis 3, section 1.7.3). It is seen in the superimposition that the long axis of D'-E' (H2a and G4: N5-K10) of G4 does not overlap the long axis for D'-E' of H2h (Figure 8-21). Thus the superimposition shows a difference in long axes existing at different angles to A helix for H2a compared to H2h.

The change in conformation was also measured using RMSD as performed as seen in section 4.1.1 for residues of D' (H2h and G4: N5-F6). The value for the RMSD for D' was 2.94 Å, which is less than the threshold for change of 3.70 (Table 8-8). This value for the RMSD does not agree with the prediction that G4 and H2h are different but this evaluation is limited to the small number of residues that was used for the RMSD calculation. These residues N5-F6 were chosen to demonstrate the change of residues showing identical helical secondary structure. Only two residues can be used as G4 has

only two residues for identical D' residues that have helical secondary structure. If all of the D' residues (H2a and G4: E1-F6) are used the RMSD was 8.94 Å (Table 8-8), very large. However, this difference in RMSD could be accounted for by the loss of secondary structure in G4 at the N-terminus. Thus this is inconclusive as to whether D' is changing based on RMSD as only two residues with defined helical structure could be evaluated.

Lastly G4 and H2h support the prediction that A helix is stationary (Hypothesis 3, section 1.7.3). This is well shown by the superimposition (Figure 8-1). Also the A helix conformation overall ((H2a and G4: P23 – T31) = 2.16 Å, Table 8-12) supports that the A helix is remaining stationary relative to the CNBD. In a more detailed evaluation on a residue to residue basis many of the individual A helix residues are shown as being more stationary than others. This is demonstrated by many of the residues having RMSD below 2 Å (Table 8-12). Thus the prediction that A helix is remaining stationary is also supported by G4 compared to H2h.

6.5 G4 vs. G3 demonstrates a difference for Thm and D' relative to the CNBD.

Based on the apo H2a versus holo H2h comparison, I hypothesized that comparison of unliganded (G4) and bound (G3) MD models would show differences in the region of interest. This comparison was used to test whether a simulation versus simulation comparison could confirm the results of unliganded model H2a compared to H2h crystal.

From the superimposition (Figure 8-22) it is seen that there are considerable differences in G4 compared to G3 in the region of interest. The CNBDs show the most

similarity (2.17 Å, Table 8-8) and the region of interest shows no similarity except in the single helix for D' and E' helices, and the match of the A helix position. The RMSD shows large differences for D' (G4 and G4: N5-F6) = 10.25 Å, Table 8-8) and E' (5.41 Å, Table 8-8) but small differences for A helix ((G4 and G3: P523-T531) = 2.07 Å, Table 8-8).

6.5.1 G4 vs. G3 confirms E' is translating in apo HCN*.

Particularly confirming the differences of H2a versus H2h is that E' has a different conformation (Hypothesis 3, section 1.7.3). From the superimposition appearance, E' K10 is closer to the A helix N-terminus for G4 than G3.

Based on the distance measurements for comparing E' relative to F' in G3 and G4, as performed in section 4.1, there is no significant difference (Distance = -0.37 Å). The superimposition and RMSD for the Thm ((G3 and G4: K510-F518) = 4.09 Å, Table 8-8) confirms that a change in the Thm is occurring. Furthermore it is changing in that E' is closer to A helix N-terminus in G4 than G3 (Figure 8-22). This with all results comparing unliganded models to H2h where E' was closer to the N-terminus of A helix in unliganded models than H2h.

6.5.2 G4 vs. G3 confirms D' is changing conformation in apo HCN*.

There is a tilt in D' comparing G3 and G4 confirming a tilt of H2a versus H2h but not confirming the direction (Hypothesis 3, section 1.7.3). The tilt is obvious from the appearance of Figure6.3 where the long axis of D'-E' of G3 is pointing away from A helix much more than in G4. This is compared to residues of A helix (G3 and G4: P523-T531) that were found to be stationary such as N524 and T531 (Table 8-12).

Also G3 compared to G4 confirms that A helix (G3 and G4: P23-T31) as a whole is stationary. This seen in the superimposition that there is a general similarity between A helices and confirmed by the RMSD for overall ((G3 and G4: P23-T31) = 2.74 Å, Table 8-8) or comparing individual residues (Table 8-12).

6.6 G3 is possibly an alternative cAMP bound HCN* structure

I had at the beginning of this study hypothesized that a simulation of H2h would give a model of apo HCN* such that the cAMP would move out of the binding site during the MD process. So I hypothesized that G3 region of interest (D'-A) would have changed conformation, from H2h, to a similar conformation to H2a. This was not possible after realizing that G3 was obviously a holo model based on the obvious presence of cAMP in the binding site (Figure 8-5).

Thus G3 versus H2a was another unliganded (H2a) and bound (G3) comparison. As was previously seen for unliganded and bound comparisons (section 4.2, 4.3, 4.4, 5.1, 5.2 and 5.3) there was low similarity in the region of interest ((G3 and H2a: N5-T31) = 4.78 Å, Table 8-8, Figure 8-23: The protein models H2a and G3 show similarity in the CNBD only.). As seen in Figure 6.4 D' N-terminus is tilted in G3 much further from A helix C-terminus than in H2a. This was unexpected based on comparisons of H2a and G4 to H2h. Also overall the whole protein does show similarity but as was previously seen this is due to similarity in the CNBDs (Whole domains (G3 and H2a: N5-T66, N69-L133) = 3.101 Å and CNBD (G3 and H2a: K34-T66, N69-L133) = 2.369 Å, Table 8-8).

I hypothesized that G3 would demonstrate similarity to H2h based on both being cAMP bound models (Figure 8-5). Based on the superimposition (Figure 8-24) it is seen

that there is considerable similarity in the CNBDs with a slight difference in position of the C-helix (CNBDs (G3 and H2h: K534-T566, N569-L633) = 2.17 Å, Table 8-13). As for the region of interest the only similarity exists in A helices as can be seen in the superimposition and further supported by the RMSD (Region of interest (G3 and H2h: P523 – T531) = 7.35 Å; D' (G3 and H2h: E501-F506) = 14.08 Å; Thm (G3 and H2h: K510-F518) = 4.95 Å; A (G3 and H2h: P523-T31) = 2.96 Å, Table 8-11). The separation distance between E' and F' was greater in H2h than G3 (difference in distances was G3-H2h: -1.47 Å, Table 8-13) accounting for the lack of similarity for the Thm (RMSD = 4.90 Å). Finally D'-E' long axis is angled away from the A helix, defined as stationary by many residues including F25 and M29 in G3 and H2h, both under 2 Å for RMSD. Thus G3 is shown to not be similar to H2h. Thus it is possible that G3 is an alternative conformation of cAMP bound HCN* fragment (E1-L133).

Also of interest was a difference in the orientation of cAMP comparing G3 and H2h. The cAMP of G3 is in a more elongated conformation than in H2h (Figure 8-24). It could be that cAMP was leaving the binding domain during the simulation and if the simulation had been extended longer it might have been in the surrounding environment at the completion of the simulation.

Thus based on cAMP position and orientation and differences in orientation of model/crystal structure components it is possible that G3 is a transition structure between cAMP bound and unliganded forms, or an alternative cAMP bound conformation.

7: Discussion

I desired to know the structural differences that could be proposed as due to cAMP binding only in HCN channels. I looked at changes in a region of interest defined as D'–A helices. These changes seen described movements of D' and E' helices, with A helix as stationary.

7.1 cAMP binding may cause the gating ring to change pulling S6 away from vertical pore axis.

The message transmission of binding of cAMP in the CNBD to the pore is an allosteric process. The binding of cAMP causes a shift in the PBC to interact with cAMP; the loop between beta strands two and three interact with the R591 to stabilize it and the C-helix after it has moved in towards the cAMP to interact via R635 (Zhou and Seigelbaum, 2007). This results in a shift in the helices N-terminal of the beta-roll (including D'-A helices) away from the beta-roll, proposed based on mechanism of other CNBD containing proteins (Kornev et al., 2008). These results have been supported by structural studies using FRET that show movements of F' upon binding cAMP (Taraska and Zagotta, 2007; Taraska et al., 2007). This change in the C-linker helices is known to cause a change in D' and A' helices (Hua and Gordon, 2005; Craven and Zagotta, 2006; Craven et al., 2008). Also, Alteri et al., 2008 demonstrated significant differences in the N-terminal helices of cAMP bound and unliganded states of the cytoplasmic domain of Mlotik1.

D' is the most C-terminal helix of the gating ring and is now believed to play an integral role in the modulation of HCN channels. D' interacts with B' of another adjacent subunit of the tetrameric HCN2 channel based on the crystal structure solved by Zagotta et al., 2003 and Flynn et al., 2007. Since I believe these interactions must be stable to allow a stable gating ring and holo cytoplasmic domain conformation, then structural changes in D' must be sending a ripple effect up to the S6 helices of the pore due to structural changes from C-terminal to D'.

C-terminal to D' is E', F' and the CNBD of both Mlotik1 and HCN2. In my thesis, it was shown that A-helix undergoes very little structural changes upon binding cAMP (Figure 8-23, Figure 8-22, Figure 8-21, Figure 8-16 and Figure 8-14). Also like other CNBDs (Clayton et al., 2005; Taylor et al., 2005) the beta strand region (beta roll) does not change significantly change in conformation but the C-helix is proposed to swing out with cAMP leaving the binding site (Taraska and Zagotta, 2007; Matuleuf et al., 1993; Varnum et al. 1999). The previously published information, and the idea that it has been previously proposed that message transmission is due to a ripple effect sequentially from CNBD to S6, meant that E' and F' were the focus of the interest in the C-linker for my thesis (Johnson and Zagotta, 2001). In addition, the idea that E' and F' are the only two helices that exist outside the gating ring and the CNBD led to my belief that they must be acting as a linker between the CNBD and the gating ring.

I postulate that D' changes in conformation to send a message to the S6 helices, so that the S6 helices must rotate inward towards the vertical axis perpendicular to the membrane when cAMP is released from the binding domain (Johnson and Zagotta, 2001;

Giorgetti et al., 2005). I am postulating that this is due to a signal from the CNBD that gives rise to a translation of E' and F' to begin the ripple effect to the S6 helices.

I am making this postulation based on the evidence from the comparative and MD models. The modeling techniques, comparative modelling and molecular dynamics, confirm that structural differences exist in apo and holo models for the C-linker. Particularly E' and D' have been demonstrated to be different in conformations comparing apo models and holo structure. Thus based on the results comparing apo models to H2h: D' is tilting away from the A helix and E' is translating parallel to A helix. Also A helix is remaining stationary as can be concluded based on all apo and holo comparisons. The direction of the structural changes in D' and E' did not agree for comparisons of G3 and G4 only.

Since structural changes are occurring in both the transmission hinge motif (E' and F') and D', after cAMP is removed, it would seem likely that E' and F' is causing D' to move. I have not ruled out the possibility that all three could be moving independent of each other in a response to cAMP, or that D' could be acting on E' and F'. Based on connectivity and the fact that the source of difference between holo crystal structure and apo models is the cAMP, it would make sense that the conformational change was being transmitted C-terminus to N-terminus.

Studies on the conformational changes of S6 helices have been done to propose that there is significant helix movement closer together when cAMP is removed from the system (Johnson and Zagotta, 2001; Clayton et al., 2004). Thus it would be reasonable to believe that the gating ring must be changing in conformation to cause this change in the pore.

I know from the comparative modeling and simulations that D', in the models, of the gating ring does exist in a different conformation for apo models versus holo structure of the cytoplasmic domains. Again since D' is the closest gating ring helix to the CNBD, which is the site of significant difference between the models of apo and holo, I postulate that D' is causing a ripple effect to give rise to closure of the pore S6 helices.

I had hypothesized that the gating ring interacting helices, of the holo HCN2 full length cytoplasmic domain (Asp443-Arg632) between adjacent subunits, would not be close enough for proper interactions. Thus I hypothesized that the helices of the tetrameric structure would no longer resemble a ring. So my predictions were that interactions of the holo HCN2 crystal structure between D' with B' and A' with C' would not exist. Also I predicted that all residues of these original interactions would have a distance greater than 4 Å between there closest atoms. Lastly I had predicted that a snapshot of the gating ring helices in the apo conformation would no longer resemble a ring but would be a "dismantled gating ring".

This would be contradictory to the current model that the crystal structure represents a resting state meaning that it may not be the channel modulating state as proposed by both Hua and Gordon, 2005 and Craven and Zagotta, 2007. Both studies demonstrated helices of the C-linker to be in different conformations in the cAMP bound state than those of the crystal structure by Zagotta et al., 2003.

I had planned to determine if this gating ring dismantling was true by using molecular dynamics, to simulate the whole holo HCN2 crystal structure as a tetramer. This would allow prediction of an apo model starting from the proposed holo structure (Zagotta et al., 2003).

7.2 Template selection was based on structural similarity

The most significant structural difference between Mlotik1 and HCN2 in the cytoplasmic domain is the lack of helices to make up a gating ring. In fact the C-linker of Mlotik1 is only two helices and a large loop that is homologous to D' – A helices of HCN2 (Clayton et al., 2004 and Alteri et al., 2008). Thus using Mlotik1 (Clayton et al., 2004) imposes the restriction of not being able to build a comparative model of HCN2 cytoplasmic domain inclusive of all helices composing the gating ring. However, due to sequence similarity, secondary structure similarity, and tertiary structure similarity of Mlotik1 to HCN2 it was a good choice. As well the high similarity in the CNBDs gives good validation that structural differences in the region of interest could be cAMP dependent.

7.3 Holo structure selection for comparison was based on availability

Currently there is only one structure for holo HCN* of HCN2. Thus the selection of target structure and raw sequence was that of the most current crystal structure (Zagotta et al., 2003).

7.4 H2h was chosen as the best holo model based on structure quality.

Two holo crystal structures were used in this research for comparison with apo models. One is a fragment of the full length holo HCN* crystal structure (H2h) and the other is a simulated holo HCN* MD model (G3). The decision to use H2h for comparison of apo versus holo differences was on structure quality from ProSA evaluation.

The energy was used to determine H2h to be the best holo model. Comparing ProSA evaluation for H2h and G3 (Figure 8-9 and Figure 8-19) it is seen that H2h is better quality than G3. This was based on the energy profile being more flat and thus lower energy for more residues of H2h than G3. This means it is overall a better model. Also confirming this was the values of the z-score (H2h: -6.99 and G3: -6.17, Table 8-4). Thus H2h was chosen as the best holo HCN* fragment to use in making conclusions.

7.5 G3 has an apo-like conformation but with cAMP in the binding site.

The simulation that produced G3 was originally intended to produce an apo conformation of HCN*. This was based on the assumption that cAMP would leave the binding site during the simulation. Since cAMP is still in the binding site (Figure 8-5) it is only proper to call it a holo conformation. However, the conformation did show some overall similarity to H2a RMSD for whole structures 3.10 Å (Table 8-8).

For H2a compared to G3 there is similarity in both the region of interest and the CNBD. The CNBD had the most similarity with the only really significant difference being a difference in position of the C-helix. As for the region of interest residues of D' and E' seem to form a continuous helix similar to H2a. Also residues of the A helix superimpose well. Thus H2a and G3 in general show some similarities.

Thus it is possible that G3 represents a "transitioning structure" of the cytoplasmic domain of HCN2. This is supported by the obvious difference in the conformation of cAMP in the binding site comparing G3 and H2h (Figure 8-24). Thus it could be that the cAMP of G3 is positioned to move out of the binding site. Had the simulation been run longer the cAMP may have left the binding site creating a new apo model.

I would have liked to determine if G3 might be a small conformational change away from the apo conformation of HCN*. I would hypothesize that G3, if used for a simulation removing cAMP, would change to become the same as the MD model G4. This is based on the number of similarities between G3 and G4 (Figure 8-22). This includes the E' and D' helix fusion in both models, lack of conformational change in A helix, and CNBD similarity. The strategy would be to perform a simulation with G3 after removing cAMP from the structure file for another 20ns. I would predict that secondary and tertiary structures would demonstrate similarity.

7.6 Mlotik1 shows structures changes in the holo to apo transition.

I compared structures of the Mlotik1 cytoplasmic domain in unliganded (Mla) and bound form (Mlh). If no structural differences existed then it would have been invalid to predict structural changes comparing cAMP bound HCN* fragment (H2h) and unliganded comparative models that were built based on Mla. As seen in the superimposition of the whole domains (Figure 5.x) (Mlh and Mla: R219-R349) there are structural differences in the region of interest (R219-A250). The residues homologous to the D' helix are tilted further away from the αA helix in Mla compared to Mlh, while the αA helix is stationary. The residues homologous to E' and F' helix are also in a different conformation in Mlh compared to Mla. All of these observations were supported by the RMSD values for each homologous H2h helix (RMSD: D' (Mlh and Mla: R219-V224) = 7.15Å; E' (Mlh and Mla: Q228-V230) = 5.05 Å; F' (Mlh and Mla: P234-V236) = 5.42 Å; αA (Mlh and Mla: P241-R249) = 1.79 Å; Thm (Mlh and Mla: Q228-F236) = 5.30 Å; region of interest (Mlh and Mla: R219-R249) = 4.81 Å; CNBDs (Mlh and Mla: R252-

R349) = 3.43 Å; whole domains (R219-R349) = 3.78 Å, Table 8-15). Thus Mlh and Mla show significant differences in the region of interest and none in the CNBD.

7.7 Secondary structure changes could be cAMP dependent.

The secondary structure of the region of interest (D', E', F') in H2a was very different than in H2h (Figure 8-14). This suggests that not only the positioning of helices is cAMP dependent but also the formation of helical secondary structure is cAMP dependent. Specifically, I wanted to know whether F' residues were truly helix-forming in H2h or if the appearance of helix structure was just an anomaly of the particular viewing software.

7.7.1 F' secondary structure is not conclusive regarding cAMP dependence.

Based on comparisons of unliganded and cAMP bound simulated models F' secondary structure is not conserved during the binding transition. This is based on comparative models of unliganded to cAMP bound transition, with loop secondary structure for F' in unliganded to helix structure for the bound state. Molecular dynamics conformation for HCN* fragment shows a unliganded to cAMP bound transition of helical structure in the unliganded and none in the bound state. A cAMP dependent conformational movement of the F' helix is backed up by studies demonstrating that F' experiences some sort of conformational change that is cAMP dependent (Taraska and Zagotta et al., 2007; Zhou and Seigelbaum, 2007; Berrera et al., 2006). The secondary structure in F' is shown to vary based on an HCN channel from sea urchins cAMP bound HCN* source crystal structure from the channel SpIH (2ptm, Flynn et al., 2007). Also soluble protein studies and crystal structure of HCN* for HCN2 show F' losing

secondary structure in the cAMP bound to unliganded transition (Taraska et al., 2009). Secondary structure has been shown to vary in Mlotik1 solved structures of the cytoplasmic domain. The recently solved NMR structure of the cAMP bound cytoplasmic domain of Mlotik1 has F' helical structure and the source crystal structures for both bound and unliganded have a loop (Schuke et al., 2009; Alteri et al., 2009). Therefore the question of F' secondary structure cAMP dependence cannot be proposed at this time.

7.8 H2a and G4 support a possibility of multiple unliganded conformations.

I hypothesized that the conformation of H2a and G4 would be conserved as a signature conformation of the unliganded form. I predicted that homologous residues of G4 would have a similar region of interest including a helix loop helix motif, as was seen for H2a. I also predicted that the orientation of the homologous residues of D' would be tilted similar to H2a, D' and E' would be a single helix, and residues homologous to A helix would be a helix in the similar position to H2a and G4.

As seen in Figure x the D' and E' helices of H2a and G4 have similar orientation and F' has a different position. The D' and E' helices are pointing in the same direction relative to the plane of the page. Although D' of G4 looks to be possibly pointing into the page and thus there maybe a difference in the orientation in the plane perpendicular to the page. Also F' of G4 primarily defined by F18 is translated into the page relative to F18 of H2a. Thus G4 and H2a have some similarities in the region of interest but enough differences to suspect that they could represent two possible conformations of the region

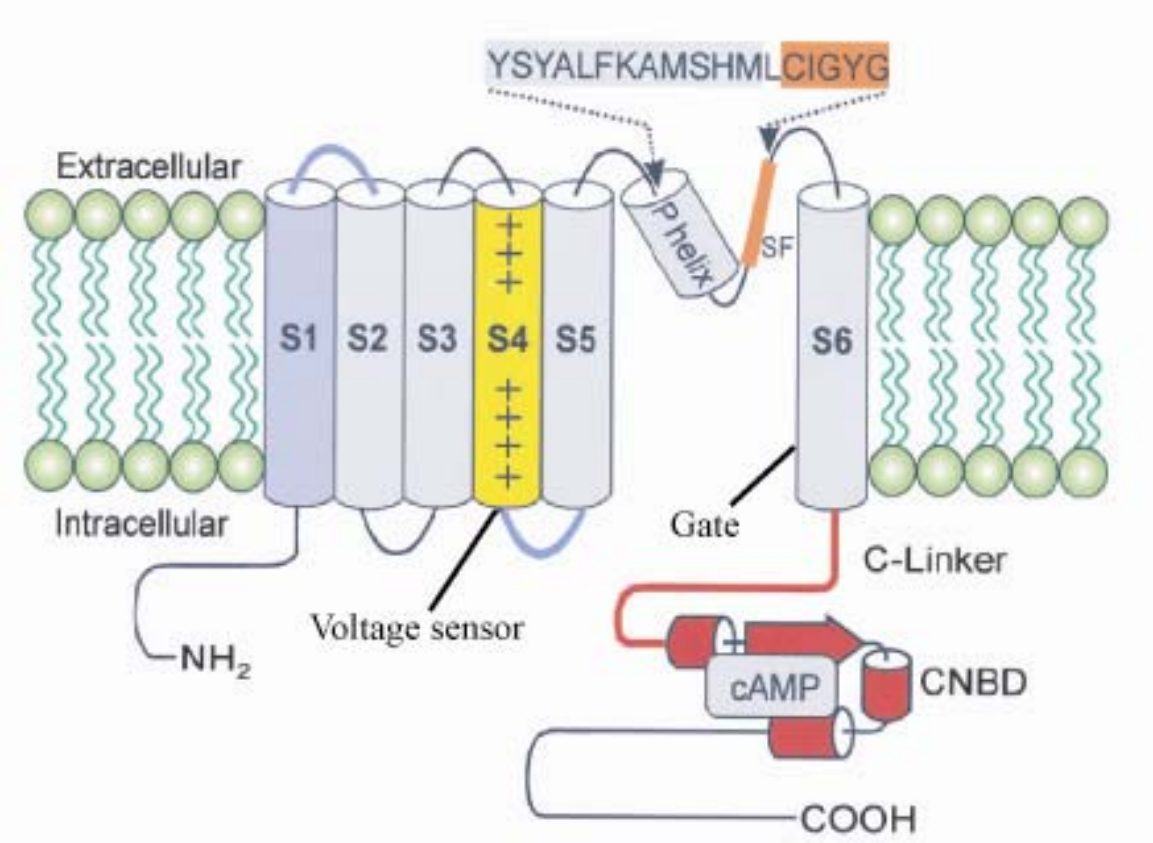
of interest in the unliganded state of the protein. F' secondary structure is not conclusive regarding cAMP dependence.

Based on comparisons of unliganded and cAMP bound simulated models F' secondary structure is not conserved during the binding transition. This is based on comparative models of unliganded to cAMP bound transition, with loop secondary structure for F' in unliganded to helix structure for the bound state. Molecular dynamics conformation for HCN* fragment shows a unliganded to cAMP bound transition of helical structure in the unliganded and none in the bound state. A cAMP dependent conformational movement of the F' helix is backed up by studies demonstrating that F' experiences some sort of conformational change that is cAMP dependent (Taraska and Zagotta et al., 2007; Zhou and Seigelbaum, 2007; Berrera et al., 2006). The secondary structure in F' is shown to vary based on an HCN channel from sea urchins cAMP bound HCN* crystal structure from the channel SpIH (2ptm, Flynn et al., 2007). Also soluble protein studies and crystal structure of HCN* for HCN2 show F' losing secondary structure in the cAMP bound to unliganded transition (Taraska et al., 2009). Secondary structure has been shown to vary in Mlotik1 solved structures of the cytoplasmic domain. The recently solved NMR structure of the cAMP bound cytoplasmic domain of Mlotik1 has F' helical structure and the source crystal structures for both bound and unliganded have a loop (Schuke et al., 2009; Alteri et al., 2009). Therefore the question of F' secondary structure cAMP dependence cannot be proposed at this time.

8: Appendices

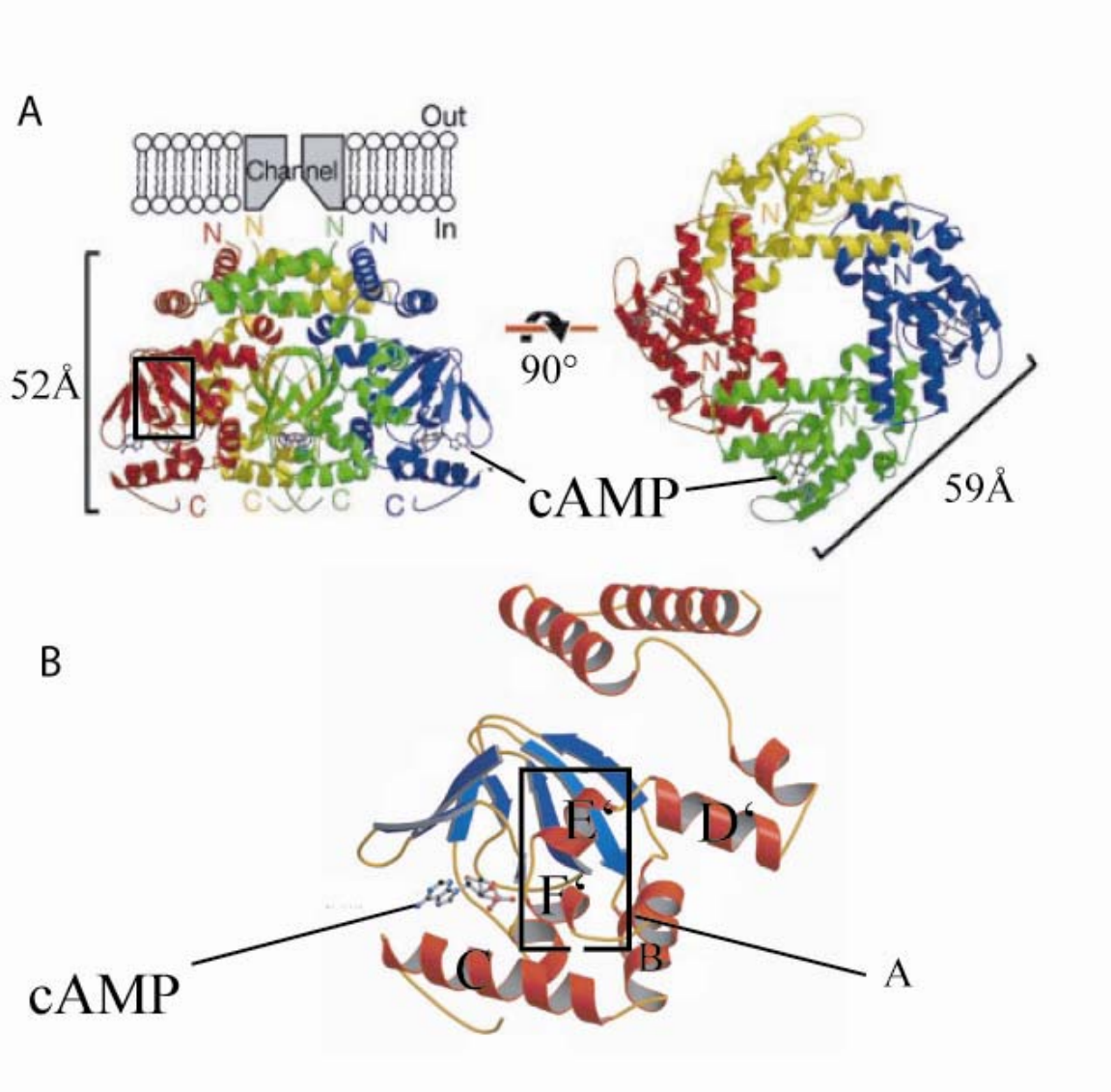
8.1 Figures

Figure 8-1: HCN is a member of family of the six transmembrane helix ion channels.



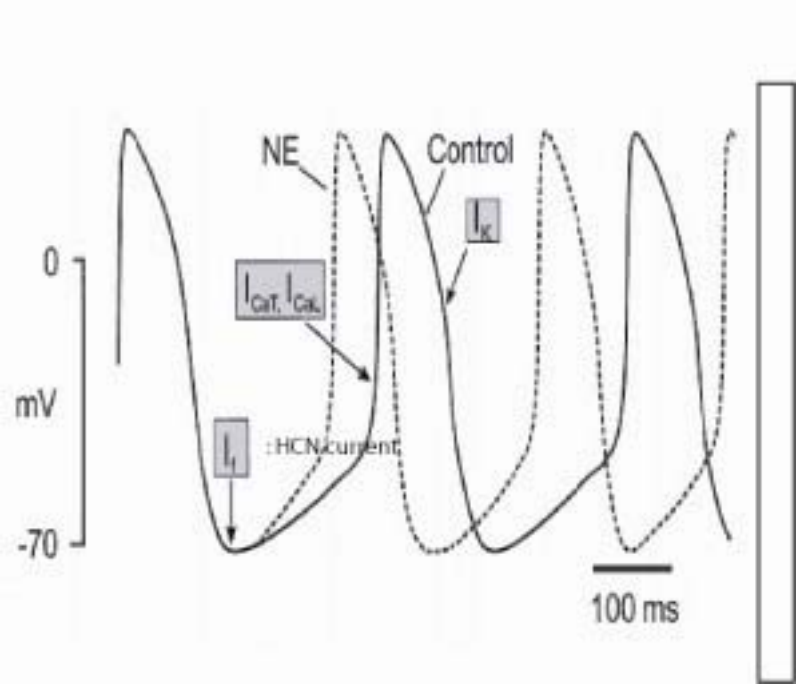
Notes: Figure shows one of four subunits with cAMP in the CNBD. SF: selectivity filter. (Swartz, 2004)

Figure 8-2: Structure of HCN2 cytoplasmic domain.



Notes: Transmission hinge motif (Thm) exists external to the gating ring and CNBD. Thm is found within the black rectangle. A. Side view and a bottom view of the tetrameric form of HCN2 cytoplasmic fragment crystal structure. The cartoon above the side view depicts the membrane portion of the channel within a lipid bilayer. B. Side view of a single subunit of the cytoplasmic fragment (Zagotta et al., 2003).

Figure 8-3: The protein HCN is activated after repolarization of the membrane.



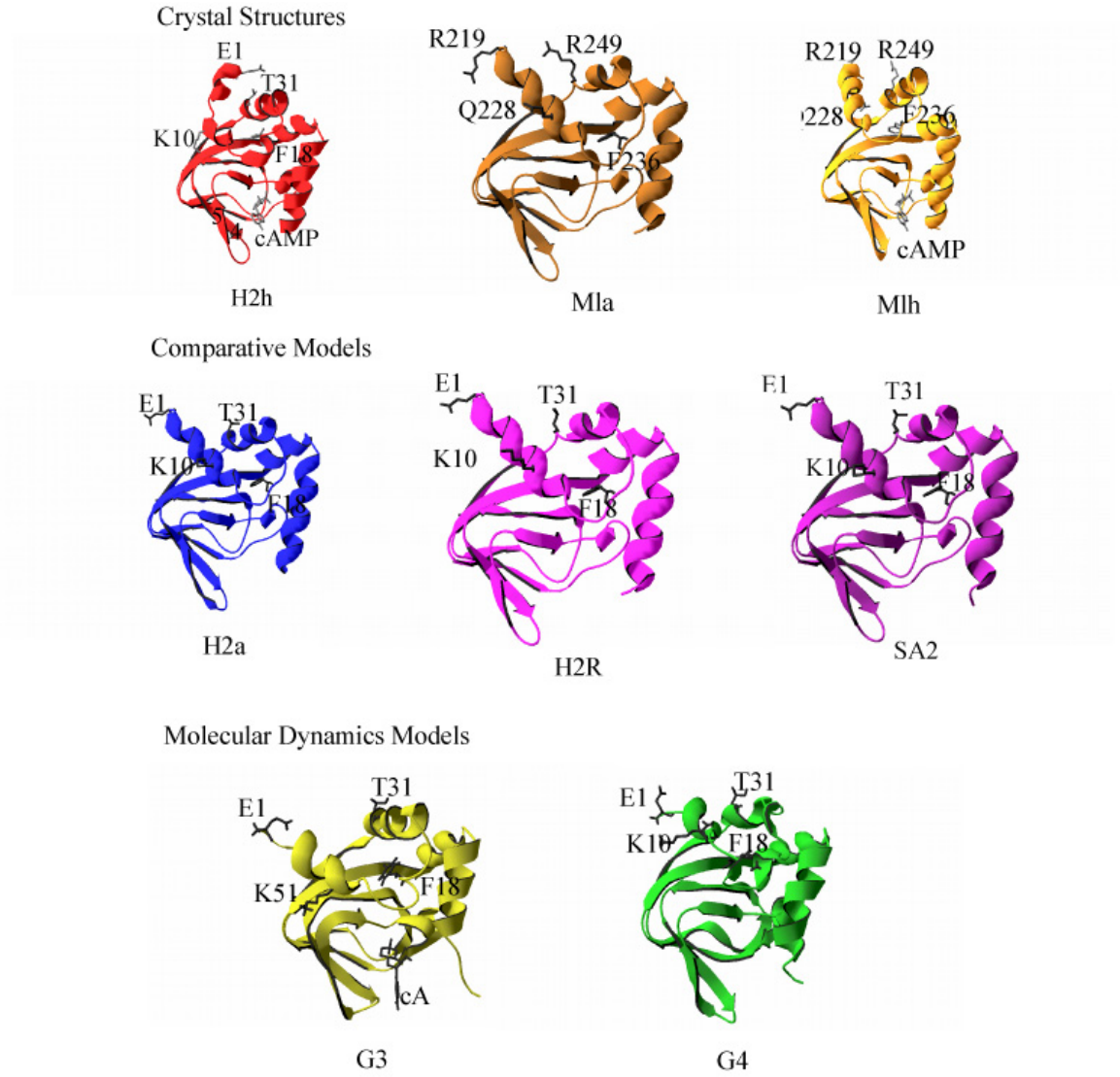
Notes: Figure shows the membrane voltages over time for a nerve impulse with norepinephrine (NE) and without (control). I_f denotes the time of activation of HCN channels. $I_{CaT, CaL}$ and I_K denote activation of calcium and potassium channels respectively (Biel et al., 2002).

Figure 8-4: Homology between HCN2 and Mlotik1 cytoplasmic domain sequences.

```
H2h
1q5o  443 DSSRRQYQEK YKQVEQYMSF HKLPADFRQK IHDYYEHRYQ
          .hhhhhhhhh hhhhhhhhhh h...hhhhhh hhhhhhhhh.
Name          A'                              B'

H2h   501                      EE IVNFNCRKLV ASMPLFANAD
1q50  482 GKMFDEDSIL GELNGPLREE IVNFNCRKLV ASMPLFANAD
          .....hhhhh h...hhhhhh hhhh....hh h ...hhh..
Name              C'          >  D'       E'      F'

H2h   523 PNFVTAMLTK LKFEVFQPGD YIIREGTIGK KMYFIQHGVV
1Q5O  523 PNFVTAMLTK LKFEVFQPGD YIIREGTIGK KMYFIQHGVV
          hhhhhhhhh. .sssss...s ss.......s ssssss..ss
Name           A        1       2             3      4

H2h   563 SVLTKGNKEM KLSDGSYFGE ICLLTRGRRT ASVRADTYCR
1q5o  563 SVLTKGNKEM KLSDGSYFGE ICLLTRGRRT ASVRADTYCR
          sss......s ss...sss.. hhhhh.....  .ssssss..ss
Name                 5      6      P          7      8

H2h   603 LYSLSVDNFN EVLEEYPMMR RAFETVAIDR
1q5o  603 LYSLSVDNFN EVLEEYPMMR RAFETVAIDR
          sssssshhhh hhhh...hhh hhhhhhhhhh
Name                    B             C
H2h       L
1q5o  633 LDRIGKKNSI L-
          hh........ .-
```

Notes: Dots in amino acid sequence represent conservation. Secondary structures are denoted with each sequence. Reference residues are in bold.

Figure 8-5: Crystal structure models of the proteins HCN and Mlotik1.



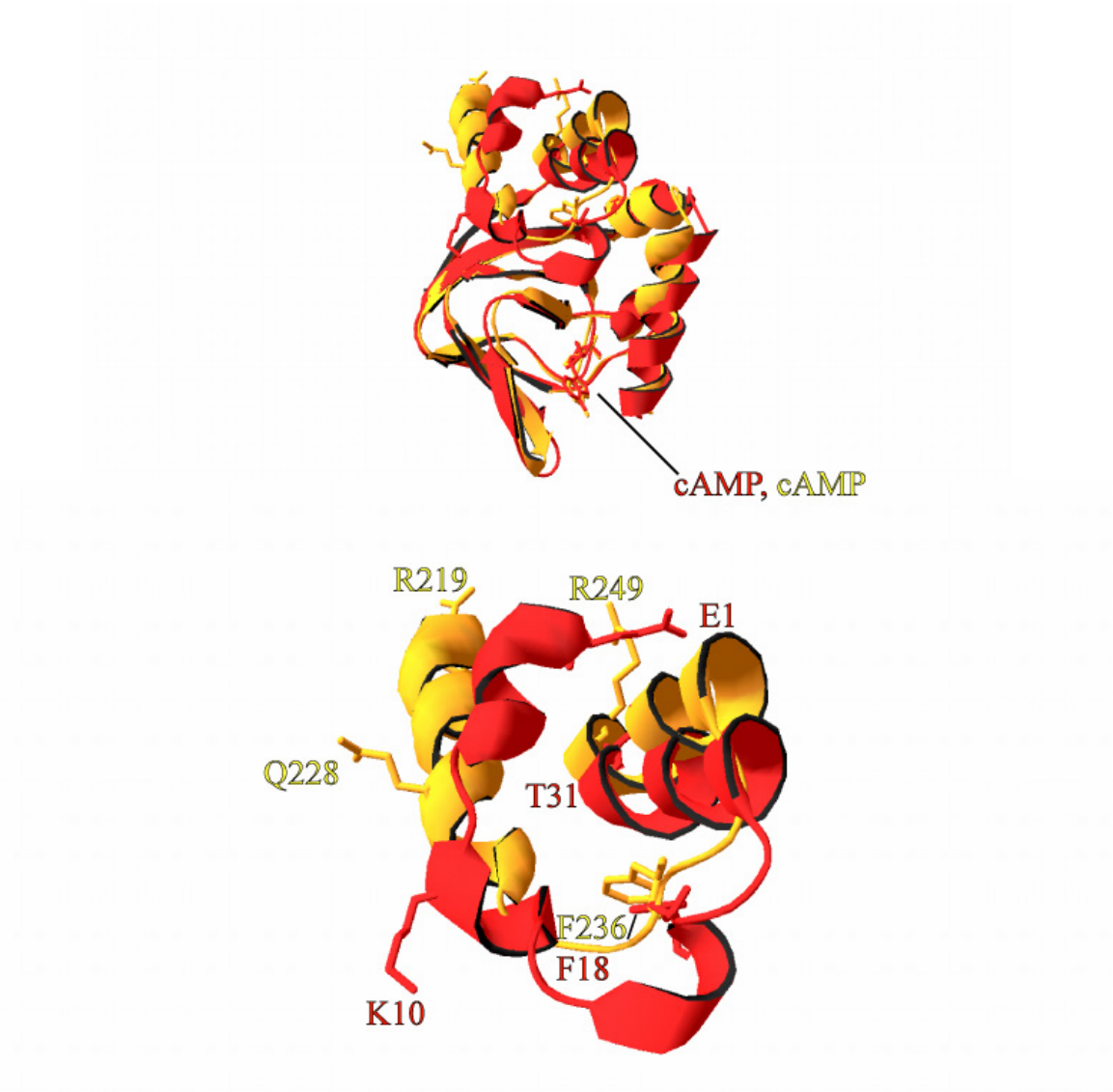
Notes: Figure includes original crystal structures, new comparative models and simulated models of HCN2 and Mlotik1 cytoplasmic fragments.

Figure 8-6: Alignment used for comparative model building.

```
Name            A'                          B'
1q5o         .hhhhhhh hhhhhhhhhh hhh...hhhh
1q5o         DSSRRQYQ EKYKQVEQYM SFHKLPADFR


Name                               C'        D'
1q5o         hhhhhhhhhh h       hhh hhh    hhhh
1q5o         QKIHDYYEHR YQGKMFDEDS ILGELNGPLR


Name                   E'      F'           A
1q5o         hhhhhh    h  hh    hhh     hhhhhhhh
1q5o         EEIVNFNCRK LVASMPLFAN ADPNFVTAML
Mla          RRGDFVRNWQ ...AV...QK LG.AVLVEIV
Mla          hhhhhhhhh hhh           hhhhhhhh
Name             αA'                     αA

1q5o       h  sssss    sss         sssssss
1q5o       TKLKFEVFQP GDYIIREGTI GKKMYFIQHG
Mla        RA.RARTVPA .AV.CRI.EP .DR.F.VVE.
Mla        hh sssss    sss         sssssss
Name            1       2            3

1q5o       sssss       sss   sss   hhhhh
1q5o       VVSVLTKGNK EMKLSDGSYF GEICLLTRGR
Mla        S...A.--PN PVE.GP.AF. ..MAL.SGEP
Mla        ssssss sss ss    sss   hhhhh
Name        4          5      6     P

1q5o          sssss   sssssssshh hhhhhh   h
1q5o       RTASVRADTY CRLYSLSVDN FNEVLEEYPM
Mla        .S.T.S.A.T VS.L..HSAD .QMLCSSS.E
Mla            ssss   sssssssshh hhhhhh   h
Name            7       8          B

1q5o       hhhhhhhhhh hhhh          -
1q5o       MRRAFETVAI DRLDRIGKKN SIL-
Mla        IAEIFRKTAL E.R
Mla        hhhhhhhhhh hh
Name            C
```

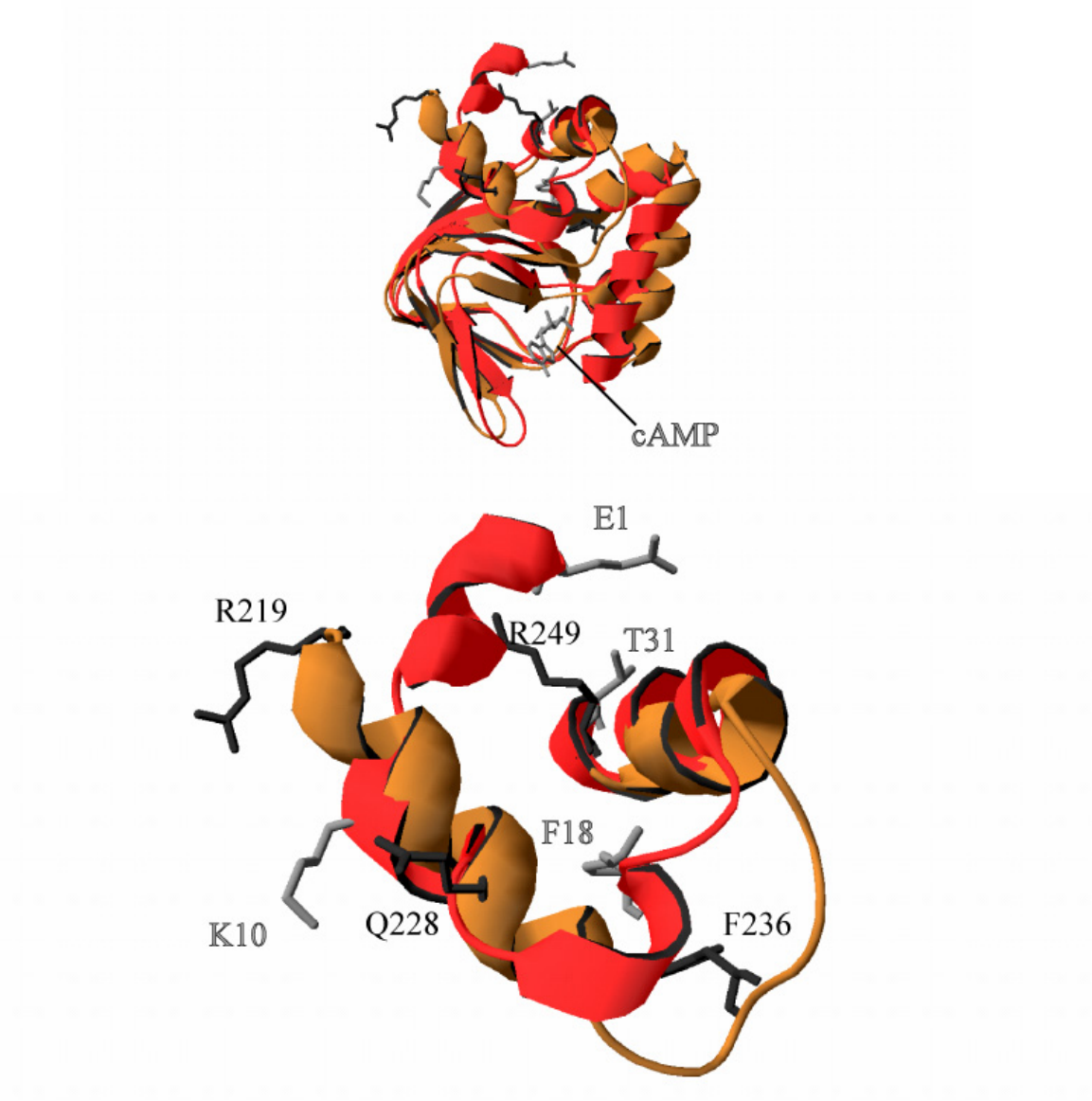
Notes: Reference residues are indicated in bold (as seen in figure 9-4 and Table 8-5)

Figure 8-7: The protein models H2h and Mlh show similarity in whole protein.



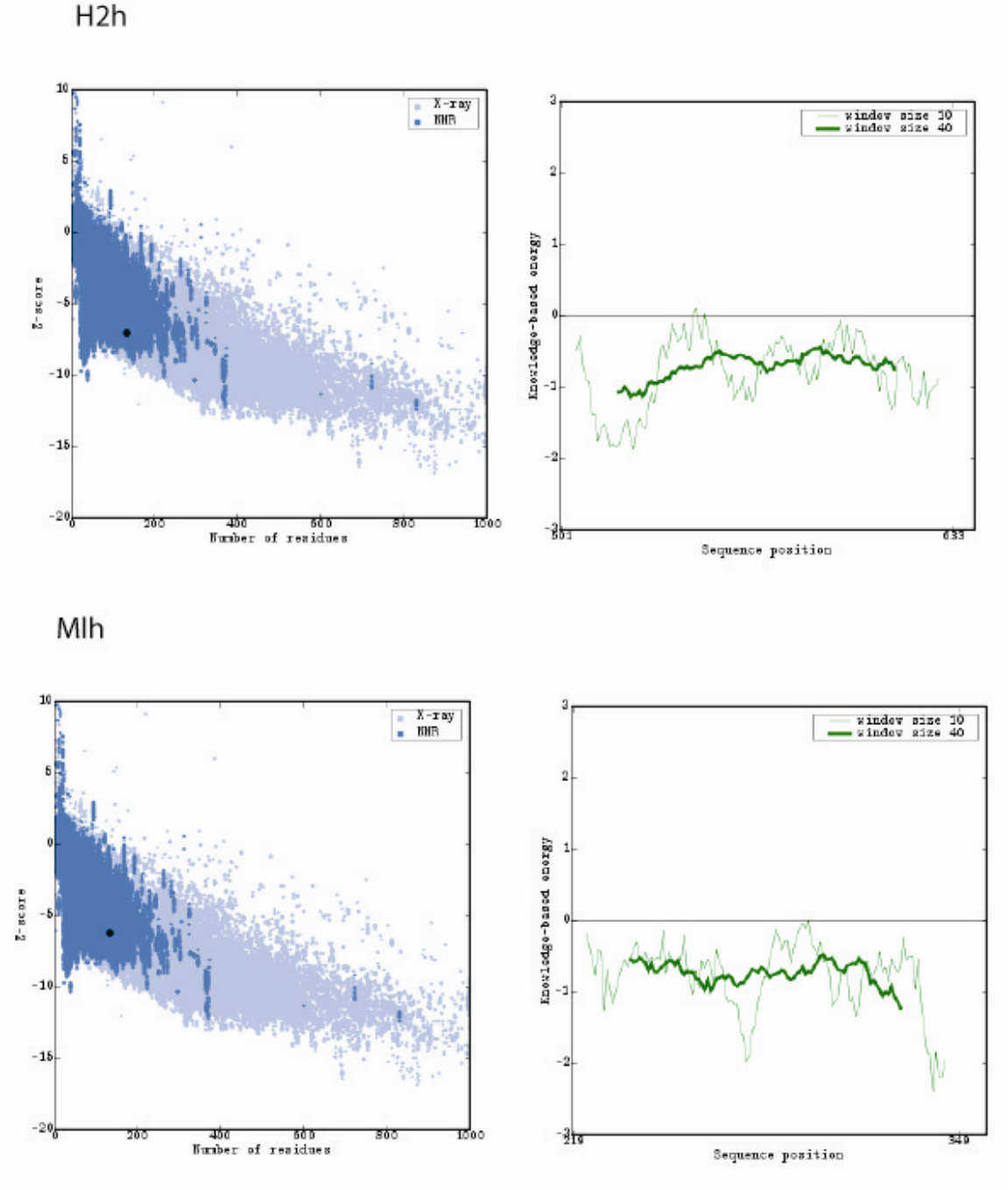
Notes: Numbering shown is based on numbering in Table 8-5. H2h is in red (E1 and cAMP) and Mlh (R219 and cAMP) in yellow. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-8: The protein models H2h and Mla show a difference in conformation of D'.



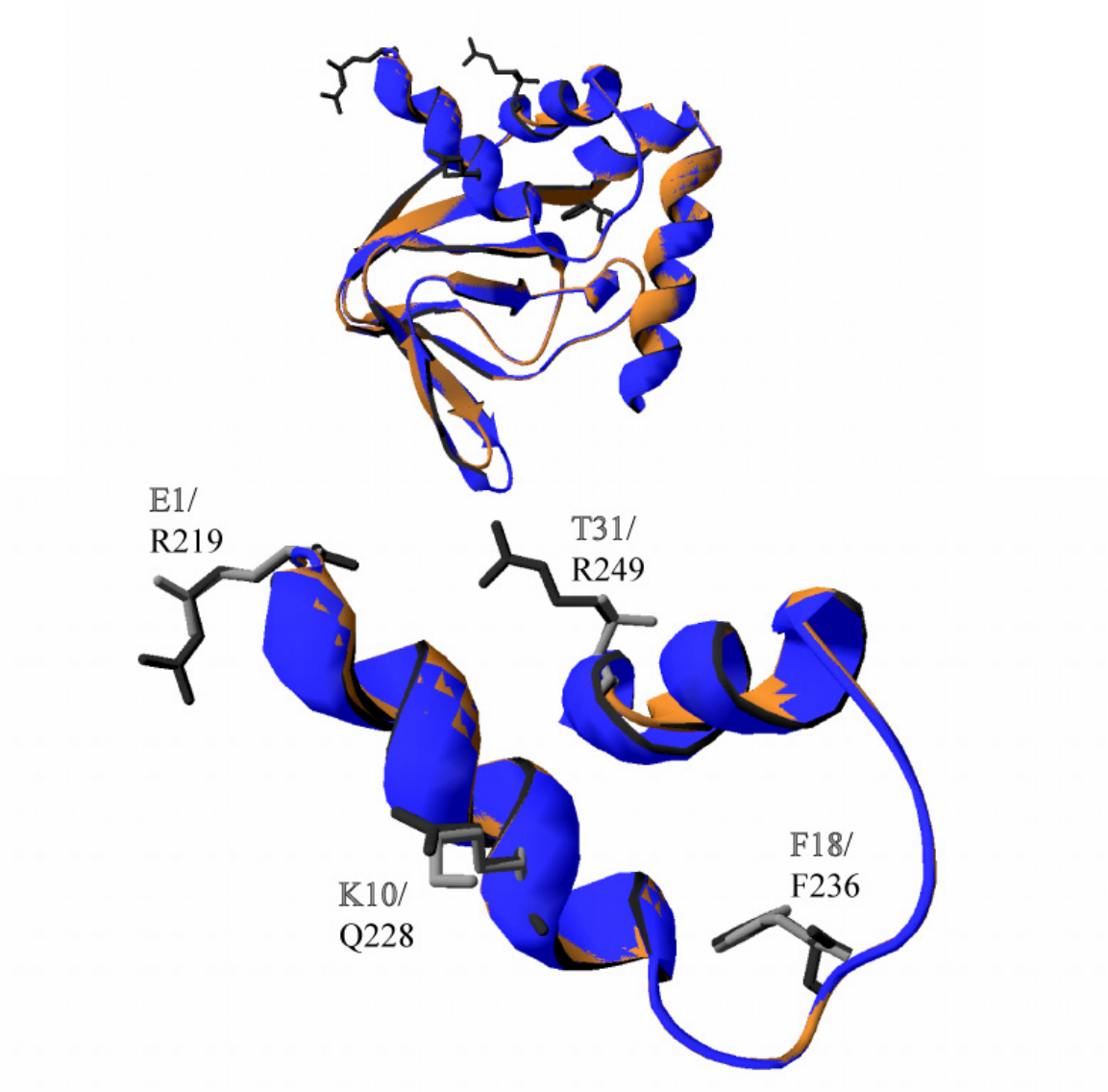
Notes: Notes: Numbering shown is based on numbering in Table 8-5. H2h is in red (E1 and cAMP) and Mla (R219) in brown. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-9: The protein models H2h and Mlh are good quality structures.



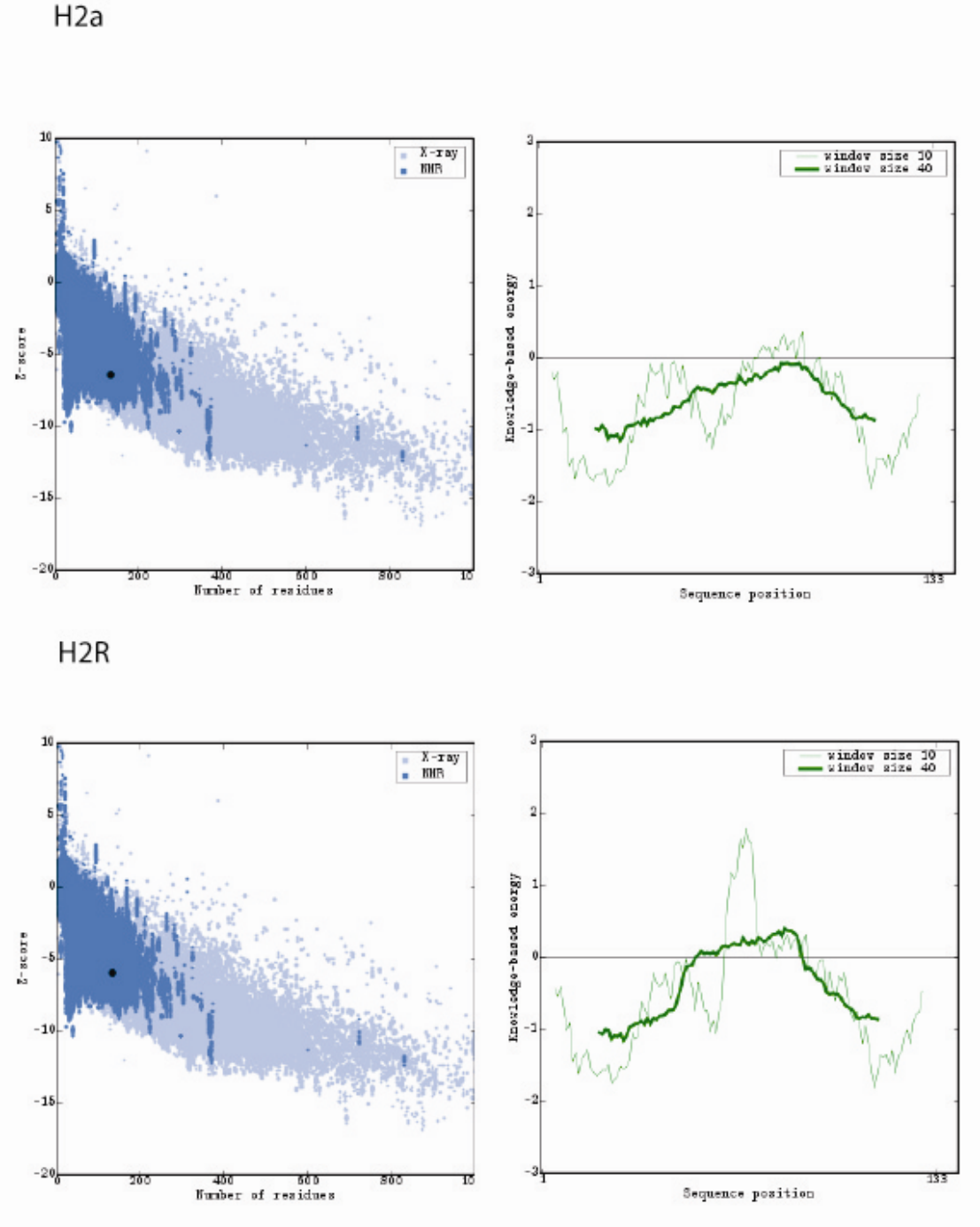
Notes: Structure quality was evaluated using z-score and energy profiles. Z-score is on the vertical axis of the graphs on the left and number of residues on the horizontal axis. Energy is on the vertical axis on the graphs on the right with sequence position on the horizontal axis. The residues number is found on the horizontal axis.

Figure 8-10: The protein models Mla and H2a show similarity for all parts of the protein.



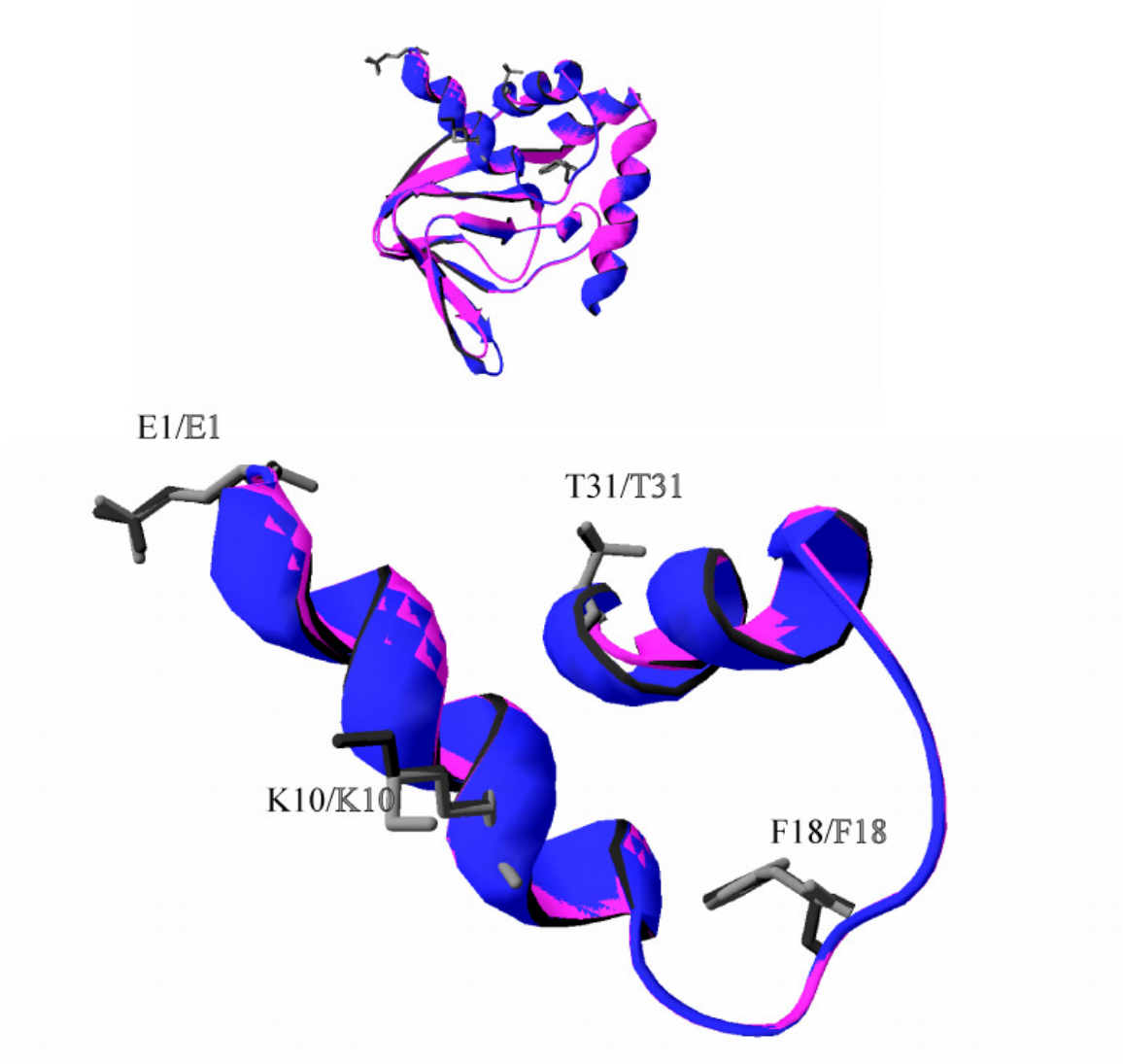
Notes: Numbering shown is based on numbering in Table 8-5. H2a is in blue (E1) and Mla (R219) in yellow. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found at top blown up for the region of interest.

Figure 8-11: The protein models H2a and H2R are good quality structures.



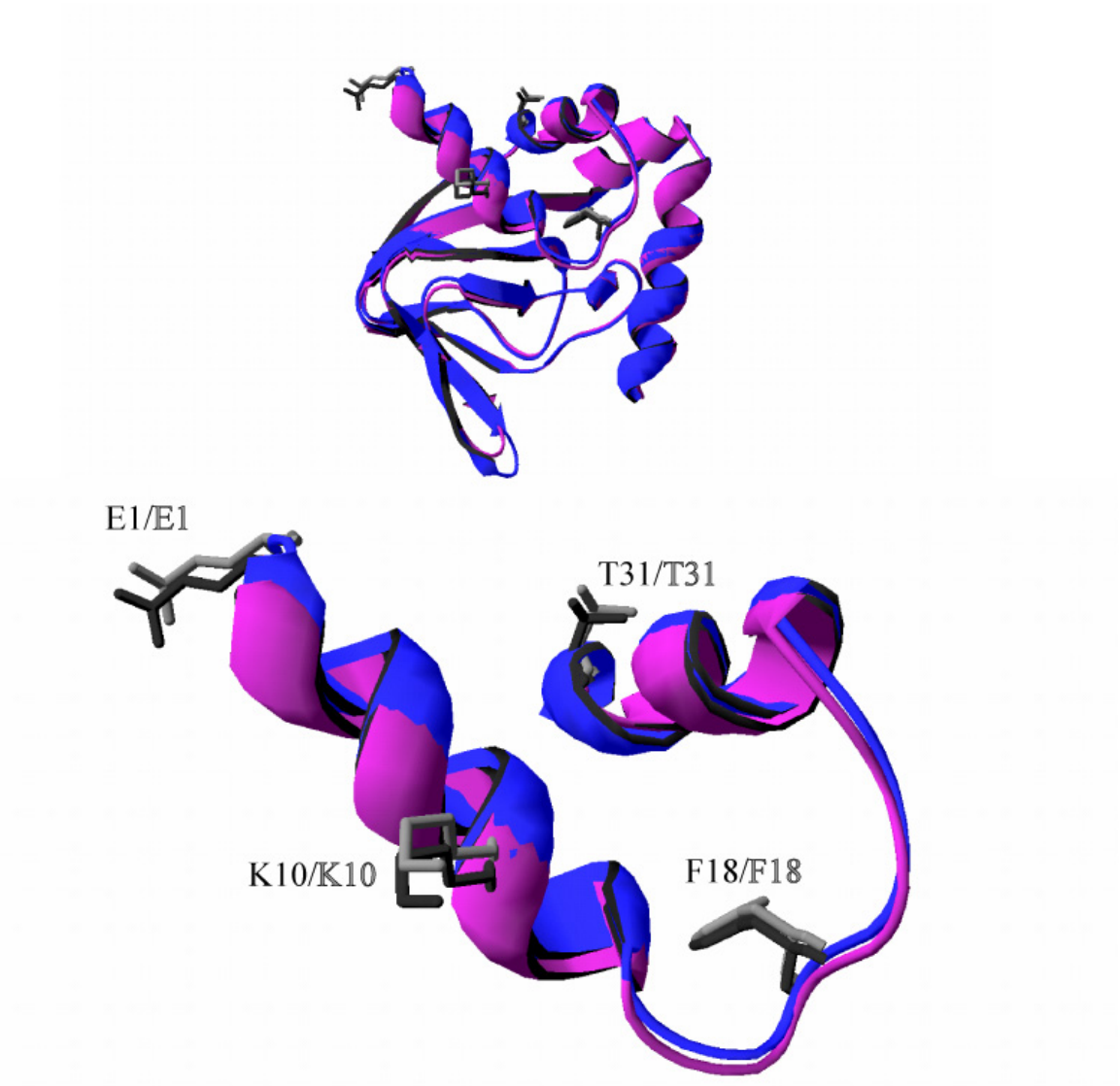
Notes: Structure quality was evaluated using z-score and energy profiles. Z-score is on the vertical axis of the graphs on the left and number of residues on the horizontal axis. Energy is on the vertical axis on the graphs on the right with sequence position on the horizontal axis. The residues number is found on the horizontal axis.

Figure 8-12: The protein models H2a and H2R show similarity in tertiary structure.



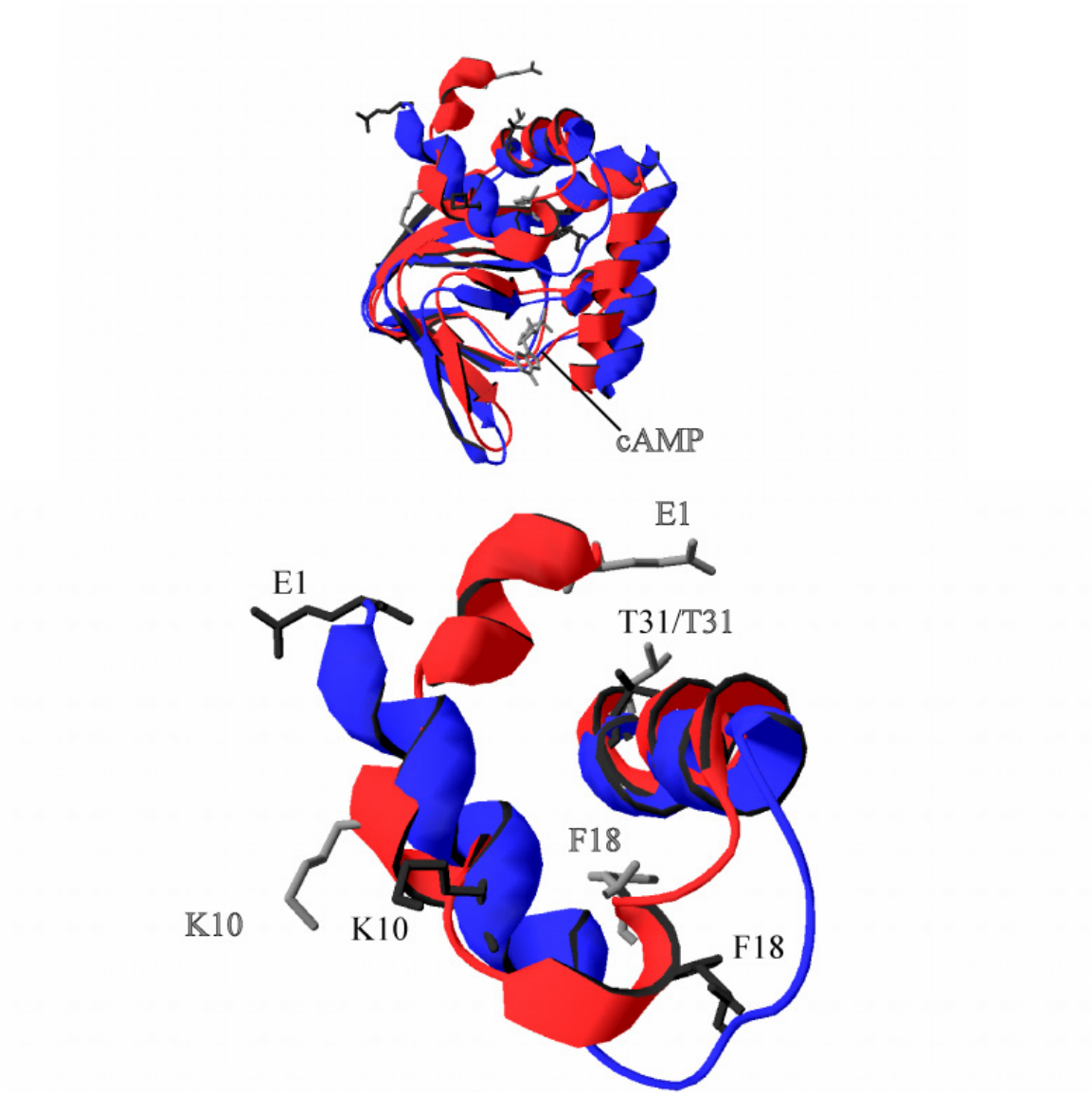
Notes: Numbering shown is based on numbering in Table 8-5. H2a is in blue (E1) and H2R (E1) in pink. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-13: The protein models H2a and SA2 show similarity in tertiary structure.



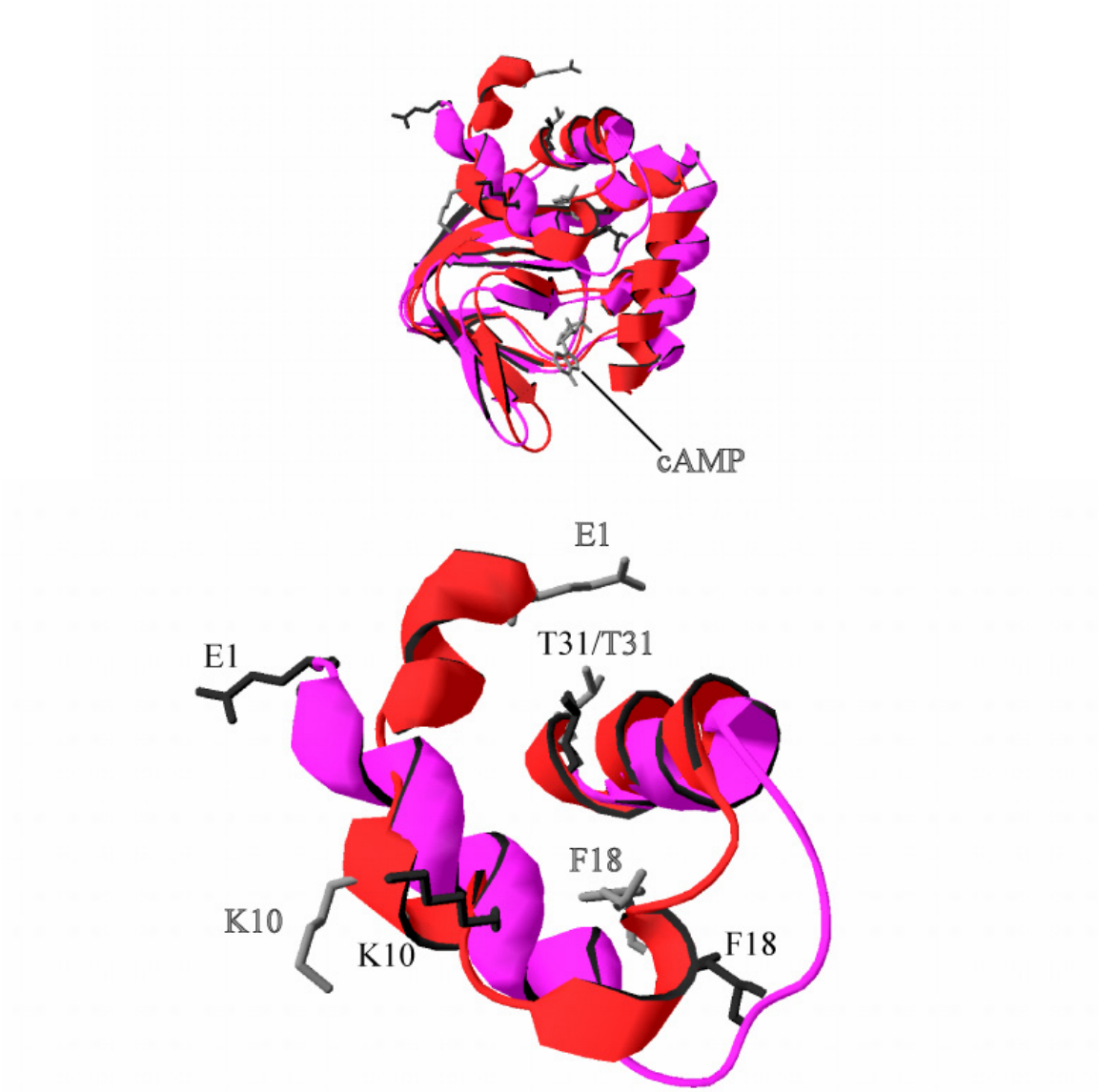
Notes: Numbering shown is based on numbering in Table 8-5. H2a is in blue (E1) and SA2 (E1) in yellow. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-14: The protein models H2h and H2a show different conformations for D'.



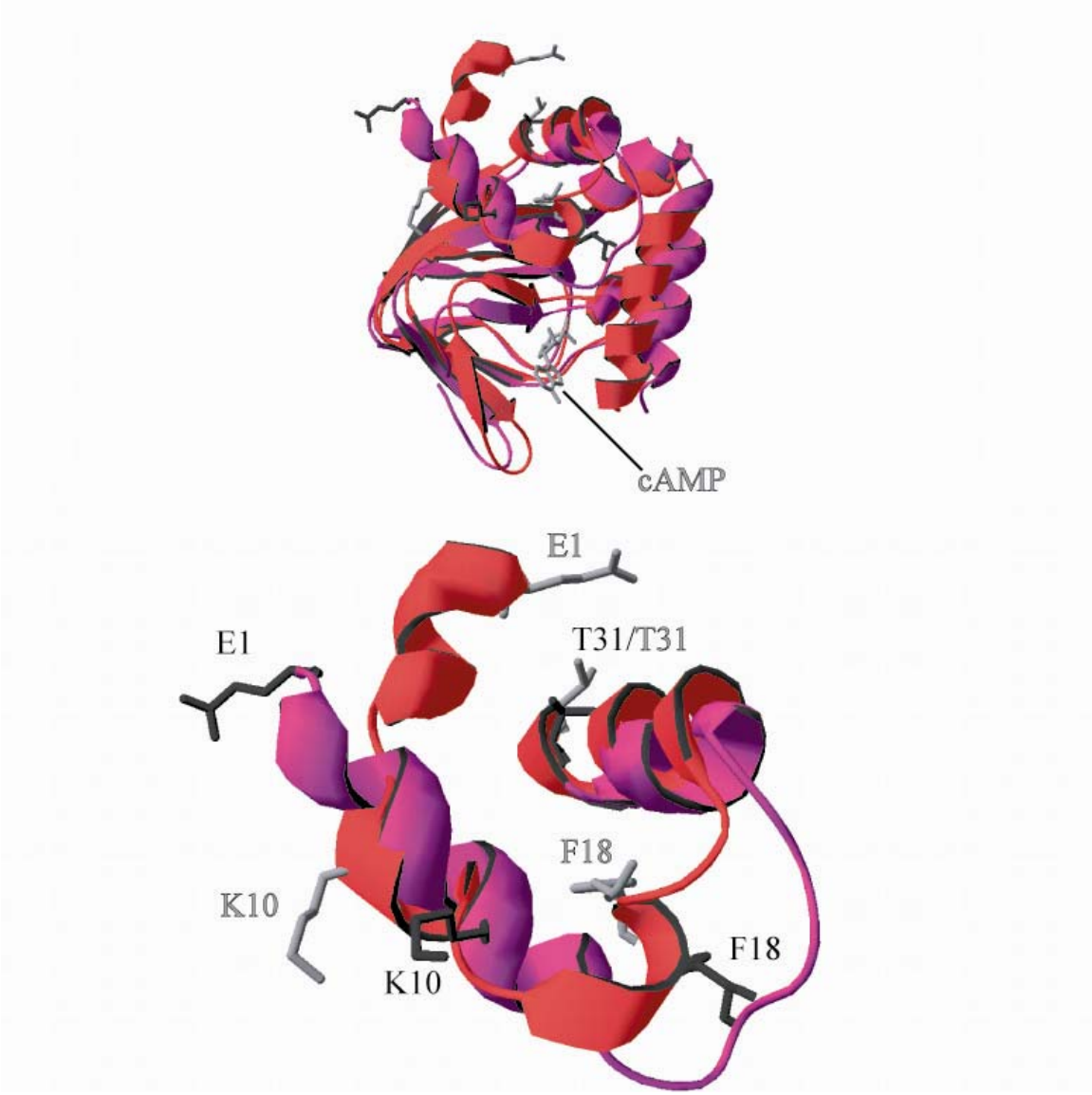
Notes: Numbering shown is based on numbering in Table 8-5. H2a is in blue (E1) and H2h (E1 and cAMP) in yellow. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-15: Protein models H2h and H2R show significant differences in residue E1.



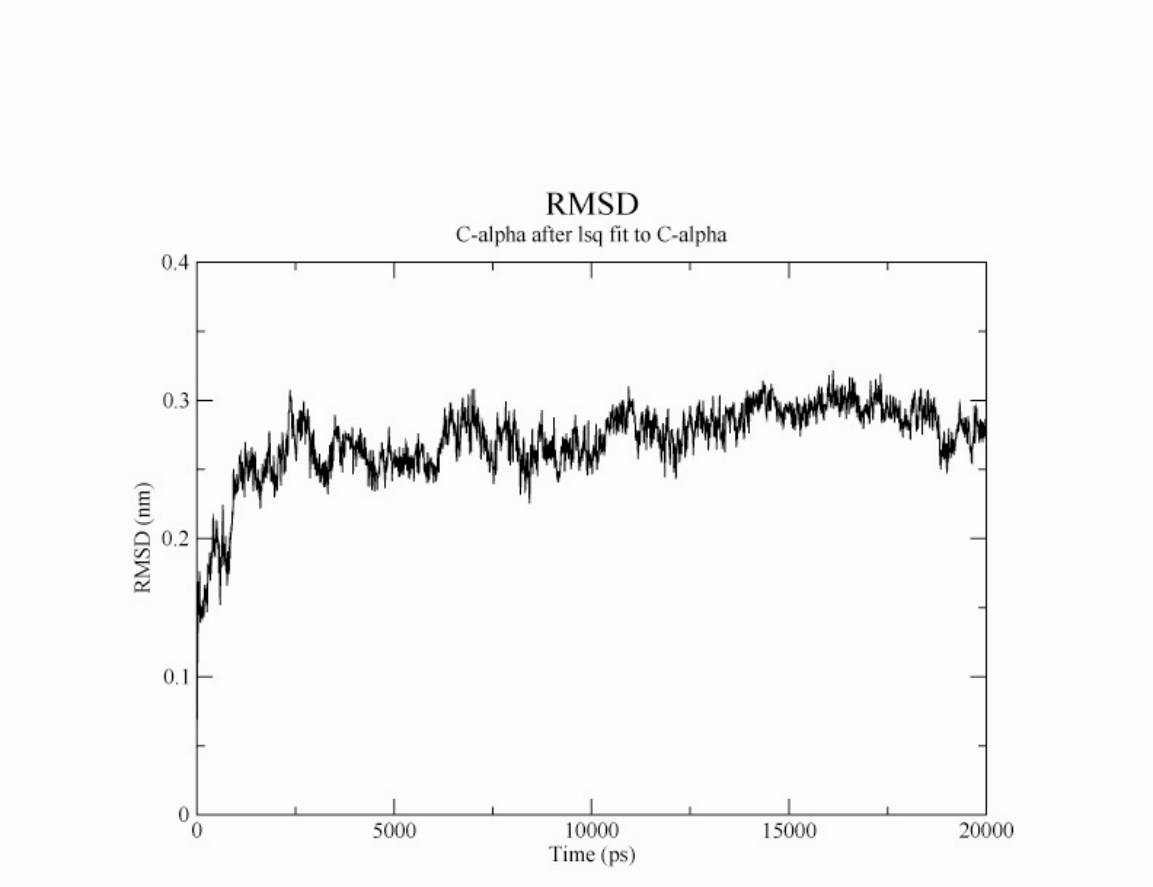
Notes: Numbering shown is based on numbering in Table 8-5. H2h is in red (E1 and cAMP) and H2R (E1) in pink. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-16: The protein models SA2 and H2h confirm that D' is tilting away from A.



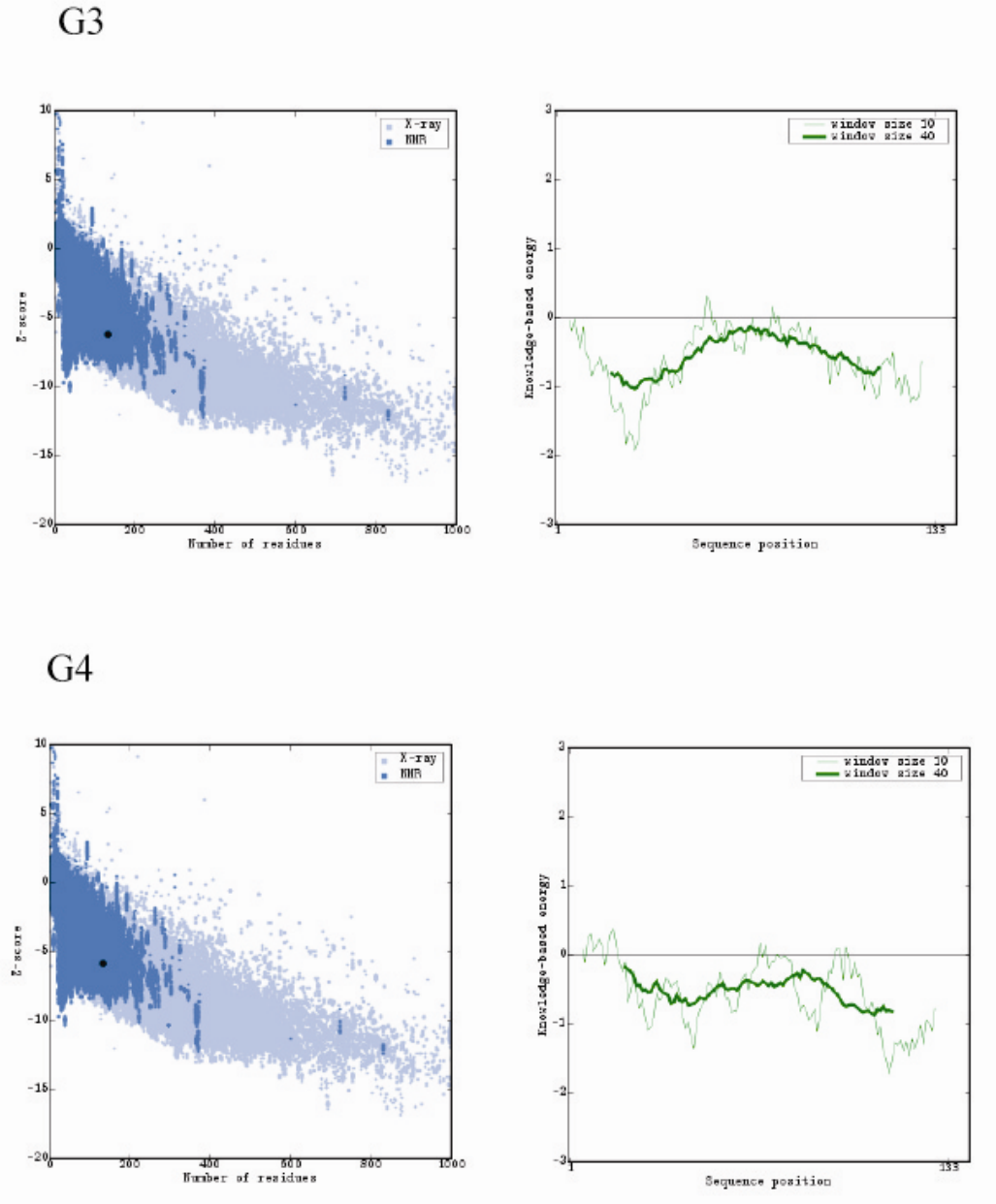
Notes: Numbering shown is based on numbering in crystal structures. SA2 is in purple (E1) and H2h (E1 and cAMP) in red. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-17: The protein model of G4 is stable after 5ns of simulation of H2h (apo).



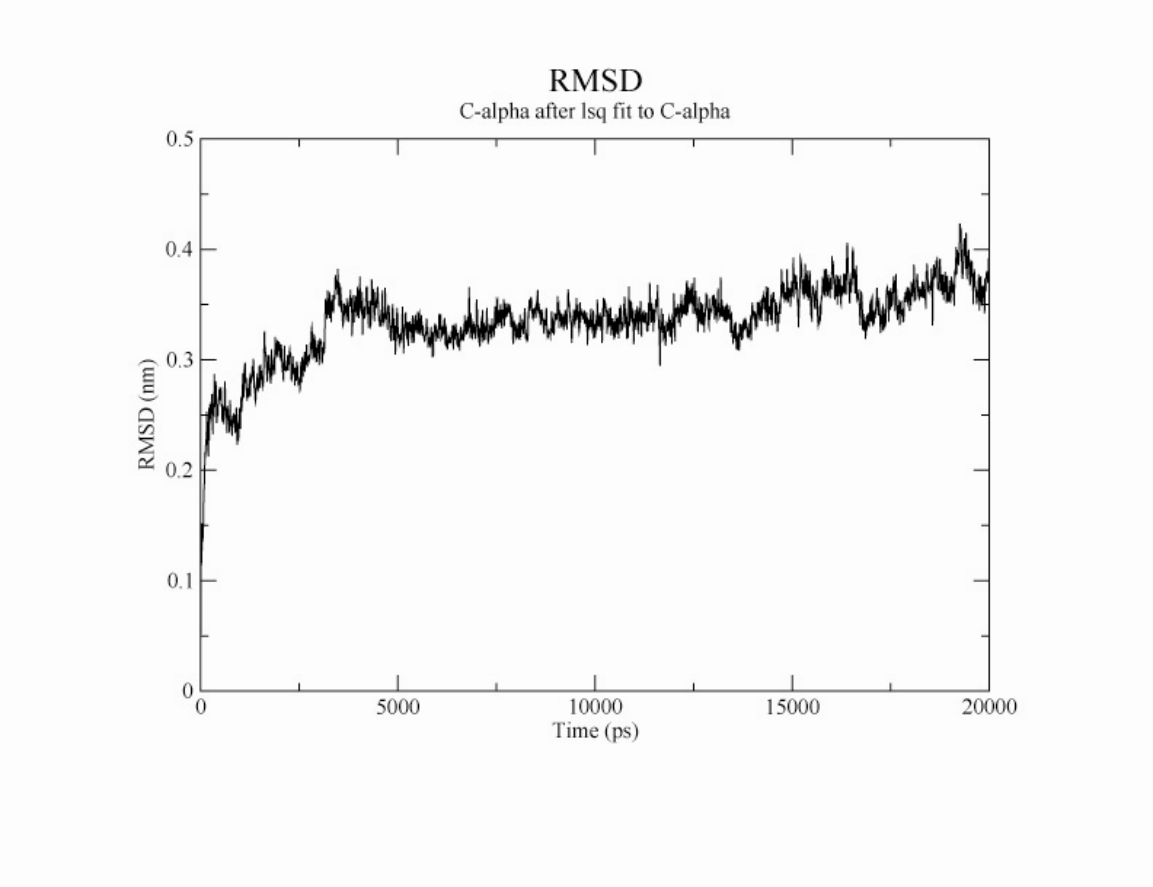
Notes: Graph reports the RMSD for superimpositions of each successive frame of the simulation on the original input structure (frame 0) using only the Calpha carbon atoms. All significant movements in the simulation have stopped after 5ns.

Figure 8-18: The protein models G3 and G4 are good quality structures.



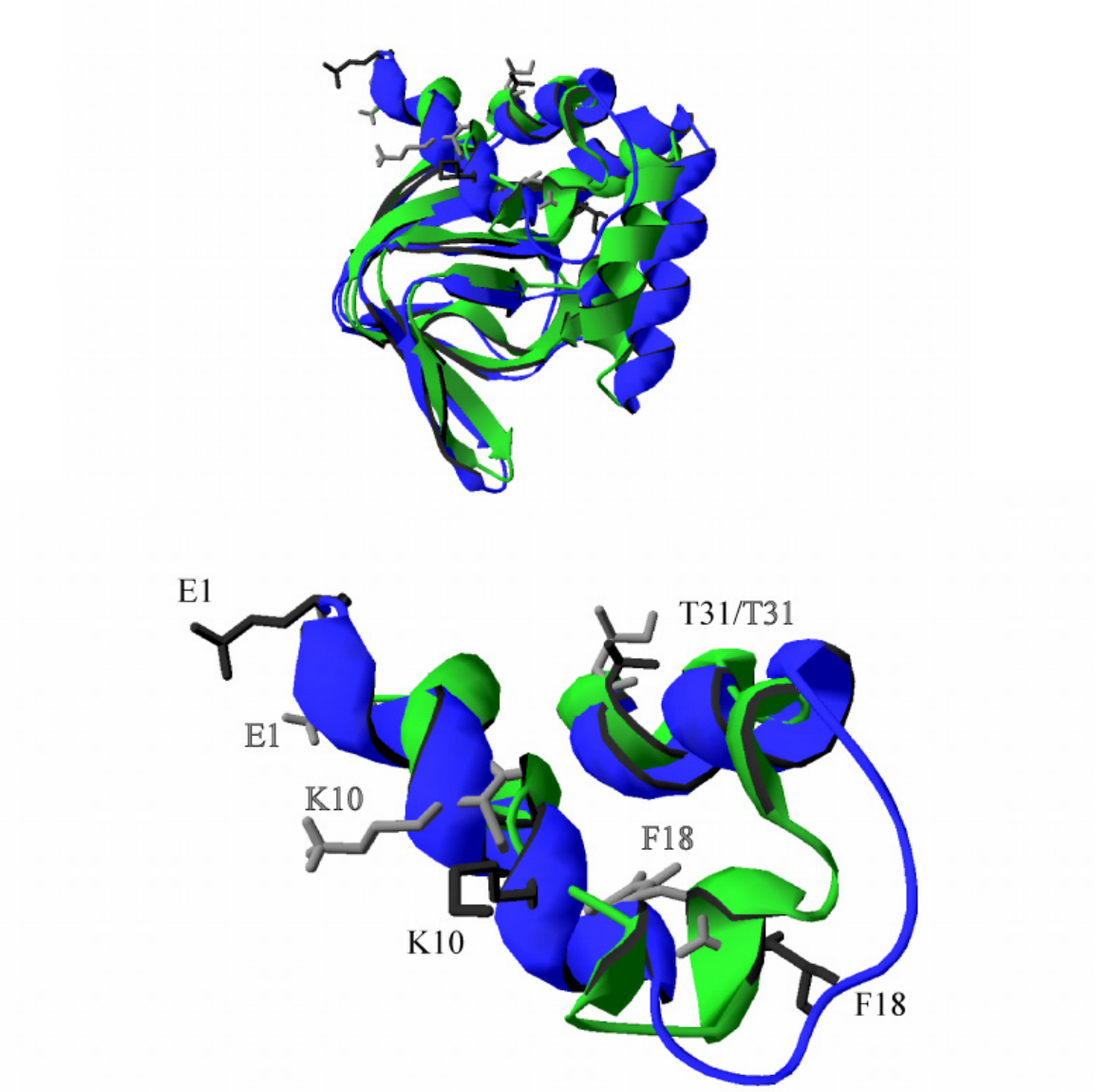
Notes: Structure quality was evaluated using z-score and energy profiles. Z-score is on the vertical axis of the graphs on the left and number of residues on the horizontal axis. Energy is on the vertical axis on the graphs on the right with sequence position on the horizontal axis. The residues number is found on the horizontal axis.

Figure 8-19: The protein model G3 is stable after 5ns of simulation.



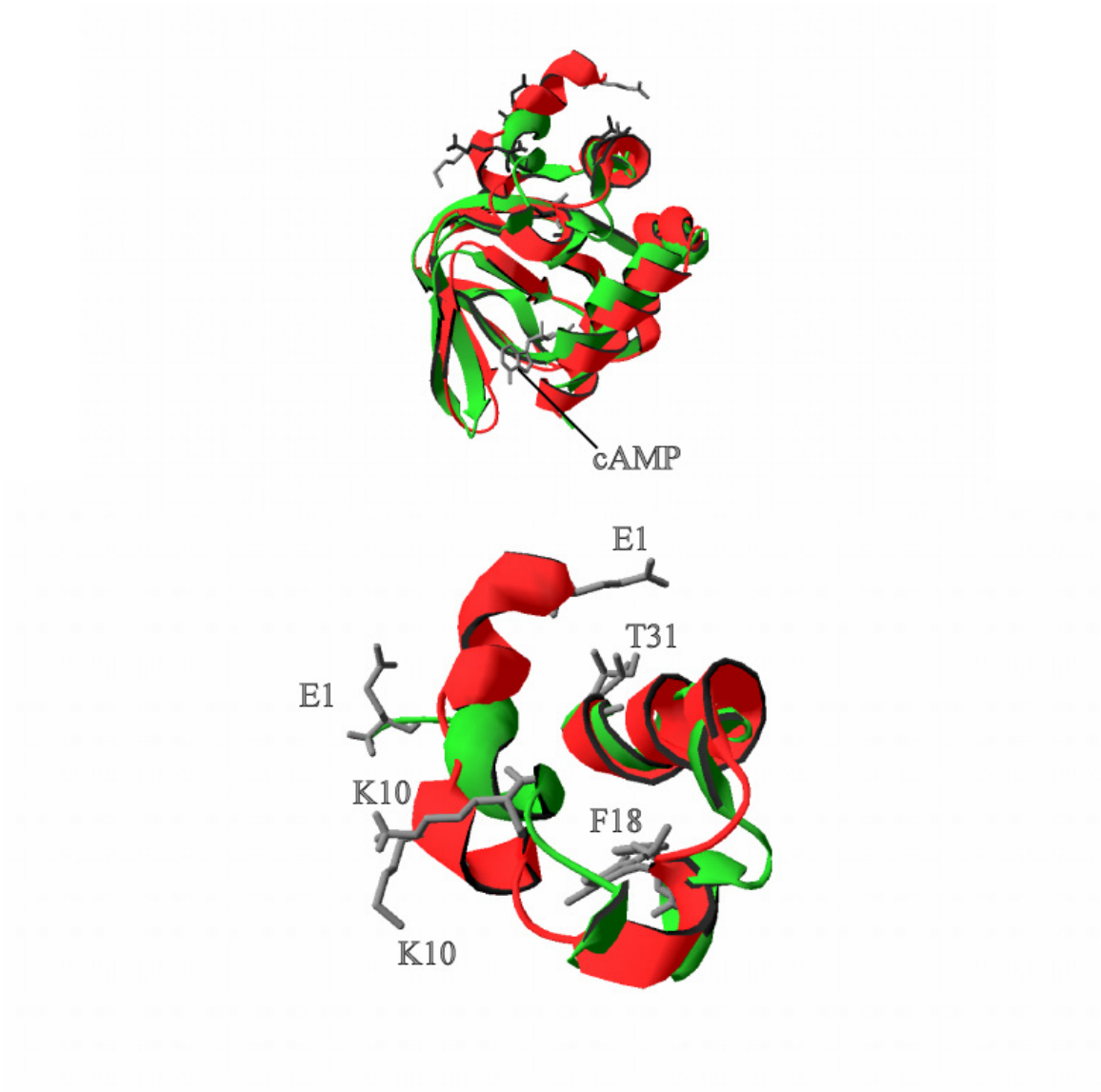
Notes: Graph reports the RMSD for superimpositions of each successive frame of the simulation on the original input structure (frame 0) using only the Calpha carbon atoms. All significant movements in the simulation have stopped after 5ns.

Figure 8-20: The protein models H2a and G4 show similarity in the region of interest.



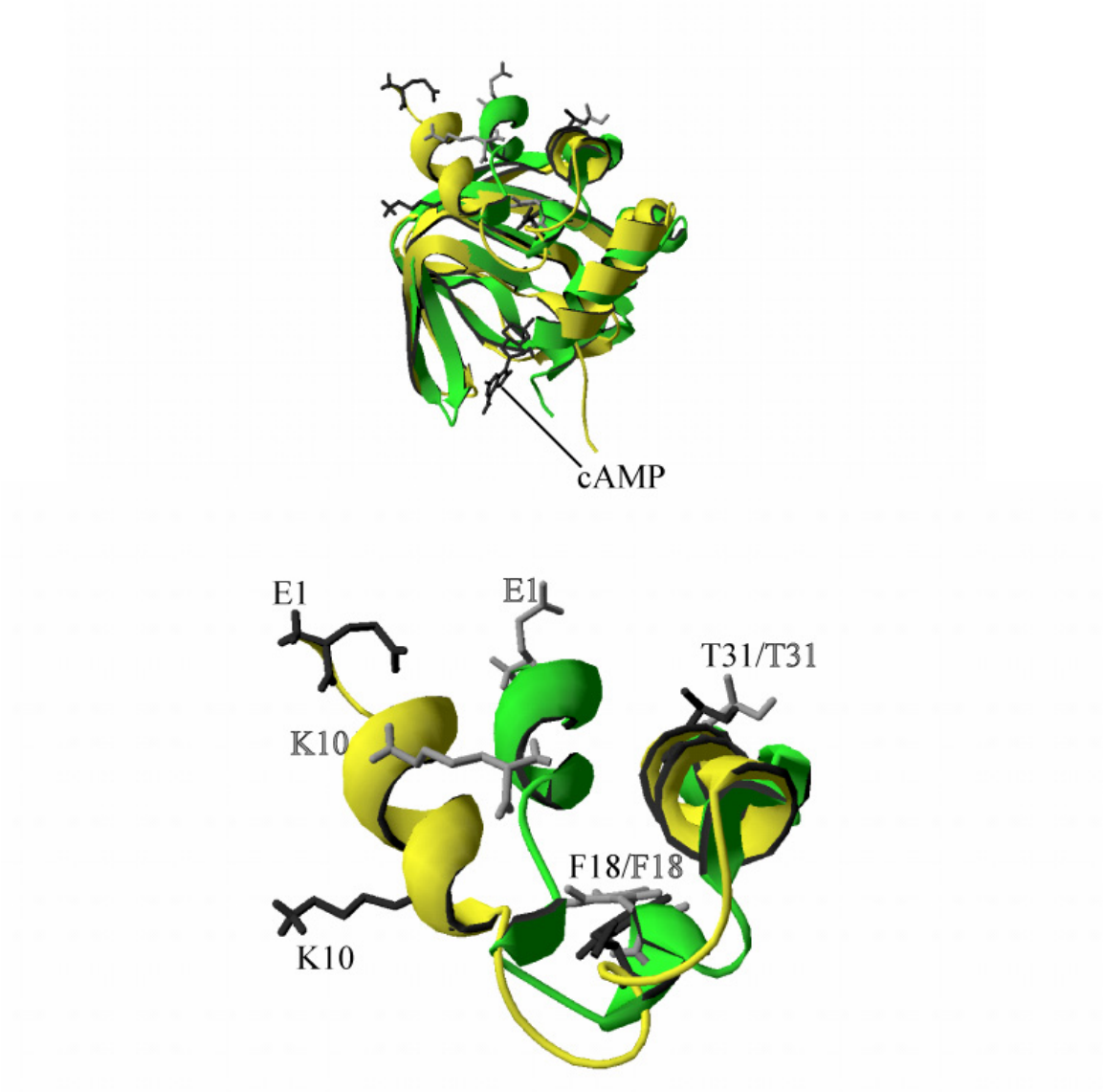
Notes: Numbering shown is based on numbering in Table 8-5. H2a is in blue (E1) and G4 (E1) in green. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-21: The protein models H2h and G4 show different conformations for helix D'.



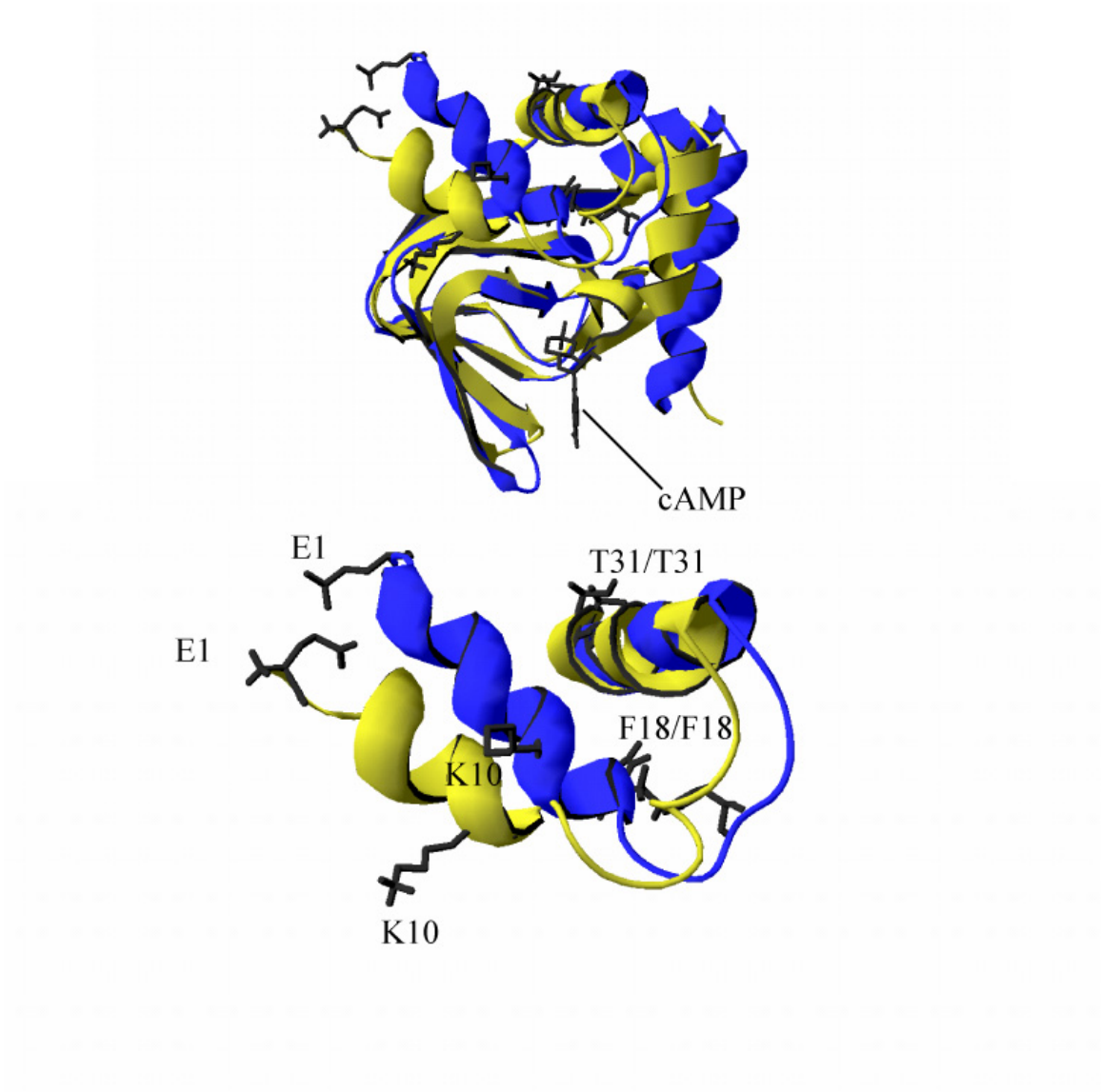
Notes: Numbering shown is based on numbering in Table 8-5. H2h is in red (E1 and cAMP) and G4 (E1) in green. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-22: The protein models G3 and G4 show a difference in the helix D'.



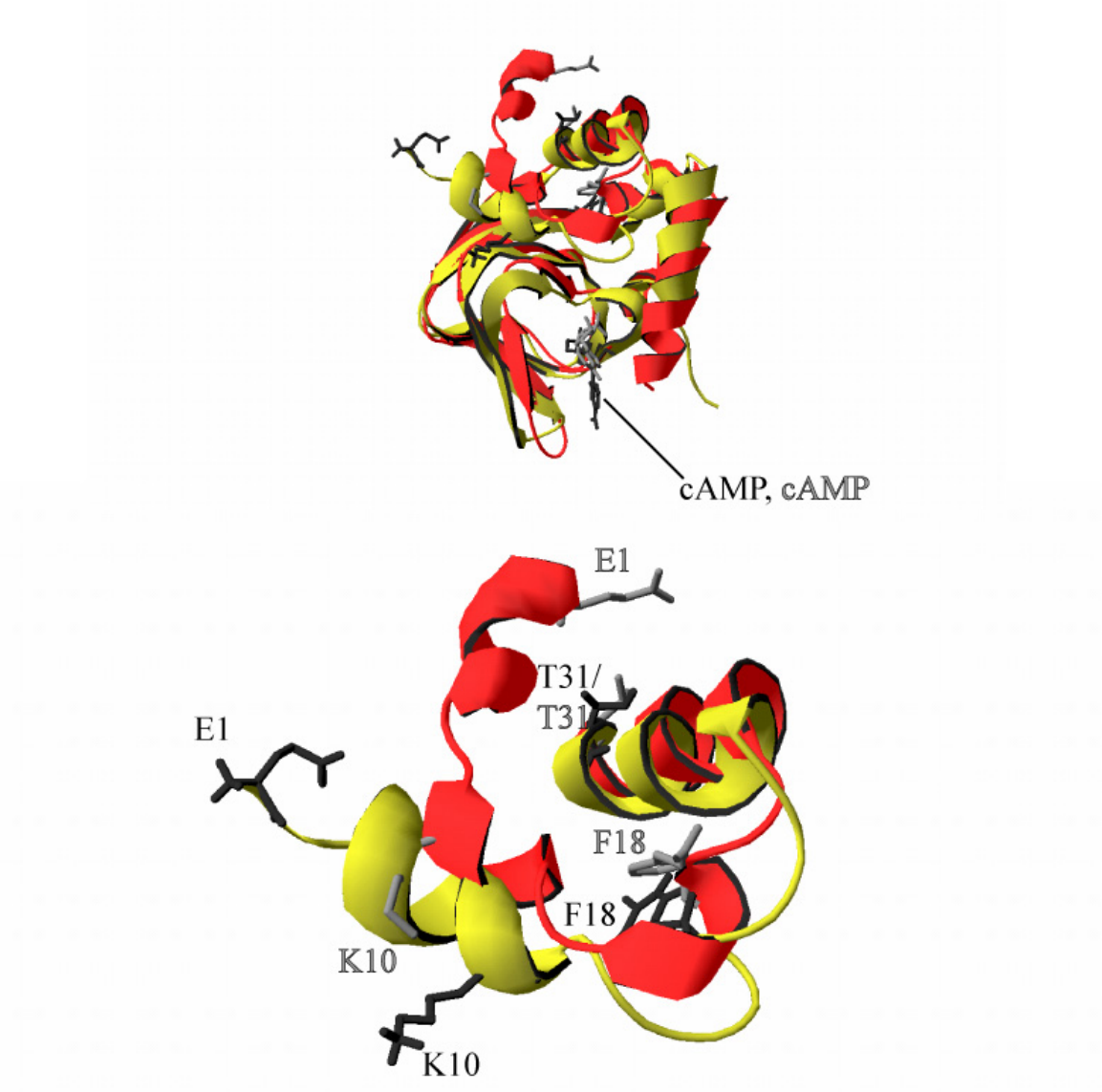
Notes: Numbering shown is based on numbering in crystal structures. G3 is in yellow (E1 and cAMP) and G4 (E1) in green. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-23: The protein models H2a and G3 show similarity in the CNBD only.



Notes: Numbering shown is based on numbering in crystal structures. H2a is in blue (E1) and G3 (E1) in yellow. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

Figure 8-24: The protein models H2h and G3 show a difference in the helix D'.



Notes: Numbering shown is based on numbering in crystal structures. H2h is in red (E1 and cAMP) and G3 (E1 and cAMP) in yellow. Top is a superimposition of whole proteins using CNBD residues only. Bottom is superimposition found in A blown up for the region of interest.

8.2 Tables

Table 8-1: Crystal structures and modified crystal structures used.

Name	Output Description (General)	Output Description (technical)
1q5o	holo cytoplasmic domain of HCN2	D443-L643 and CMP201 (cAMP); E501-L633 is equivalent to E1-L133 in models seen below.
1q5omod	1q5o with MSE-->MET	E501-L633 and d201(cAMP), MSE -->MET: 460, 485, 515, 529, 554, 572, 620, 621
H2h	Fragment of 1q5omod	E501-L633 and d201
1vp6	holo cytoplasmic domain of Mlotk1	A, C: V218-G350, CMP360, CMP368
Mlh	1vp6 fragment	R219-R349 and d360 (cAMP) of 1vp6 chain C
1u12	apo cytoplasmic domain dimer	A: R219-R349 and B: G221-E347
Mla	1u12 monomer	R219-R349 of 1u12 chain A
Models		
H2a	H2h in the apo form built with Swiss-Model.	E1-L133 of H2h sequence.
H2R	H2h in the cAMP unliganded state built with Fit Raw Sequence.	E1-L133 of H2h sequence.
SA2	cAMP unliganded state of H2h built with Build Preliminary model (from selected layers)	E1-L133 of H2h sequence.
G4	cAMP unliganded state of H2h built via simulation of H2h without cAMP for 20ns.	E1-L133 of H2h sequence. 2 Na+ and 4 Cl-
G3	cAMP bound state of H2h built via simulation of H2h with cAMP for 20ns.	E1-L133 of H2h sequence. 2 Na+ and 3 Cl- and d134 (cAMP).

Notes: The letters preceeding the residues indicates the chain name if more than one. All structure were modified by saving the correct residues in pdbviewer except for 1q5o. 1q5o was modified in the text of the pdb file.

Table 8-2: 1vp6 compared to 1q5o using RMSD

Chain ID	CNBD	Region of Interest
A of 1vp6	1.16	19.83
C of 1vp6	1.17	6.01

Notes: CNBD residues from 1q5o (K34-T66 and N69-L133) were superimposed on the CNBD residues of each Mlotik1 structure (R250-R349). The RMSD was calculated for specific subregions.

Table 8-3: 1u12 compared to 1q5o using RMSD (Angstroms)

Chain ID	CNBD	Region of Interest
A of 1u12	2.022	5.361
B of 1u12	3.178	5.849

Notes: 1q5o was superimposed onto 1u12 using the CNBD (With chain A of 1u12 1q5o: K34-T66, N69-L133, 1u12a: R250-R349; with chain B of 1u12 1q5o: K34-T66, N69-L133, 1u12B: R250-G347).

Table 8-4: ProSA values for models and structures.

Model	Z-score
H2h	-6.99
Mla	-6.30
Mlh	-6.25
H2R	-5.91
SA2	-6.32
H2a	-6.37
G3	-6.17
G4	-5.81

Notes: Z-score of <-3.57 indicates good quality structure.

Table 8-5: Dictionary for residues of models and structures in region of interest.

Standard name	H2h, SA2	H2a, G3, G4, H2R	Mla and Mlh
1	**E501**	**1**	**219**
2	E502	2	220
3	I503	3	221
4	V504	4	222
5	N505	5	223
6	F506	6	224
7	N507	7	225
8	C508	8	226
9	R509	9	227
10	**K510**	**10**	**228**
11	L511	11	229
12	V512	12	230
13	A513	13	231
14	S514	14	232
15	M515	15	233
16	P516	16	234
17	L517	17	235
18	**F518**	**18**	**236**
19	A519	19	237
20	N520	20	238
21	A521	21	239
22	D522	22	240
23	P523	23	241
24	N524	24	242
25	F525	25	243
26	V526	26	244
27	T527	27	245
28	A528	28	246
29	M529	29	247
30	L530	30	248
31	**T531**	**31**	**249**

Notes: Standard name is the one referred to in the text. H2h, Mla and Mlh numbers are those from the crystal structures. Continuation of the numbering used for residues C-terminal of T531 (refer to Figure 8-4)

Table 8-6: RMSD (Angstroms) of H2h relative to Mlotik1 models after superimposing CNBDs.

Protein	Region of Interest	CNBD	Whole Protein
Mlh	5.49	1.90	3.13
Mla	5.84	2.38	3.52

Notes: CNBD residues from 1q5o (K34-T66 and N69-L133) were superimposed on the CNBD residues of each Mlotik1 structure (R250-R349). The RMSD was calculated for specific subregions.

Table 8-7: RMSD (Angstroms) of apo structures compared to H2a after CNBD superimposition.

Superimposed apo model	ROI	CNBD	Whole
Mla	0.10	0.26	0.23
H2R	0.63	1.29	1.16
SA2	0.01	0.26	0.23
G4	4.89	2.65	3.30

Notes: RMSD (Å) calculated as found in Table 8-7. ROI: region of interest, HCN*: E1-T532); CNBD: HCN* (K34-L133) excluding K67-G68. For corresponding Mla residues see Table 8-3.

Table 8-8: RMSD (Angstroms) of apo versus holo HCN* structures after CNBD superimposition.

Superimposed apo model with H2h	**D'**	**E'**	**F'**	**Thm**	**A**	**RoI**	**CNBD**	**Whole**
H2a	7.72	4.98	5.85	5.43	2.02	5.45	2.03	3.19
H2R	7.53	4.98	5.78	5.36	1.90	5.36	2.02	3.15
SA2	8.09	5.11	6.00	5.86	2.23	5.36	2.02	3.15
G4	2.94*, 8.94	3.99	2.04	3.36	2.16	4.96	1.46	2.72

Superimposed apo model with G3	D'	E'	F'	Thm	A	RoI	CNBD	Whole
H2a	8.02	3.92	3.54	3.60	1.82	4.78	2.37	3.10
G4	10.25*, 11.29	5.41	2.70	4.09	2.74	6.20	3.56	2.17

Notes: See notes for table 8-7. *D' RMSD (Å) for G4 done also done with residues 5-6 only. D': E1-F6, E': K10-V12, F': P16-F18, Thm: K10-F18 , A: P23-T31, ROI: E1-T31, CNBD: K34-T66, N69-L133, whole: E1-T66, N69-L133

Table 8-9: Distances (Angstroms) between residues 10 and 18.

holo models or crystal structures	**Distance**
H2h	11.06
Mlh	12.20
G3	9.59

apo models or crystal structures	**Distance**
Mla	13.21
H2a	11.18
SA2	11.18
H2R	11.18
G4	9.22

Table 8-10: Distance (Angstroms) comparisons for H2h (holo) and apo models.

apo model or crystal structure	K10-F18 distance apo (A):	K10-F18 distance H2h (B):	Difference
Mla	13.21	11.06	2.15
SA2	11.18	11.06	0.12
H2R	11.18	11.06	0.12
G4	9.22	11.06	-1.84

Notes Distances were measured: between residues 10 and 18 of H2h (holo) and apo models. Difference = A– B.

Table 8-11: Distance (Angstroms) comparisons for H2h (holo) and holo models.

holo model or crystal structure	K10-F18 distance apo (A):	K10-F18 distance H2h (B):	Difference
G3	9.59	11.06	-1.47
Mlh	12.20	11.06	1.14

Notes: See notes for table 8-8.

Table 8-12: Distance (Angstroms) between A helix residues of holo versus apo models.

Residue	H2h and H2a	H2h and G4	G3 and G4
23	2.33	3.41	1.58
24	2.35	3.19	0.88
25	1.48	2.31	2.43
26	2.11	2.76	3.69
27	2.22	1.60	4.27
28	1.39	0.70	3.85
29	1.51	1.08	2.54
30	2.08	1.13	2.25
31	1.93	0.72	1.80

Notes: Distances were measured for the same residues of the A helix within a superimposition of the two models or crystal structures in question.

Table 8-13: RMSD (Å) of H2h (holo) relative to the CNBD.

Superimposed holo model	**D'**	**E'**	**F'**	**Thm**	**A**	**RoI**	**CNBD**	**Whole**
G3	14.08	4.95	3.93	4.9	2.96	7.35	2.17	4.05

Notes: See notes for table 8-5.

Table 8-14: Distance (Angstroms) comparisons for H2a (apo) and other apo models.

apo model or structure	K10-F18 distance apo (A):	K10-F18 distance H2a (B):	Difference
Mla	11.18	11.18	0.00
H2R	11.18	11.18	0.00
SA2	11.18	11.18	0.00
G4	9.22	11.18	-1.96

Notes: See notes for table 8-8.

Table 8-15: RMSD (Angstroms) for Mlh superimposed on Mla.

H2h structural component	RMSD
D'	7.15
E'	5.05
A	5.42
Thm	1.79
Region of interest	4.81
CNBD	3.43
Whole domain	3.78

Notes: D': R219-V224, E': Q228-V230, F': P234-V236, A: P241-R249, Thm: Q228-F236, region of interest: R219-R249, CNBDs: R252-R349, whole domain: R219-R349)

References

Accili, E.A., C. Proenza, M. Baruscotti and D. DiFrancesco. 2002. From funny current to HCN channels: 20 years of excitation. *News Physiol Sci*. 17:32-37.

Altieri, S.L., G.M. Clayton, W.R. Silverman, A.O. Olivares, E.M. De la Cruz, L.R. Thomas and J.H. Morais-Cabral. 2008. Structural and energetic analysis of activation by a cyclic nucleotide binding domain. *J Mol Biol*. 381:655-669.

Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res*. 28:235-242.

Berrera, M., S. Pantano and P. Carloni. 2006. cAMP Modulation of the cytoplasmic domain in the HCN2 channel investigated by molecular simulations. *Biophys J*. 90:3428-3433.

Biel, M., A. Schneider and C. Wahl. 2002. Cardiac HCN channels: structure, function, and modulation. *Trends Cardiovasc Med*. 12:206-212.

Boron, W.F. and E.L. Boulpaep. 2005. Medical physiology : a cellular and molecular approach. Vol. Updated edition. Elsevier Saunders, Philadelphia, Pa. xiii, 1319 pp.

Clayton, G.M., W.R. Silverman, L. Heginbotham and J.H. Morais-Cabral. 2004. Structural basis of ligand activation in a cyclic nucleotide regulated potassium channel. *Cell*. 119:615-627.

Craven, K.B. and W.N. Zagotta. 2004. Salt bridges and gating in the COOH-terminal region of HCN2 and CNGA1 channels. *J Gen Physiol*. 124:663-677.

Craven, K.B. and W.N. Zagotta. 2006. CNG and HCN channels: two peas, one pod. *Annu Rev Physiol*. 68:375-401.

DiFrancesco, D. and P. Tortora. 1991. Direct activation of cardiac pacemaker channels by intracellular cyclic AMP. *Nature*. 351:145-147.

Doyle, D.A., J. Morais Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait and R. MacKinnon. 1998. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science*. 280:69-77.

Flynn, G.E., K.D. Black, L.D. Islas, B. Sankaran and W.N. Zagotta. 2007. Structure and rearrangements in the carboxy-terminal region of SpIH channels. *Structure*. 15:671-682.

Giorgetti, A., P. Carloni, P. Mistrik and V. Torre. 2005. A homology model of the pore region of HCN channels. *Biophys J*. 89:932-944.

Guex, N. and M.C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 18:2714-2723.

Hefti, M.H., K.J. Francoijs, S.C. de Vries, R. Dixon and J. Vervoort. 2004. The PAS fold. A redefinition of the PAS domain based upon structural prediction. *Eur J Biochem*. 271:1198-1208.

Heginbotham, L., Z. Lu, T. Abramson and R. MacKinnon. 1994. Mutations in the K+ channel signature sequence. *Biophys J*. 66:1061-1067.

Hess, B., H. Bekkerr, H.H.J.C. Berendsen and J.G.E.M. Fraaije. 2007. LINCS: a linear constraint solver for molecular simulations. . *Journal of Computational Chemistry*. 18:1463-1472.

Hua, L. and S.E. Gordon. 2005. Functional interactions between A' helices in the C-linker of open CNG channels. *J Gen Physiol*. 125:335-344.

Humphrey, W., A. Dalke and K. Schulten. 1996. VMD: visual molecular dynamics. *J Mol Graph*. 14:33-38, 27-38.

Johnson, J.P., Jr. and W.N. Zagotta. 2001. Rotational movement during cyclic nucleotide-gated channel opening. *Nature*. 412:917-921.

Johnson, J.P., Jr. and W.N. Zagotta. 2005. The carboxyl-terminal region of cyclic nucleotide-modulated channels is a gating ring, not a permeation path. *Proc Natl Acad Sci U S A*. 102:2742-2747.

Kim, C., N.H. Xuong and S.S. Taylor. 2005. Crystal structure of a complex between the catalytic and regulatory (RIalpha) subunits of PKA. *Science*. 307:690-696.

Kornev, A.P., S.S. Taylor and L.F. Ten Eyck. 2008. A generalized allosteric mechanism for cis-regulated cyclic nucleotide binding domains. *PLoS Comput Biol*. 4:e1000056.

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. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23:2947-2948.

Matulef, K., G.E. Flynn and W.N. Zagotta. 1999. Molecular rearrangements in the ligand-binding domain of cyclic nucleotide-gated channels. *Neuron*. 24:443-452.

Nimigean, C.M., T. Shane and C. Miller. 2004. A cyclic nucleotide modulated prokaryotic K+ channel. *J Gen Physiol*. 124:203-210.

Noma, A. and H. Irisawa. 1976. Effects of calcium ion on the rising phase of the action potential in rabbit sinoatrial node cells. *Jpn J Physiol*. 26:93-99.

Proenza, C., N. Tran, D. Angoli, K. Zahynacz, P. Balcar and E.A. Accili. 2002. Different roles for the cyclic nucleotide binding domain and amino terminus in assembly and expression of hyperpolarization-activated, cyclic nucleotide-gated channels. *J Biol Chem*. 277:29634-29642.

Robinson, R.B.a. and S.A. Siegelbaum. 2003. Hyperpolarization-activated cation currents: from molecules to physiological function. *Annu Rev Physiol*. 65:453-480.

Schwarzenbacher, R., A. Godzik and L. Jaroszewski. 2008. The JCSG MR pipeline: optimized alignments, multiple models and parallel searches. *Acta Crystallogr D Biol Crystallogr*. 64:133-140.

Schwede, T., J. Kopp, N. Guex and M.C. Peitsch. 2003. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*. 31:3381-3385.

Swartz, K.J. 2004. Towards a structural view of gating in potassium channels. *Nat Rev Neurosci*. 5:905-916.

Taraska, J.W. and W.N. Zagotta. 2007. Structural dynamics in the gating ring of cyclic nucleotide-gated ion channels. *Nat Struct Mol Biol*. 14:854-860.

Thompson, J.D., D.G. Higgins and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673-4680.

Varnum, M.D., K.D. Black and W.N. Zagotta. 1995. Molecular mechanism for ligand discrimination of cyclic nucleotide-gated channels. *Neuron*. 15:619-625.

Wiederstein, M. and M.J. Sippl. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 35:W407-410.

Zagotta, W.N., N.B. Olivier, K.D. Black, E.C. Young, R. Olson and E. Gouaux. 2003. Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature*. 425:200-205.

Zhou, L. and S.A. Siegelbaum. 2007. Gating of HCN channels by cyclic nucleotides: residue contacts that underlie ligand binding, selectivity, and efficacy. *Structure*. 15:655-670.