Towards an accurate alignment of G-protein-coupled receptors from different classes.

J.M.Farr

A thesis submitted for the degree of MSc by dissertation

Department of Biological Sciences

University of Essex

October 2016

Abstract

G-protein coupled receptors (GPCRs) are a large and diverse family of cell surface receptors involved in signal transduction. They share a general structure of a seven transmembrane domain, an extracellular N-terminus and an intracellular C-terminus domain responsible for coupling to the G-protein. GPCRs are split into 6 families (class A to F) based on sequence homology. GPCRs are associated with a number of diseases such as Schizophrenia, Parkinson's, Alzheimer's, anxiety and some cancers making them attractive targets for drug design. There is huge potential for further development of drugs that target GPCRs and the more that is understood about the receptors and their structures, the better the chance of discovering a successful drug compound. An accurate way to model the structure of other GPCRs using known structures from X-ray crystallography would therefore be very useful. To do this, sequences from the different classes of GPCR will need to be aligned as accurately as possible. Half sphere exposure data can be used to more accurately identify the transmembrane regions which are the most conserved across the different classes. The method used then compared ungapped alignments of defined helical regions. An ungapped pair-wise alignment was carried out to compare two helix sequences at a time (one from class C and the other from class A, B, E or F) and to score how well they align at each position. This data was then collected, each top scoring alignment from the pair-wise alignments is noted and the alignment that has the highest number of votes is considered to be the best. The criteria that were scored at each alignment were the BLOSUM matrix scores, hydropathy, entropy, amino acid volume and variability.

Acknowledgements

I would like to start by saying a big thank you to my supervisor Christopher Reynolds for his continued patience and support throughout the completion of this masters dissertation. He has given me so much help and advice throughout the project even through the broken foot that postponed my studies for a year. I cannot thank him enough for everything that he has done for me throughout my time at university.

I would also like to thank my family and particularly my parents for their continued support over the last three years. I would not have been able to complete my studies without the help they have given me.

Finally, I would like to thank all the other lecturers that have helped me and supported me during my time at the University of Essex both as an undergraduate and postgraduate student.

Table of Contents

Chapter 1

Overview of GPCRs

1.1. General introduction to the GPCR superfamily

G-protein coupled receptors (GPCR's) are the largest superfamily of cell surface receptors (Chun et al. 2012). They are a very diverse family of receptors and their main function is cell signal transduction (Kroeze et al. 2003). GPCR's are present in most eukaryotic organisms from mammals to insects and even plants. This shows that GPCR's originated very early on in evolution. GPCR's are characterised by having a common domain of seven transmembrane helices (7TM region). These helices are connected by intracellular and extracellular loops and pass through the cell membrane, anchoring the GPCR to the plasma membrane. Ligands for some GPCR families bind directly to this 7TM region. Some of the families of receptors, however, also have a large, extracellular N-terminus region that can consist of a number of different domains with a variety of functions including ligand binding (Fredriksson and Schiöth, 2005). GPCRs also have an intracellular C-terminus region. These regions can vary widely in shape and length between the main classes and have different functions depending on the class of receptor (Kroeze et al. 2003).

There is a wide variety of receptors in the GPCR family and also a diverse range in the type of ligand they respond to, including amino acids, light, neurotransmitters, hormones and large proteins (Kroeze et al. 2003). The ligand

may bind to the receptor differently in the different classes but always causes a conformational change in the receptor. This conformational change helps to activate the G-protein coupled to the receptor. There are a number of types of G-protein that can be coupled to the receptor and that will determine which signalling pathway is activated (Kroeze et al. 2003 and Watkins et al. 2016).

The G-proteins that associate with GPCRs are heterotrimers. They consist of an α, β and a γ subunit (Wall et al. 1995). Heterotrimeric G-proteins are very important in GPCR cell signalling as they are the mediating step (Baltoumas et al. 2013). The role of the G-protein is to transmit the signal from an activated (usually ligand bound) GPCR to the enzymes and channels etc. (Neer and Smith, 1996). G-proteins have an active and an inactive state that they can switch between, depending on the state of the GPCR they are coupled with (Kleuss et al. 1993). The β and the γ subunits are bound very tightly together and are commonly known as the βγ subunit (Neer and Smith, 1996). When this βγ subunit is bound to the α subunit in the heterotrimeric complex they are inactive and do not interact with the effector molecules. In this inactive state the α subunit is tightly bound to a GDP molecule and this interaction is further stabilised by the attached βγ component (Wall et al. 1995). The G-protein becomes active when a ligand is bound to the GPCR as it causes conformational changes to the GPCR which then causes a conformational change in the α subunit (Kleuss et al. 1993). This conformation change allows the GDP molecule to be replaced with a GTP molecule. This switch causes further conformational changes that allows the α-GTP complex to dissociate from the βγ subunit (Wall et al. 1995). These separated subunits can then go on to interact with different cell effector molecules and cause the cell to respond to the external ligand (Kleuss et al. 1993). The α subunit contains GTPase that will cause

the GTP to be slowly hydrolysed into GDP (Wall et al. 1995). This then allows the two components to reassemble into the inactive state. This mechanism is very important and it is vital for GPCR signal regulation (Neer and Smith, 1996). There are a number of different possible isoforms for each of the three subunits in heterotrimeric G-proteins (over 20 have been discovered for the α subunit for example). The type of subunit present in the G-protein determines the type of GPCRs it will interact with and the effector molecules it is able to activate. These are important for ensuring signalling specificity (Baltoumas et al. 2013).

GPCR's have been found to play important roles in a large variety of organs and cells and have been associated with a number of diseases. They are therefore very important to pharmaceutical companies and have been the focus of lots of research as drug targets. There is a high demand for potential new drugs to combat the important diseases associated with GPCRs and so there has been significant research into searching and screening for new drug leads in this area (Chun et al. 2012).

There is a downside to using GPCR's as drug targets, however. One issue is the fact that GPCR's are seen in such a wide variety of locations in the body. This potentially makes it difficult to obtain high specificity in the target sites. Low specificity in the target site leads to a number of problems including tolerance and potentially hazardous side effects (Chun et al. 2012).

Despite problems with using GPCRs as drug targets they are such an important group that the benefits outweigh potential risks. This, however, means that determining as much as possible about the structure, activation method and function of the different types of GPCR is more important than ever in the search for

drugs with the optimum outcome (Chun et al. 2012), hence this investigation into the class A – class C alignment.

1.2. GPCR families

GPCR's are classified into families based on the conservation of important amino acids in the transmembrane helical regions of the receptors (Chun et al. 2012). There are a number of different classification systems that have been used for the GPCR superfamily. One of these classification systems is the GRAFS system that splits GPCR's into 5 main families. These are the glutamate receptors, the rhodopsin receptors, adhesion receptors, frizzled receptors and the secretin receptors. Another system is the Class A-F system as described below (Fredriksson and Schiöth, 2005).

1.2.1. Class A

Class A receptors (also known as the rhodopsin-like receptors) are the largest class of GPCR containing approximately 670 receptor proteins. This group has the largest number of drugs targeting them on the market, for example, antihistamines, cardiovascular drugs and antipsychotics. As this group is so large it can be further divided into the α, β, γ and δ groups. Class A receptors have had the most research carried out on them as they are the largest group. The first solved GPCR crystal structure was of the class A receptor bovine rhodopsin (Palczewski et al., 2000).

Rhodopsin is found in the rod cells of the retina and is involved in vision. Light is absorbed by these GPCRs causing them to activate and trigger visual signal transduction (Maeda et al. 2003).

Class A receptors mostly have a very short N-terminus region that does not usually bind to the ligands. However, the N-terminus is involved in binding for class A peptide receptors such as neurotensin receptors (White et al. 2012). There is quite a range in the ligands that bind to class A receptors including light, peptides, amines and purines. These ligands bind directly to the transmembrane and the extracellular loop regions. The TM regions form a pocket for the ligand and the interactions formed cause a conformational change, thus activating the G protein (Lagerström and Schiöth, 2008).

A large number of crystal structures have been found for other GPCRs in class A, also. Examples include the beta1-Adrenergic Receptor (Warne et al. 2008); A2A Adenosine Receptor (Jaakola et al. 2008); Protease activated receptor 1 (Zhang et al. 2012); CXCR4 Chemokine Receptor (Wu et al. 2010).

1.2.2. **Class B**

The class B GPCR receptors are often split into two groups; the secretin receptors and the adhesion receptors (Lagerström and Schiöth, 2008). The secretin receptors are found in many different organs and can activate a wide variety of pathways. Secretin receptors are involved in many different cellular processes for example appetite regulation, regulation of bile secretion and homeostasis etc. (Siu et al. 2006). There are drugs currently on the market that target the secretin receptor and these are used to treat, for example, hypercalcaemia, hypoglycaemia and osteoporosis. Due to the involvement these receptors have in the regulation of appetite they are a potential future drug target for combating type 2 diabetes. Secretin receptors have an extracellular ligand binding domain that binds to peptide hormones (Lagerström and Schiöth, 2008).

Adhesion receptors are known to be involved in cell growth and the immune system. They are also found to be present in the central nervous system but their function here is not yet known. The activation mechanism involves the cleaving of the extracellular domain by a protease revealing a 7TM stalk that activates the receptor (Stoveken et al. 2015). Adhesion receptors have very long and diverse N-terminus regions that are highly glycosylated and contain several domains. These N-terminus regions form rigid structures outside the cell. There is not much known about the adhesion receptor ligands so far but the ones that are known are all large and membrane bound (Lagerström and Schiöth, 2008).

Both secretin and adhesion receptors have conserved cysteine residues in the first two transmembrane regions. Secretin also has conserved cysteines in the N-terminus region that form cysteine bridges and stabilise the structure of the receptors (Lagerström and Schiöth, 2008). While the extracellular regions of secretin and adhesion receptors are quite different, the TM regions align well, and so in our analysis (Chapter 3), these two families are combined.

There have been two crystal X-Ray structures found so far for the class B receptors. These are for the Glucagon Receptor (Siu et al. 2013) and for the Corticotropin-Releasing Factor Receptor 1 (Hollenstein et al. 2013).

1.2.3. **Class C**

Class C receptors (also known as the glutamate receptors) are a family of receptors found mainly in the CNS. They have been connected to a number of CNS disorders and also to disorders in calcium regulation. They consist of 22 human proteins and can be further split into subgroups including the metabotropic glutamate receptors (mGLu), GABAb, calcium sensing receptors (CASRs) and taste receptors. The N-

terminus region in class C is very large and is folded into two lobes or domains. The two lobes operate in a venus fly trap (VFT) mechanism in which the ligand binds in the cleft between the two lobes and stabilises a closed conformation, activating the cell signalling pathway. So far there have been drugs developed that successfully target certain receptors in this class. For example, a drug has been developed that targets CASR's and treats hyperparathyroidism (Lagerström and Schiöth, 2008).

Allosteric ligands have been discovered that bind directly to the TM regions of the GPCR. The discovery of these allosteric ligands may aid drug design as, if drugs that mimic these with high affinity are developed, it may be possible to develop drugs with a much better specificity for the target receptor (Lagerström and Schiöth, 2008).

At the start of the research program no structures had been found for class C GPCRs. Since then structures have been found for metabotropic glutamate receptor 5 (Christopher et al. 2015 and Doré et al. 2014) and for metabotropic glutamate receptor 1 (Wu et al. 2014).

1.2.4. **Class D**

Class D receptors are fungal mating pheromone receptors. Pheromones are a type of molecule that facilitates communication between different individuals in a species. Pheromones can relate to a variety of things such as food, alarm and mating for example. The behaviour of the individual receiving the pheromone is usually altered in some way depending on the message. Mating pheromones are detected in fungi by class D GPCR's (Lock et al. 2014). Once a pheromone binds to the GPCR a signal pathway is activated to initiate mating (Xue et al. 2008).

There are currently no X-Ray crystal structures for class D GPCRs.

1.2.5. **Class E**

The class E GPCRs are also known as the cyclic AMP (cAMP) receptors. These receptors are present in Dictyostelium organisms and are vital in their development. When the receptors are activated by a cAMP molecule they allow single cells to aggregate together and allows them to form a multicellular organism. Without the presence of this receptor the cells do not develop but remain as single cells (Johnson et al. 1993). These receptors align well to plant GPCRs and in this thesis the two groups are treated as one group (Taddese et al., 2014). Currently there are no X-Ray crystal structures for receptors in this class.

1.2.6. **Class F**

The frizzled/smoothened family of GPCR's (Class F) have been associated with a number of types of cancer development. Therefore development of drugs to target these receptors would be very significant for medicine and drug design. So far there have not been any successful drugs discovered for this class. The frizzled receptors bind to wnt glycoprotein ligands. The ligands bind to cysteine rich regions in the extracellular loops. The smoothened receptor has been found to function without the presence of a ligand. Small molecules have been found to be able to bind directly to the TM regions and affect the signalling of the GPCR (Lagerström and Schiöth, 2008).

An X-ray crystal structure for the smoothened receptor has been found by Wang et al. (2013).

1.3. Class C specific structure

Like all GPCRs, receptors in class C all contain seven highly conserved transmembrane helical regions with short loop regions both intracellular and extracellular joining them. These loop regions vary much more widely in length and amino acid residues but have been seen to contain some conserved residues and motifs. Unlike some other classes of GPCR, the 7TM region itself does not directly bind to the ligands, however it does still contain a conserved binding pocket (Chun et al. 2012).

Large extracellular region

Class C GPCR's are different from other classes in that they have a very large extracellular domain involved in binding the ligands. This domain contains a number of orthosteric sites. The extracellular domain of class C GPCR's is described as a venus fly trap (VFT) module due to the fact that it folds into two lobes or domains with a central cleft that acts as a binding pocket for the ligands. The VFT domain has an open and a closed configuration which it switches between in the absence of a bound ligand. When the ligand binds it stabilises the closed conformation activating a signalling pathway. This VFT domain is connected to the seven transmembrane (7TM) region by a cysteine rich domain. This connecting section consists of approximately 80 amino acids and has 9 cysteines that remain

completely conserved and have been found to be vital to the activation of the receptors (Chun et al. 2012). In the metabotropic glutamate (mGlu)–like receptors, a highly conserved disulfide bridge is seen to form between the VFT region and the cysteine rich domain. This interaction was discovered to be vital for the formation of interactions between the VFT and 7TM regions. Mutations were carried out to prevent the formation of these activations and it was found that without them, activation cannot occur (Chun et al. 2012).

Dimerisation

Class C GPCR's have been shown to undergo mandatory dimerisation to go into their active state. This dimerisation can be either homodimerisation (as in mGlu and calcium sensing receptors) or heterodimerisation (as in GABAb). (Chun et al. 2012, Marshall and Foord 2010 and Marshall 2007).

Complex activation process

The activation process for class C GPCRs is seen to be very complex. The ligand for class C receptors binds directly to the extracellular venus fly trap (VFT) domain. This binding causes conformational changes in all of the receptors domains and activates the coupled G-protein. Class C receptors are dimers and as such have four or six independent domains. These domains however are all connected by allosteric interactions so that a conformational change in one causes the change in the other regions (Chun et al. 2012).

1.4. Class C is further split into subgroups

Metabotropic glutamate receptors

One subgroup of class C GPCR receptors are the metabotropic glutamate receptors (mGlu). These receptors are activated by the neurotransmitter glutamate. Glutamate is the main excitatory neurotransmitter in the central nervous system (Xu et al. 2014). The metabotropic glutamate family of receptors is further divided into three groups (I, II and III) of receptors that share a high level of sequence homology. Figure 1.1 shows the three different groups and their members. Group I contains mGluR1 and mGluR5 (Speyer et al. 2014). This group, when activated by a ligand, increases the level of inositol 1,4,5-triphosphate and diacylglycerol (Julio-Pieper et al. 2011). Group I mGlu receptors induce a very strong presynaptic stimulatory response causing excitation. The signalling cascades triggered by receptors in this group are involved in cell proliferation. Group I also modulates some ion channels (Speyer et al. 2014). Group II contains mGluR2 and mGluR3 and group III contains mGluR4, mGluR6, mGluR7 and mGluR8. Both of these groups decrease the levels of intracellular cAMP when bound to a ligand (Julio-Pieper et al. 2011).

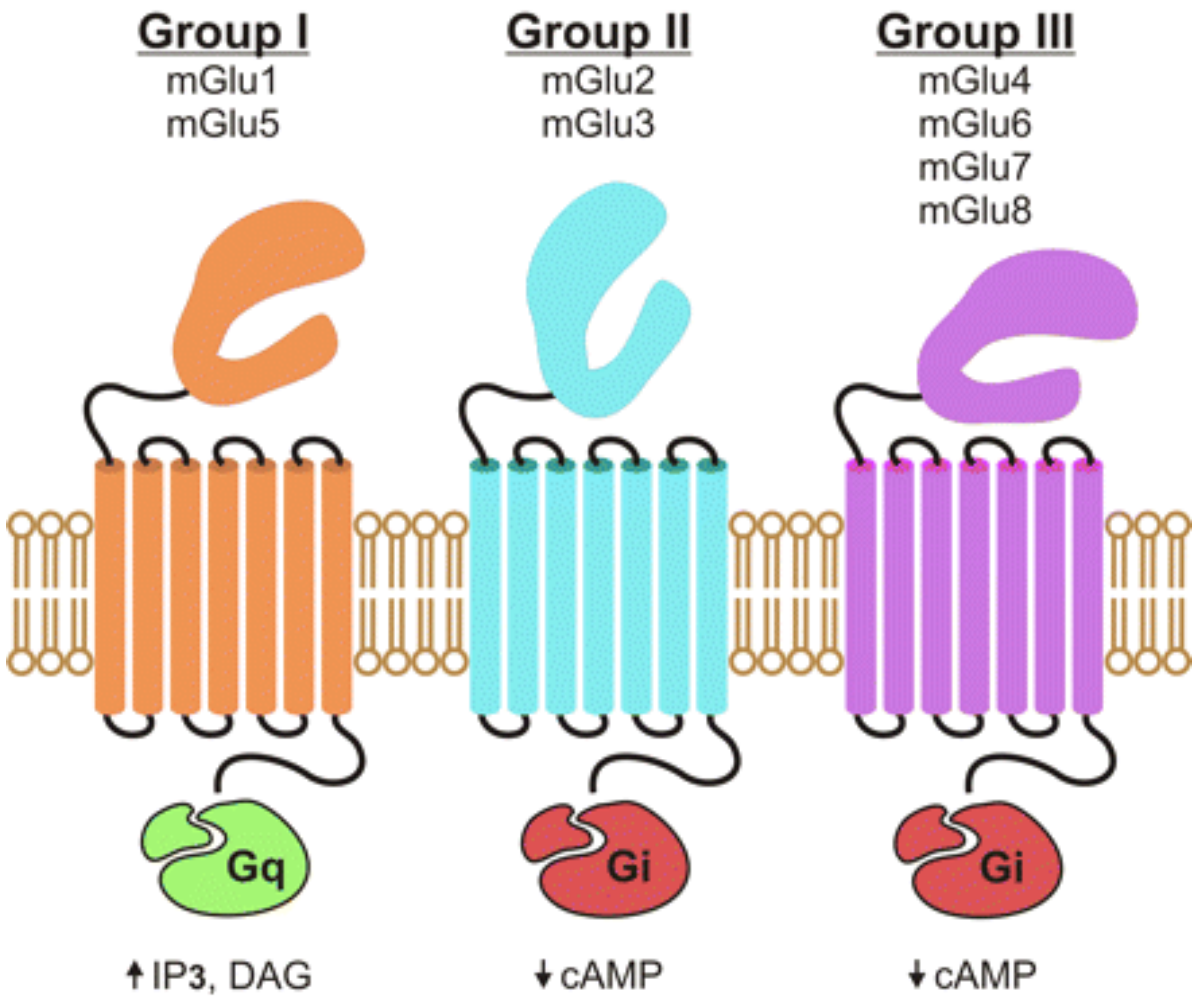

Figure 1.1: The three subgroups of mGlu receptors. Taken from (Julio-Pieper et al. 2011).

Metabotropic glutamate receptors are found in large numbers throughout the central nervous system (CNS) (Kang et al. 2014). In particular they are found in the synaptic cleft of neurons and glial cells. They have also been found in many other locations in the body outside of the CNS (Julio-Pieper et al. 2011). These receptors are very widespread in the brain and as such are involved in many CNS functions. This means that when the receptors start to function incorrectly it can cause many neurological problems. It has been shown that mGlu receptors play a role in depression, anxiety, pain, epilepsy, schizophrenia, Alzheimer's and Parkinson's disease. The severity of some of the conditions linked to mGlu receptors makes

studying them and finding suitable drugs to target them a highly important area of medical research (Kang et al. 2014). It is not only neurological conditions that mGlu receptors have been linked with. They have also been found to play a role in certain types of cancer including melanomas and gliomas. This is because the receptors are involved with cell proliferation and are known to contribute to tumour growth in breast cancer (Speyer et al. 2014).

There are eight subtypes of mGlu receptor and these are named mGluR1 through to mGluR8 (Xu et al. 2014).

Metabotropic glutamate receptors are homodimeric meaning they have two subunits of the same type working together to produce a cell response. The receptors, as they belong to class C, have large N-terminus regions folded into the venus fly trap conformation. As the receptor is a homodimer the ligand binding site is identical in both of the subunits. The glutamate molecule binds to one of the VFT modules and causes it to change to its closed conformation. The binding of a ligand also causes the two units to move closer together. It is unclear what the true active state of mGlu receptors is. Some believe that the receptor is active when one subunit is bound to the ligand and in the closed conformation while the other is unbound and open (closed-open conformation). Others believe that receptors in the closed-open conformation are only partly activated however and they believe that a second ligand has to bind to the other subunit and cause it to change to its closed conformation (closed-closed conformation) for the receptor to be truly active (Kubo and Tateyama, 2005). Figure 1.2 below shows the different possible conformations of the dimer depending on the presence of zero, one or two glutamate molecules. The diagram also shows that the space between the two subunits is decreased with the addition of each glutamate molecule.

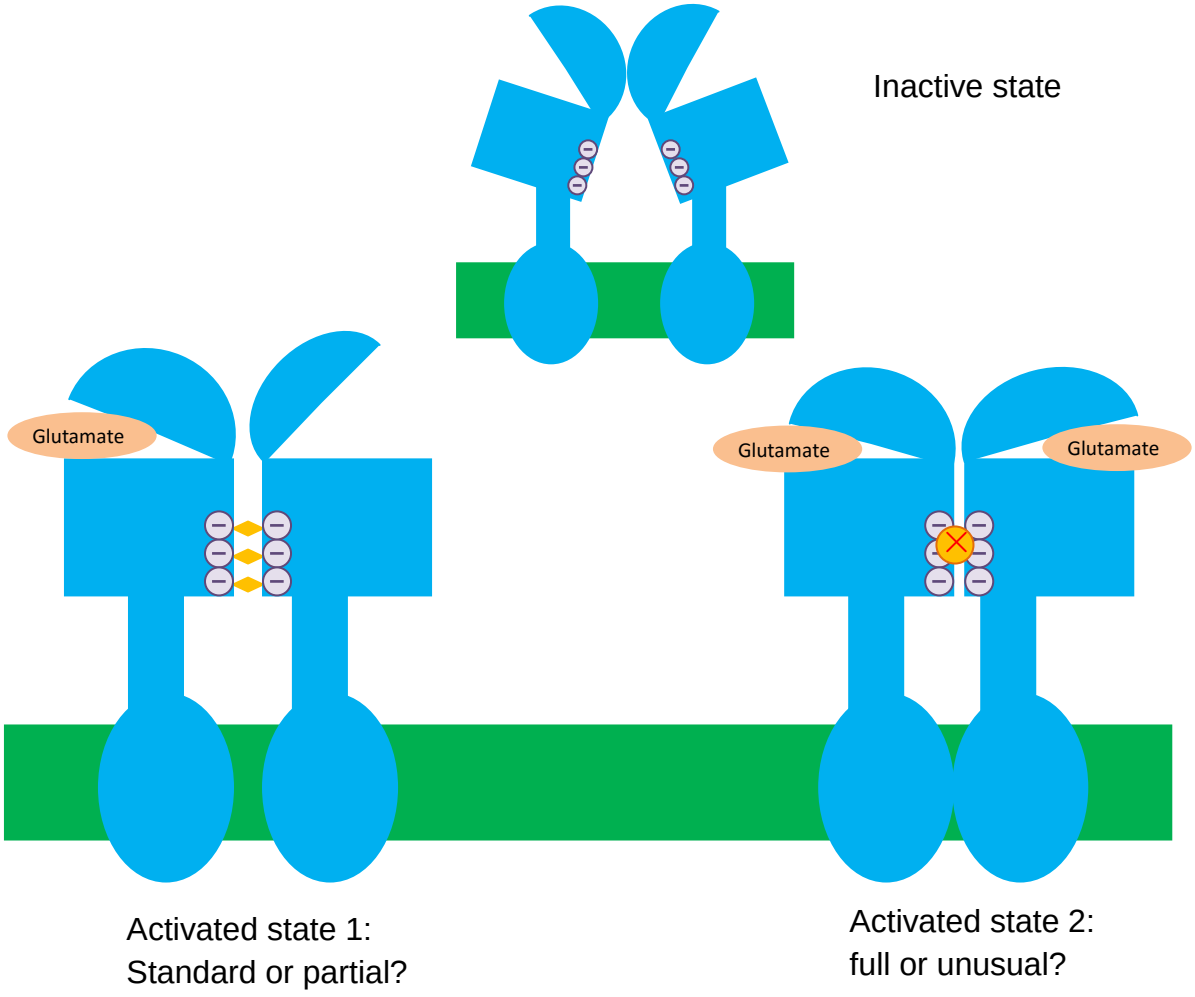

Figure 1.2: The three possible activation states of mGlu receptors. Top: inactive with no bound ligand. Bottom left: Activated state 1 with one bound glutamate molecule (Closed-Open). Bottom right: Activated state 2 with two bound glutamate receptors (Closed-Closed). Adapted from (Kubo and Tateyama, 2005).

The fact that mGlu receptors are seen to be so widespread across the body outside of the CNS indicates that they have a variety of roles elsewhere that are currently unknown. Continued research into the other functions of mGlu receptors outside of the CNS could prove to be very important. The more that we know about these receptors and the ligands that they respond to, the easier it will be to design more efficient and specific drugs to combat associated diseases (Julio-Pieper et al. 2011).

GABAb

Another main subfamily of class C GPCRs are the GABAb receptors. These receptors are found mainly in the central nervous system (CNS) and can be seen both pre-synaptically and post-synaptically. The main role of GABAb receptors is to inhibit neurotransmission in the brain and CNS (Geng et al. 2013). Pre-synaptically, when activated by a ligand, these receptors will inhibit the action of calcium channels in the neuron, which in turn inhibits the release of neurotransmitters. Post-synaptically, an activated GABAb receptor will cause neuron hyperpolarisation by triggering the opening of potassium channels (Bettler et al. 2004).

As these receptors are involved with the brain and neurotransmission, when there are problems with GABAb receptors this can lead to a number of different neurological disorders (Geng et al. 2013). Examples of diseases and disorders that GABAb receptors are thought to play a role in include schizophrenia, epilepsy, spasticity, addiction, anxiety, depression and other neurological conditions. They are also known to play a role in cell survival, position of neurons and migration in the CNS (Xu et al. 2014). There is only one drug currently on the market that specifically targets GABAb receptors. It is a selective agonist for the receptor and is

used for the treatment of muscle spasticity in disorders such as cerebral palsy and multiple sclerosis and also for spinal cord injuries (Geng et al. 2013).

Like all class C receptors, the GABAb receptors are obligatory dimers (Geng et al. 2013). They can also be seen as tetramers and other oligomers with a higher number of subunits. However, the more subunits, the weaker the interaction between them seem to be (Xu et al. 2014). GABAb receptors are heterodimers so they have two different subunits that cause the activation of a signalling pathway (Geng et al., 2013). Studies have shown that there is only one single type of GABAb receptor even though more than one function is observed. The same type of GABAb receptor is seen in the CNS both pre and post-synaptically (Xu et al., 2014). The two subunits of GABAb receptors are known as GBR1 and GBR2 (Geng et al., 2013). These subunits do not function by themselves (Marshall and Foord 2010, Marshall 2007 and Bettler et al., 2004). They have different functions but both are required to be present and functional for successful activation. Both subunit's N-terminus regions are folded into the venus fly trap (VFT) module that is common among class C GPCR's. In the inactive state the VFT module on both subunits are in the open conformation. The ligand binds only to the subunit GBR1. It binds in the cleft of the VFT module and in the case of an agonist, causes the subunit to adopt the closed conformation. Antagonists bind to the same place but in that case they cause the VFT to remain in the open conformation thus leaving it inactive (Geng et al., 2013). The GBR2 subunit does not appear to bind to any known ligands, but even though the GBR2 subunit is not involved directly in the ligand binding it does help to increase the binding affinity for the ligand by stabilising the agonist bound closed conformation of the VFT module (Xu et al., 2014). The transmembrane region of the GBR2 subunit is also important as it binds to the

heterotrimeric g-protein inside the cell (Geng et al., 2013). Another function of the GBR2 subunit is seen in the binding of positive allosteric modulators (PAM) to its TM region. These interactions increase the effect of the bound agonist (Xu et al., 2014).

The GBR1 subunit alone cannot leave the endoplasmic reticulum (ER) and move to the cell membrane. This is because it has a very high ER retention signal. The C-terminal regions of the two subunits interact with each other to form a coiled-coil (see Figure 1.3 below). This interaction between the GBR1 and the GBR2 molecule masks this signal and the molecules are able to reach the cell surface membrane as a heterodimer. Therefore both subunits are required to be present for the correct function of the GABAb receptor to be carried out (Xu et al., 2014).

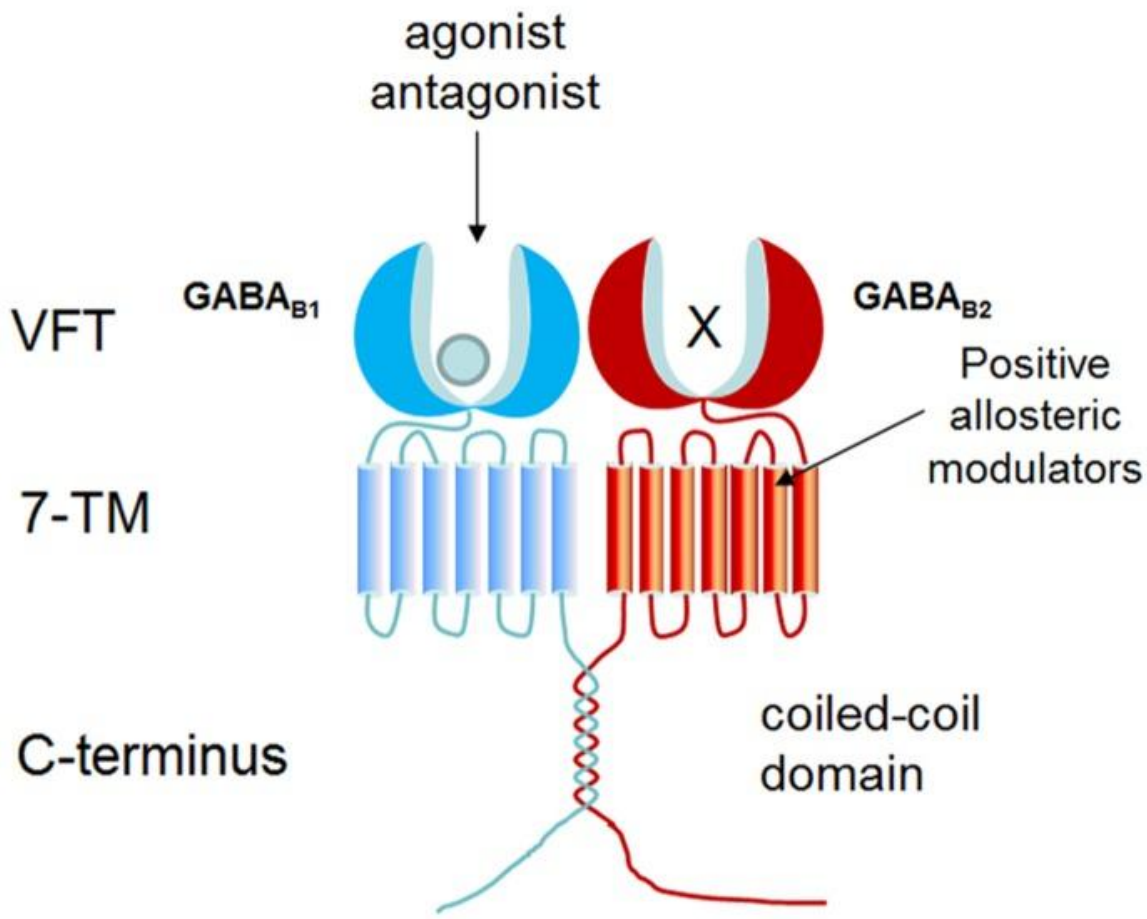

Figure 1.3: The structure of the GABAb dimer with the ligand bound to the GABAb1subunits VFT module. Taken from (Xu et al., 2014).

Although the GBR1 subunit oscillates between the open and closed conformation depending on if it is or isn't bound to an agonist ligand, the conformation of the GBR2 subunit is not significantly affected by the presence of a ligand. As this subunit does not interact directly with the ligand it remains in the open conformation regardless of whether or not a ligand is present. All of the ligands that bind to and activate the GABAb receptors are derivatives of the GABA molecule. As such, they all share the structure of a γ-amino acid. Once the agonist is in the binding pocket several chemical interactions are formed. The main ligand-

receptor interactions are hydrogen bonds but some Van der Waals forces are also in effect (Geng et al., 2013). Although there is only one type of GABAb receptor there are a number of isoforms that have been discovered. These isoforms may be the cause of any functional diversity seen in the receptors (Xu et al., 2014).

Calcium sensing

The third main subfamily of GPCR's in the class C family is the calcium sensing receptors (CASRs). The main role of the calcium sensing receptors is in calcium homeostasis. It is very important to maintain the level of calcium within a narrow range as small changes putting the level either side of this range can be life-threatening. These receptors are found all over the body but are of particular importance in the parathyroid gland. In the parathyroid gland the CASRs are responsible for regulating the synthesis and the release of parathyroid hormone (PTH). Parathyroid hormone is a calcium regulating hormone (Tfelt-Hansen and Brown, 2005).

In the kidneys, calcium sensing receptors further aid the regulation of calcium levels by regulating the excretion of calcium and influencing the movement of water and electrolytes to maintain the necessary range of calcium in the body. Other organs where the calcium receptors are important for signal transduction include the placenta, brain and the pancreas (Goodman 2004).

The importance of the role the calcium sensing receptors play in calcium homeostasis is seen in the consequences of naturally occurring mutations. When these mutations cause either loss or gain in function of the receptors a number of diseases are seen to occur. The severity of the disease depends on whether the mutation in the receptor is heterozygous or homozygous. Heterozygous mutations

are not as severe as homozygous mutations. For example, familial hypocalciuric hypercalcaemia (FHH) also known as familial benign hypocalciuric hypercalcaemia (FBHH) is the result of a heterozygous mutation and is often asymptomatic. A homozygous mutation causes neonatal severe primary hyperparathyroidism (NSHPT). Unlike FHH, this disease is quite severe and it can be fatal if left untreated. Gain of function mutations have been seen to cause autosomal dominant hypoparathyroidism (ADH) which generally results in asymptomatic hypocalcaemia (Tfelt-Hansen and Brown, 2005).

These receptors are not only found in the cells of the parathyroid gland; they have other functions when present in other types of cells. Some other roles that CASRs have include modulating the cell cycle and the regulation of ion channels etc. (Tfelt-Hansen and Brown, 2005).

As is common among class C GPCRs, CASRs have a very large extracellular domain folded into a venus fly trap conformation (Trivedi et al. 2008). It is believed that extracellular calcium ions interact with acidic portions of the binding domain allowing the receptor to measure the level of calcium. The exact binding locations in the extracellular domain are not known (Goodman 2004).

Activation of the CASRs occurs when the level of extracellular calcium either increases or decreases outside of the narrow normal range. An increase in the extracellular calcium causes a reduction in PTH release. This will in turn decrease the calcium levels. A decrease in the extracellular calcium causes inactivation of the CASR and triggers the release of PTH and therefore the calcium levels will also increase (Goodman 2004).

The extracellular domain in CASRs is so large that there is an issue of promiscuity. There are a number of sites that allow the binding of other charged molecules. This makes creating drugs with a high level of selectivity difficult (Trivedi et al. 2008).

CASRs are functional as homodimers (Tfelt-Hansen and Brown, 2005). The two subunits are linked by disulfide bridges arising from interactions between cysteine residues in the extracellular region of the receptor (Goodman 2004). Other non-covalent interactions also occur to strengthen the interaction between the two subunits (Tfelt-Hansen and Brown, 2005).

The function of the CASR means that they are attractive as drug targets for altering the function of the parathyroid gland (Goodman 2004). There have been drugs approved to help treat hyperparathyroidism (HPT) by targeting CASRs (Tfelt-Hansen and Brown, 2005). The approved drug is a calcimimetic drug. It lowers the levels of PTH and therefore extracellular calcium by activating the receptor. A potential drug for the treatment of osteoporosis also targets CASRs. These drugs are calcilytics and they stimulate PTH release under normal calcium levels. These have not yet been approved (Trivedi et al. 2008).

1.5. Class C GPCR role in Schizophrenia

Schizophrenia is one of the main diseases that has been directly linked to class C GPCRs (Foster and Goa, 1998). It is a disorder of the central nervous system that affects approximately 1% of the population worldwide (Vinson and Conn, 2012). The age of onset in patients that have schizophrenia is between the late teens to

mid 30's. There is no cure for schizophrenia and it is a chronic disease that is usually very disabling. It often involves extended hospitalisation of patients meaning it is a very expensive disease to treat (Foster and Goa, 1998).

The symptoms of schizophrenia are often split into three categories; the positive, the negative and the cognitive symptoms. Positive symptoms are when the patient perceives non-existent stimulation. This results in hallucinations that can be visual and/or audio and also paranoia. Negative symptoms of schizophrenia include depression and social withdrawal and finally cognitive symptoms may include disorganised thoughts and a reduced ability to process information. So far, the treatment available for schizophrenia has been much more effective on the positive symptoms than on the negative or cognitive symptoms. Therefore the search for new drugs should focus on treating all three groups of symptoms (Vinson and Conn, 2012).

Metabotropic glutamate receptors 5, 2 and 3 are seen to play an important role in schizophrenia and are potential targets for drugs to treat the disease. Activation of group II mGlu (mGlu2 and mGlu3) receptors by agonists is seen to reduce excitatory neurotransmission by decreasing glutamate release. The antipsychotic effects seen with these agonists make them a very promising drug lead. Selective activators of mGlu5 have also shown promising potential in the treatment of both positive and negative symptoms as well as showing cognitive enhancing effects (Vinson and Conn, 2012).

Many class C GPCRs have been associated with important diseases but only a few currently have approved drugs available that target them. There is not as much known about this class of GPCRs as there is, for example, with class A.

Further increasing our knowledge of the structure and function of class C will make finding a suitable drug much easier and mean that it is likely to be a much more efficient drug with fewer side effects.

Chapter 2

An approach to identifying the optimal helical range for alignment

2.1 Introduction

Several studies have sought to generate homology models of GPCRs based on the structure of a GPCR in a different class where the alignment is not straightforward (Bissantz et al., 2004; Conner et al., 2005; Coopman et al., 2011; deGraaf et al., 2011; Dong et al., 2007; Donnelly, 1997; Frimurer and Bywater, 1999; Sheikh et al., 1999; Taylor et al., 2003). For example, Vohra et al. (2013) have showed that there is agreement on the class A - class B GPCR alignment when the receptors share conserved aromatic residues in a given helix but not otherwise. These studies carried out an explicit alignment on each helix. However, in the absence of structure, it is not clear which residues are helical – if the helix is too short, there may not be enough information to determine the alignment but if it is too long, the lack of a common secondary structure may contaminate the alignment results. One approach that can be taken when there are a number of structures of a given class, e.g. class A is to determine the consensus from a range of structures. Thus, the program used to align the helices from different classes of G-protein coupled receptor (GPCR) requires a range of residues thought to be in the helical region of the protein. The more accurate this region is, the more accurate the results will be from the analysis program. Here we explore a new approach to identifying this common helical region.

Half sphere exposure, HSE, (Hamelryck 2005) measures how buried or exposed a particular residue in a protein sequence is. The method is based on counting the number of neighbouring residues within a half-sphere of radius 13Å centred on the Cα carbon atom of the amino acid, as shown in Figure 2.1.

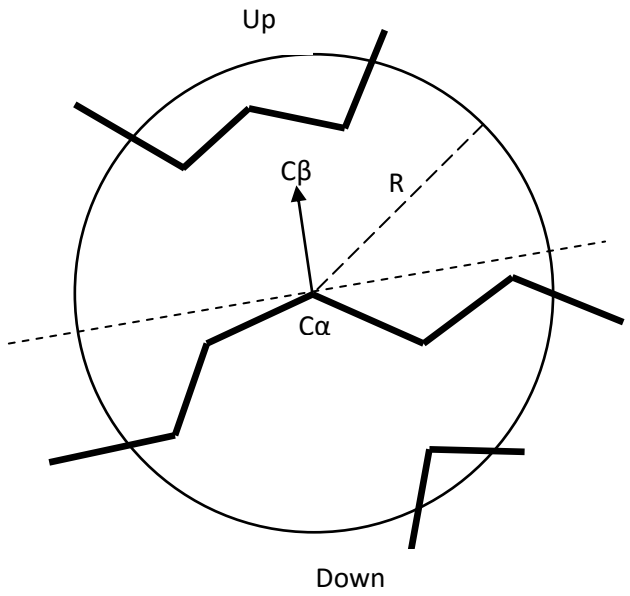

Figure 2.1: A diagram to illustrate HSE adapted from Hamelryck (2005). A sphere with radius of 13 Å surrounding the Cα atom of the amino acid is divided in two, perpendicular to the Cα-Cβ vector, and the residues within the half sphere are counted. Since this is a count of residues, there are no units.

The residues of an alpha helical structure will all follow a similar pattern for each transmembrane (TM) region for GPCR's, therefore, measuring the HSE is useful when trying to determine the alpha helical regions of the GPCR. When a number of proteins have been analysed the results can be plotted to give a good

prediction of the common areas of alpha helical structure across the range of GPCR classes.

For each transmembrane region, a figure was produced to show the HSE results for a number of GPCR's. Namely the beta1-Adrenergic Receptor (2VT4 - chain A) (Warne et al. 2008); Rhodopsin (1F88) (Palczewski et al. 2000); A2A Adenosine Receptor (3EML) (Jaakola et al. 2008); Protease activated receptor 1 (3VW7) (Zhang et al. 2012); CXCR4 Chemokine Receptor (3ODU) (Wu et al. 2010); Smoothened Receptor (4JKV) (Wang et al. 2013); Glucagon Receptor (4L6R) (Siu et al. 2013); Corticotropin-Releasing Factor Receptor 1 (4K5Y) (Hollenstein et al. 2013). The area where the pattern of results matches well is a good indication of where the common helical regions are.

2.2 Methods

The half-sphere exposure, HSE was determined using the method of Hamelryck (2005). Eight GPCRs were chosen as representative structures to test the method, namely receptors with the following PDB codes: 2VT4, 1F88, 3EML, 3VW7, 3ODU, 4JKV, 4L6R and 4K5Y. These include structures from class A (2VT4, 1F88, 3EML, 3VW7, 3ODU), class B (4K5Y, 4L6R) and class F (4JKV). Class A, class B and class F GPCRs are remote homologues sharing ~10-18% identity (Taddese et al. 2014) and so cannot be easily aligned using traditional sequence alignment. The alignment was therefore taken from Taddese et al. (2014), which was generated by structural alignment using the salign module of Modeller (Eswar et al. 2007) and also given in the original GPCR X-ray crystal structure publications

(Siu et al. 2013 and Wang et al. 2013). This alignment was used to align the HSE results.

2.3. Results and Discussion

2.3.1. TM1

Figure 2.2 shows the HSE results for the TM1 region of the 8 GPCR's. The further away from the alpha helical TM1 region, the less the data matches up. The area in which the data from the different GPCR's matches very well will be the predicted TM1 region used for alignment. The figure shows that the TM1 region is likely to be from residue 38-64 (as numbered in the figure). The old range of residues being used was from 43-65 (Vohra et al. 2012 and Taddese et al. 2014). This new window increases the region by 4 residues. This increase will give more data for use in the next stage and should result in a more accurate alignment. In addition, the top values of each peak and the bottom of each peak are fairly consistent for this helix. This is in contrast to TM3 in which the peaks can vary greatly from each other. The consistency in the peaks shows that the environment is fairly consistent across the whole of the helix and that the helix is similarly exposed across all the residues.

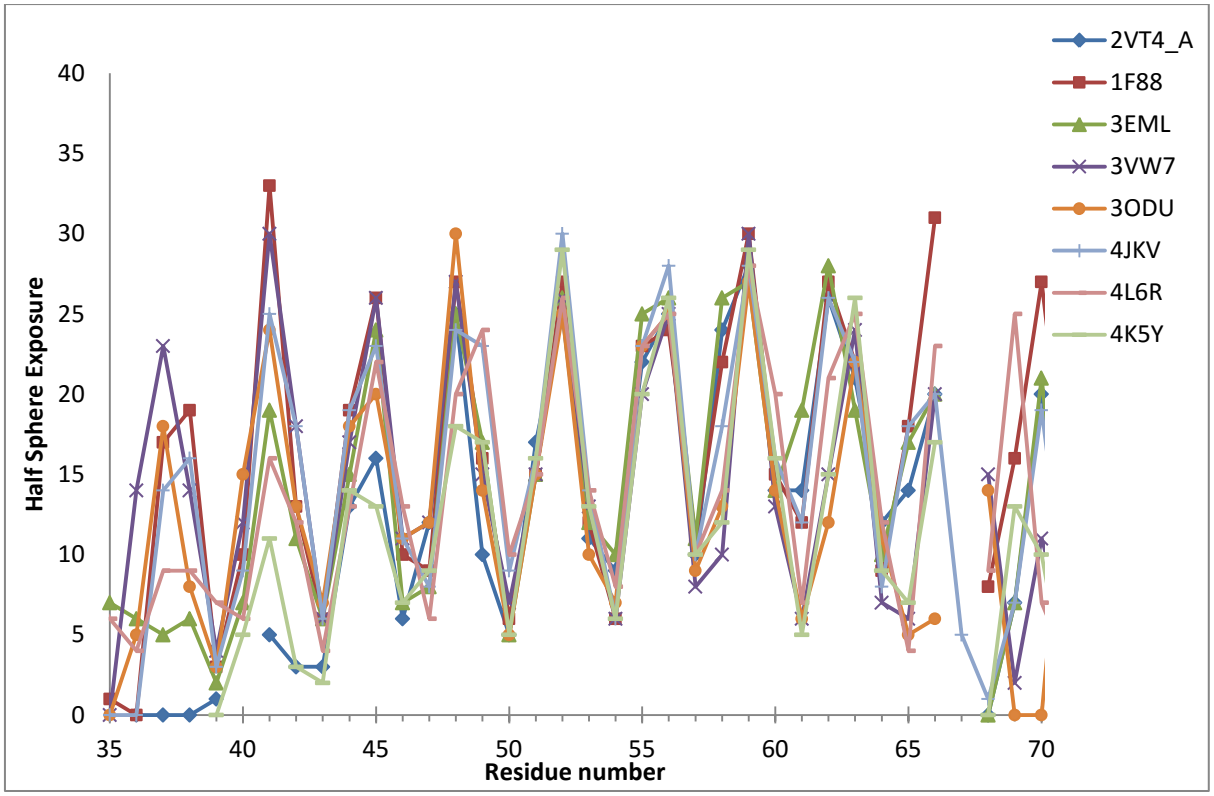

Figure 2.2: Half Sphere Exposure results for TM1. The low HSE regions correspond to exposed residues; high HSE regions correspond to the buried (internal) residues. The class A GN motif is at positions 55-56.

2.3.2. TM2

Figure 2.3 shows the HSE results from the TM2 region. The figure shows that the area between residues 72-92 match up very well and therefore this is the new predicted TM2 alpha helical region. The old region was between residues 79-95 (Vohra et al. 2012 and Taddese et al. 2014), so this shows an increase of 4 residues as well as a shift of 3 residues. The increase in window length increases the data available for use in the alignment especially in the lower residue numbers. The shift of 3 residues removes the area in the higher residue numbers where the pattern does not match and should improve the accuracy by not including the areas of the GPCR where the structures start to differ. The peptide receptors (3ODU –

CXCR4 Chemokine receptor and 3VW7 – Protease activated receptor 1) contain a bulge in TM2 (Kant and Vriend 2014), which should result in a gap / insertion in the alignment. This bulge does not affect the HSE results as all receptors for this region match the pattern, regardless as to whether they have a bulge or not. This indicates that the HSE results may make one contribution towards improving the alignment, but it does not address other issues such as bulges in alpha helices.

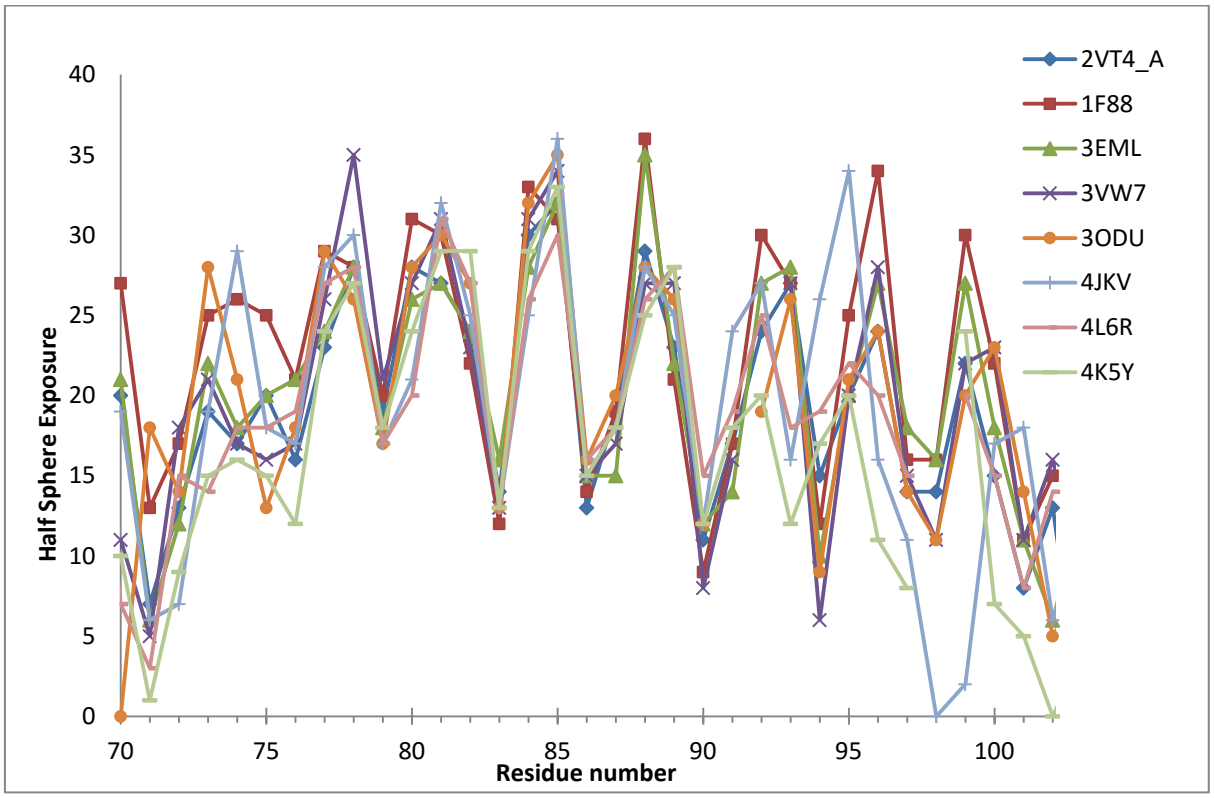

Figure 2.3: Half Sphere Exposure results for TM2. The low HSE regions correspond to exposed residues; high HSE regions correspond to the buried (internal) residues. The class A LxxxD motif is residues 81-85.

2.3.3. TM3

Figure 2.4 shows the HSE results of the TM3 region. These results do not line up quite as clearly in this TM region. This is likely due to the TM3 helix passing behind TM4 making all the residues in this region (~132-142) buried, rather than following

the typical alternating pattern of exposed and buried residues. The HSE results for this region are all high which confirms that these residues are buried so although the pattern does not line up as neatly, the results for these residues are consistent with the structure. In contrast to TM1, the high points and low points of the peaks vary greatly from each other as the helix passes behind TM4 and back out again. The region that matches up the best on Figure 4 is between residues 123-155. This is an increase of 6 residues from the old region which was between residues 125-151 (Vohra et al. 2012 and Taddese et al. 2014). This increase in the window length should allow for a more accurate alignment even though the results in the middle of the helix do not line up as well.

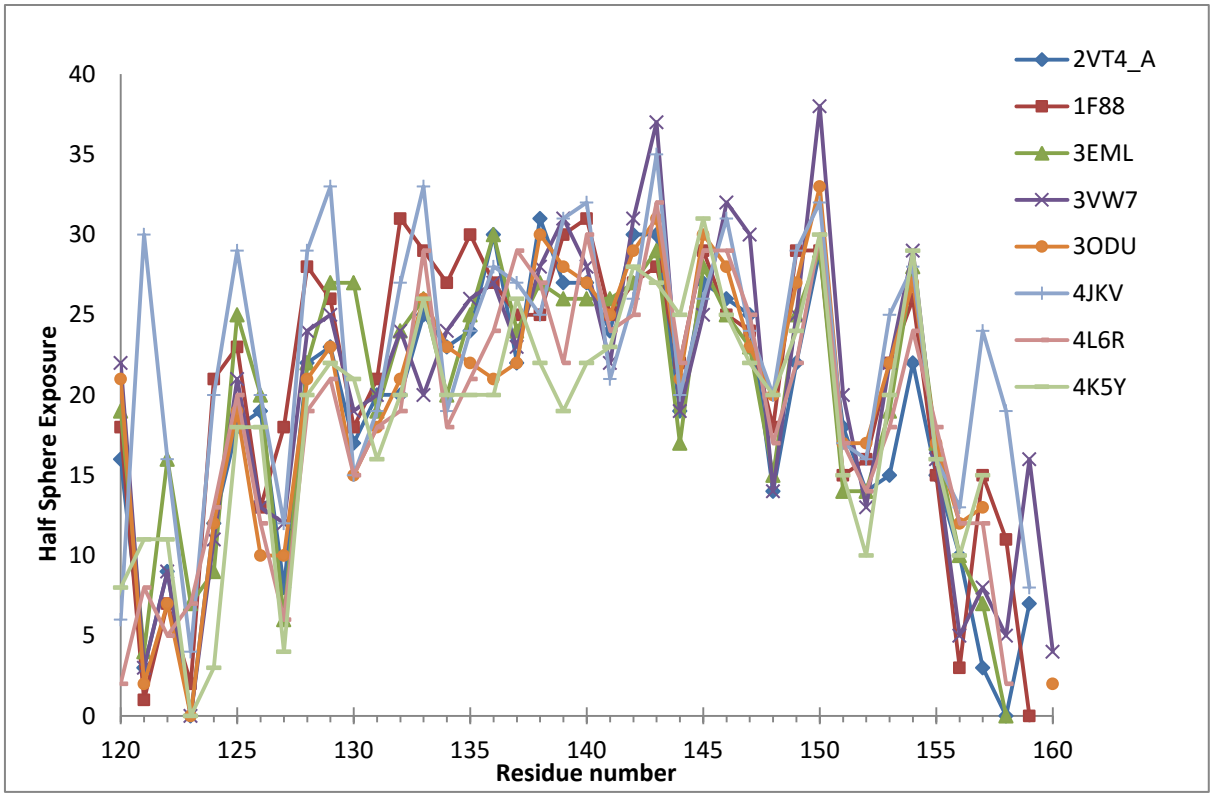

Figure 2.4: Half Sphere Exposure results for TM3. The low HSE regions correspond to exposed residues; high HSE regions correspond to the buried (internal) residues. The class A DRY motif is residues 149-151.

2.3.4. TM4

Figure 2.5 shows the results from the HSE analysis of the TM4 region. The section of this graph that lines up the best is between residues 173-191. This new window is a 1 residue increase from the old window of 169-186 (Vohra et al. 2012 and Taddese et al. 2014). It also shifts the analysis window of TM4 over by 4 residues which should increase the accuracy of the alignment. Since the analysis, a newer structure has been released for 4L6R by a different group. The structure for TM4 has some errors that are fixed in the newer version. However, this does not seem to affect the HSE results as seen in the figure below.

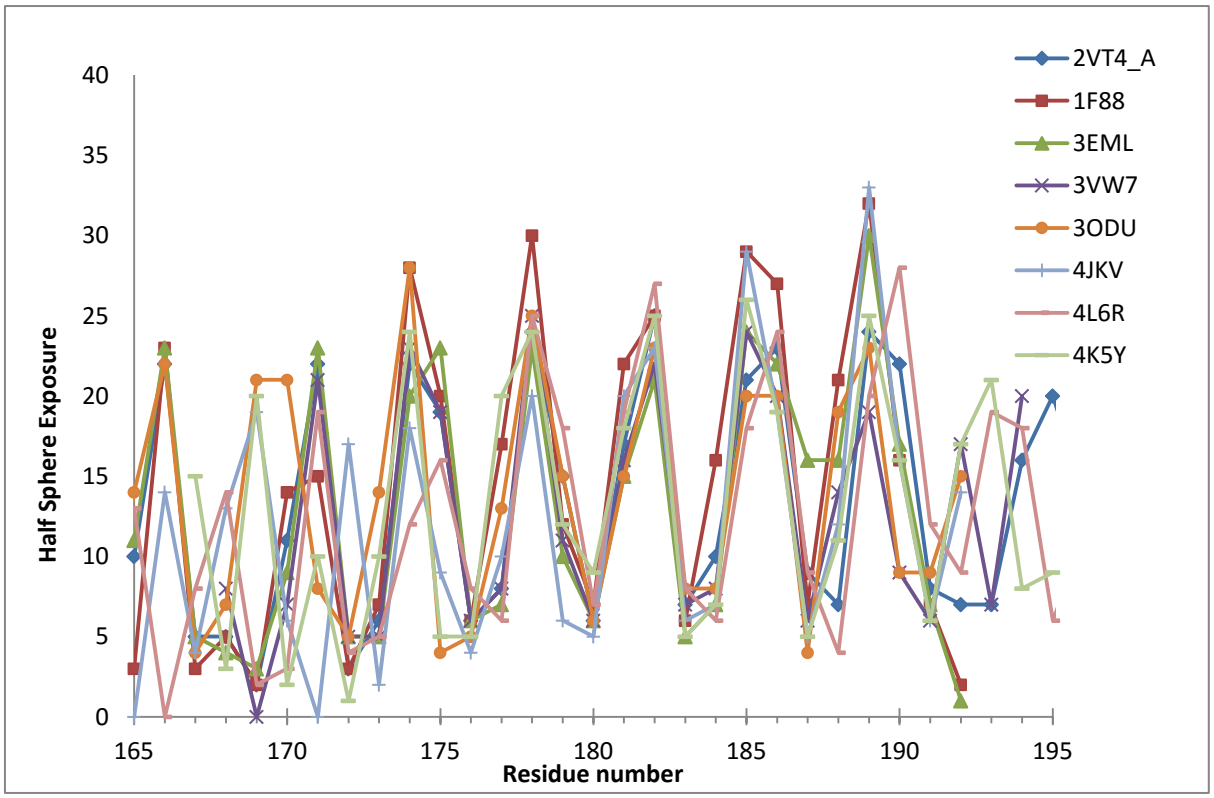

Figure 2.5: Half Sphere Exposure results for TM4. The low HSE regions correspond to exposed residues; high HSE regions correspond to the buried (internal) residues. The class A W motif is residue 179.

.

2.3.5. TM5

Figure 2.6 shows the HSE results from the TM5 region of the 8 GPCR's tested. There is a very clear region in which the results line up. This is between residues 233-258 which is a 5 residue increase from the old window of 239-259 (Vohra et al. 2012 and Taddese et al. 2014). This 5 residue increase in the size of the window gives more data for the analysis of the alignment and should lead to more accurate results. However, the published structural alignment for TM5 for the class A – class B and class A – class F alignment (Siu et al. 2013 and Wang et al. 2013) indicates gaps in the vicinity of the proline, creating a situation similar to that arising through the bulge in TM2, indicating that additional improvements, besides those arising from HSE may be required to improve the reliability of the alignment, but these are beyond the scope of this work.

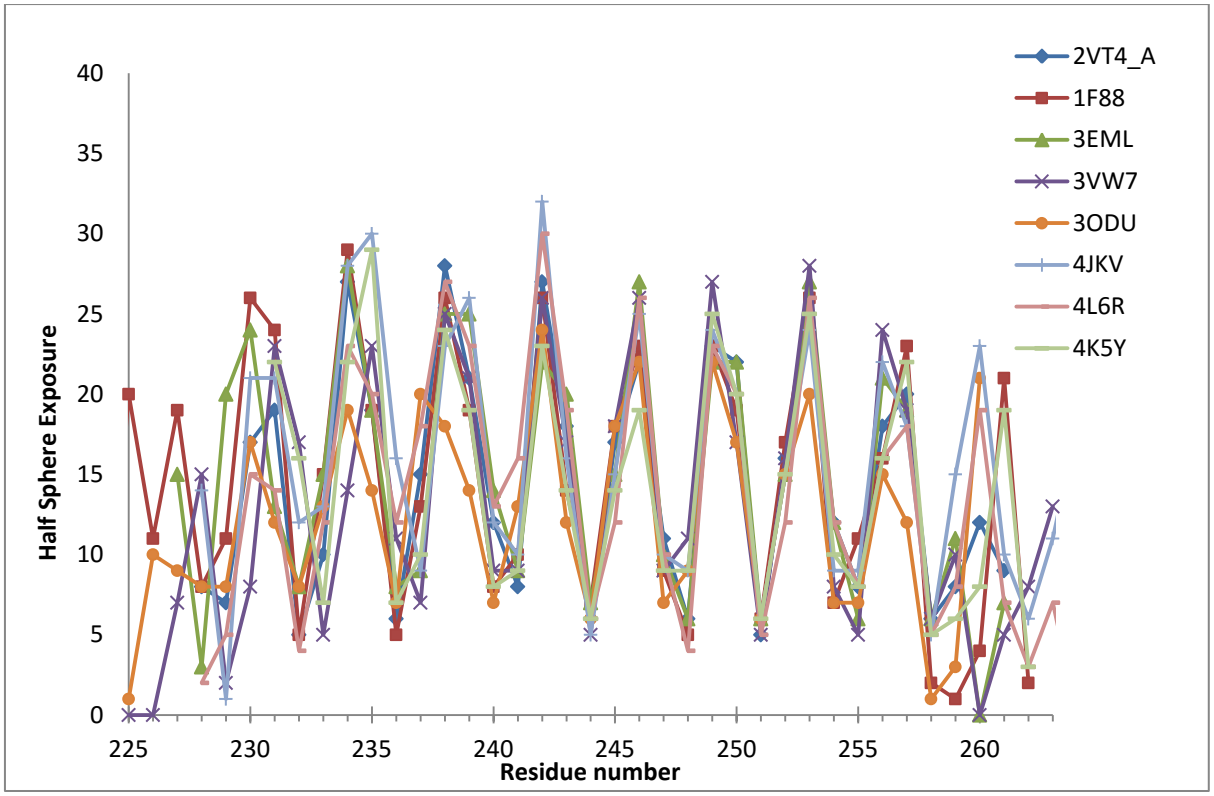

Figure 2.6: Half Sphere Exposure results for TM5. The low HSE regions correspond to exposed residues; high HSE regions correspond to the buried (internal) residues. The class A FXXP motif is residues 239-242.

2.3.6. TM6

Figure 2.7 shows the HSE results from TM6. The graph shows that the results between residues 286-304 match up the best, making this the new predicted TM6 alpha helical region. The only receptor that does not fit well with the pattern in the whole of this region is 4K5Y. This divergence from the pattern is due to the bend in the top of TM6 and TM7 for this receptor so the results from this section are expected and the new window should not be affected by it (Hollenstein et al. 2013). The original region used was between residues 284-305 (Vohra et al. 2012 and Taddese et al. 2014), so this shows a decrease by 3 residues. Although this is a smaller window and therefore uses less data for the alignment, it should still provide

a more accurate result as it does not include results from areas that do not match in structure as well that may be affecting the quality of the alignment.

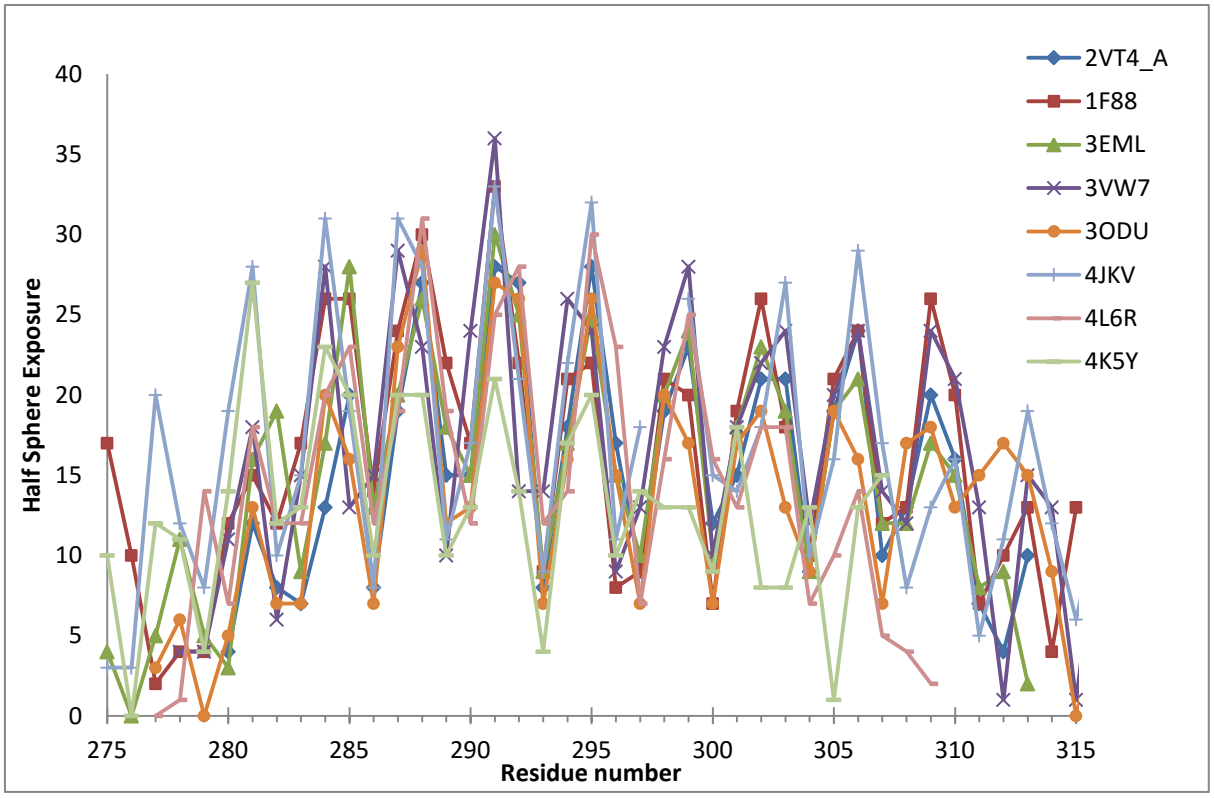

Figure 2.7: Half Sphere Exposure results for TM6. The low HSE regions correspond to exposed residues; high HSE regions correspond to the buried (internal) residues. The class A CWLP motif is residues 301-304.

2.3.7. TM7

Figure 2.8 shows the HSE results from the TM7 region of 8 GPCR's. Some of the GPCR results do not match up well at the beginning of this region but generally, between residues 337-358 has the best match between most of the receptors. This new window is the same length as the old window between residues 336-357 (Vohra et al. 2012 and Taddese et al. 2014), but shifted by 1 residue. This shift should provide a more accurate region for the alignment.

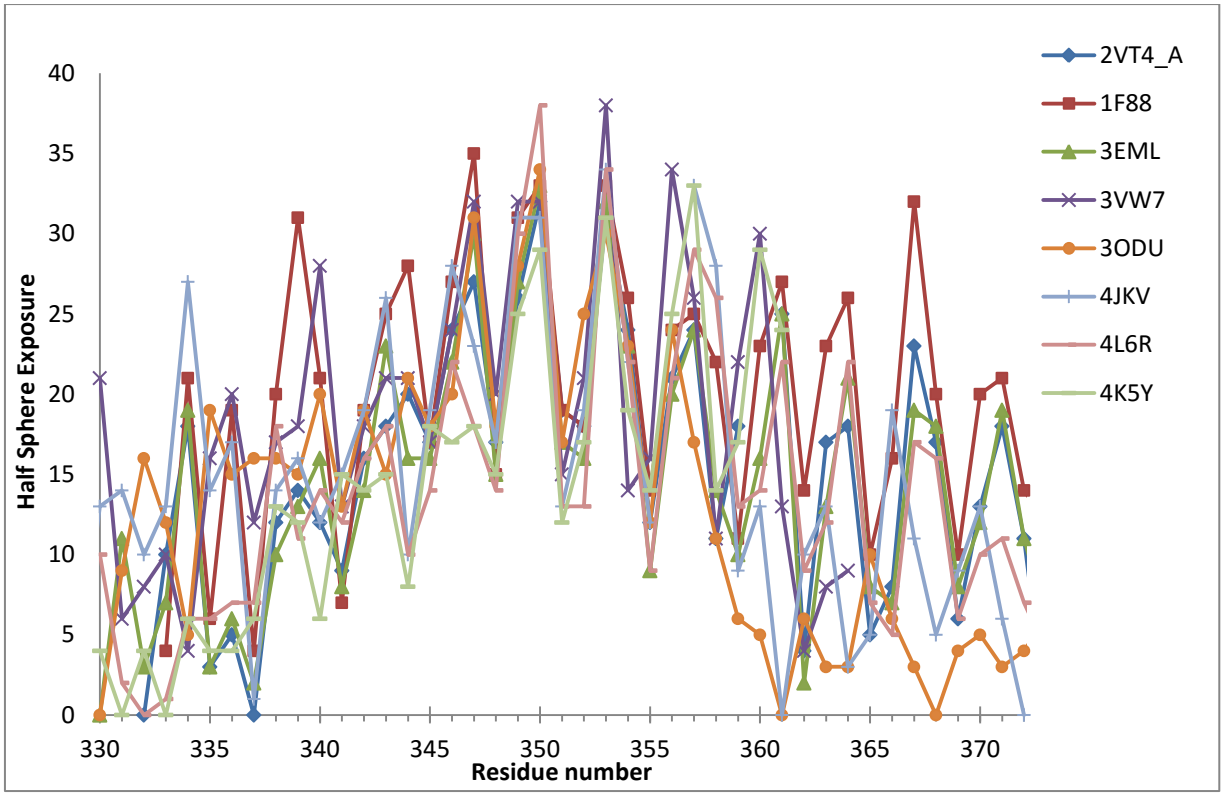

Figure 8: Half Sphere Exposure results for TM7. The low HSE regions correspond to exposed residues; high HSE regions correspond to the buried (internal) residues. The class A NPXXY motif is residues 353-357.

2.4. Conclusions

HSE analysis has shown clear patterns for the alpha helical transmembrane regions. The extra information provided on the environments and positions of the residues should improve accuracy. The results have shown changes are needed in the windows of all of the TM regions over those used in previous publications (Vohra et al. 2013 and Taddese et al. 2014). Some are small changes such as a simple 1 residue shift for TM7, but even this should mean clearer and more accurate results in the alignment. The new helix positions will be investigated in the next chapter for the purpose of predicting the alignment of class C to class A, class B and class F.

Chapter 3

Alignment of the new optimal helical ranges of Class C GPCRs with other classes

3.1. Introduction

Aligning different classes of G-protein-coupled receptors is a difficult task due to the lack of homology between the different classes. There have been many methods developed over the years to generate an alignment between class A and B GPCRs (Bissantz et al., 2004; Conner et al., 2005; Coopman et al., 2011; deGraaf et al., 2011; Dong et al., 2007; Donnelly, 1997; Frimurer and Bywater, 1999; Sheikh et al., 1999; Taylor et al., 2003). The method outlined by Vohra et al. (2013) to align class A and B GPCRs is the basis for the method used below. For this method to be as accurate as possible it is important to have the best window for the alpha helical regions. The optimal helical regions found from the half sphere exposure analysis in Chapter 2 have been used for the alignment of class C with GPCRs from classes A, B, E and F.

3.2. Methods

The method outlined by Vohra et al. (2013) has been modified to carry out the alignment between class C and the other classes of GPCR. This method compared ungapped alignments of defined helical regions (in this case the windows were determined using half sphere exposure data). An ungapped pair-wise alignment was then carried out to compare two helix sequences at a time and to score how well

they align at each position; one sequence of the pair was taken from say class A and the other from class C. The data are then collected, each top scoring alignment from the pair-wise alignments is noted and the alignment that has the highest number of votes is considered to be the best.

The method scores each pair-wise alignment based on a number of different criteria over the residues in the sliding window. The BLOSUM matrix is a substitution matrix created by Henikoff and Henikoff (1992) that scores alignments based on similarity of amino acids, as judged by their propensity to substitute for each other at given positions within a multiple sequence alignment.

The hydrophobicity of the residues was also measured using the Kyte-Doolittle hydropathy scores (Kyte and Doolittle, 1982). The best pair wise alignment is determined by a linear regression between the hydrophobicity values in the given sliding window on say the class A helix and those on the class C helix: the highest R value indicates the best alignment.

Other properties that were compared and scored are the amino acid volume, entropy and variability; which are also used via linear regression. Entropy is a measure of conservation. The more amino acids there are at a given position, the higher the entropy score. The variability is also a measure of conservation but within very similar subsets. It can be useful to know the areas where the variability is particularly high or low in similar sequences as this can give an indication of the areas that are the most important for function. The entropy and variability for TM1 of the class C GPCRs is shown in Figure A1 in the appendix. Each property was scored at each alignment and scaled between 0 and 1 and a product of all of these scores was used to determine the best overall alignment. The advantage of scaling between 0 and 1 is that the overall score for low scoring alignments tends to

decrease and the overall score for high scoring alignments tends to increase. Because of this, it is not necessary to identify the best alignment as the highest scoring one, but rather it just needs to be a high scoring alignment over all the measures used in the product. Further products were found excluding one or more of the properties to check that an extreme outlier was not skewing the results. These results were plotted on two graphs to show the best alignment for class C with each of the other classes.

3.3. Results and Discussion

3.3.1. TM1

3.3.1.1. Class A – class C

Figure 1 shows the product scores for the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the alignment of class C with class A receptors. The Figure also shows the product of the alignment scores with K, Vo and E removed one at a time. This is to see if one of the criteria is much lower for an alignment and may be affecting the products of the results of the best alignment. There is a percentage identity of 13.8 between these two classes at the optimal alignment. Percentage identity is important in the quality of the alignment. The higher the % ID the more accurate the alignment will be.

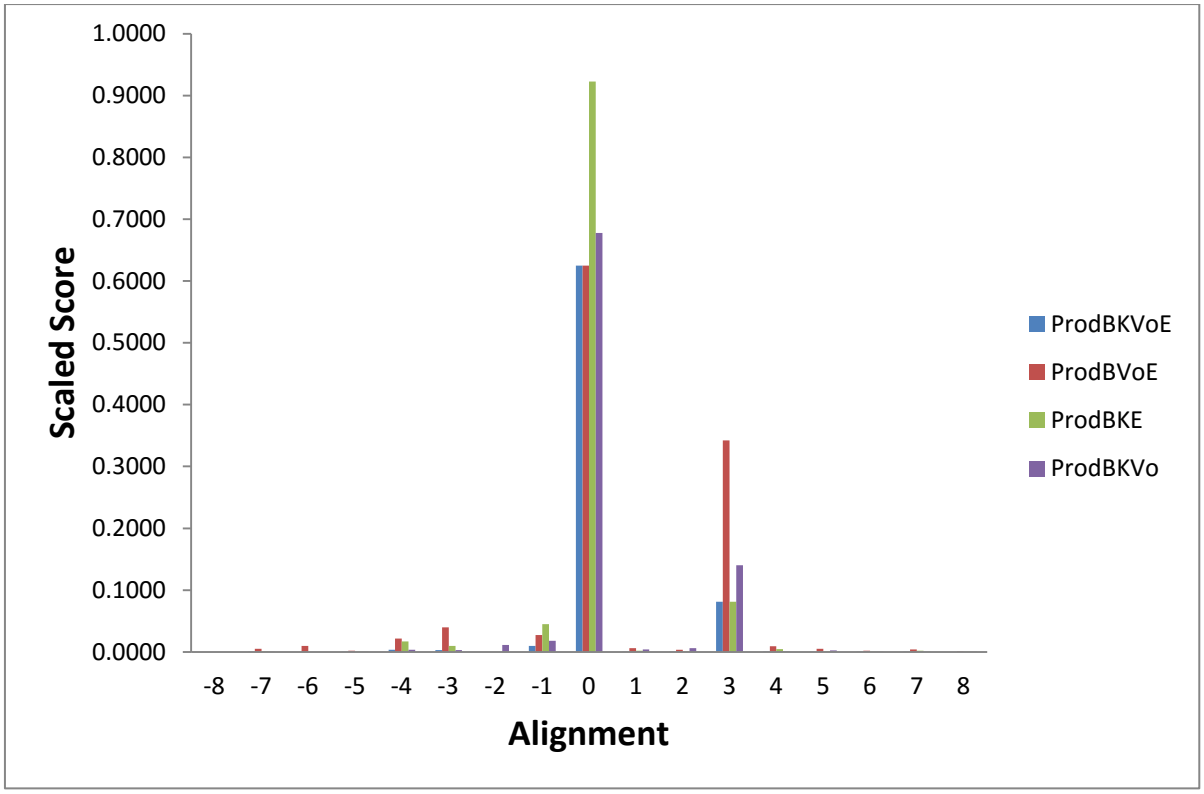

Figure 1: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the A-C TM1 region alignment.

Figure 2 shows further products of B, K, Vo, and E along with variability (Va) rather than entropy as a measure of conservation that are measured to give a further look into the accuracy of the alignment between class A and class C GPCRs. The results seen in Figures 1 and 2, show that the best alignment between classes A and C is the 0 alignment with a second best alignment of +3.

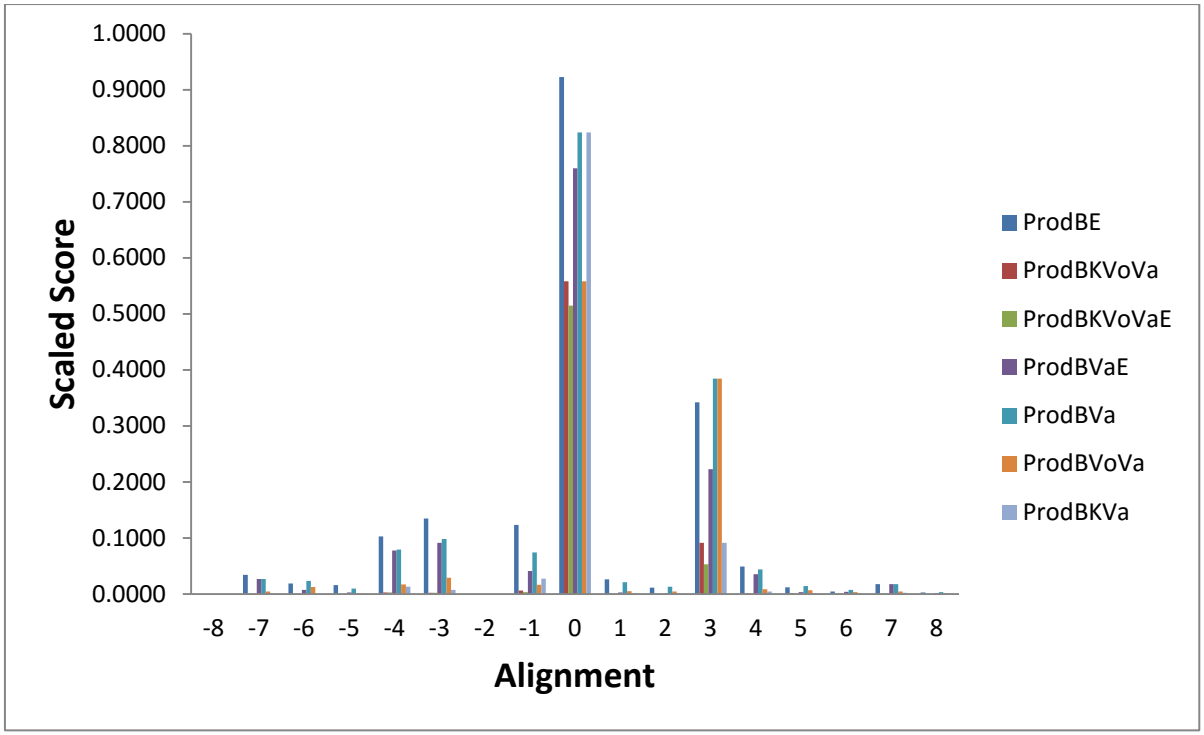

Figure 2: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the A-C TM1 alignment.

3.3.1.2. Class B – class C

Figure 3 shows the product scores for the alignment of class C with class B receptors. There is a percentage identity of 11.7 between these two classes in the 0 alignment.

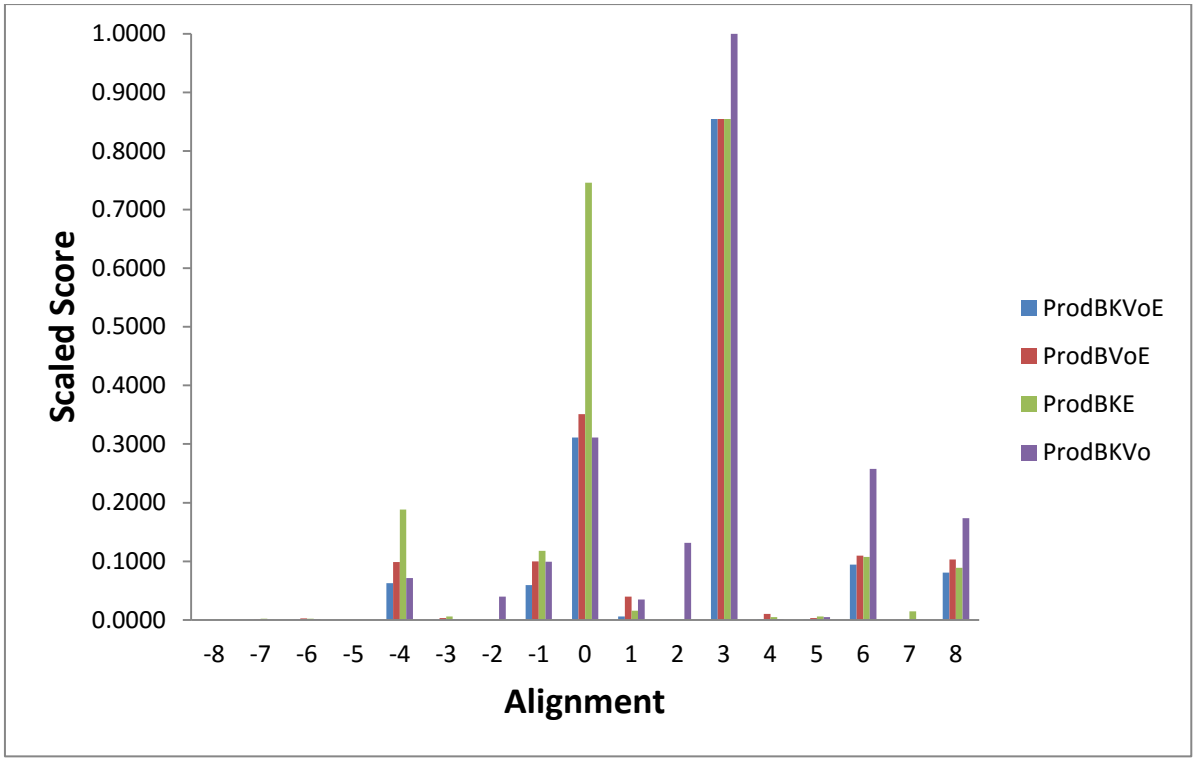

Figure 3: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the B-C TM1 region alignment.

Figure 4 shows the products of B, K, Vo, E and Va for the B-C alignment. The results in Figures 3 and 4 show the +3 alignment to be the best with a second best alignment of 0. This is the opposite to the results of the A-C alignment.

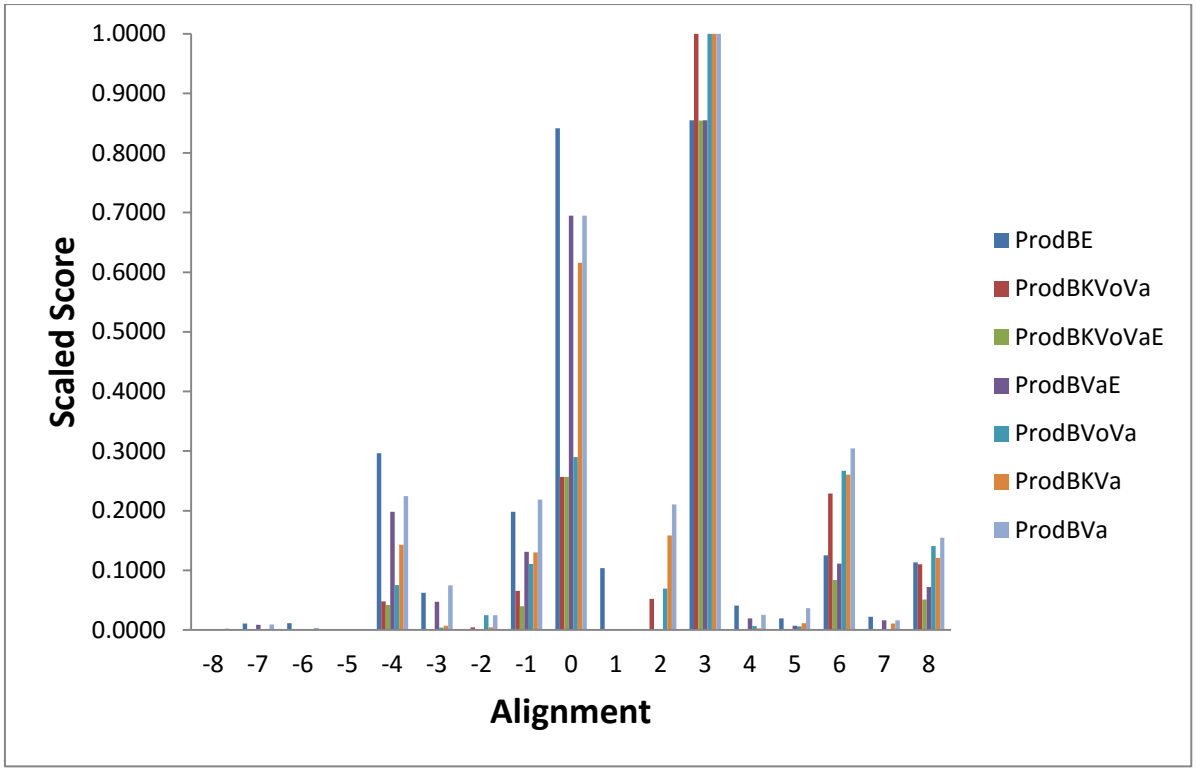

Figure 4: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the B-C TM1 alignment.

3.3.1.3. Class E – Class C

Figure 5 shows the products of the scores of B, K, Vo and E for the alignment of the TM1 region of E-C. There is a percentage identity of 14.5 between these two classes for the 0 alignment.

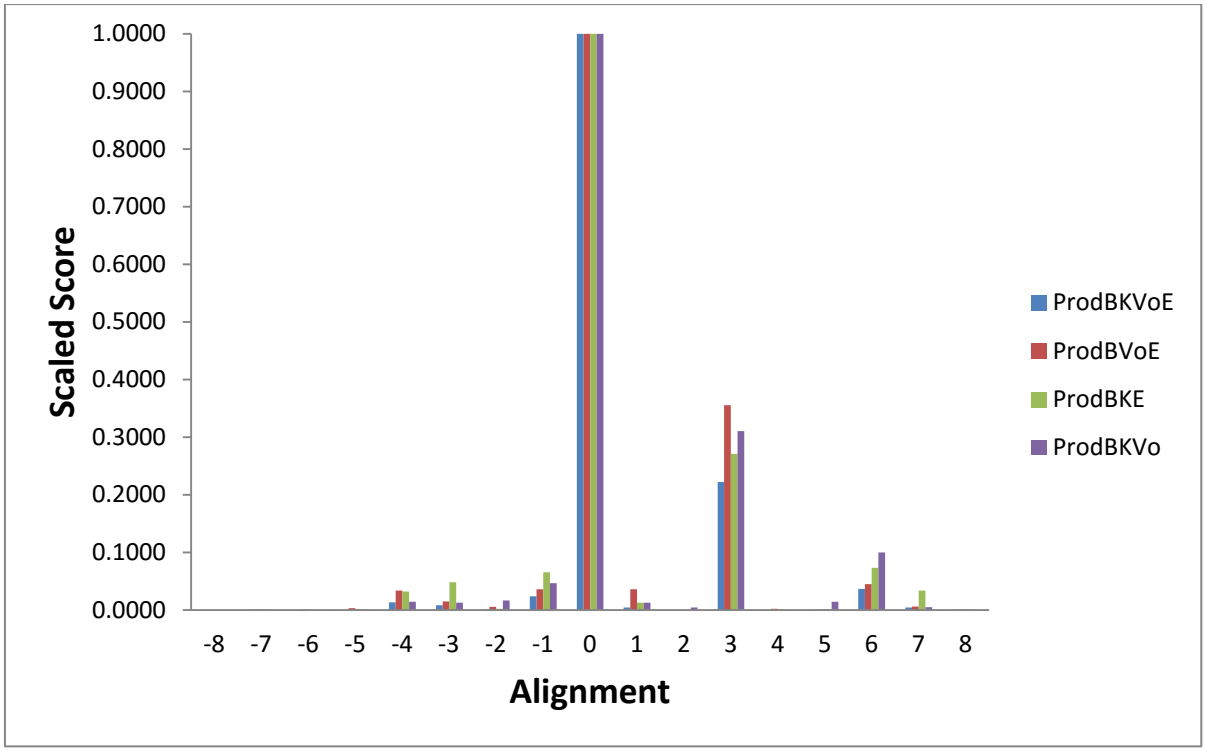

Figure 5: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the E-C TM1 region alignment.

Figure 6 shows the products of B, K, Vo, E and Va for the B-C alignment. Figures 5 and 6 show the 0 alignment to clearly be the preferred option, with a second smaller peak at the +3 alignment. This result is in agreement with A-C for this TM region.

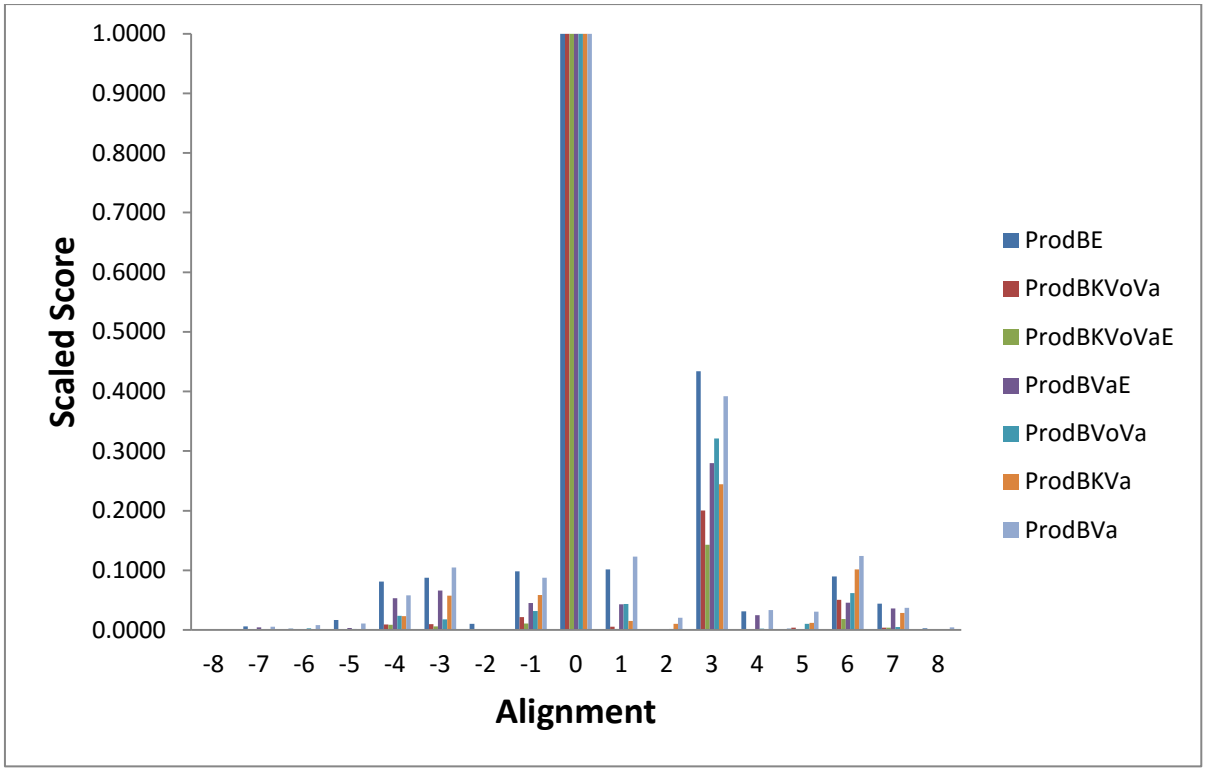

Figure 6: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the E-C TM1 alignment.

3.3.1.4. Class F – class C

Figure 7 shows the product scores for B, K, Vo and E for the F-C alignments of TM1.There is a percentage identity of 15.7 between these two classes for the 0 alignment. From this Figure we can see that the K score does not agree with the others for the best alignment. When the K score is removed from the product the alignment score for the 0 alignment goes up significantly higher than all the other results even though with it included it looks as though +3 is the best alignment.

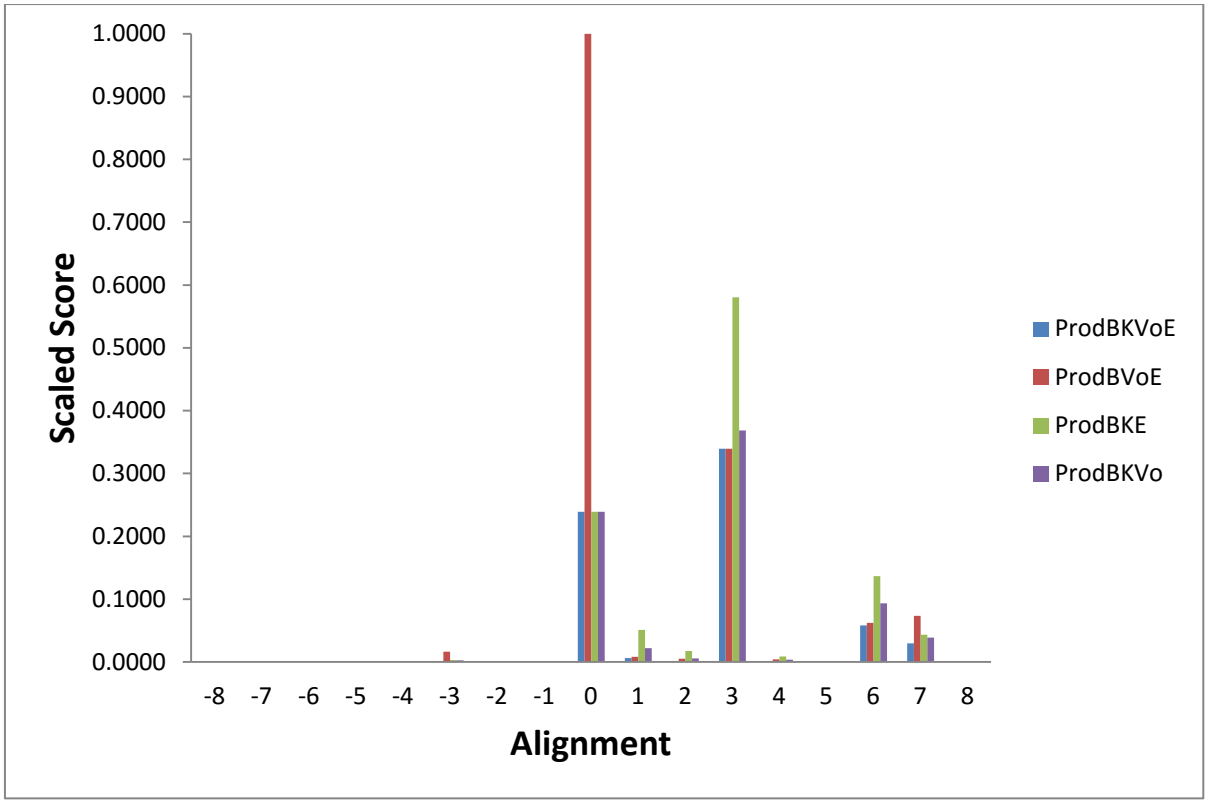

Figure 7: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the F-C TM1 region alignment.

Figure 8 shows the products of B, K, Vo, E and Va for the F-C alignment. Again the products that do not include the K values give the 0 alignment significantly higher results. If the K result is included both figures 7 and 8 show the +3 alignment to be the best but without they show the 0 to be the best alignment.

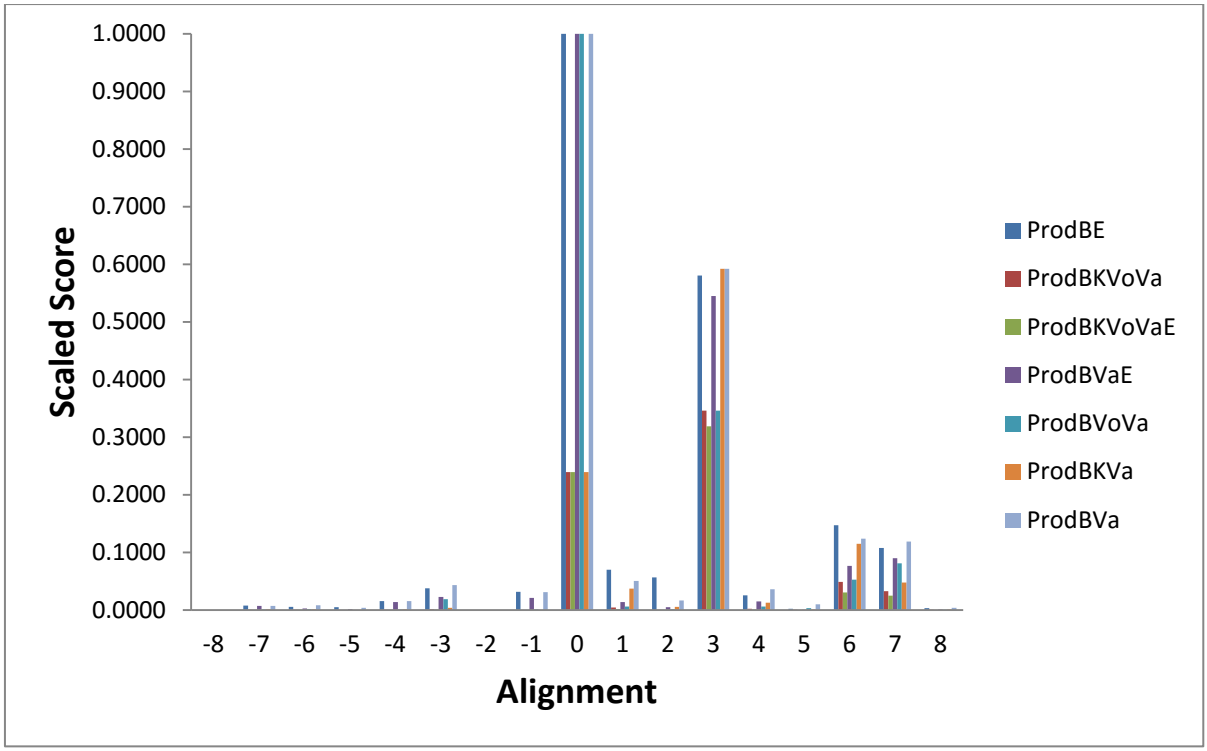

Figure 8: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the F-C TM1 alignment.

The 0 alignment is given in Chart 1.

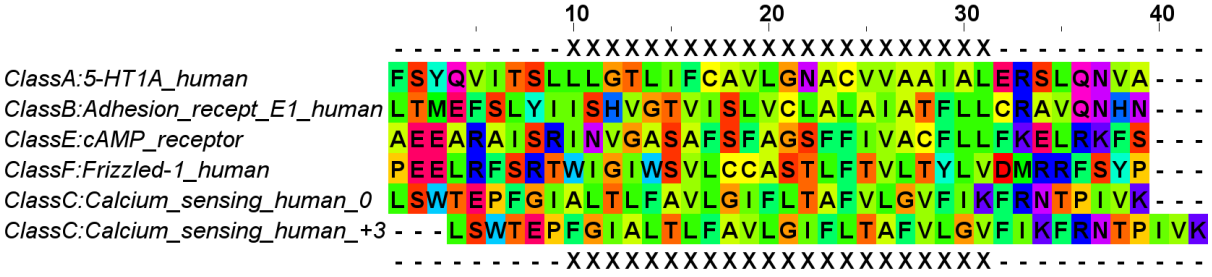

Chart 1: The alignment of the class C TM1 against TM1 for class A, class B, class E and class F. The class A - class B - class E - class F alignment is defined by the structural alignment of the relevant X-ray crystal structures. The correct 0 alignment (as defined by structural alignment) is given, along with the alternative of +3 indicated by Figures 1, 4, 5, 7 and 8. The Xs mark the window over which the alignment was determined. The colour scheme for the residues is the default Taylor colour scheme used by Jalview.

Analysis of the figures in Tadesse et al., (2014), similar to Figures 1-8 here, shows that the most reliable method for determining the correct alignment is the Blosum scores, with entropy similar to variability and hydrophobicity and volume less reliable. Consequently, in Chart 2 we show that a composite score can be obtained for the alignment of class C to class A, class B, class E and class F by scaling the 4 blosum scores given in Figures 1, 3, 5 and 7 between 0 and 1 and multiplying together. The resulting composite multi-reference alignment score is given in Chart 2 and this confirms that the method gives the correct result for TM1, since the major peak is alignment 0, in both bar charts. This shows that the method has given the correct results for TM1, despite the low percentage identity of 13.0 – 15.7, which is well within the twilight zone.

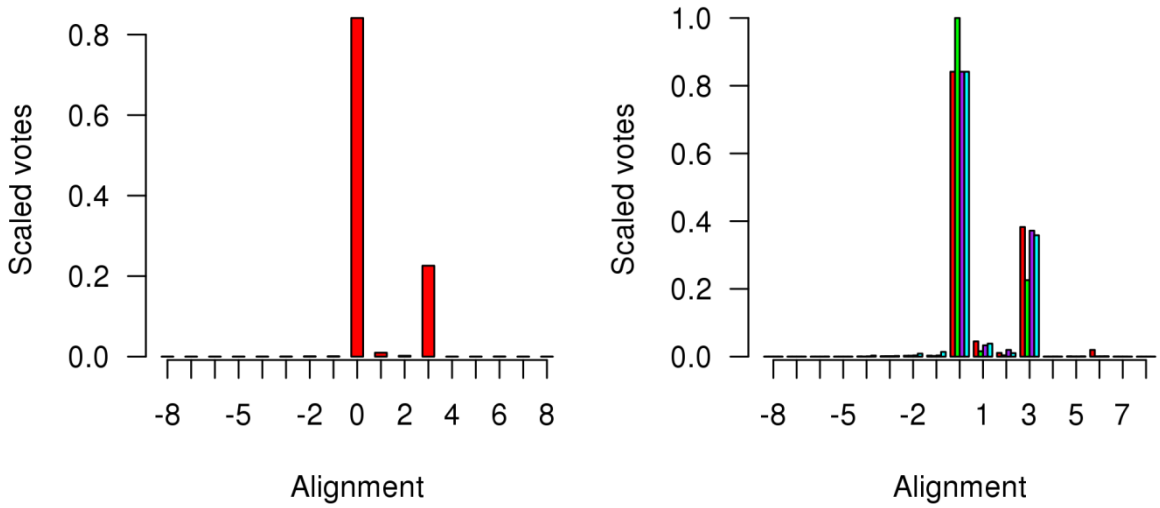

Chart 2: The multi-reference alignment scores. On the left the score is given for multiplying the Class A, class B, class E and class F scores together; on the right the score is given for missing out class A (red), class B (green), class E (purple) and class F (cyan).

3.3.2.TM2

3.3.2.1. Class A – class C

Figure 9 shows the product scores for the alignment of A-C of the TM2 region. There is a percentage identity of 11.4 between these two classes for the 0 alignment.

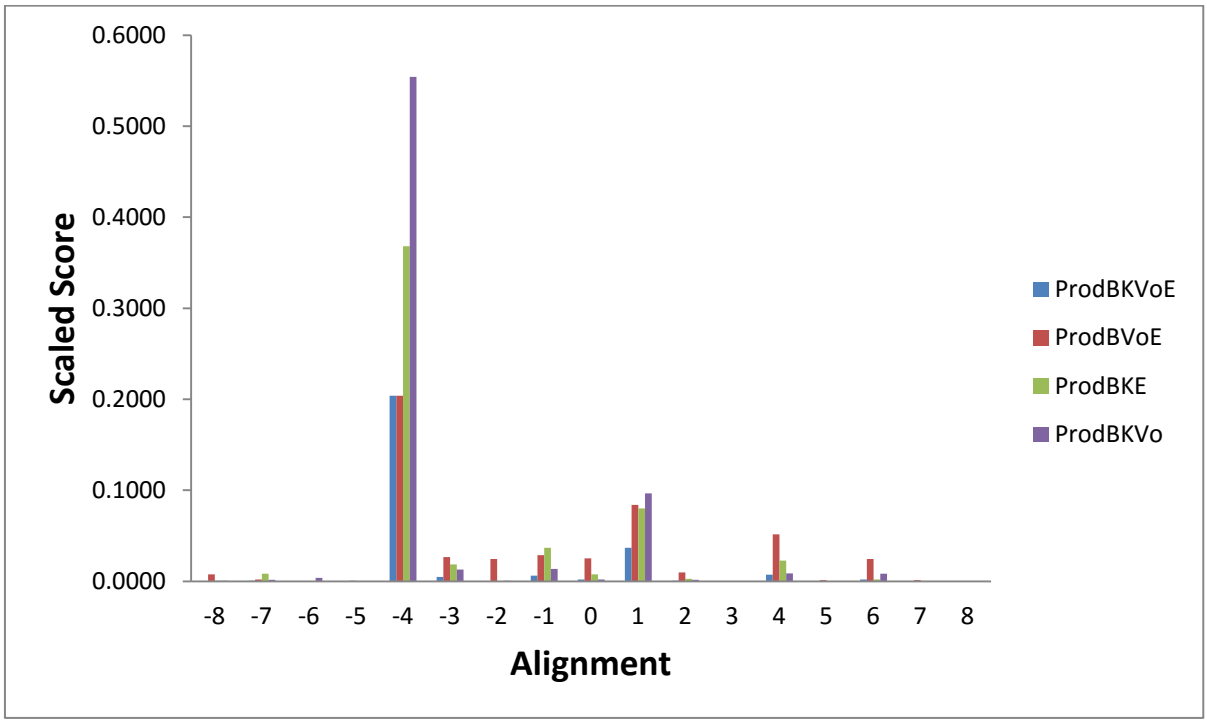

Figure 9: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the A-C TM2 region alignment.

Figure 10 shows the products of B, K, Vo, E and Va for the A-C alignment. The results seen in Figure 10 do not have as clear results as the others and shows no real clear peak as the best alignment. Figure 9 however, shows the -4 alignment to be the best with a small peak at +1 as second best. The product scores for these however are still much lower than the scores seen in the TM1 alignments, even for

the highest peak. The lower scores indicate the results may not be as reliable for this helix.

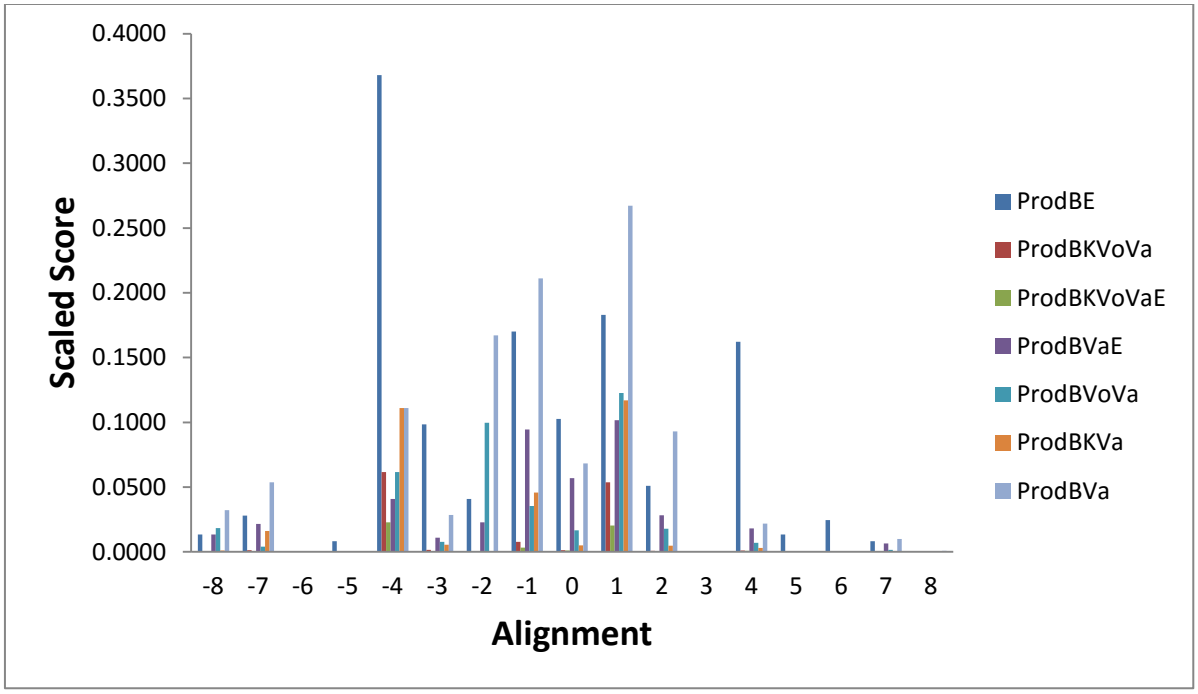

Figure 10: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the A-C TM2 alignment.

3.3.2.2. Class B – Class C

Figure 11 shows the product scores for the alignment of B-C of the TM2 region. There is a percentage identity of 13.0 between these two classes. For this alignment the 0 alignment is given the best score for all products apart from when the hydrophobicity score is removed. In that case the -1 alignment is given the best score. The score for the 0 alignment is particularly strong when the amino acid volume is removed from the product.

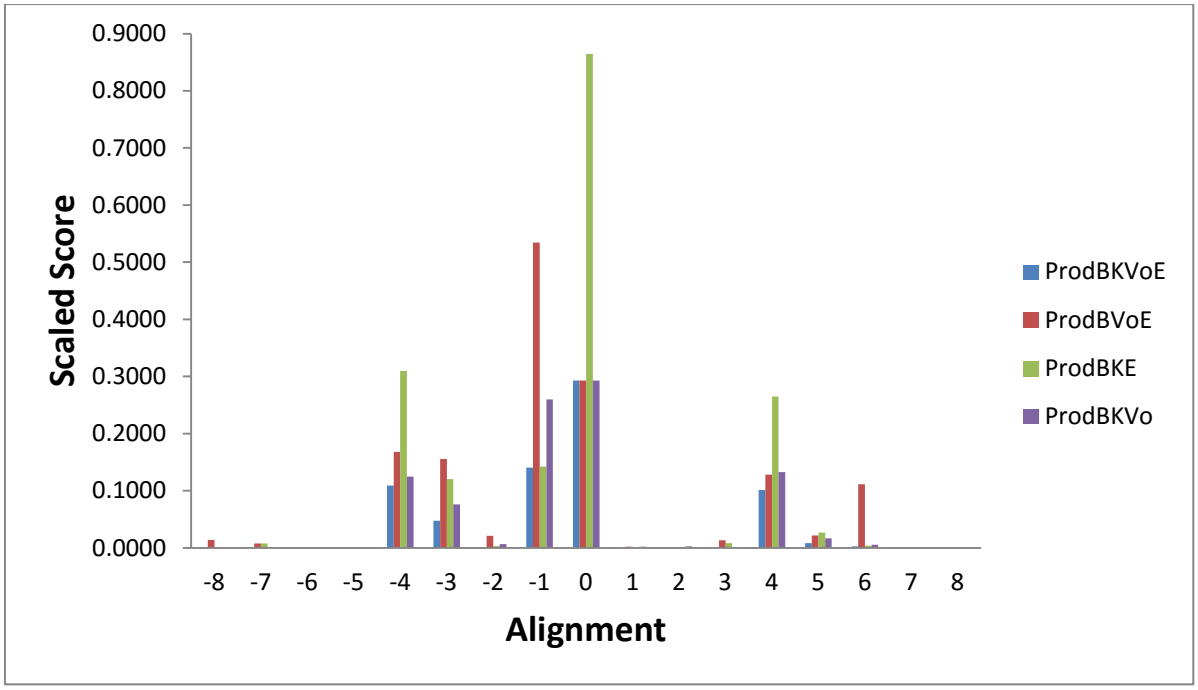

Figure 11: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the B-C TM2 region alignment.

Figure 12 shows the products of B, K, Vo, E and Va for the B-C alignment. The results seen in Figure 12 are again not particularly clear here, but the 0 alignment is always the highest scoring apart from the product of B and Va in which the -1 alignment score is slightly higher.

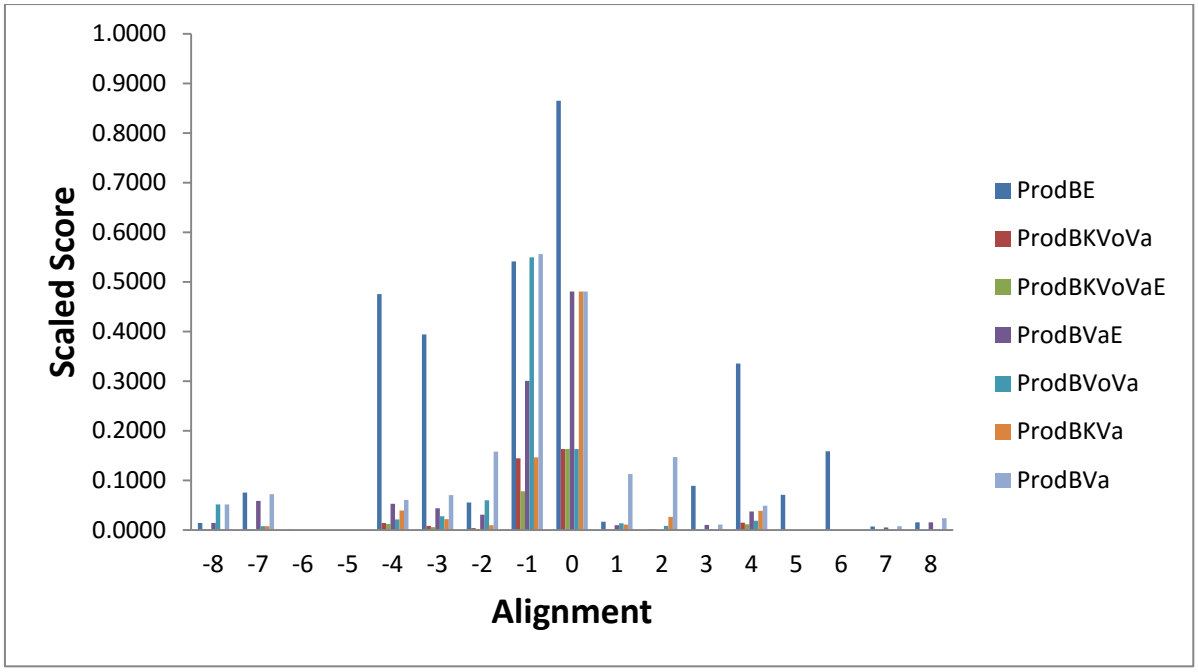

Figure 12: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the B-C TM1 alignment.

3.3.2.3. Class E – class C

Figure 13 shows the product scores for B, K, Vo and E. There is a percentage identity of 13.2 between these two classes for the 0 alignment.

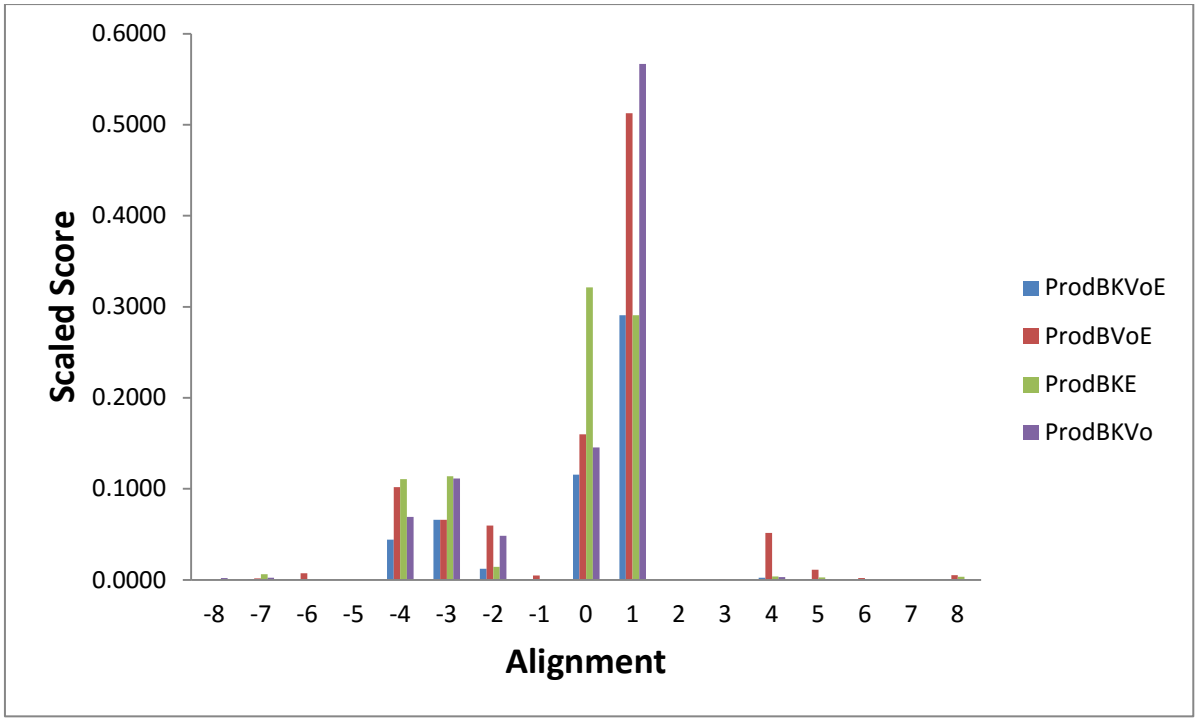

Figure 13: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the E-C TM2 region alignment.

Figure 14 shows the products of B, K, Vo, E and Va for the B-C alignment. Again the results from this alignment are not as clear as in others but Figures 13 and 14, show the highest values for the +1 alignment quite closely followed by the 0 alignment or, when amino acid volume is removed, the 0 alignment slightly overtakes the +1 alignment.

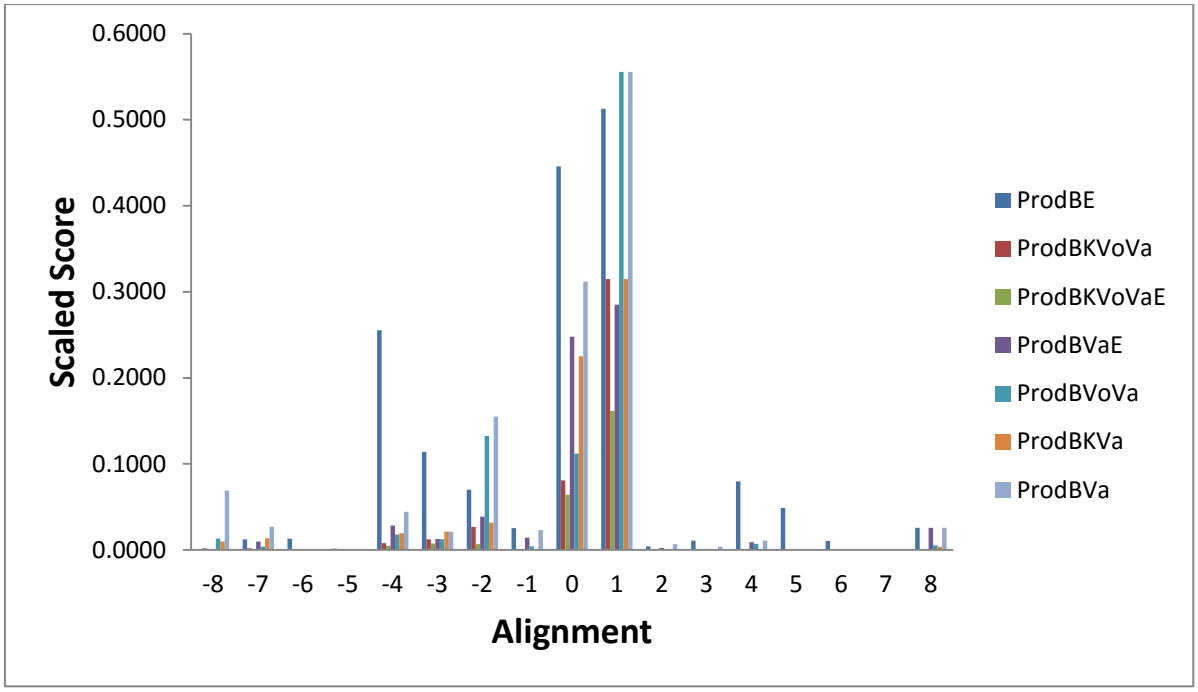

Figure 14: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the E-C TM2 alignment.

3.3.2.4. Class F – class C

The product scores for B, K, Vo and E for the TM2 region of F-C are shown in Figure 15. There is a percentage identity of 7.9 between these two classes for the 0 alignment, which is particularly low.

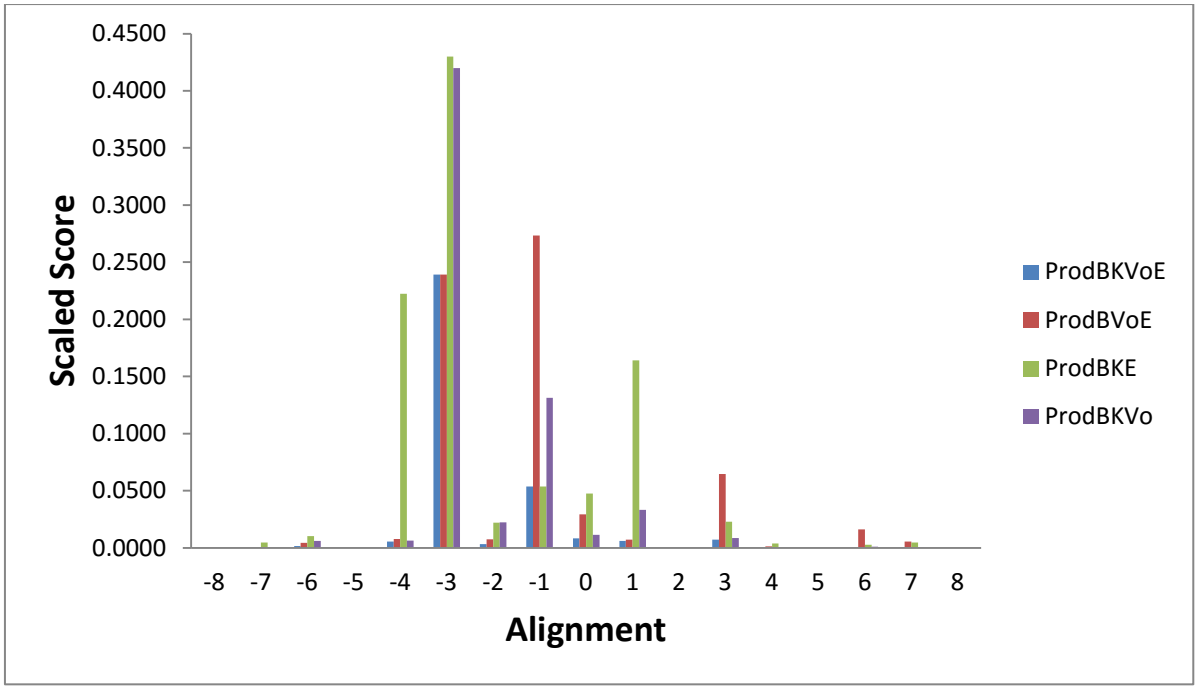

Figure 15: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the F-C TM2 region alignment.

Figure 16 shows the results of the of B, K, Vo, E and Va for the F-C alignment.

These results so not show a clear peak for the best alignment but +3, -1 and +1 are mainly the highest scorers.

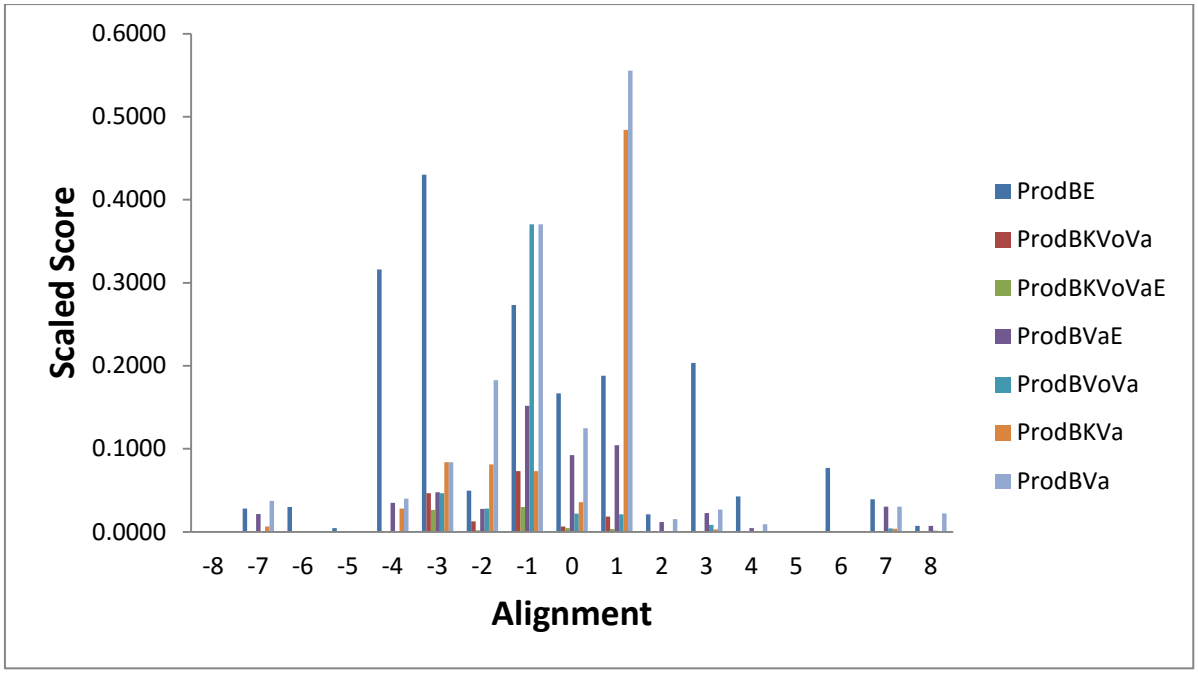

Figure 16: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the F-C TM2 alignment.

Chart 3 shows that the multi-reference alignment gives ambiguous results, both of which are wrong relative to the structural alignment. The alignments given are +1 and -4, with the correct 0 score having a very low score, despite getting reasonable score for class B and class E. The -4, 0 and +1 alignments are given for reference in Chart 4. It is not clear why TM1 gives good results and TM2 gives poor results, but one reason may be that the helices are not uniformly alpha-helical. Thus, it is known that many class A GPCRs have a bulge in TM2 (van der Kant and Vriend 2014) and this may be the reason why the method does not work for TM2, as an underlying assumption is that the helices are essentially ideal alpha-helices and at present we do not know how much deviation from ideality can be tolerated.

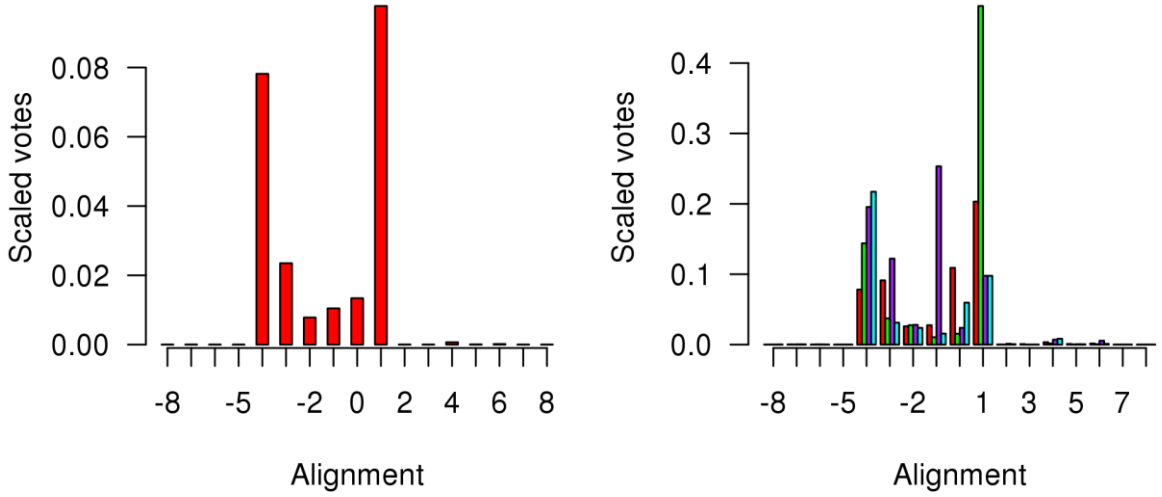

Chart 3: The multi-reference alignment scores. On the left, the score is given for multiplying the class A, class B, class E and class F scores together; on the right the score is given for missing out class A (red), class B (green), class E (purple) and class F (cyan).

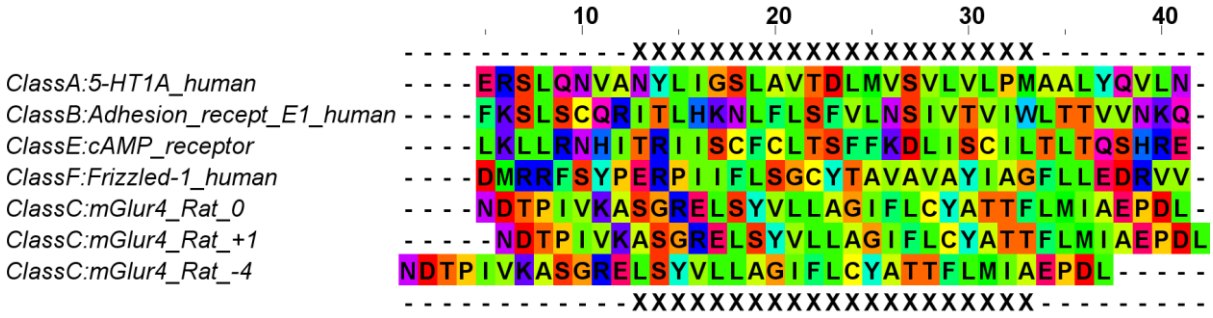

Chart 4: The alignment of the class C TM2 against TM2 for class A, class B, class E and class F. The class A - class B - class E - class F alignment is defined by the structural alignment of the relevant X-ray crystal structures. The correct 0 alignment (as defined by structural alignment) is given, along with the alternatives indicated by Figures 9, 10, 13 and 14, and chart 3. The Xs mark the window over which the alignment was determined. The colour scheme for the residues is the default Taylor colour scheme used by Jalview.

3.3.3. TM3

3.3.3.1. Class A – Class C

Figure 17 shows the product scores of B, K, Vo and E for the TM3 region of A-C. There is a percentage identity of 12.4 between these two classes. The hydrophobicity scores for this region are really having a big effect on the product scores as with them no real peaks are seen but without the 0 alignment is the clear best followed by a smaller peak at -3.

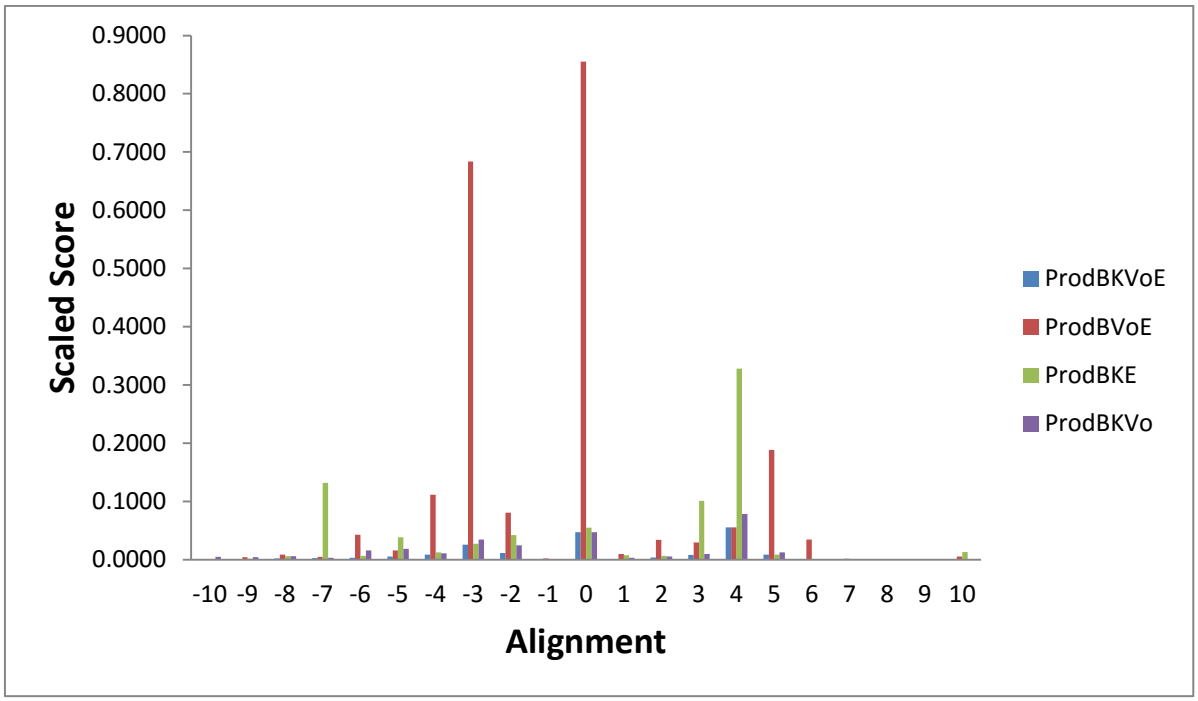

Figure 17: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the A-C TM3 region alignment.

The peaks in figure 18 are not as well defined but still show the best alignment to be either 0 or -3 depending on what results are included, i.e. if hydrophobicity is excluded then 0 is the preferred alignment. Since TM3 largely runs through the

middle of the TM bundle, hydrophobicity may not be used to determine the orientation of the helix anywhere near as much as is used for the other helices.

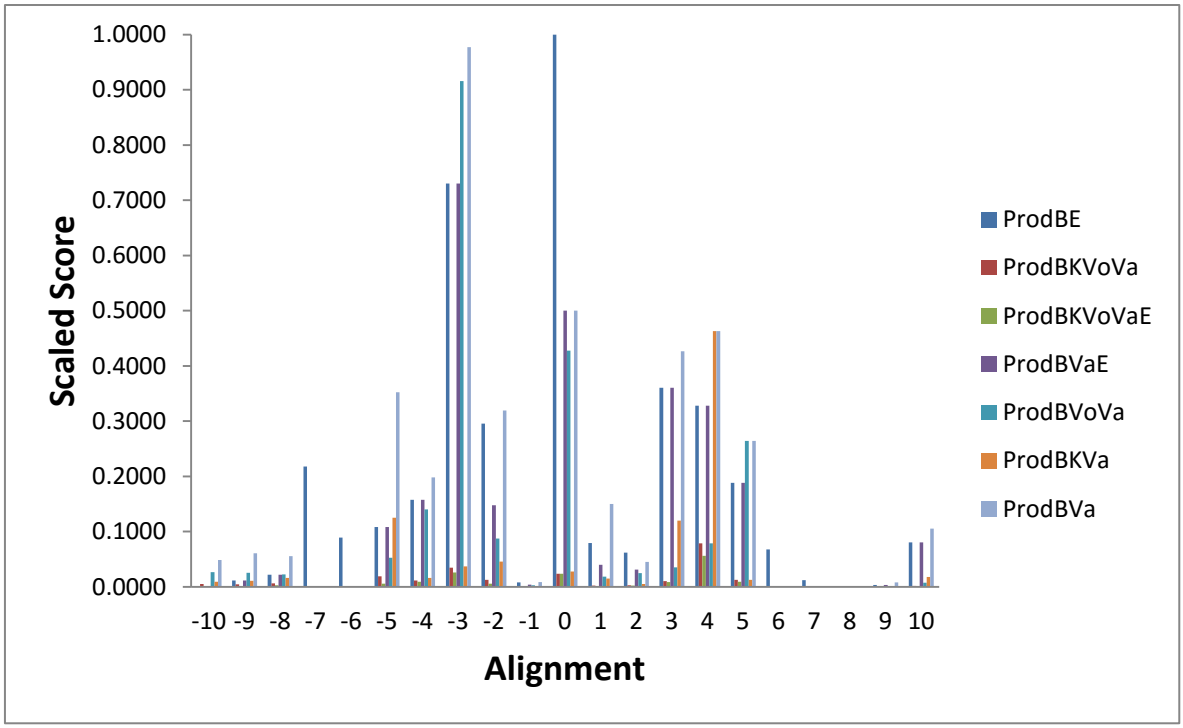

Figure 18: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the A-C TM3 alignment.

3.3.3.2. Class B – Class C

Figure 19 shows the product scores of B, K, Vo and E. There is a percentage identity of 12.3 between these two classes. The product scores for this alignment are not very high even for the best peaks and most agree that the best peak is the +5 alignment. When the amino acid volume is not included however the best alignment becomes 0.

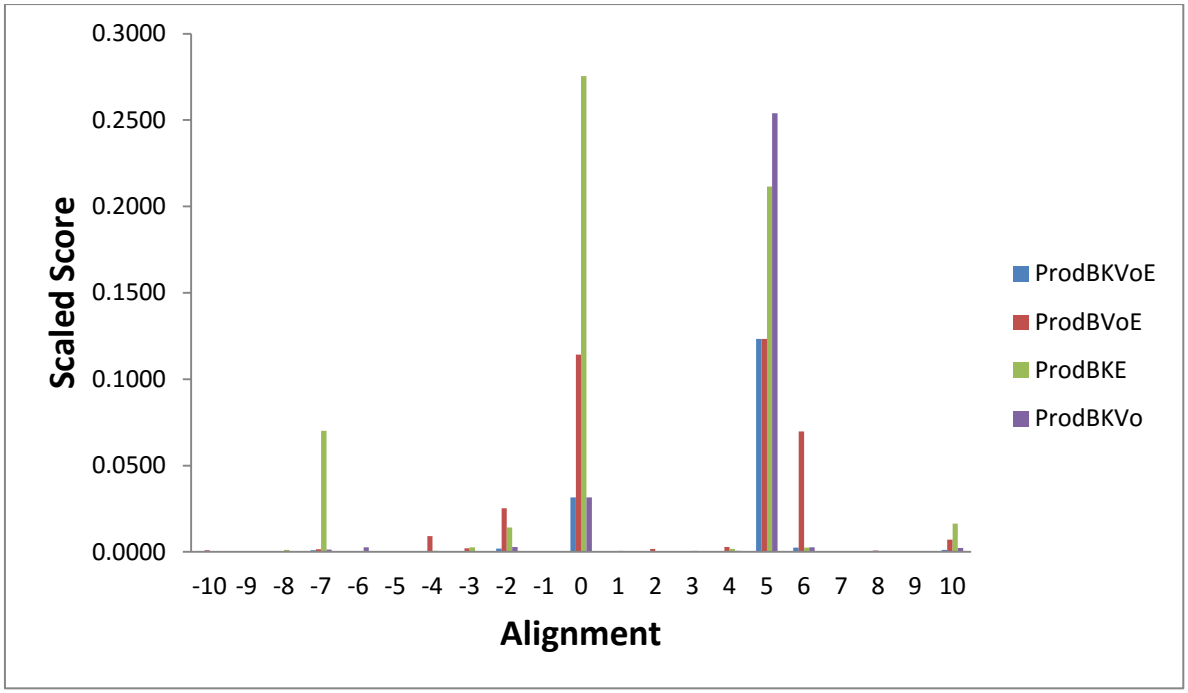

Figure 19: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the B-C TM3 region alignment.

The product scores, shown in figure 20 show a very high peak for the 0 alignment in the product scores of B and E. Other product scores for this alignment also show 0 to be the best alignment but the +5 alignment is still shown as the best by many product scores. The highest peaks for the +5 alignment are not as high as for the 0 alignment however, even though there are more of them.

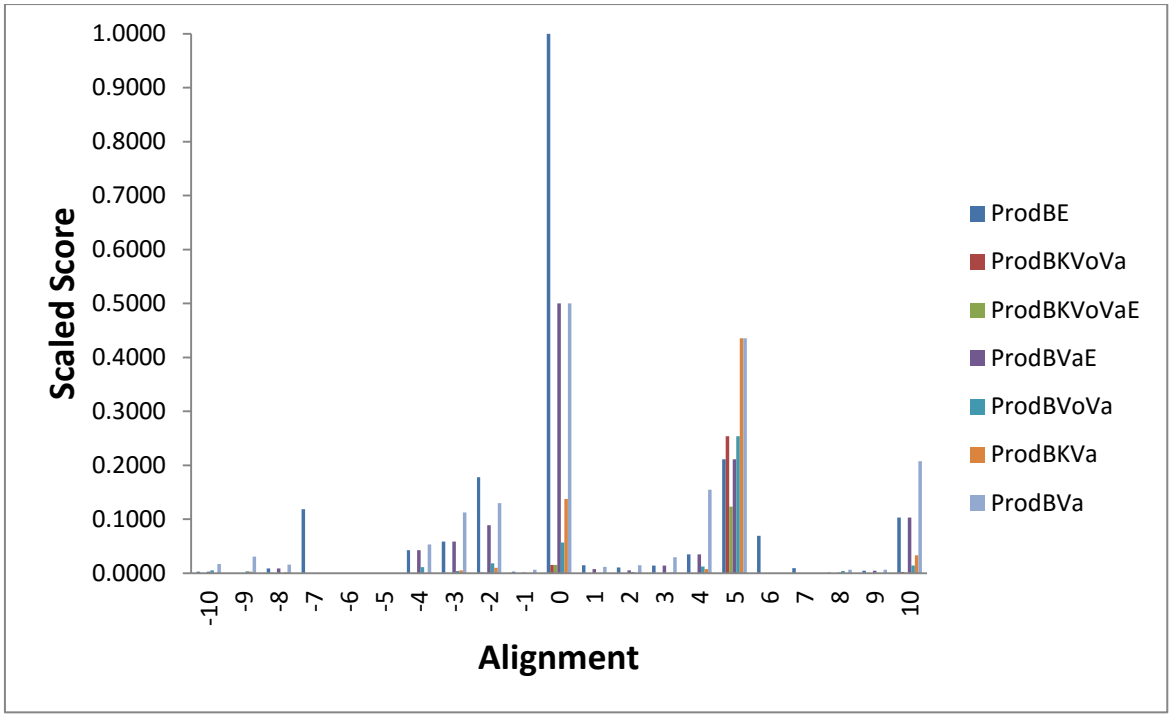

Figure 20: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the B-C TM3 alignment.

3.3.3.3. Class E – Class C

There is a percentage identity of 11.9 between these two classes. The best product score seen in figure 21 is for the 0 alignment when the score for the amino acid volume is removed. The product score when the hydrophobicity score is removed indicates the +6 alignment to be the best but this peak is not as high as the highest peak at the 0 alignment.

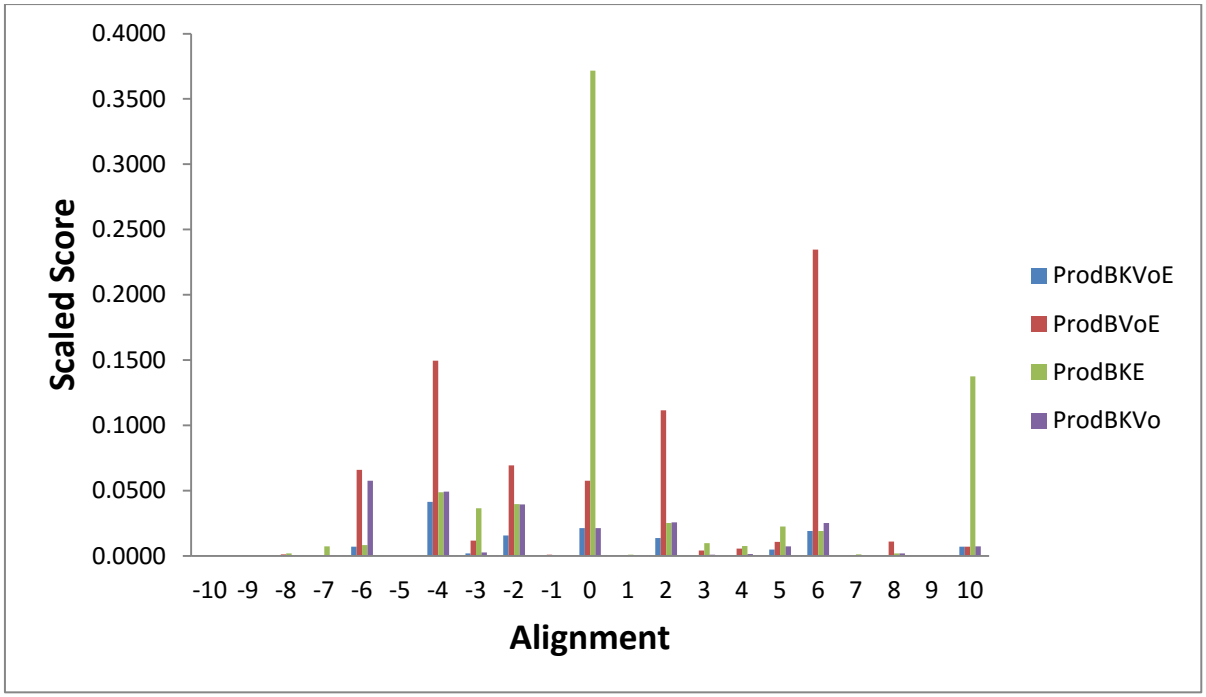

Figure 21: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the E-C TM3 region alignment.

The product scores in figure 22 show the best alignment to be the 0 alignment by quite a large amount, particularly for the BE product score.

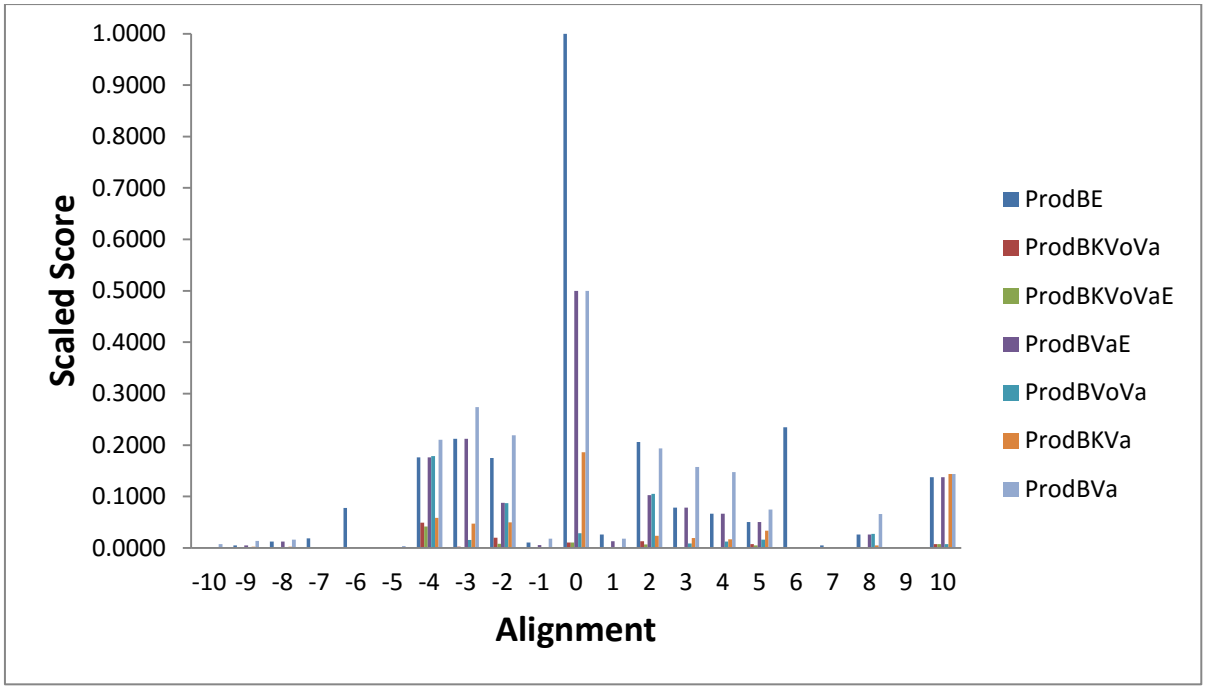

Figure 22: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the E-C TM3 alignment.

3.3.3.4. Class F – Class C

The product scores of B, K, Vo and E are shown in figure 23. There is a percentage identity of 10.4 between these two classes. The results seen in figure 23 are very low even for the highest peaks and the best alignments are indicated as -7 and -3.

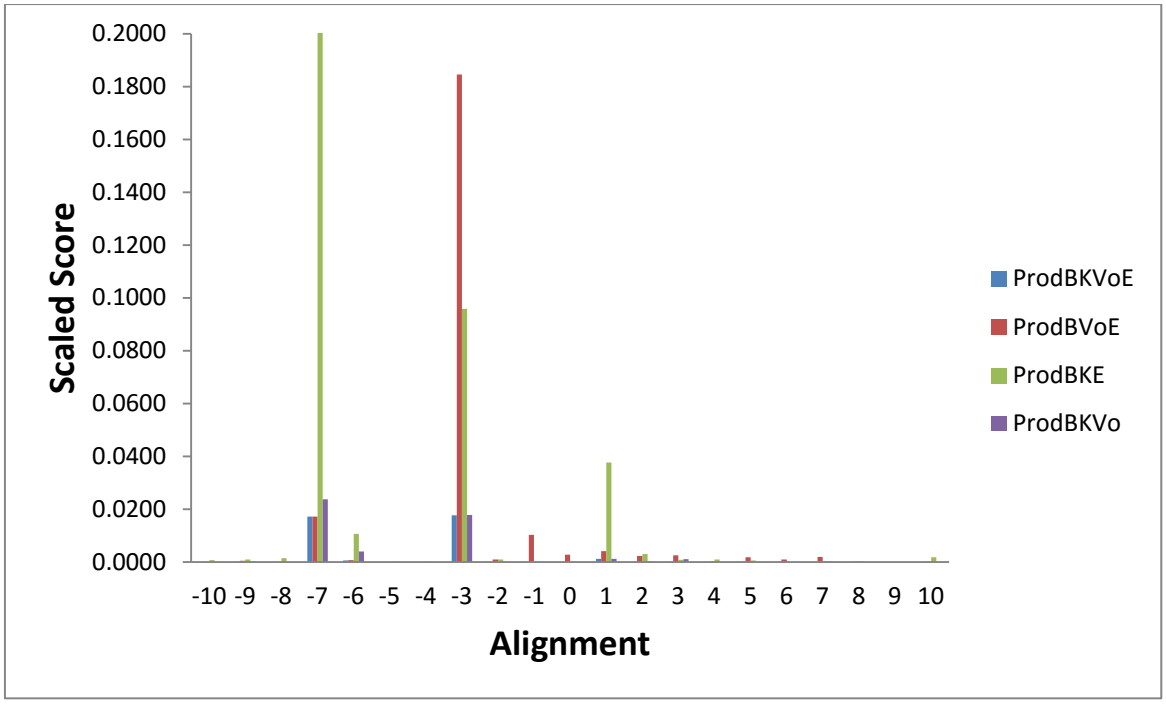

Figure 23: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the F-C TM3 region alignment.

A number of the product scores are much stronger for F-C TM3 in figure 24. These results indicate the -3 alignment to be the best.

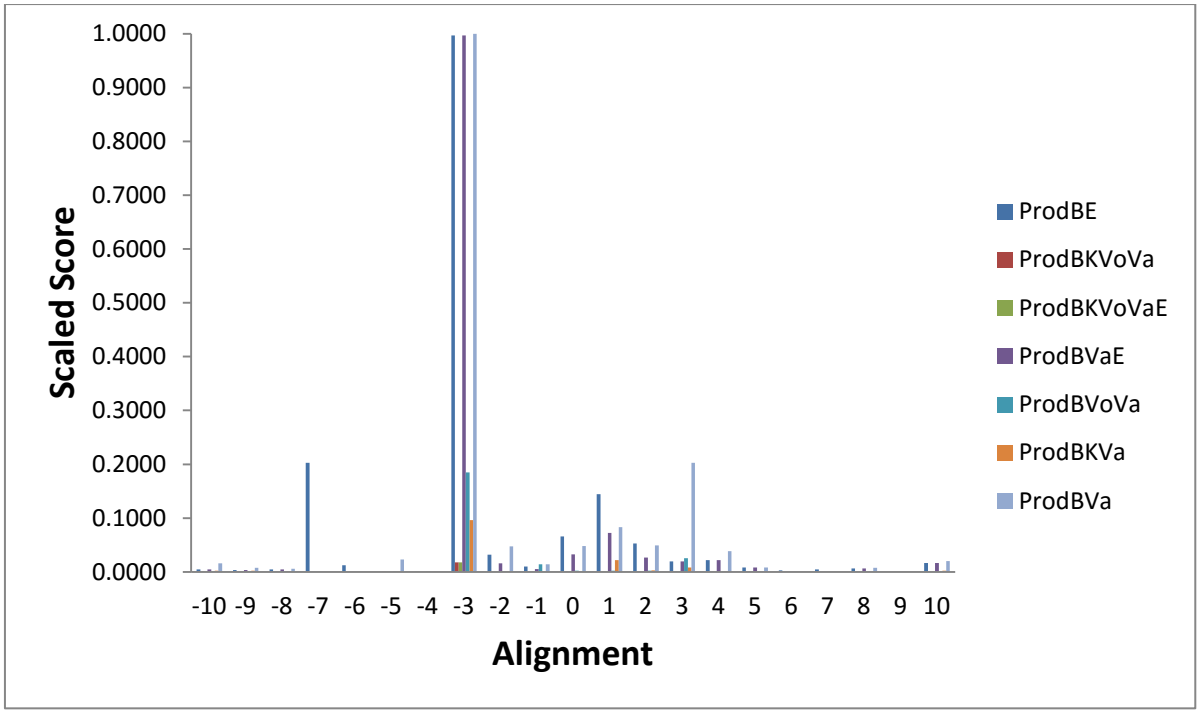

Figure 24: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the F-C TM3 alignment.

The multi-reference alignment is given in Chart 5. The Blosum matrix score is overwhelmingly in favour of alignment 0 for class A, class B and class E and this determines the preference for the correct 0 alignment. This suggests that the hydrophobicity and volume, which mitigate against alignment 0 for TM3 (e.g. Fig. 17, 19) are less accurate than the Blosum matrix scores. The conserved cysteine on TM3 is correctly aligned in the 0 alignment. The percentage ID for class A, class B, class E and class F at the 0 alignment to class A is 12.4, 12.3, 11.9 and 10.4, indicating that as the %ID approaches 10% that the results of the Blosum alignment start to become less reliable. The alignment is given in Chart 6.

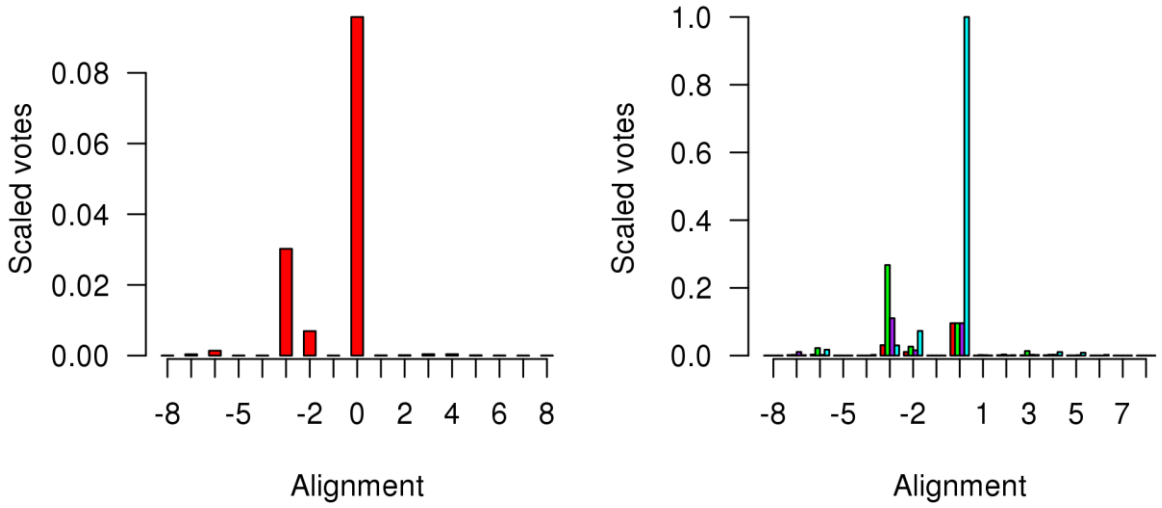

Chart 5: The multi-reference alignment scores. On the left, the score is given for multiplying the class A, class B, class E and class F scores together; on the right the score is given for missing out class A (red), class B (green), class E (purple) and class F (cyan).

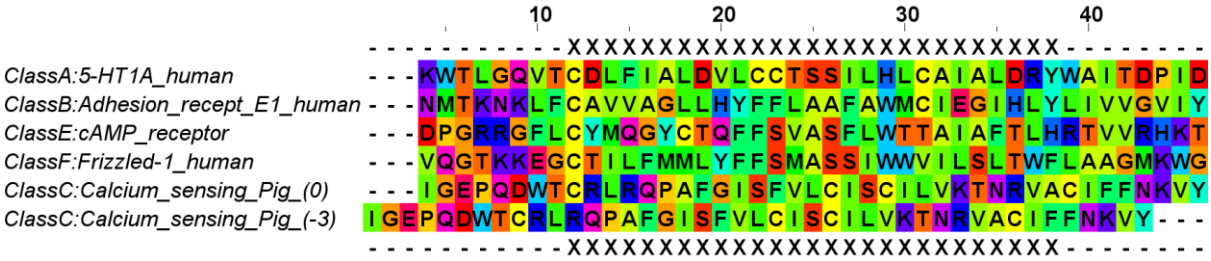

Chart 6: The alignment of the class C TM3 against TM3 for class A, class B, class E and class F. The class A - class B - class E - class F alignment is defined by the structural alignment of the relevant X-ray crystal structures. The correct 0 alignment (as defined by structural alignment) is given, along with the alternative of -3 indicated by Figures 17, 18, 23 and 24,. The Xs mark the window over which the alignment was determined. The colour scheme for the residues is the default Taylor colour scheme used by Jalview.

3.3.4. TM4

3.3.4.1. Class A – Class C

Figure 25 is the product of the scores of B, K, Vo and E for the A-C alignment of

TM4. The results show the 0 alignment to have the best score. A second, smaller

peak seen at the -10 alignment favoured when the hydrophobicity scores are excluded from the product. There is a percentage identity of 10.2 between these two classes at the 0 alignment.

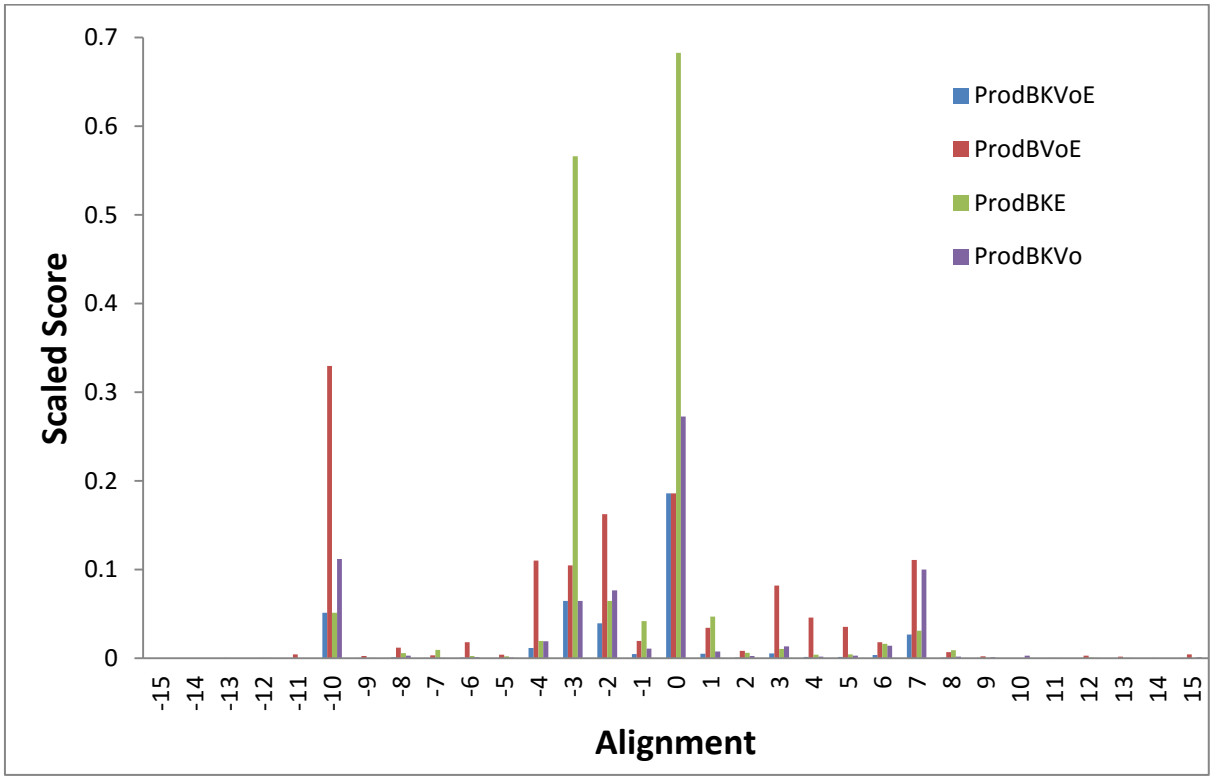

Figure 25: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the A-C TM4 region alignment.

The results in figure 26 are not as clear but many of them still indicate 0 to be the best alignment. When hydrophobicity is removed there are strong product scores for the -3 alignment. Finally, the -10 alignment has quite good product scores, though not as strong as for the 0 and -3 alignment.

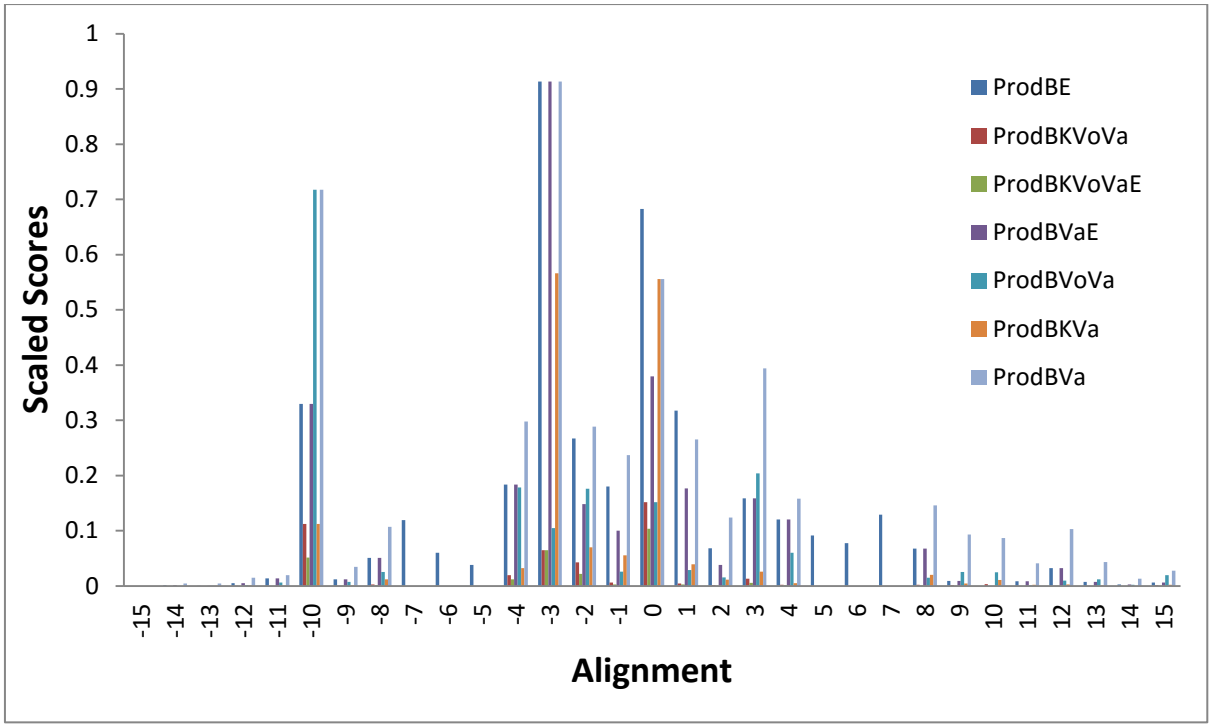

Figure 26: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the A-C TM4 alignment.

3.3.4.2. Class B – Class C

The product scores of B, K, Vo and E are given in figure 27 for the B-C alignment of TM4. There is a percentage identity of 8.4 between these two classes at the 0 alignment. There are two relatively strong peaks at the 0 and +7 alignment. The 0 alignment is preferred alignment when all of the data is included and when only Vo is excluded and the +7 is preferred for when K and also E are excluded.

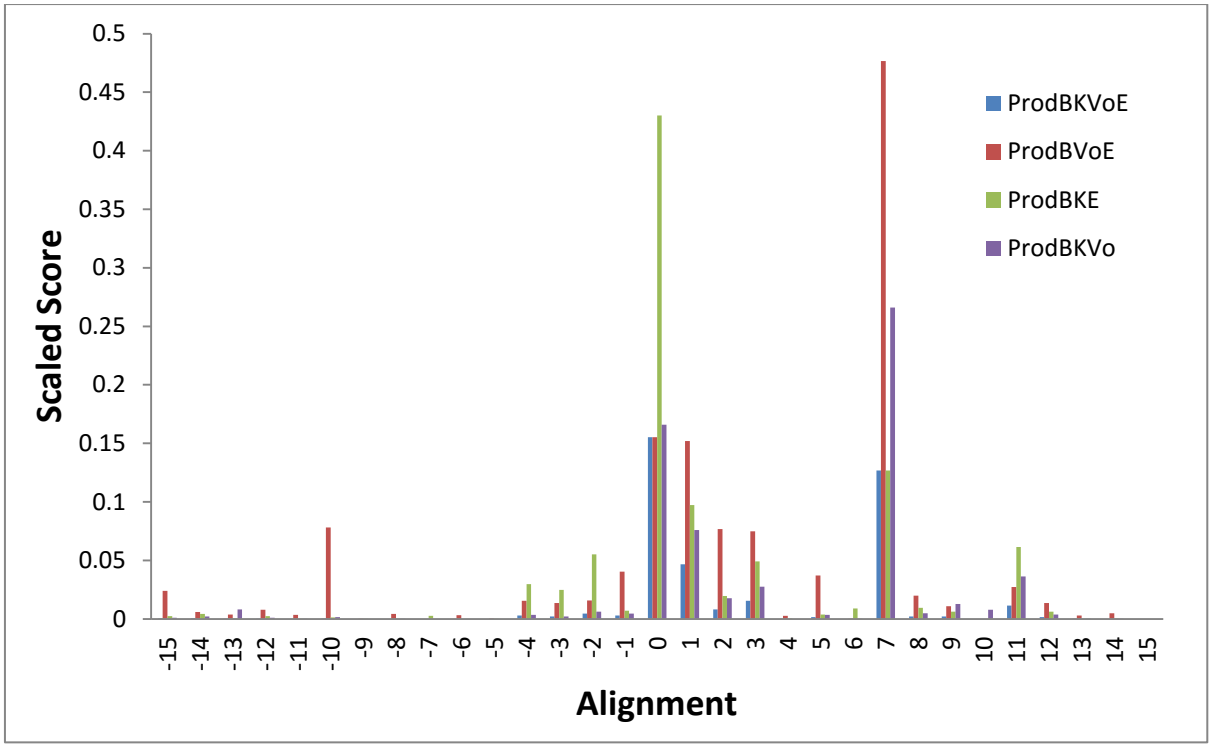

Figure 27: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the B-C TM4 region alignment.

The results from figure 28 are not as clear but indicate the 0 alignment as preferred in most cases. This second figure helps to strengthen the case for the 0 alignment and weaken the +7 as the +7 only has one high scoring result in figure 28. This means it is much less likely to be the correct B-C alignment for the TM4 region.

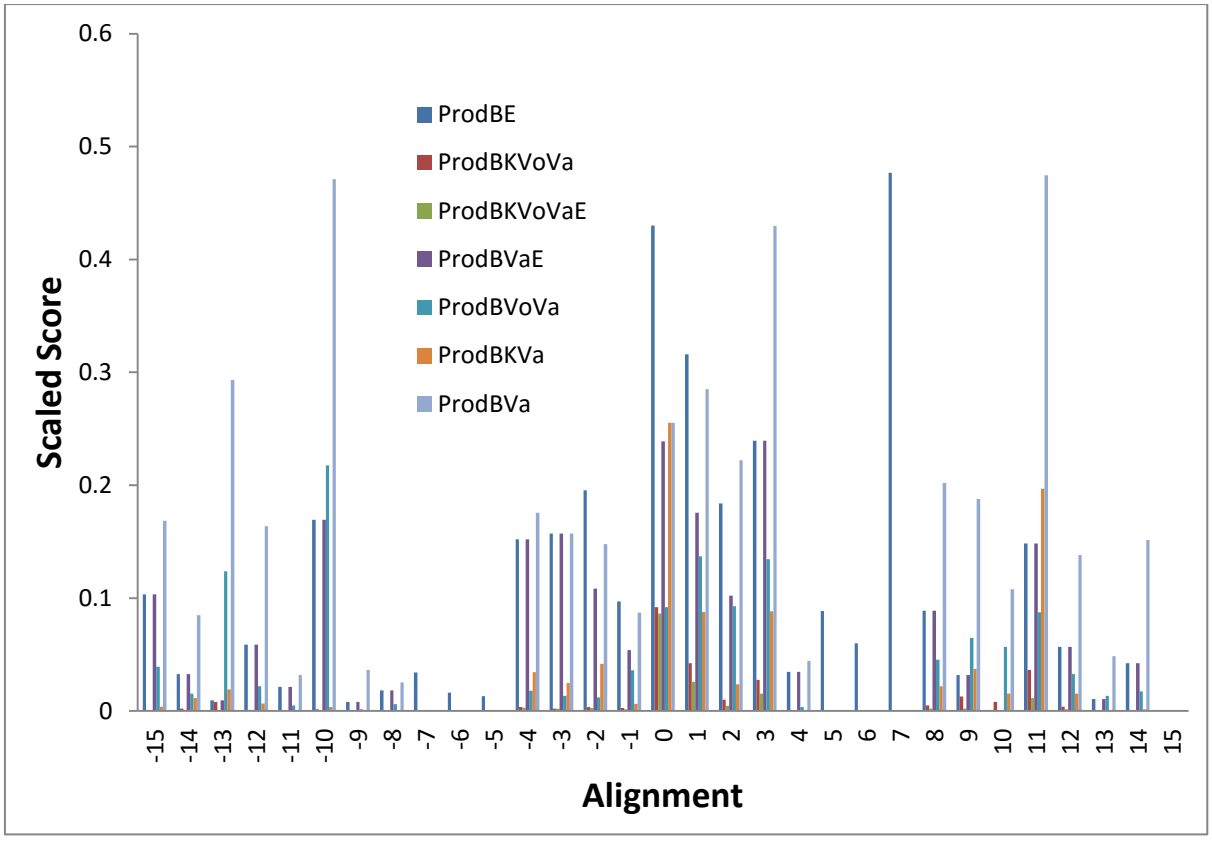

Figure 28: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the B-C TM4 alignment.

3.3.4.3. Class E – Class C

Figure 29 is the product of the scores of B, K, Vo and E for the E-C alignment of TM4. There is a percentage identity of 8.8 between these two classes for the 0 alignment. The product of the scores quite clearly indicate the 0 alignment to be the best for these classes. The result is particularly strong when the Vo score is excluded.

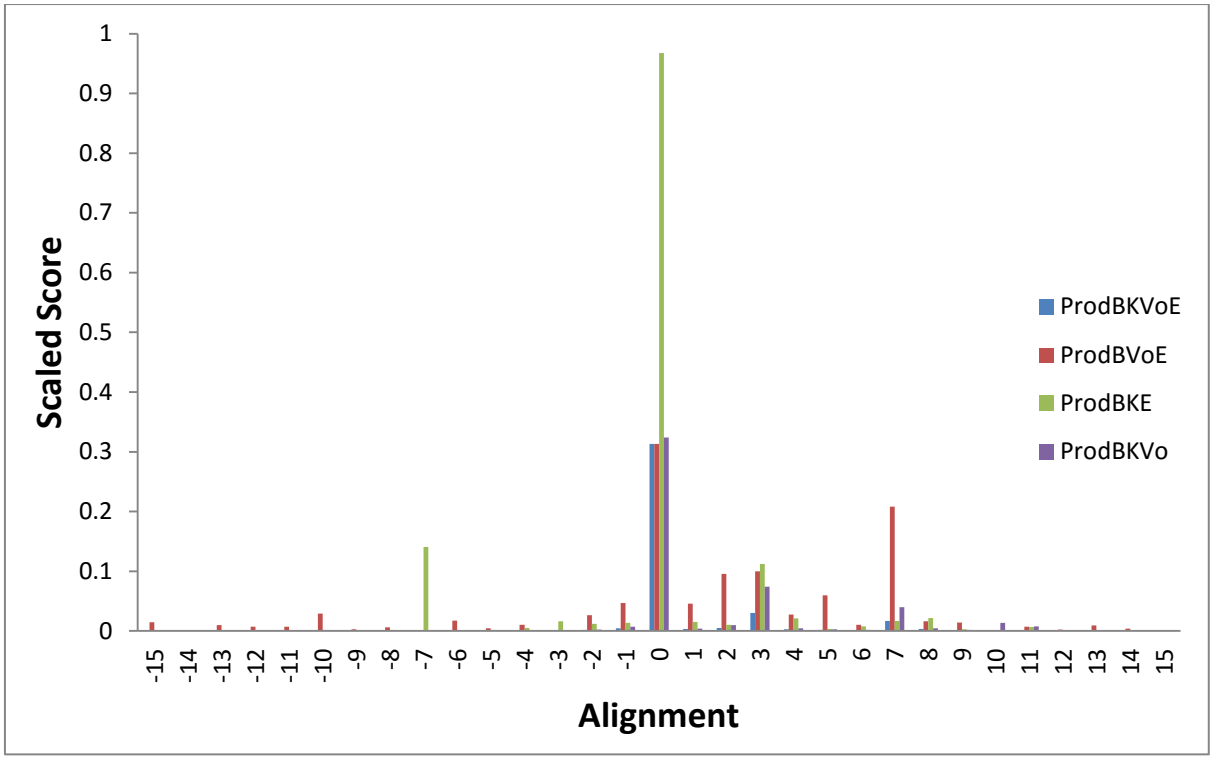

Figure 29: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the E-C TM4 region alignment.

The product scores in figure 30 strengthen the peak at the +3 alignment which was only small in figure 29 but still show the 0 alignment giving the strongest result of all indicating that this is the preferred alignment for E-C in the TM4 region.

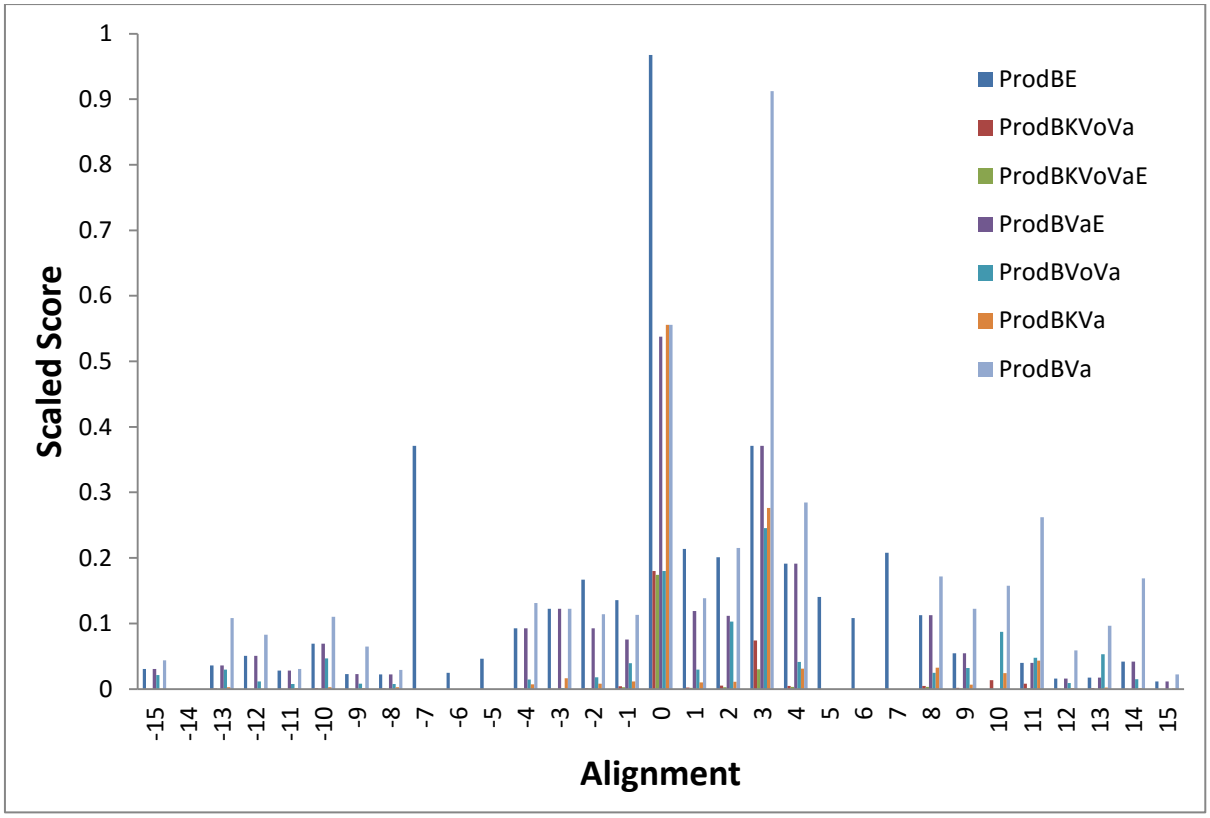

Figure 30: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the E-C TM4 alignment.

3.3.4.4. Class F – Class C

Figure 31 shows the product of scores for the F-C alignment of TM4. There is a percentage identity of 9.7 between these two classes at the 0 alignment. The scaled product scores are all quite weak for this alignment but the alignments that consistently got the highest peaks are the +3 alignment followed by the 0 and -10 alignments. The peak at -10 when K is excluded is much higher than most of the other scores but when the K result is taken into account the alignment does not score very well.

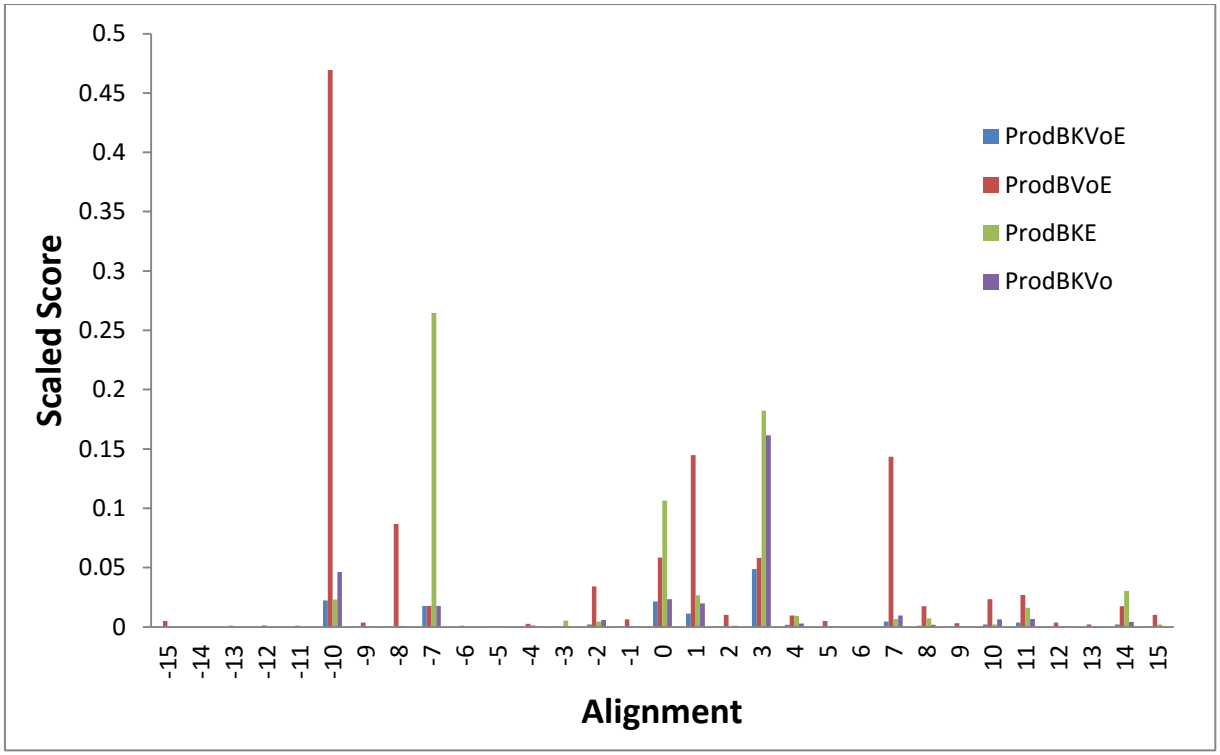

Figure 31: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the F-C TM4 region alignment.

Again, the results seen in figure 32 show the +3 alignment to be consistently the best. The highest peaks overall are seen at the -10 alignment when the K scores are not included in the product but the other scores are significantly lower than at the +3 alignment making the +3 the preferred alignment in this case.

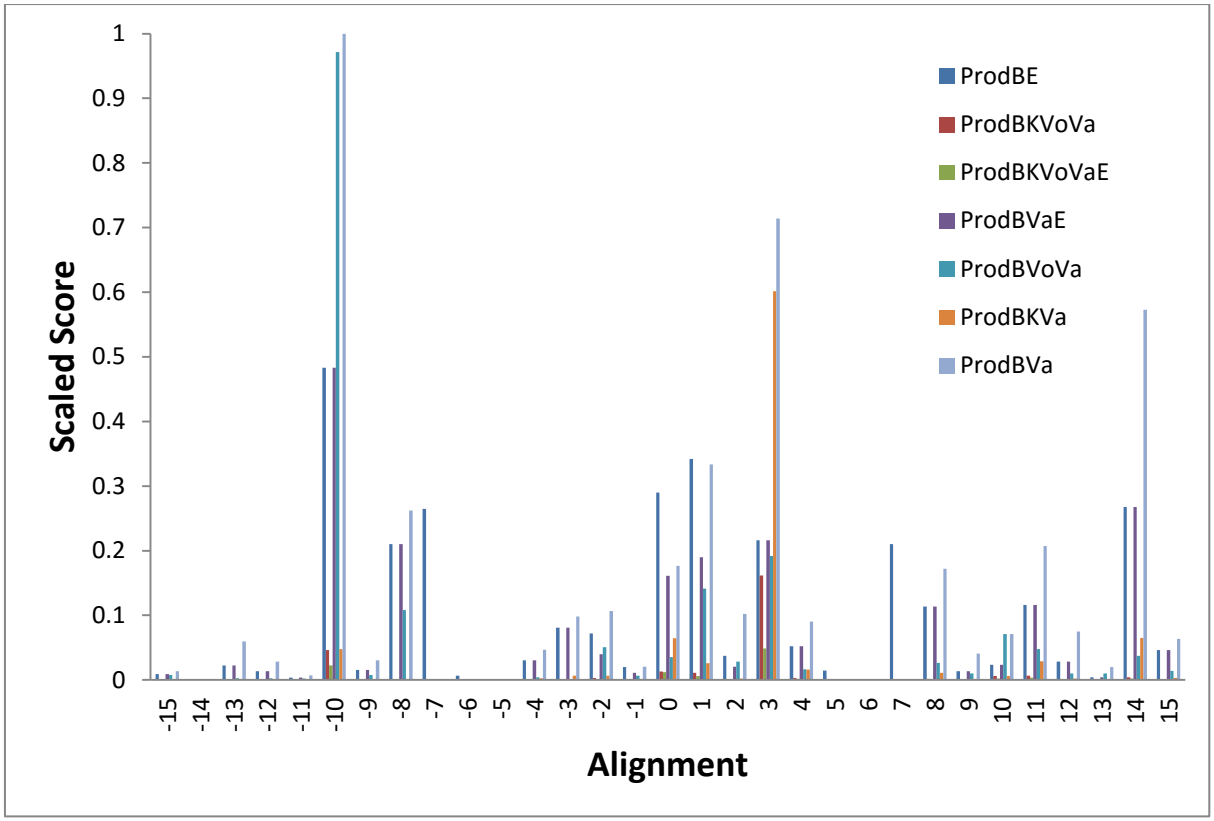

Figure 32: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the F-C TM4 alignment.

The multi-reference alignment is given in Chart 7. Class B favours alignment +7 and class F favours alignment +3, but both give a reasonable score to the 0 alignment and so this has the highest score. The percentage ID for class A, class B, class E and class F at the 0 alignment to class A is 10.4, 8.4, 8.8 and 9.7, again indicating that as the %ID approaches 10% that the results of the Blosum alignment start to become less reliable. The alignment is given in Chart 8.

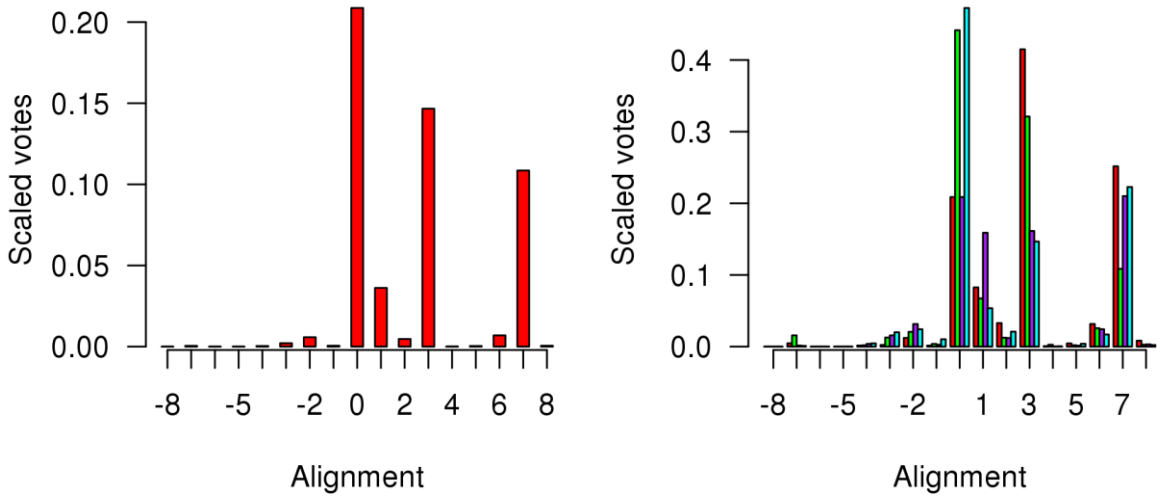

Chart 7: The multi-reference alignment scores. On the left, the score is given for multiplying the class A, class B, class E and class F scores together; on the right the score is given for missing out class A (red), class B (green), class E (purple) and class F (cyan).

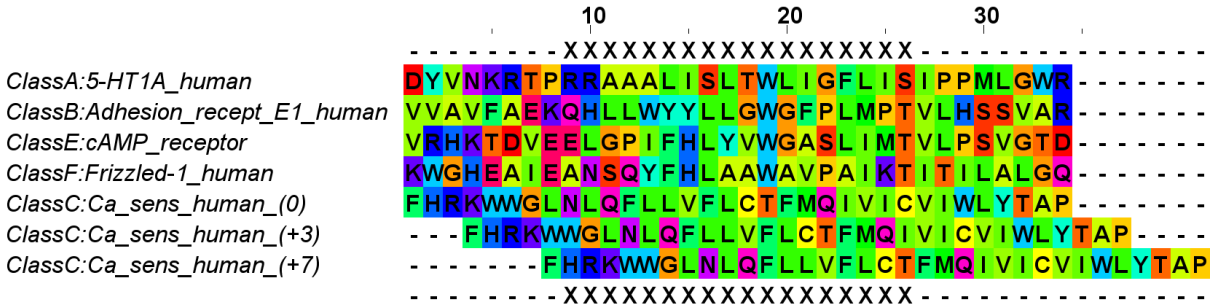

Chart 8: The alignment of the class C TM4 against TM4 for class A, class B, class E and class F. The class A - class B - class E - class F alignment is defined by the structural alignment of the relevant X-ray crystal structures. The correct 0 alignment (as defined by structural alignment) is given, along with the alternatives of +3 and +7 indicated by Figures 31 and 27, and Chart 7. The Xs mark the window over which the alignment was determined. The colour scheme for the residues is the default Taylor colour scheme used by Jalview.

3.3.5. TM5

3.3.5.1. Class A – Class C

Figure 33 shows the product of scores for the A-C alignment of TM5. There is a percentage identity of 6.7 between these two classes. The results show that the +4 alignment is the best followed by smaller peaks at the +7 alignment.

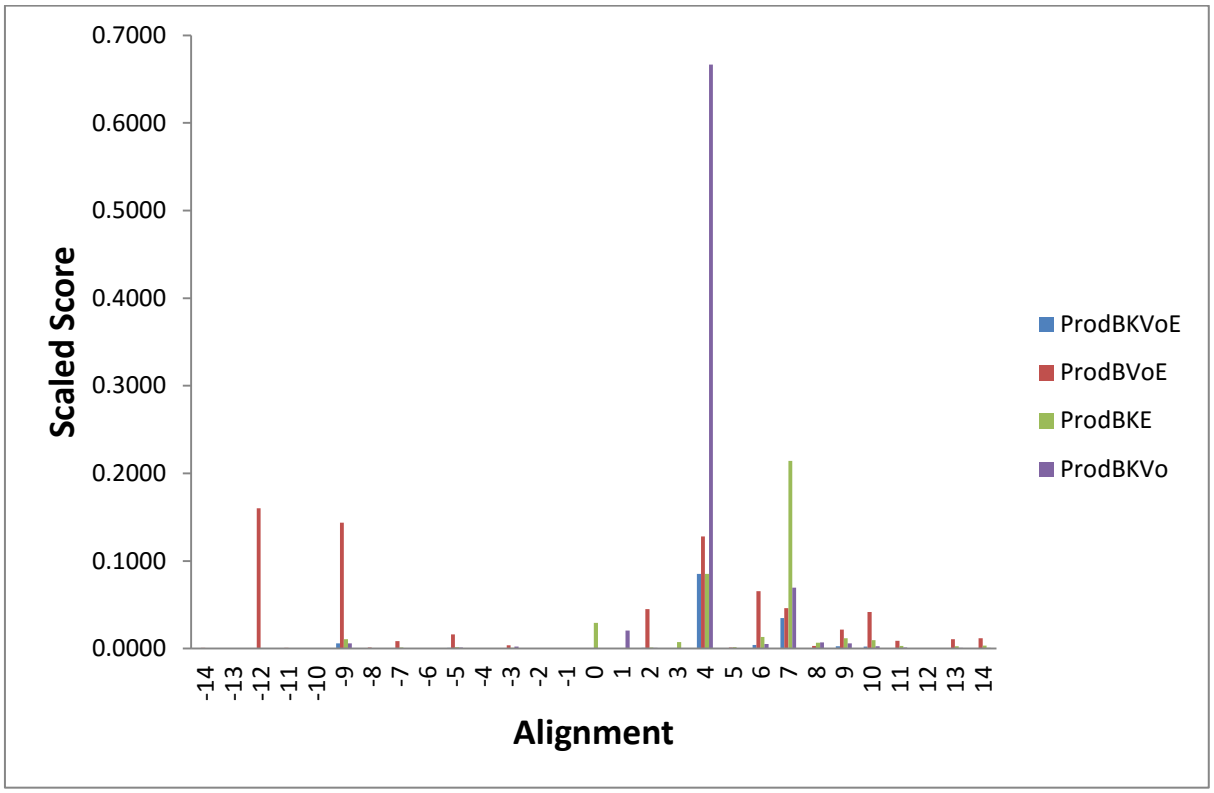

Figure 33: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the A-C TM5 region alignment.

The results in figure 34 are not as clear but show the alignment at +7 to be the best with similarly high peaks at +6.

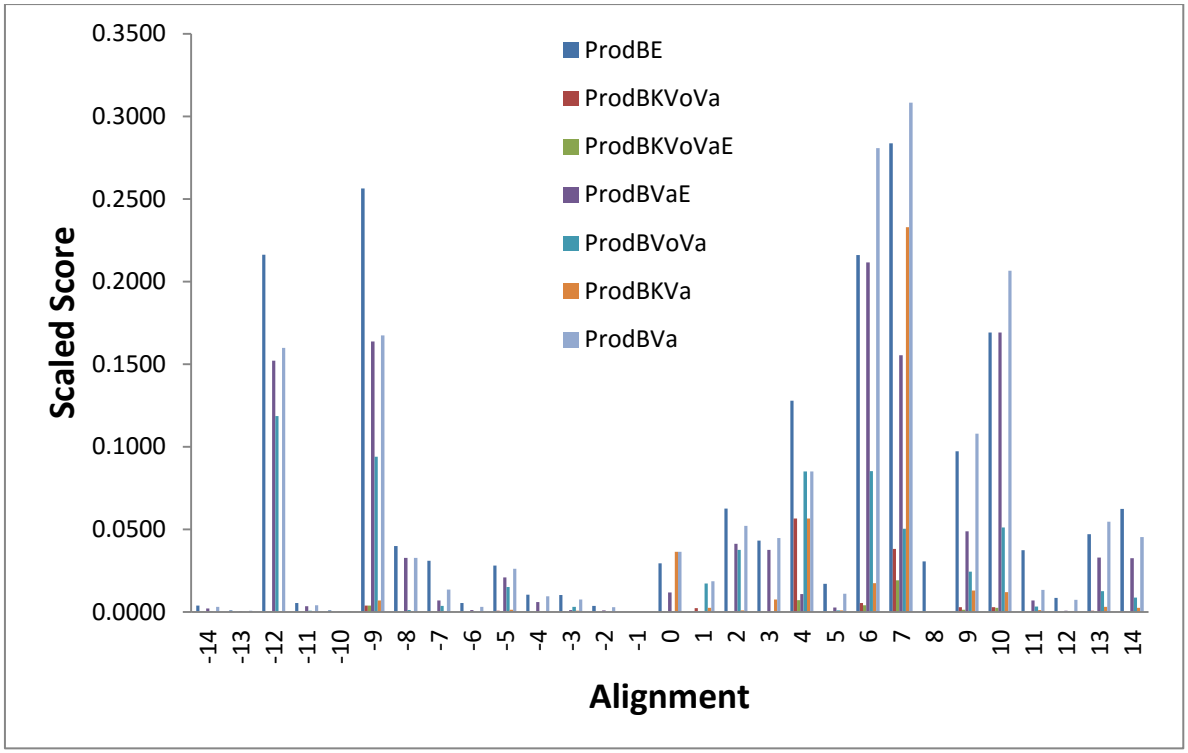

Figure 34: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the A-C TM5 alignment.

3.3.5.2. Class B – Class C

The product scores for the TM5 region of B-C are shown in figure 35. There is a percentage identity of 6.3 between these two classes. The results show the best alignment to be +4 but the peaks are not very strong, particularly when the entropy scores are included.

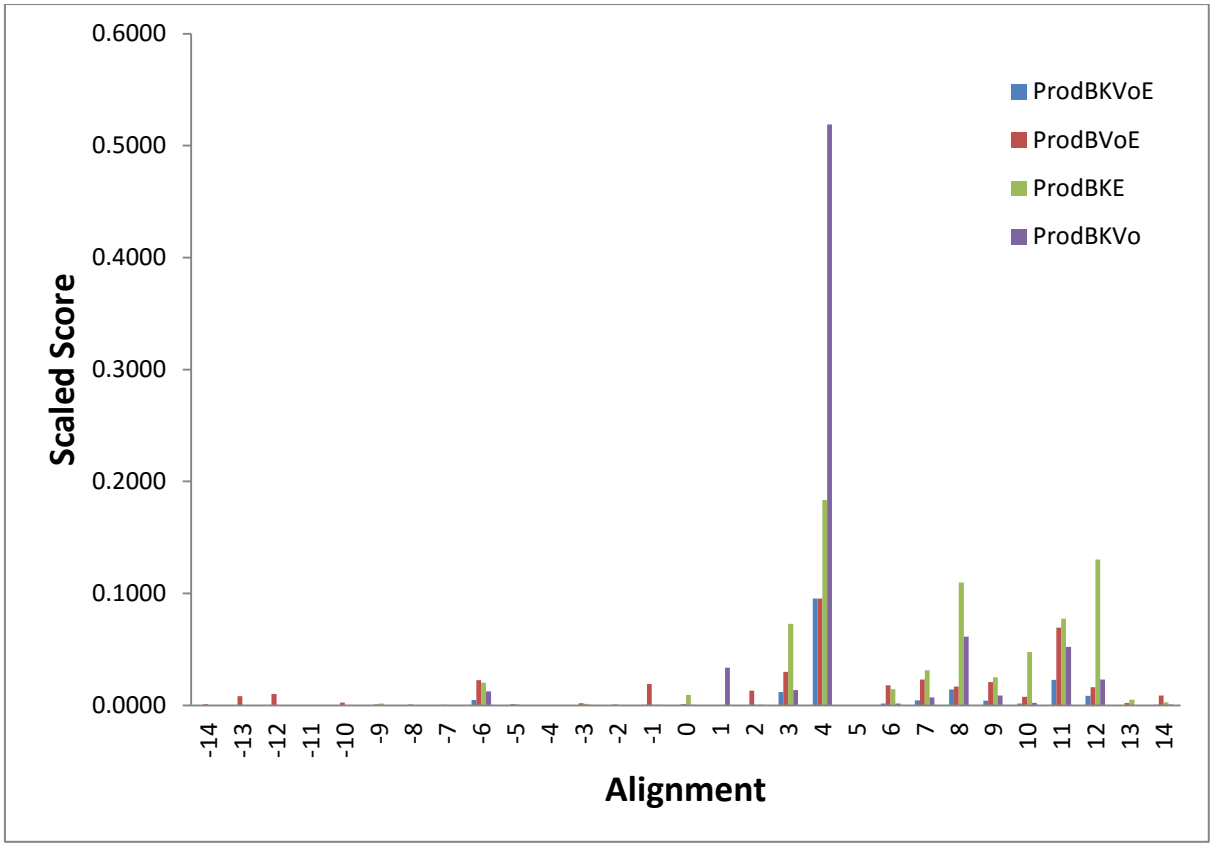

Figure 35: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the B-C TM5 region alignment.

Figure 36 does not have any particularly clear peaks but as above the alignment for +4 is one of the best seen. All the scores are quite low though.

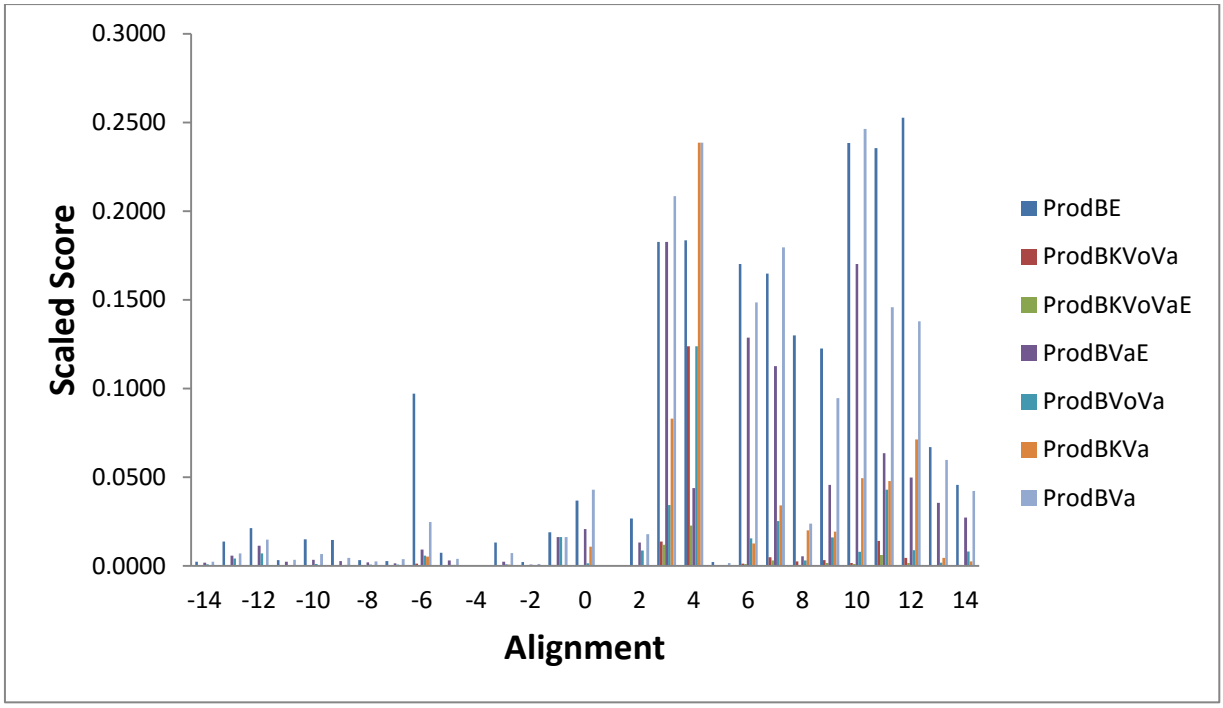

Figure 36: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the B-C TM5 alignment.

3.3.5.3. Class E – Class C

Figure 37 shows the product scores for E-C. There is a percentage identity of 4.7 between these two classes. These show a clear result for the +8 alignment being the best scoring with hardly any other significant peaks.

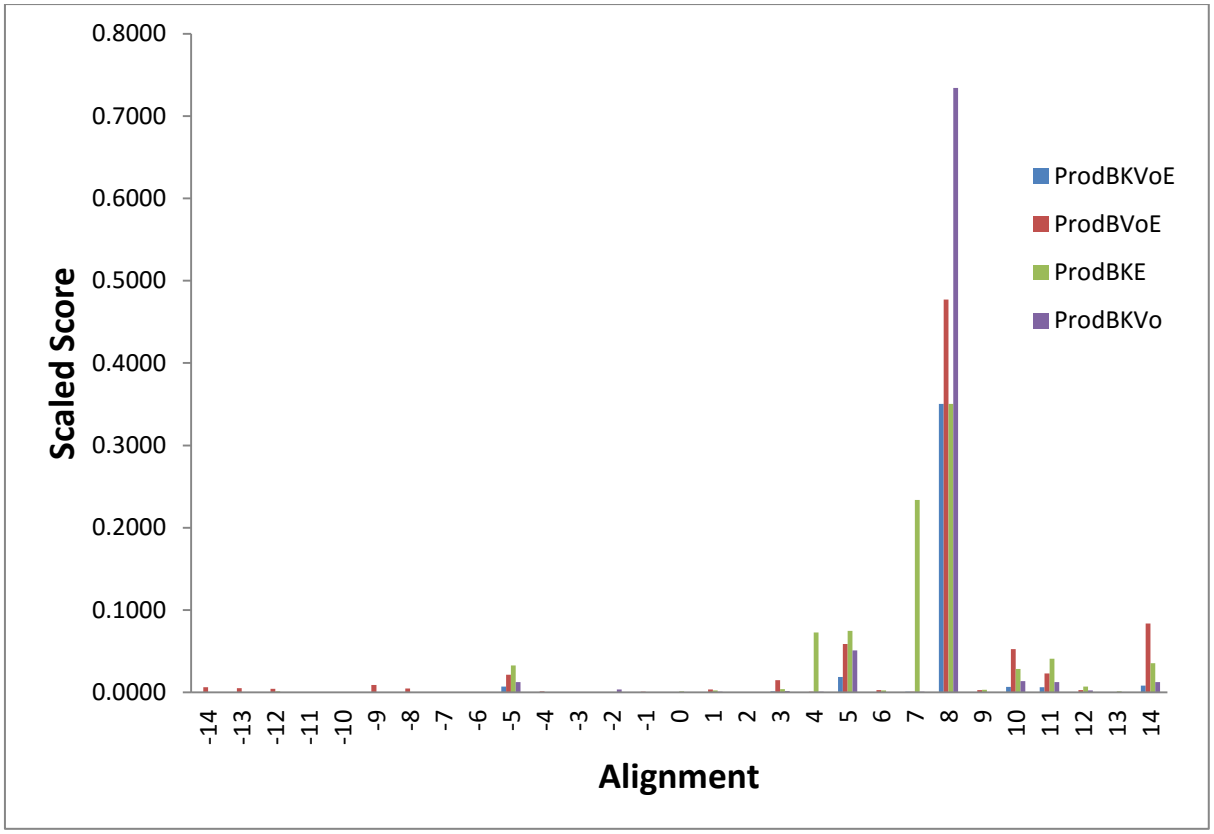

Figure 37: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the E-C TM5 region alignment.

The results from figure 38 are not as clear as in 37 but still show the +8 alignment as being the best.

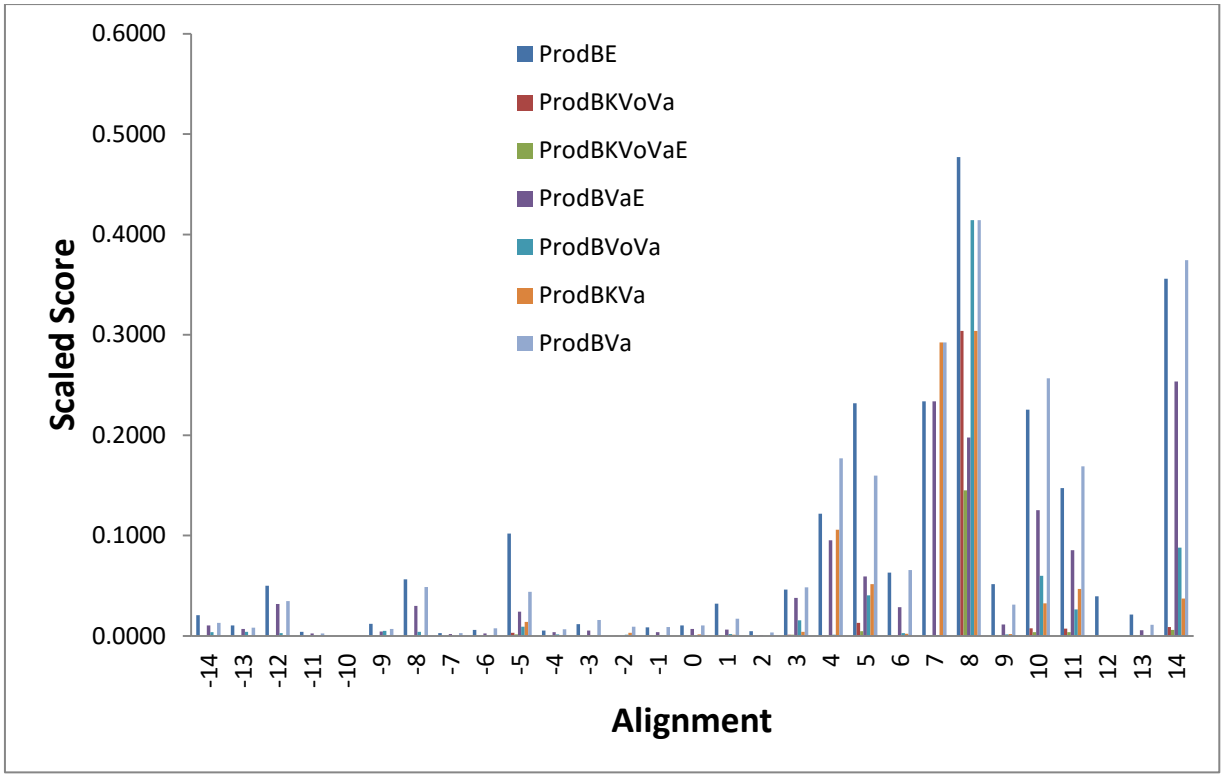

Figure 38: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the E-C TM5 alignment.

3.3.5.4. Class F – Class C

The product scores for F-C are shown in figure 39. There is a percentage identity of 8.2 between these two classes. The results show clearly a significant peak at +10 only.

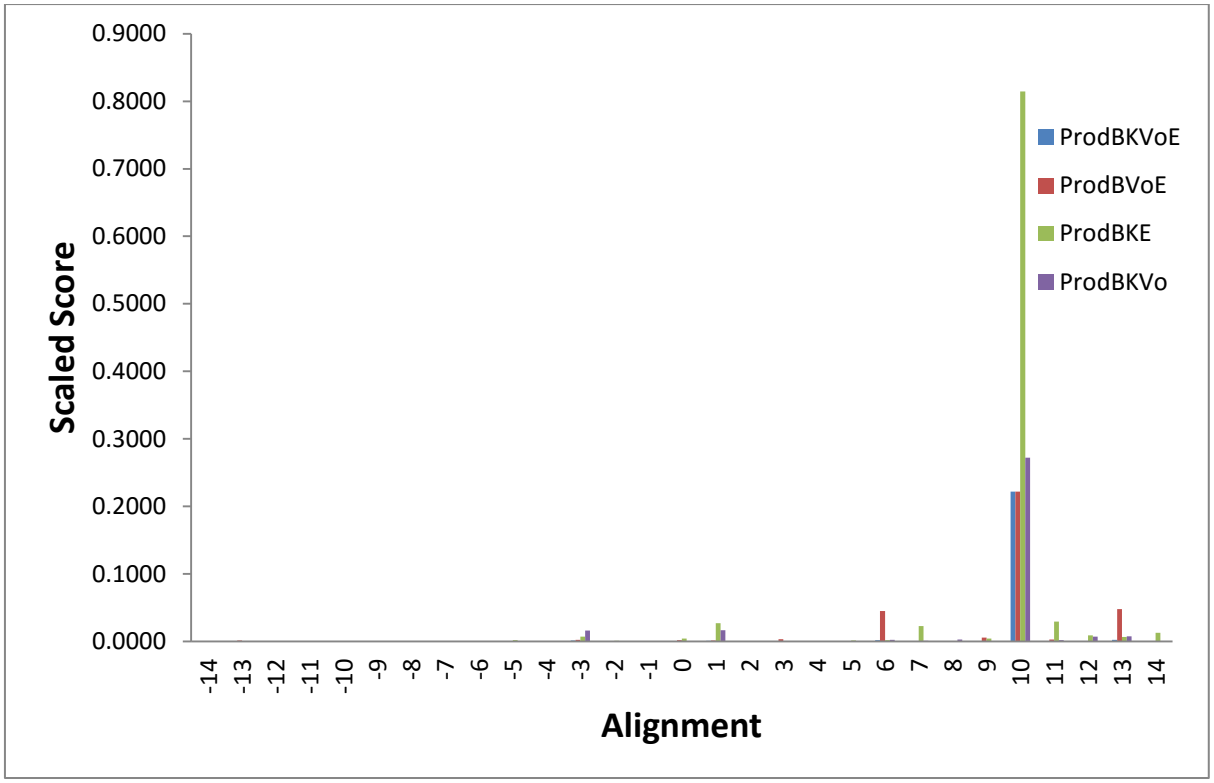

Figure 39: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the F-C TM5 region alignment.

Figure 40 also shows the alignment with the highest product score to be the +10 alignment.

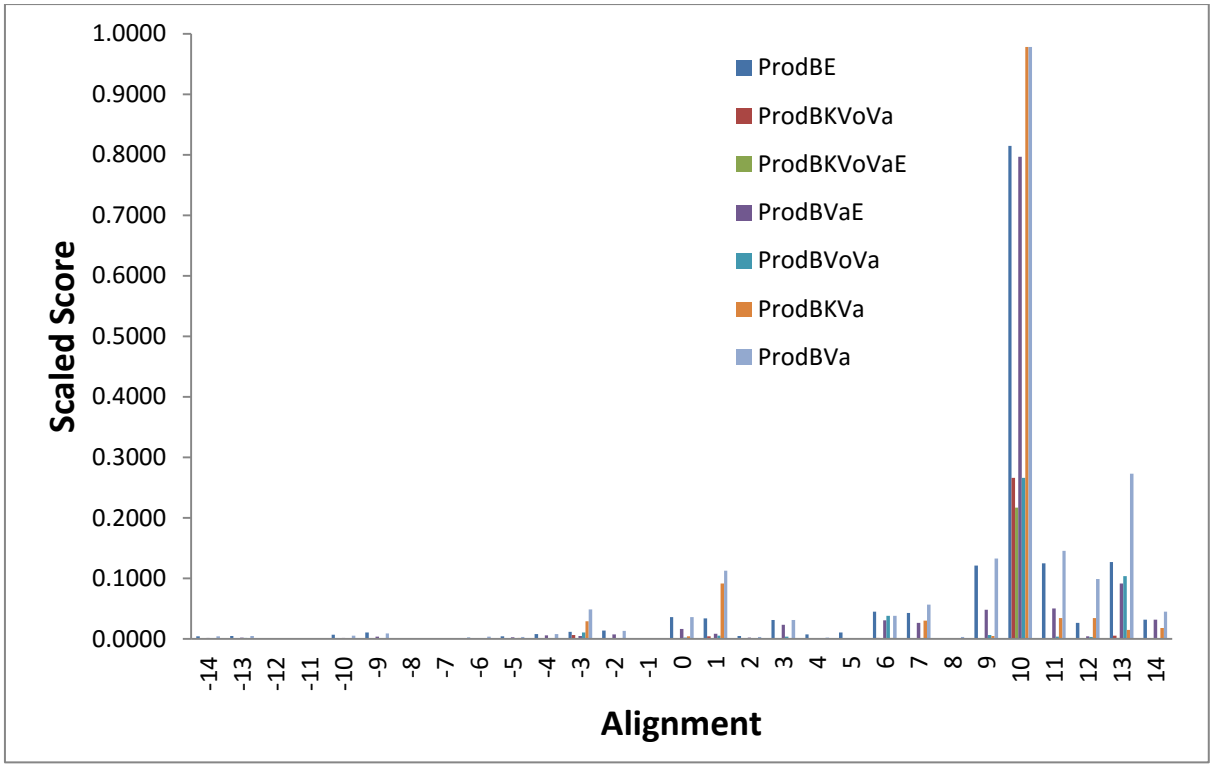

Figure 40: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the F-C TM5 alignment.

The multi-reference alignment is given in Chart 9. The Blosum matrix score favours alignment +1 for class A and class B, +5 for class E and +7 for class F. The percentage ID for class A, class B, class E and class F at the 0 alignment to class A is 6.9%, 9.7%, 6.3% and 7.0%. These figures are very low and it is unreasonable to expect the alignment method to work at such low percentage identities. The alignment, and the alternatives suggested by the methods are given in Chart 10.

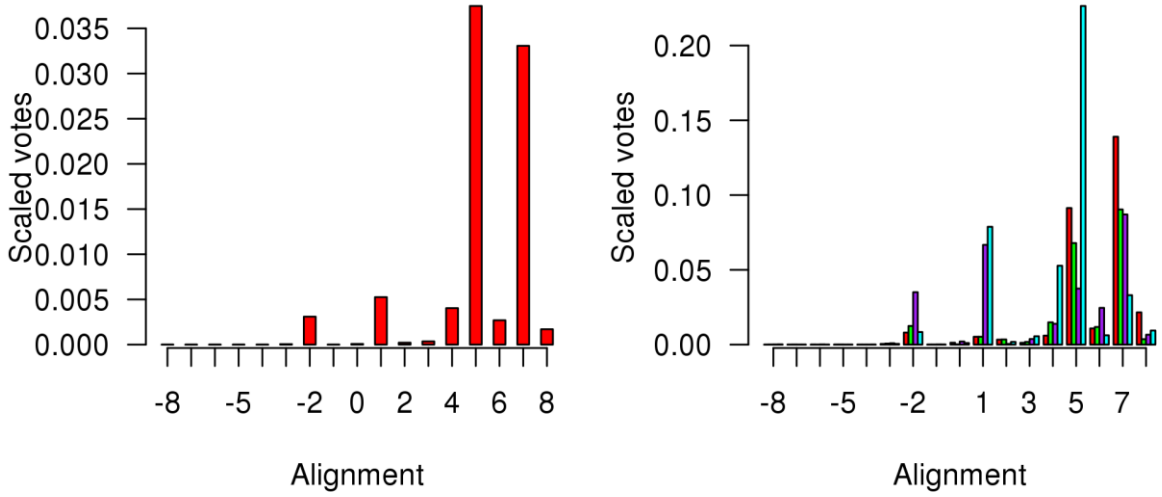

Chart 9: The multi-reference alignment scores. On the left, the score is given for multiplying the class A, class B, class E and class F scores together; on the right the score is given for missing out class A (red), class B (green), class E (purple) and class F (cyan).

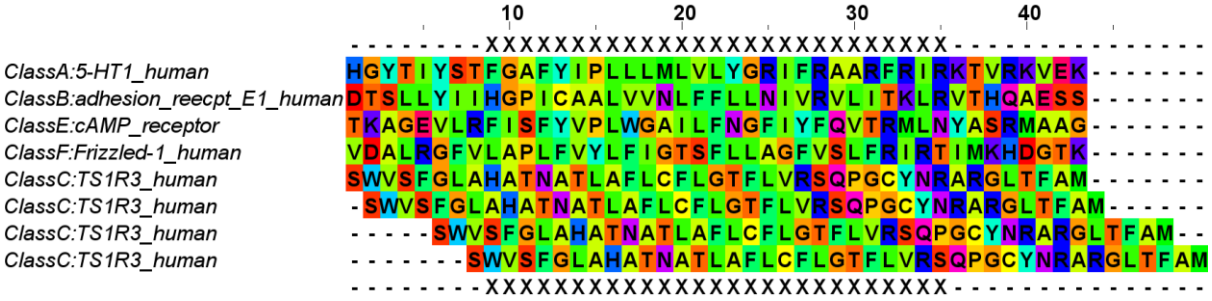

Chart 10: The alignment of the class C TM4 against TM4 for class A, class B, class E and class F. The class A - class B - class E - class F alignment is defined by the structural alignment of the relevant X-ray crystal structures. The correct 0 alignment (as defined by structural alignment) is given, along with the alternatives of +1, +5 and +7 indicated by chart 9. The Xs mark the window over which the alignment was determined. The colour scheme for the residues is the default Taylor colour scheme used by Jalview.

3.3.6. TM6

3.3.6.1. Class A – Class C

Figure 41 shows the product scores of the TM6 region of A-C. There is a

percentage identity of 14.9 between these two classes.

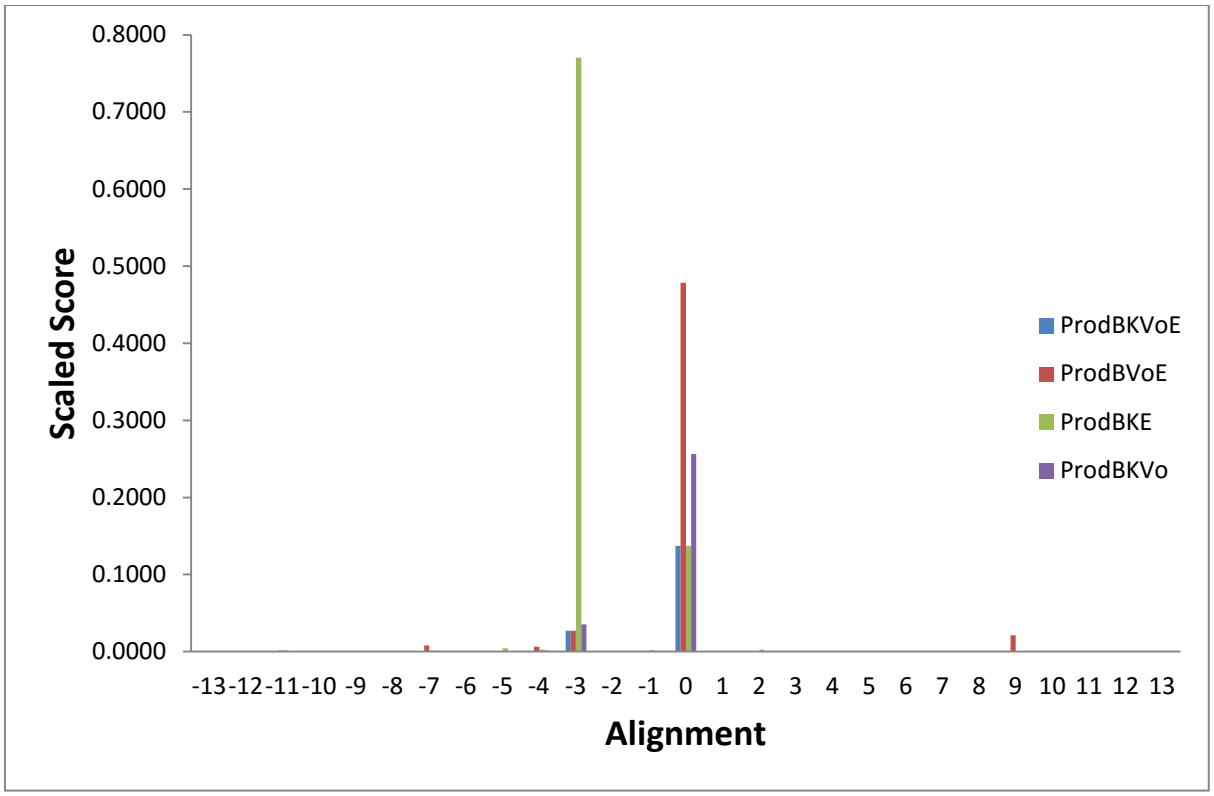

Figure 41: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the A-C TM6 region alignment.

The product scores seen in figure 41 and 42 show 0 to be the best alignment in all of the cases where the amino acid volume score is not included. When it is included the alignment -3 is seen as better.

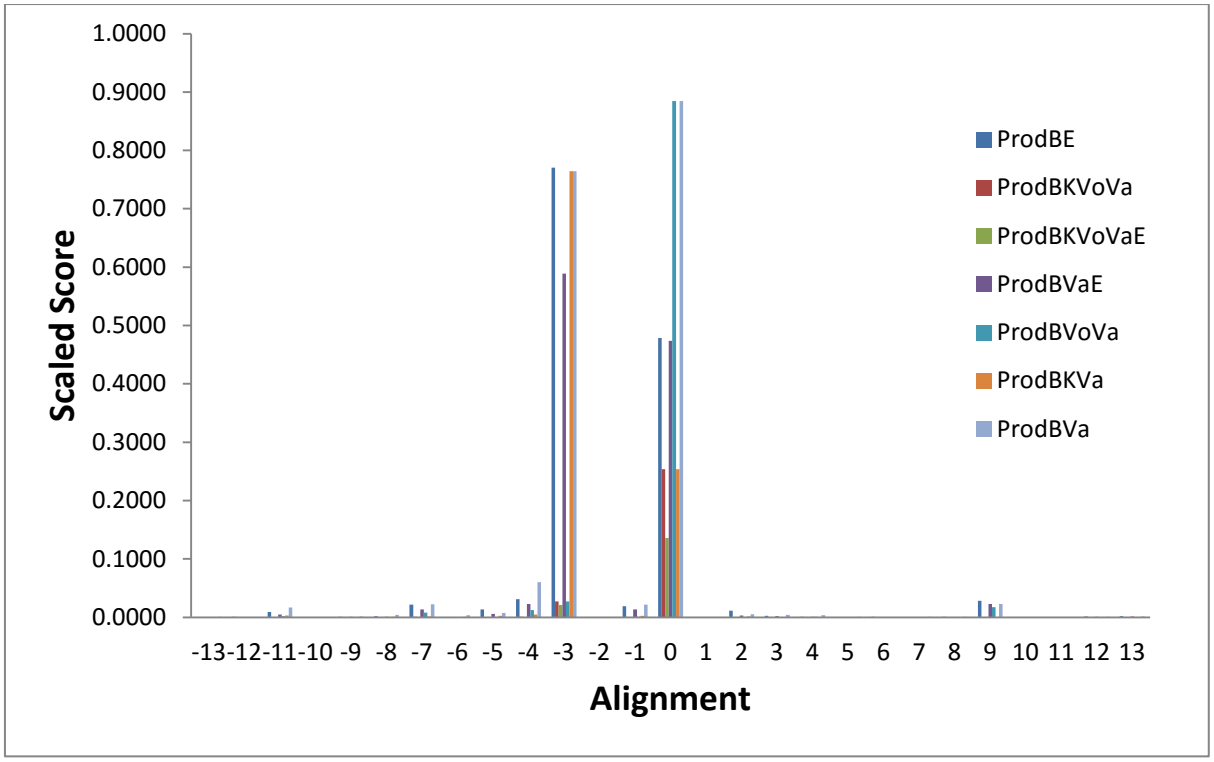

Figure 42: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the A-C TM6 alignment.

3.3.6.2. Class B – Class C

The product scores for B-C for TM6 are shown in figures 43 and 44. There is a percentage identity of 12.6 between these two classes.

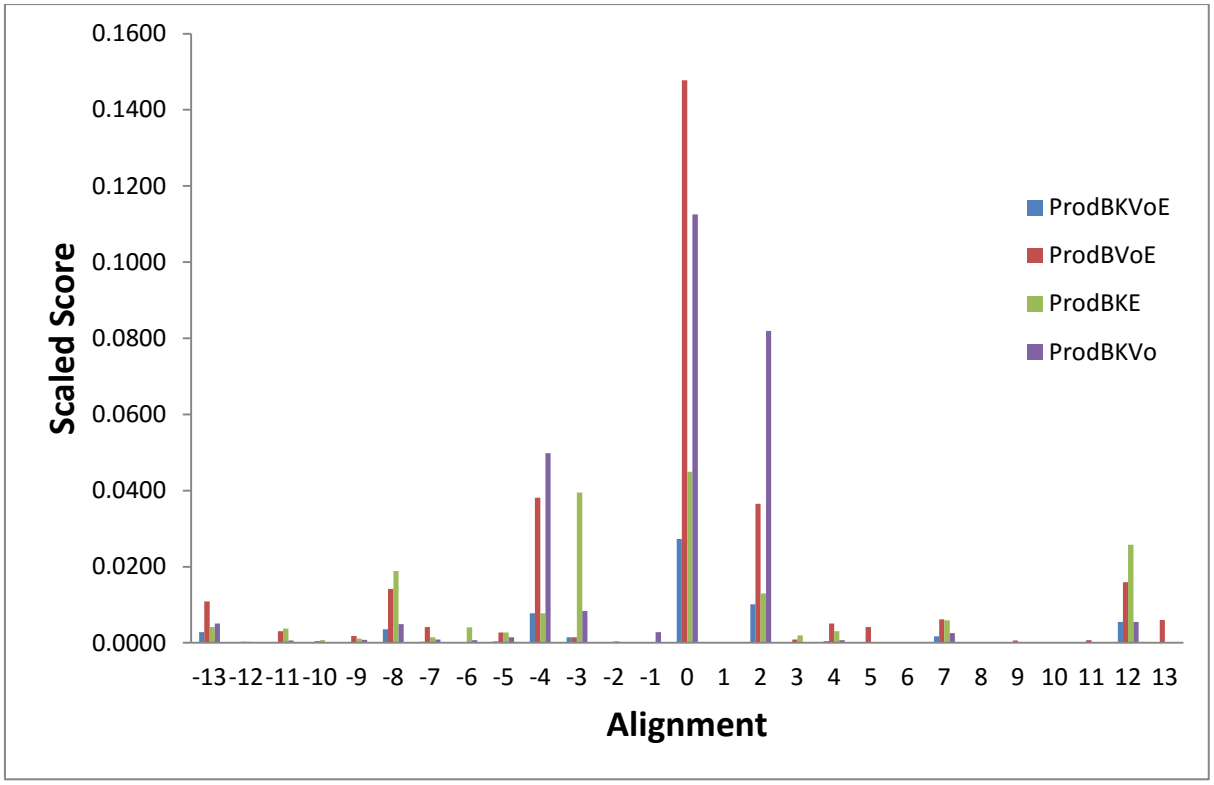

Figure 43: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the B-C TM6 region alignment.

The results, seen in figure 43 and figure 44 show that the best alignment between these two classes for TM6 is the 0 alignment. Although the results in figure 43 are very low the product scores in figure 44 are much higher and the result clearer.

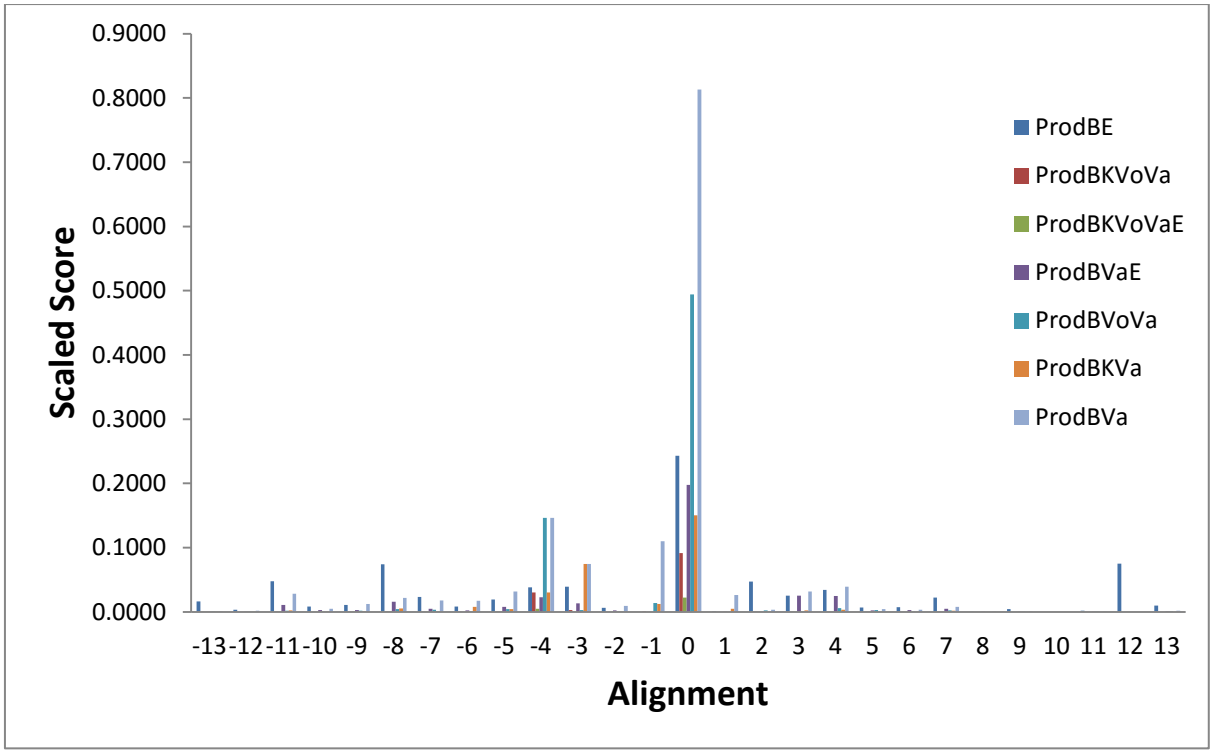

Figure 44: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the B-C TM6 alignment.

3.3.6.3. Class E – Class C

Figure 45 and 46 show product scores for the E-C alignment of TM6. There is a percentage identity of 11.5 between these two classes.

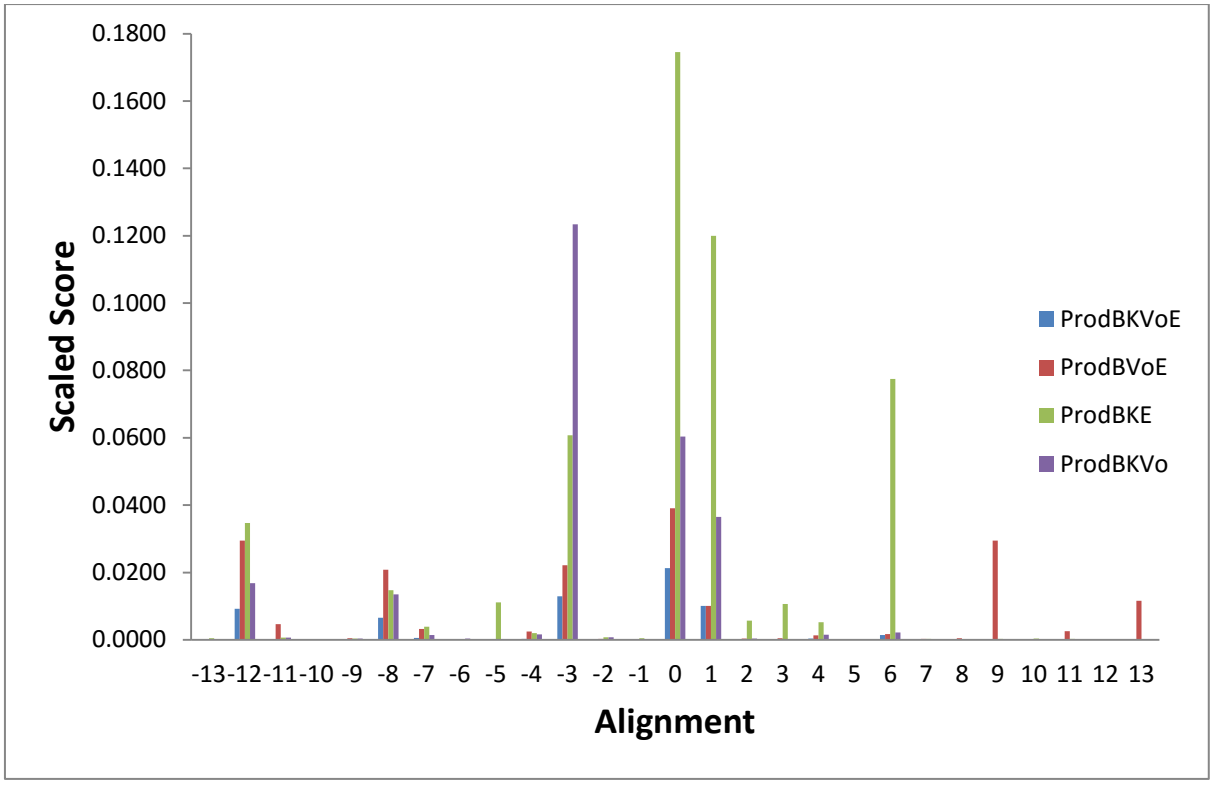

Figure 45: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the E-C TM6 region alignment.

The highest product scores seen in figures 45 and 46 are for the 0 alignment but again they are not particularly strong in figure 45. Stronger results are seen in figure 46.

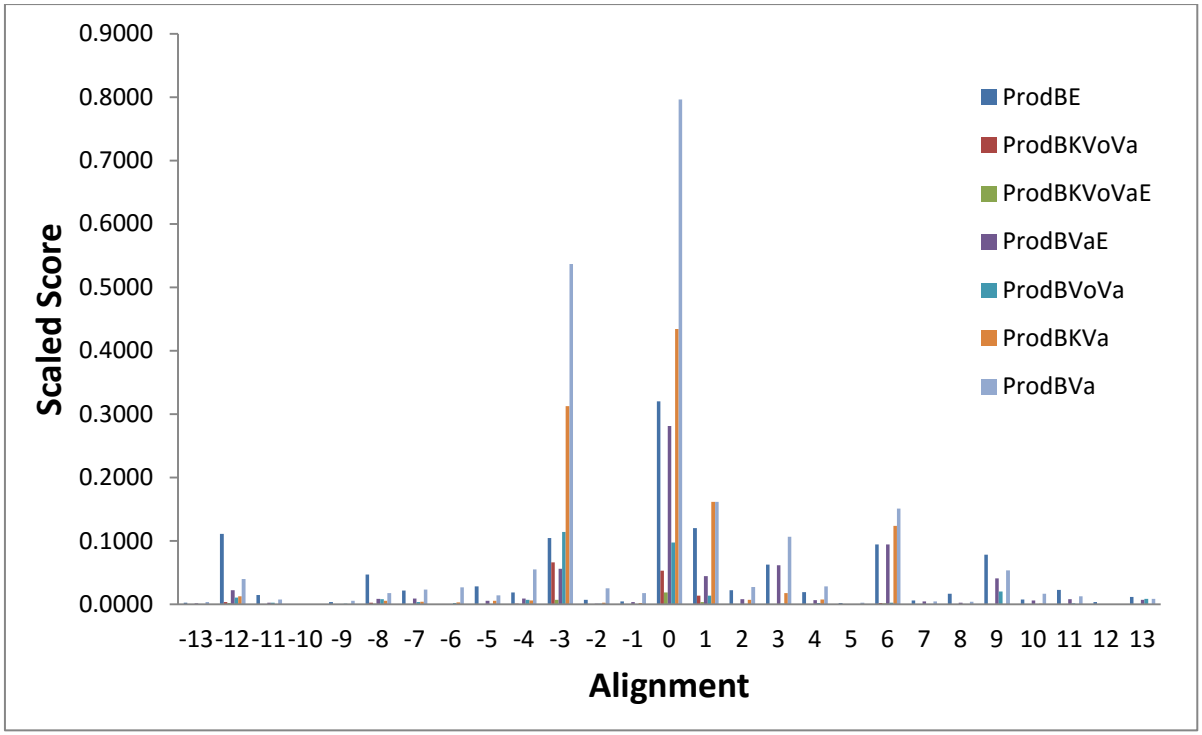

Figure 46: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the E-C TM6 alignment.

3.3.6.4. Class F – Class C

The product scores for the F-C alignment of TM6 are seen in figures 47 and 48 below. There is a percentage identity of 7.4 between these two classes.

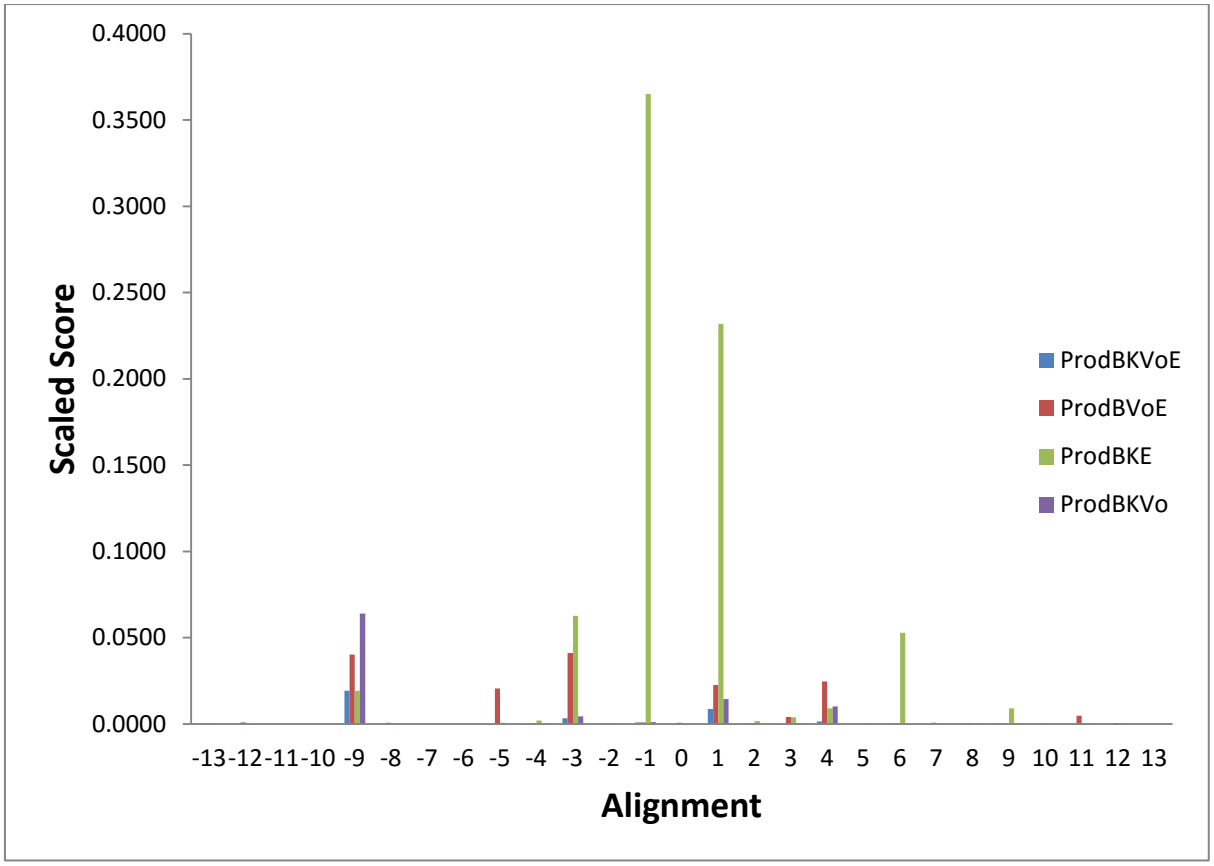

Figure 47: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the F-C TM6 region alignment.

The results for this are not particularly clear but show the highest product score at the alignment -1 and 1 in figure 47. Figure 48 shows significantly higher peaks at the alignments -3, +1 and +4.

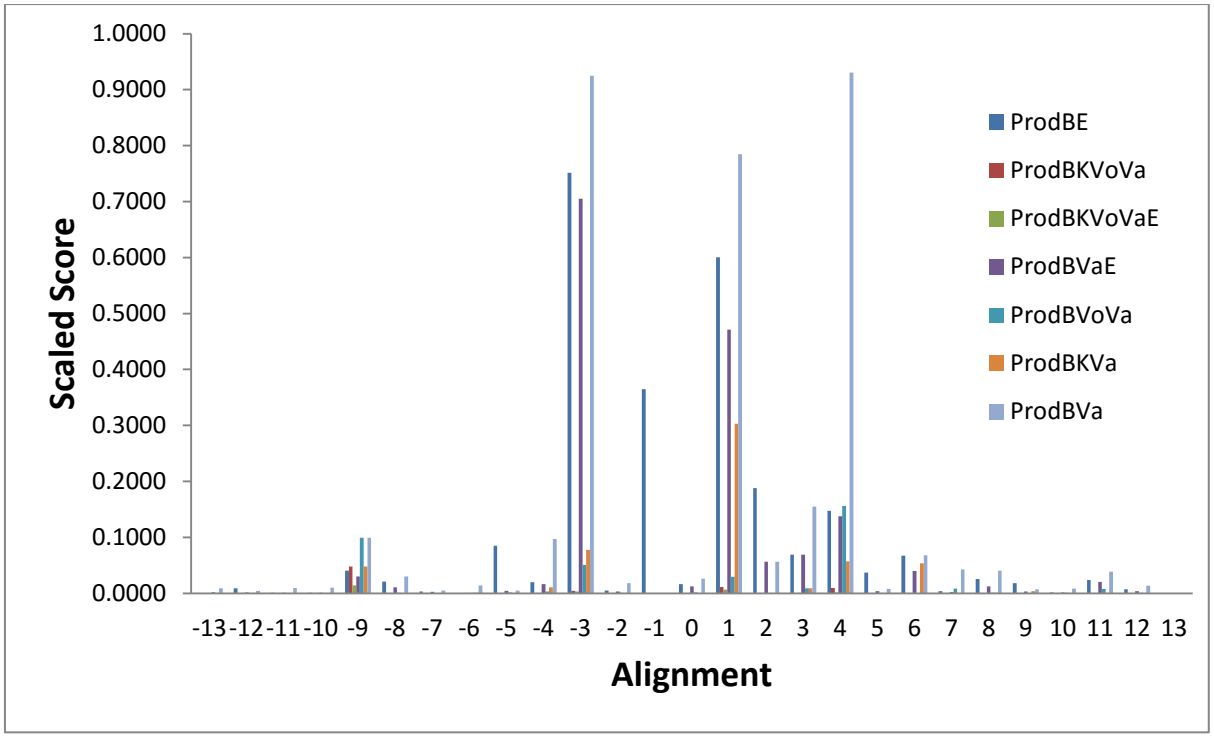

Figure 48: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the F-C TM6 alignment.

The multi-reference alignment is given in Chart 11. The Blosum matrix score is high for alignment 0 for class A, class B and class E, but class F has a low score for the correct (0) alignment and a high score for +1 and -3. Since class A and class E also give a high score for -3, this would indicate that -3 is the correct alignment. However, the percentage ID for class A, class B, class E and class F at the 0 alignment to class A is 14.9, 12.6, 11.5 and 7.4. The low percentage %ID for class F indicates that this ought to be ignored, as in the right hand part of chart 11; under these circumstances the method gives the correct result. The alignment is given in Chart 12.

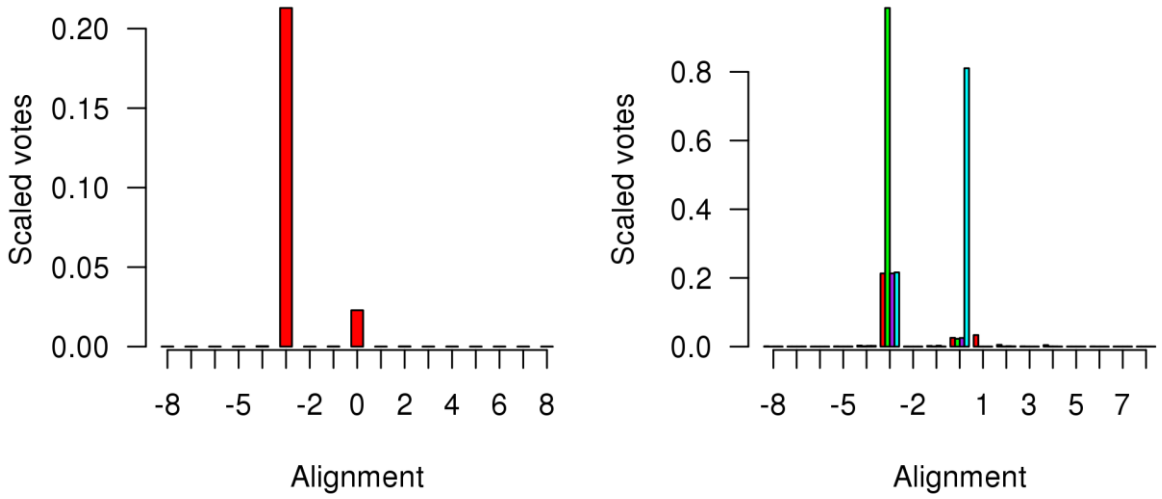

Chart 11: The TM6 multi-reference alignment scores. On the left, the score is given for multiplying the class A, class B, class E and class F scores together; on the right the score is given for missing out class A (red), class B (green), class E (purple) and class F (cyan).

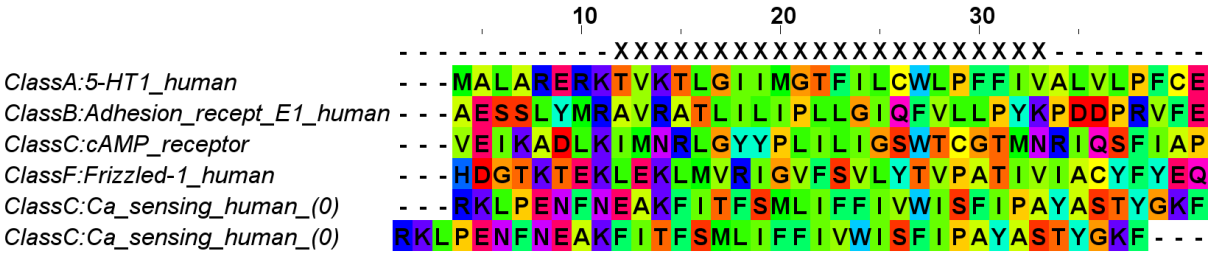

Chart 12: The alignment of the class C TM6 against TM6 for class A, class B, class E and class F. The class A - class B - class E - class F alignment is defined by the structural alignment of the relevant X-ray crystal structures. The correct 0 alignment (as defined by structural alignment) is given, along with the alternative of -3 indicated by Figures 41, 42 and 48, and chart 11. The Xs mark the window over which the alignment was determined. The colour scheme for the residues is the default Taylor colour scheme used by Jalview.

3.3.7. TM7

3.3.7.1. Class A – Class C

Figures 49 and 50 shows the product scores for B, K, Vo, Va and E for the TM7 region of the A-C alignment. There is a percentage identity of 14.5 between these

two classes at the 0 alignment. The results clearly show the 0 alignment to be the best alignment.

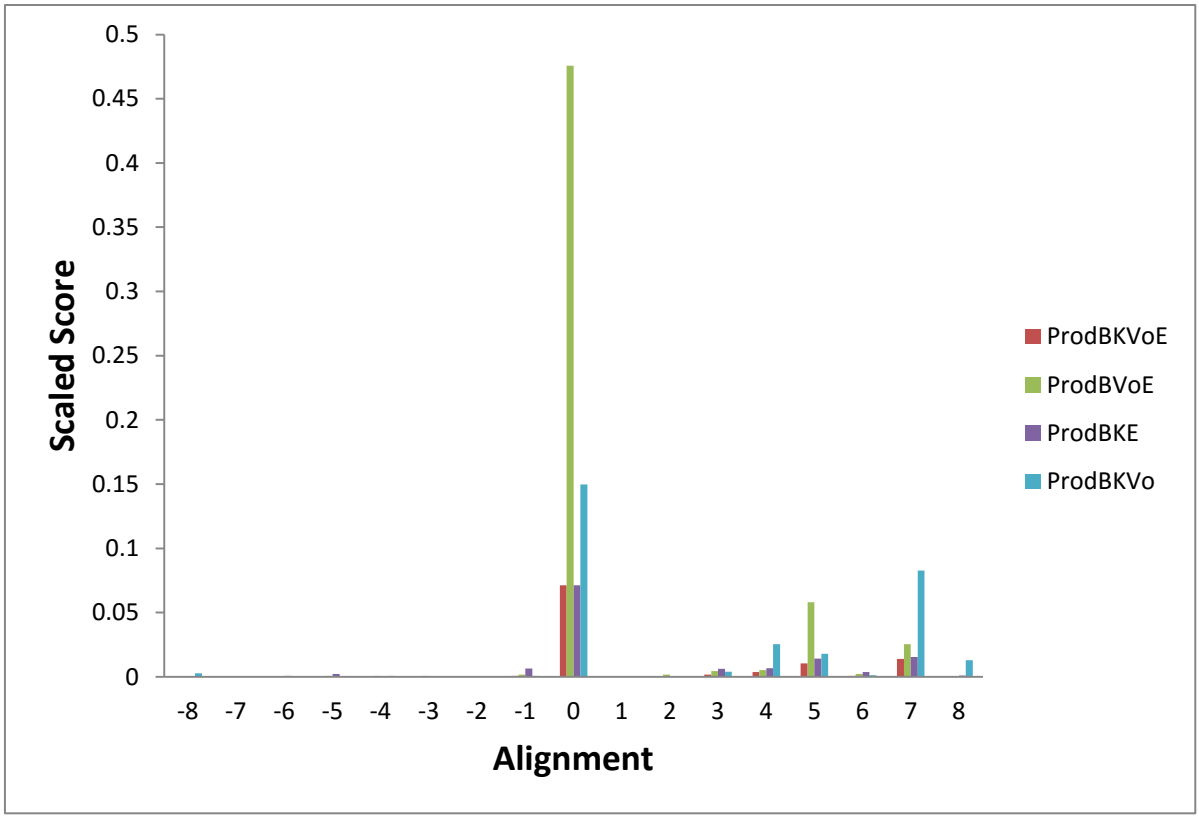

Figure 49: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the A-C TM7 region alignment.

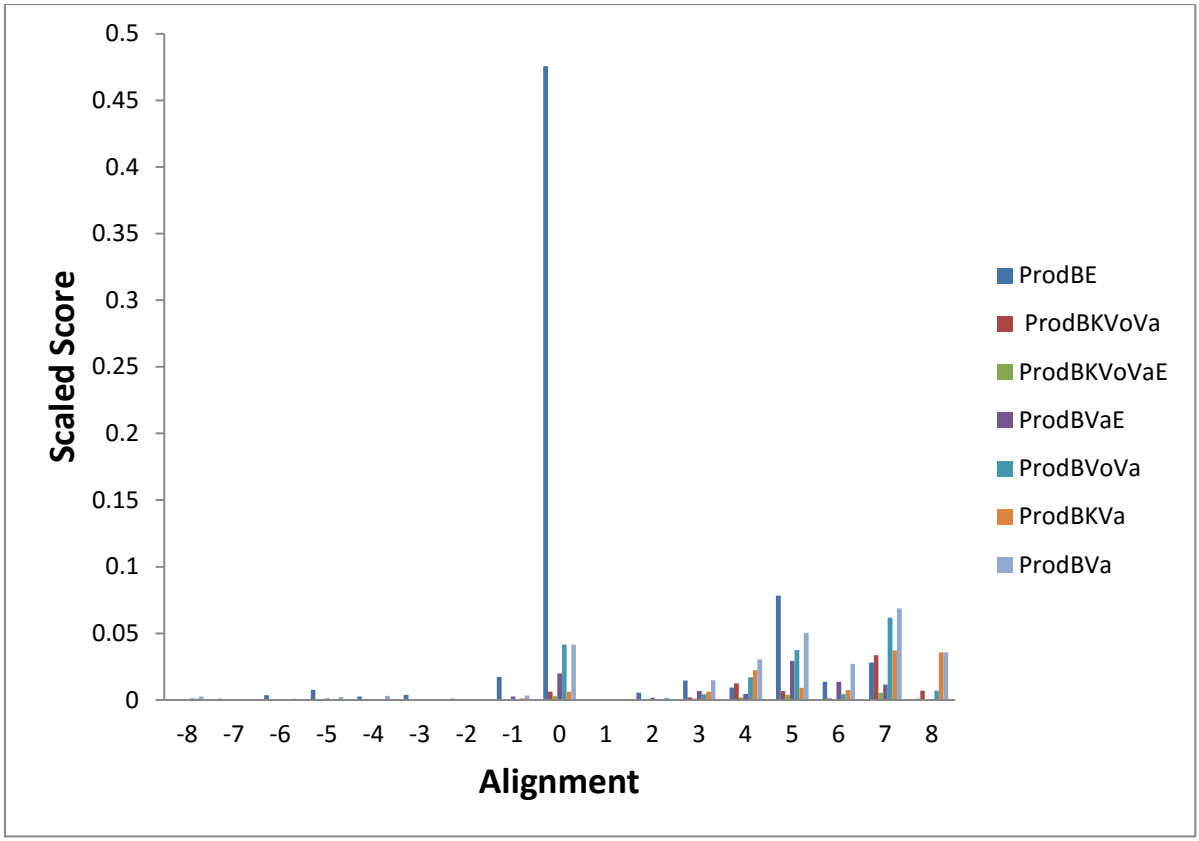

Figure 50: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the A-C TM7 alignment.

3.3.7.2. Class B – Class C

Figures 51 and 52 show the product scores of the TM1 B-C alignment. There is a percentage identity of 11.4 between these two classes at the 0 alignment. Figure 1 shows the +4 alignment to be the best with a small peak also at the +6 alignment. The results from figure 52 are not as clear but still show the +4 alignment as the best one which does not agree with the alignment of class C against class A.

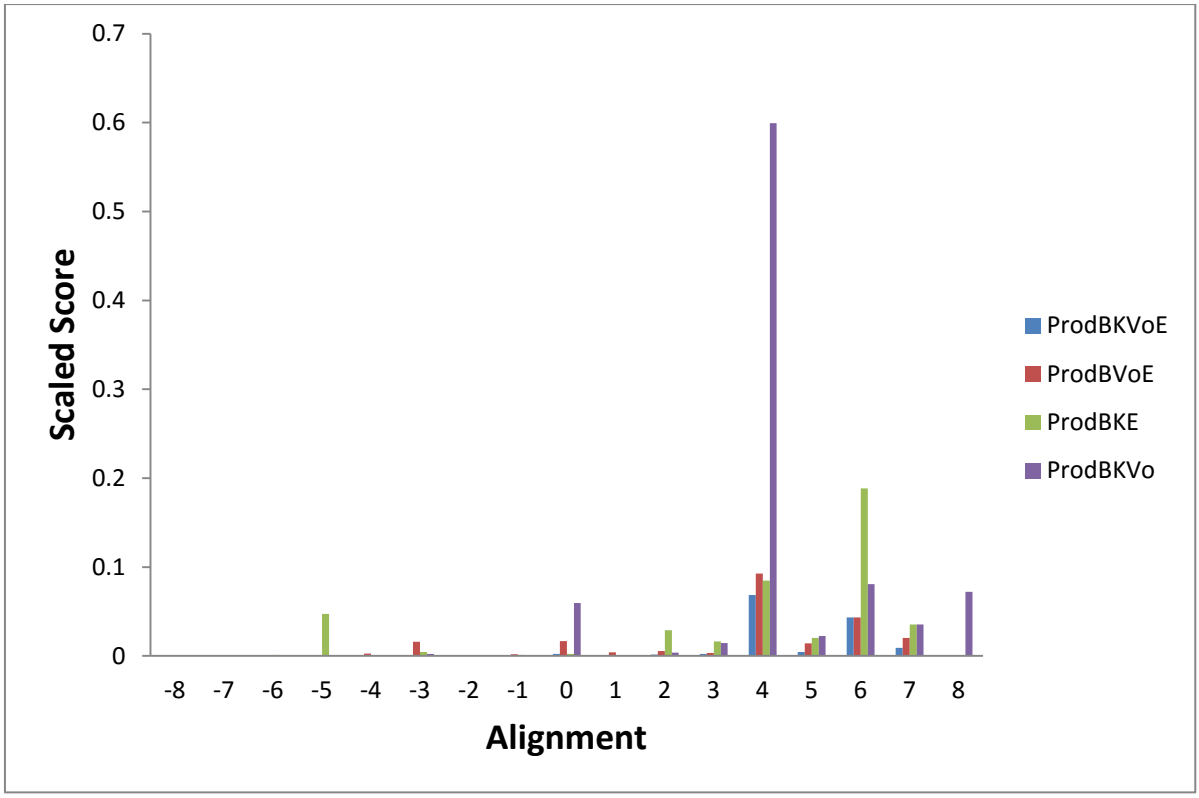

Figure 51: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the B-C TM7 region alignment.

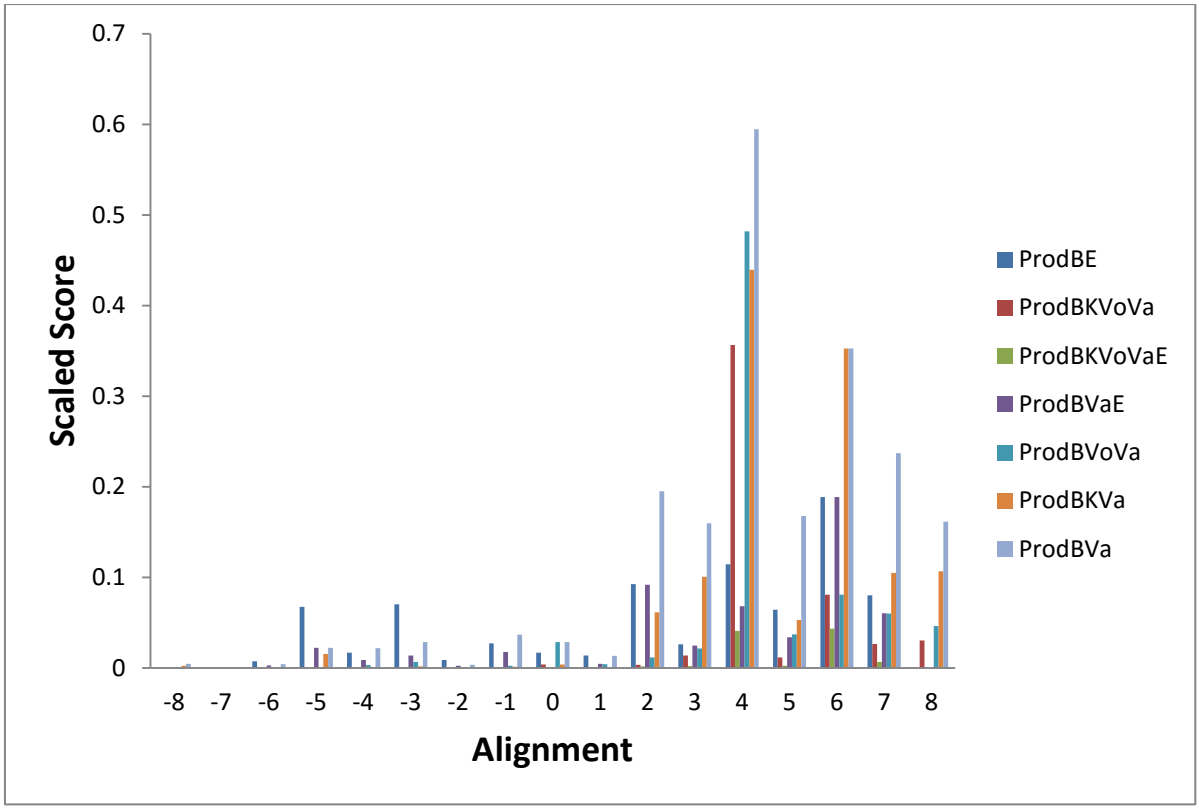

Figure 52: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the B-C TM7 alignment.

3.3.7.3. Class E – Class C

The product scores for the TM1 region of E-C are given in figures 53 and 54. There is a percentage identity of 12.8 between these two classes at the 0 alignment. The results clearly show the 0 alignment to be the best which agrees with the results of class C against class A.

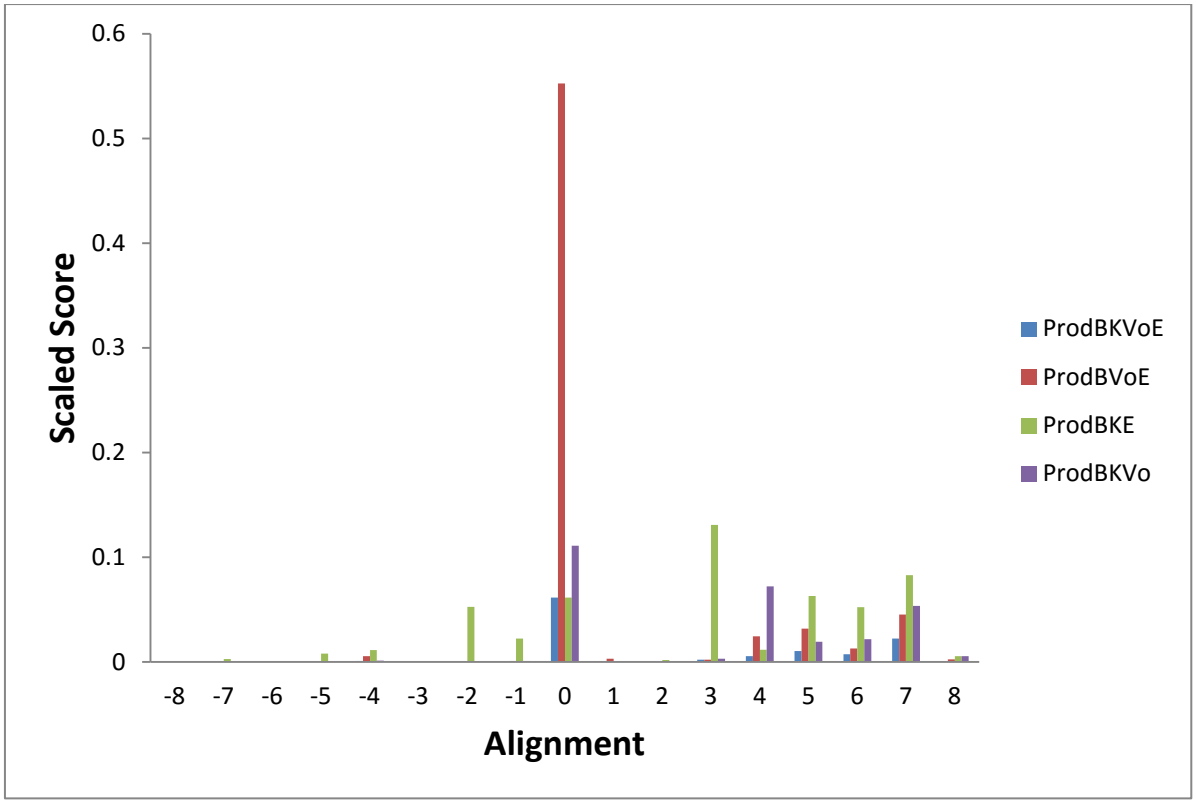

Figure 53: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the E-C TM7 region alignment.

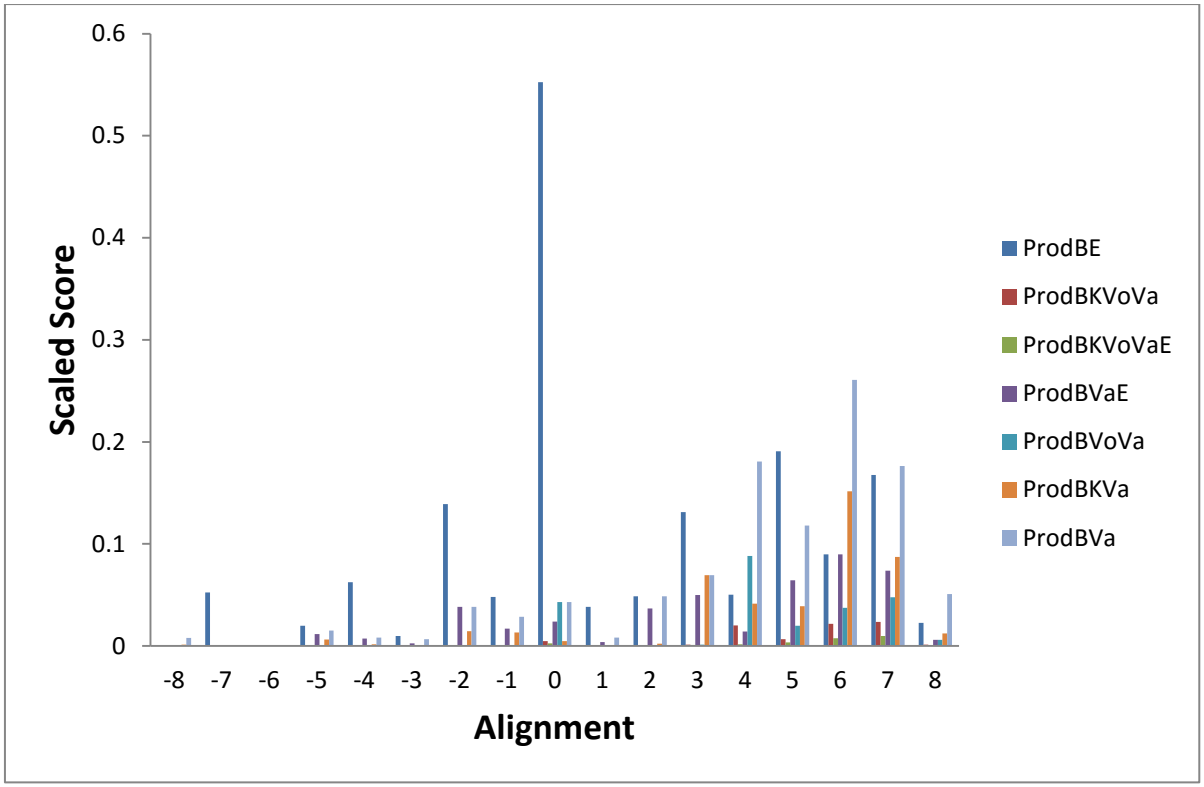

Figure 54: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the E-C TM7 alignment.

3.3.7.4. Class F – Class C

Figures 55 and 56 show the product of the scores of the TM7 alignment of class F and class C. The preferred alignment from these results is the+4 alignment. There is a percentage identity of 7.9 between these two classes at the 0 alignment. This is quite low which indicates the results of the alignment are not as reliable as the others.

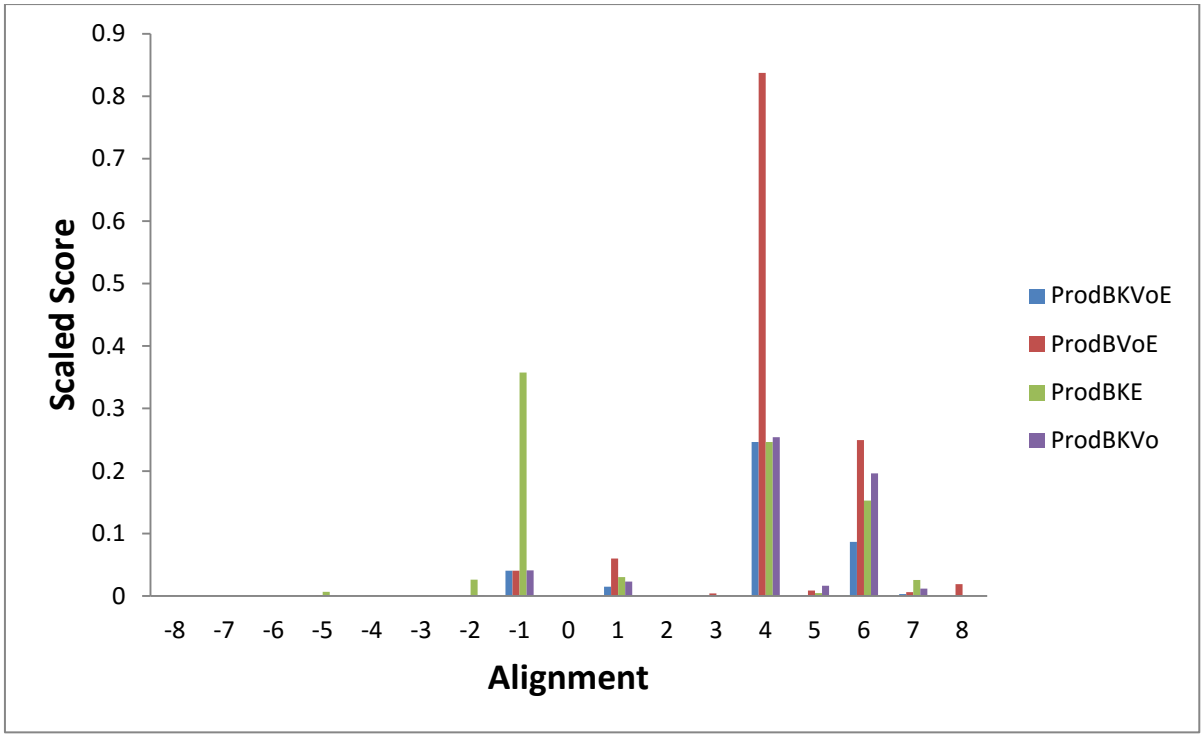

Figure 55: The product of the scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo) and entropy (E) for the F-C TM7 region alignment.

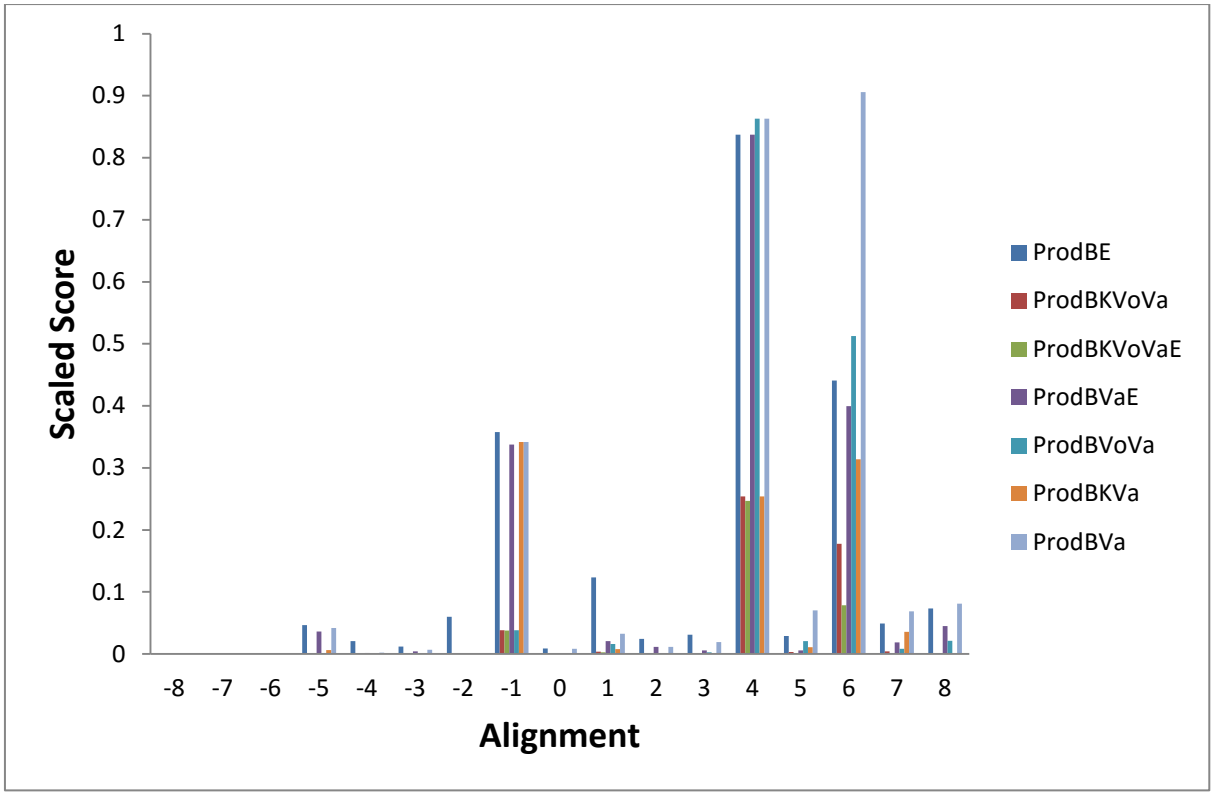

Figure 56: The product of scores of the BLOSUM substitution matrix (B), the Kyte-Doolittle hydrophobicity (K), amino acid volume (Vo), entropy (E) and Variability (Va) for the F-C TM7 alignment.

The multi-reference alignment is given in Chart 13. The Blosum matrix score favours alignment +4 with the 0 alignment also scoring well. The percentage ID for class A, class B, class E and class F at the 0 alignment to class A is 14.5%, 11.4%, 12.8% and 7.9%. These figures are acceptable apart from those for class F; when class F is ignored (rhs of Chart 13), the method gives the right answer, as the correct alignment gets a reasonable score for class A, class B and class E. The alignment, and the +4 alternative alignment as suggested by the methods, are given in Chart 14.

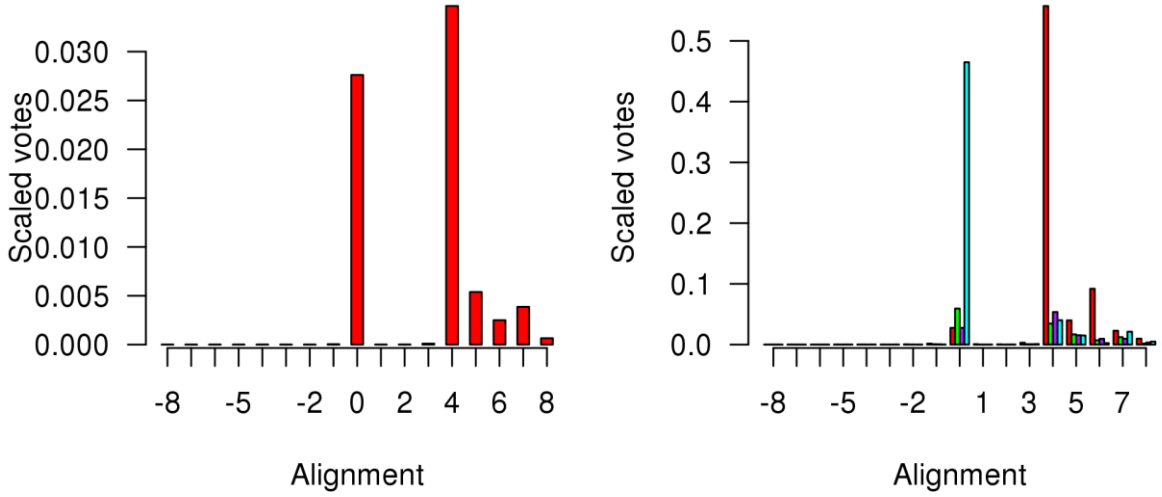

Chart 13: The multi-reference alignment scores. On the left, the score is given for multiplying the class A, class B, class E and class F scores together; on the right the score is given for missing out class A (red), class B (green), class E (purple) and class F (cyan).

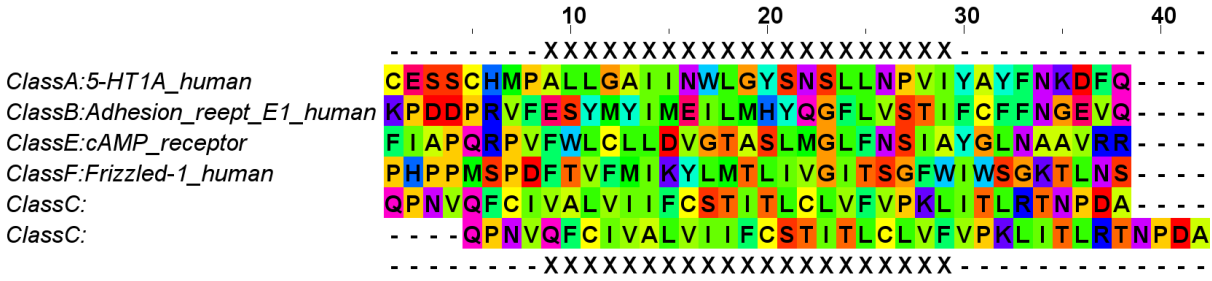

Chart 14: The alignment of the class C TM7 against TM7 for class A, class B, class E and class F. The class A - class B - class E - class F alignment is defined by the structural alignment of the relevant X-ray crystal structures. The correct 0 alignment (as defined by structural alignment) is given, along with the alternative of +4 indicated by chart 13. The Xs mark the window over which the alignment was determined. The colour scheme for the residues is the default Taylor colour scheme used by Jalview.

Concluding Remarks

The work carried out in this thesis started out as a way to predict the structure of Class C GPCRs as no structure had yet been found. Since the start of the project some Class C structures have been found so the focus of the study was altered to focus more on the accuracy of the method used.

If I had more time available it would have been useful to compare the results I gained with published results seen in the Vohra et al. (2013) paper to assess the difference made by using the HSE data. This would show how useful it is to determine the alignment window using this method. Using HSE data reduces the chance of including regions where the GPCR structures start to deviate from one another making the alignment more difficult and less accurate. It also indicates the TM regions for which the window length could be increased. A longer window length can give more accurate results as it uses more data to compare and determine the correct alignment. The results should therefore be clearer than when HSE data is not considered.

The results from the alignment generally agreed with the published structures but this is not always the case. The results for some TM regions were not as clear or as strong as expected. From the results seen the alignments where the %ID for the 0 alignment drops below 10% are the ones where this method does not give the correct alignment. However, traditional sequence alignment methods usually require %ID of 40% or higher to give accurate results. This method gives the correct answer in most cases where the %ID is over 10% which is a big improvement.

The alignment with class F GPCR's was the reason for incorrect results seen in more than one transmembrane regions and was usually the class with the lowest %ID at the 0 alignment. This indicates that class F has quite a low sequence homology with class C and the results from this alignment are less accurate than with the other classes. If I had more time I would investigate this further and determine whether class F should be included in this alignment method or not.

Also, if more time was available I would have liked to have created models for the structure of Class C GPCRs from the alignment results and compared these to the ones published to check the accuracy of the alignment results.

References

Baltoumas, F.A., Theodoropoulou, M. C. and Hamodrakas, S. J. (2013) Interactions of the α-subunits of heterotrimeric G-proteins with GPCRs, effectors and RGS proteins: a critical review and analysis of interacting surfaces, conformational shifts, structural diversity and electrostatic potentials. *J Struct Biol,* **182**, 209-218

Bettler, B., Kaupmann, K., Mosbacher, J. and Gassmann, M. (2004) Molecular structure and physiological functions of GABA(B) receptors. *Physiol Rev,* **84,** 835-867

Bissantz, C., Logean, A., and Rognan, D. (2004). High-Throughput Modeling of Human G-Protein Coupled Receptors: Amino Acid Sequence Alignment, Three-Dimensional Model Building and Receptor Library Screening. *Journal of Chemical Information and Modeling ,* **44**, 1162-1176.

Christopher, J. A., Aves, S. J., Bennett, K. A., Dore, A. S., Errey, J. C., Jazayeri, A., Marshall, F. H., Okrasa, K., Serrano-Vega, M. J., Tehan, B. G., Wiggin, G. R. and Congreve, M. (2015) Fragment and Structure-Based Drug Discovery for a Class C GPCR: Discovery of the $mGlu_5$ Negative Allosteric Modulator HTL14242 (3-Chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzonitrile). *J. Med. Chem.*, **58**, 6653-6664.

Chun, L., Zhang, W. H. and Liu, J. F. (2012) Structure and ligand recognition of class C GPCRs. *Acta Pharmacol Sin,* **33**, 312-323

Conner, A. C., Hay, D. L., Simms, J., Howitt, S. G., Schindler, M., Smith, D. M., Wheatley, M. and Poyner, D. R. (2005). A Key Role for Transmembrane Prolines in Calcitonin Receptor-Like Receptor Agonist Binding and Signalling: Implications for Family B G-Protein-Coupled Receptors. *Molecular Pharmacology ,* **67**, 20-31

Coopman, K., Wallis, R., Robb, G., Brown, A. J., Wilkinson, G. F., Timms, D. And Willars, G. B. (2011). Residues Within the Transmembrane Domain of the Glucagon-Like Peptide-1 Receptor Involved in Ligand Binding and Receptor Actication: Modelling the Ligand-Bound Receptor. *Molecular Endocrinology ,* **25**, 1804-1818

de Graaf, C., Rein, C., Piwnica, D., Giordanetto, F. and Rognan, D. (2011). Structure-Based Discovery of Allosteric Modulators of Two Related Class B G-Protein-Coupled Receptors. *ChemMedChem ,* **6**, 2159-2169

Dong, M., Lam, P. C., Gao, F., Hosohata, K., Pinon, D. I., Sexton, P. M., Abagyan, R. and Miller, L. J. (2007). Molecular Approximations between Residues 21 and 23 of Secretin and Its Receptor: Development of a Model for Peptide Docking with the Amino Terminus of the Secretin Receptor. *Molecular Pharmacology ,* **72**, 280-290

Donnelly, D. (1997). The Arrangement of the Transmembrane Helices in the Secretin Receptor Family of G-Protein-Coupled Receptors. *FEBS Letters ,* **409**, 431-436

Doré, A. S., Okrasa, K., Jayesh C. Patel, J. C., Serrano-Vega, M., Bennett, K., Cooke, R. M., Errey, J.C., Jazayeri, A., Khan, S., Tehan, B., Weir, M., Wiggin, G. R. and Marshall, F. H. (2014) Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature.* **511**, 557–562

Eswar, N., Eramian, D., Webb, B., Shen, M. Y., and Sali, A. (2008). Protein Structure Modeling with MODELLER. *Structural Proteonomics ,* **426**, 145-159.

Foster, R. H. and Goa, K. L. (1998) Risperidone. A pharmacoeconomic review of its use in schizophrenia. *Pharmacoeconomics,* **14**, 97-133

Fredriksson, R. and Schioth, H. B. (2005) The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol Pharmacol,* **67**, 1414-1425

Frimurer, T. and Bywater, R. (1999). Structure of the Integral Membrane Domain of the GLP1 Receptor. *Proteins: Structure, Function and Bioinformatics ,* **35**, 375-386.

Geng, Y., Bush, M., Mosyak, L., Wang, F. and Fan, Q. R. (2013) Structural mechanism of ligand activation in human GABA(B) receptor. *Nature,* **504**, 254-259

Goodman, W. G. (2004) Calcium-sensing receptors. *Semin Nephrol,* **24**, 17-24

Hamelryck, T. (2005). An Amino Acid has Two Sides: A New 2D Measure Provides a Different View of Solvent Exposure. *Proteins: Structure, Function, and Bioinformatics*, **59**, 38-48

Henikoff, S. and Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *PNAS*, **89**, 10915-10919

Hollenstein, K., Kean, J., Bortolato, A., Cheng, R. K., Dore, A. S., Jazayeri, A., Cooke, R. M., Weir, M. and Marshall, F. H. (2013). Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature ,* **499**, 438-443.

Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Lizerman, A. P. and Stevens, R. C. (2008). The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science ,* **322**, 1211-1217.

Johnson, R. L., Saxe, C. L., Gollop, R., Kimmel, A. R. and Devreotes, P. N. (1993) Identification and targeted gene disruption of cAR3, a cAMP receptor subtype expressed during multicellular stages of Dictyostelium development. *Genes Dev,* **7**, 273-282

Julio-Pieper, M., Flor, P. J., Dinan, T. G. and Cryan, J. F. (2011) Exciting times beyond the brain: metabotropic glutamate receptors in peripheral and non-neural tissues. *Pharmacol Rev,* **63**, 35-58

Kang, S. G., Das, P., McGrane, S. J., Martin, A. J., Huynh, T., Royyuru, A., Taylor, A., Jones, P. G.and Zhou, R. (2014) Molecular Recognition of Metabotropic Glutamate Receptor Type 1 (mGluR1): Synergistic Understanding with Free Energy Perturbation and Linear Response Modeling. *J Phys Chem B*, **118**, 6393-6404

Kleuss, C., Scherübl, H., Hescheler, J., Schultz, G. and Wittig, B. (1993) Selectivity in signal transduction determined by gamma subunits of heterotrimeric G proteins. *Science,* **259**, 832-834

Kroeze, W. K., Sheffler, D. J. and Roth, B. L. (2003) G-protein-coupled receptors at a glance. *J Cell Sci,* **116**, 4867-4869

Kubo, Y. and Tateyama, M. (2005) Towards a view of functioning dimeric metabotropic receptors. *Curr Opin Neurobiol,* **15**, 289-295

Kyte, J. and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105-132

Lagerstrom, M. C. and Schioth, H. B. (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov,* **7**, 339-357

Lock, A., Forfar, R., Weston, C., Bowsher, L., Upton, G. J. G., Reynolds, C. A., Ladds, G. and Dixon, A. M. (2014). One motif to bind them: A small-XXX-small motif affects transmembrane domain 1 oligomerization, function, localization, and cross-talk between two yeast GPCRs. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, **1838**, 3036-3305

Maeda, T., Imanishi, Y. and Palczewski, K. (2003) Rhodopsin phosphorylation: 30 years later. *Prog Retin Eye Res,* **22**, 417-434

Marshall, F. H. and Foord, S. M. (2010) Heterodimerization of the GABAb Receptor – Implications forGPCR signaling and drug discovery. *Adv Pharmacol*, **58**, 63-91

Marshall, F. H. (2007) The role of GABAb receptors in the regulation of excitatory neurotransmission. *Results Probl Cell Differ*, **44**, 87-98

Neer, E. J. and Smith, T. F. (1996) G protein heterodimers: new structures propel new questions. *Cell,* **84**, 175-178

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., Miyano, M. (2000) Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science ,* **289**, 739-745

Sheikh, S. P., Vilardarga, J. P., Baranski, T. J., Lichtarge, O., Liri, T., Meng, E. C., Nissenson, R. A. and Bourne, H. R. (1999). Similar Structures and Shared Switch Mechanisms of the beta2-Adrenoceptor and the Parathyroid Hormone Receptor. Zn(II) Bridges Between Helices III and VI Block Activation. *The Journal of Biological Chemistry ,* **274**, 17033-17041

Siu, F. K., Lam, I. P., Chu, J. Y.and Chow, B. K. (2006) Signaling mechanisms of secretin receptor. *Regul Pept,* **137**, 95-104

Siu, F. Y., He, M., de Graaf, C., Han, G. W., Yang, D., Zhang, Z., Zhou, C., Xu, Q., Wacker, D., Joseph, J. S., Liu, W., Lau, J., Cherezov, V., Katritch, V., Wang, M. W. and Stevens, R. C. (2013). Structure of the human glucagon class B G-protein-coupled receptor. *Nature ,* **499**, 444-449.

Speyer, C. L., Hachem, A. H., Assi, A. A., Johnson, J. S., Devries, J. A.and Gorski, D. H. (2014) Metabotropic glutamate receptor-1 as a novel target for the antiangiogenic treatment of breast cancer. *PLoS One,* **9**, e88830

Stoveken, H. M., Hajduczok, A. G., Xu, L. and Tall, G. G. (2015) Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *PNAS*, **112**, 6194-6199

Taddese, B., Upton, G. J., Bailey, G. R., Jordan, S. R., Abdulla, N. Y., Reeves, P. J. and Reynolds, C. A. (2014). Do Plants Contain G Protein-Coupled Receptors. *Plant Physiology ,* **164**, 287-307.

Taylor, W. R., Munro, R. E., Petersen, K. and Bywater, R. P. (2003). Ab Initio Modelling of the N-Terminal Domain of the Secretin Receptors. *Computational Biology and Chemistry ,* **27**, 103-114.

Tfelt-Hansen, J. and Brown, E. M. (2005) The calcium-sensing receptor in normal physiology and pathophysiology: a review. *Crit Rev Clin Lab Sci,* **42,** 35-70

Trivedi, R., Mithal, A. and Chattopadhyay, N. (2008). Recent updates on the calcium-sensing receptor as a drug target. *Curr Med Chem,* **15**, 178-186

van der Kant, R. and Vriend, G. (2014). Alpha-Bulges in G Protein-Coupled Receptors. *Int. J. Mol. Sci.,* **15**, 7841-7864

Vinson, P. N. and Conn, P. J. (2012) Metabotropic glutamate receptors as therapeutic targets for schizophrenia. *Neuropharmacology,* **62**, 1461-1472

Vohra, S., Taddese, B., Conner, A. C., Poyner, A. C., Hay, D. L., Barwell, J., Reeves, P. J., Upton, G. J. and Reynolds, C. A. (2013). Similarity between class A and class B G-protein-coupled recepors exemplified through calcitonin gene-related peptide receptor modelling and mutagenesis studies. *J R Soc Interface*, **10**, 20120846.

Wall, M. A., Coleman, D. E., Lee, E., Iñiguez-Lluhi, J. A., Posner, B. A., Gilman, A. G.and Sprang, S. R. (1995) The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell,* **83**, 1047-1058

Wang, C., Wu, H., Katritch, V., Han, G. W., Huang, X.-P., Liu, W., Siu, F. Y., Roth, B. L., Cherezov, V. and Stevens, R. C. (2013). Structure of the human smoothened receptor bound to an antitumour agent. *Nature ,* **497**, 338-343.

Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G. and Schertler, G. F. (2008). Structure of a Beta1-Adrenergic G-Protein-Coupled Receptor. *Nature ,* **454**, 486-491.

Watkins, H. A., Chakravarthy, M., Abhayawardana, R. S., Gingell, J. J., Garelja, M., Pardamwar, M., McElhinney, J. M. W. R., Lathbridge, A., Constantine, A., Harris, P. W. R., Yuen, T., Brimble, M. A., Barwell, J., Poyner, D. R., Woolley, M. J., Conner, A. C., Pioszak, A. A., Reynolds, C. A. and Hay, D. L. (2016) Receptor activity-modifying proteins 2 and 3 generate adrenomedullin receptor subtypes with distinct molecular properties. *J Biol Chem,* **291**, 11657-11675.

White, J. F., Noinaj, N., Shibata, Y., Love, J., Kloss, B., Xu, F., Gvozdenovic-Jeremic, J., Shah, P., Shiloach, J., Tate, C. G. and Grisshammer, R. (2012) Structure of the agonist-bound neurotensin receptor. *Nature*, **490**, 508-513

Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M. Cherezov, V. and Stevens, R. C. (2010). Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science ,* **330**, 1066-1071.

Wu, H., Wang, C., Gregory, K. J., Han, G. W., Cho, H. P., Xia, Y., Niswender, C. M., Katritch, V., Meiler, J., Cherezov, V., Conn, P. J. and Stevens, R. C. (2014) Structure of a Class C GPCR Metabotropic Glutamate Receptor 1 Bound to an Allosteric Modulator, *Science,* **344**, 58-64.

Xu, C., Zhang, W., Rondard, P., Pin, J. P. and Liu, J. (2014) Complex GABAB receptor complexes: how to generate multiple functionally distinct units from a single receptor. *Front Pharmacol,* **5**, 12

Xue, C., Hsueh, Y. P. and Heitman, J. (2008) Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in fungi. *FEMS Microbiol Rev,* **32**, 1010-1032

Zhang, C., Srinivasan, Y., Arlow, D. H., Fung, J. J., Palmer, D., Zheng, Y., Green, H. F., Pandey, A., Dror, R. O., Shaw, D. E., Weis, W. I., Coughlin, S. R. And Kobilka, B. K. (2012). High-resolution crystal structure of human protease-activated receptor 1. *Nature ,* **492,** 387-392

Appendix

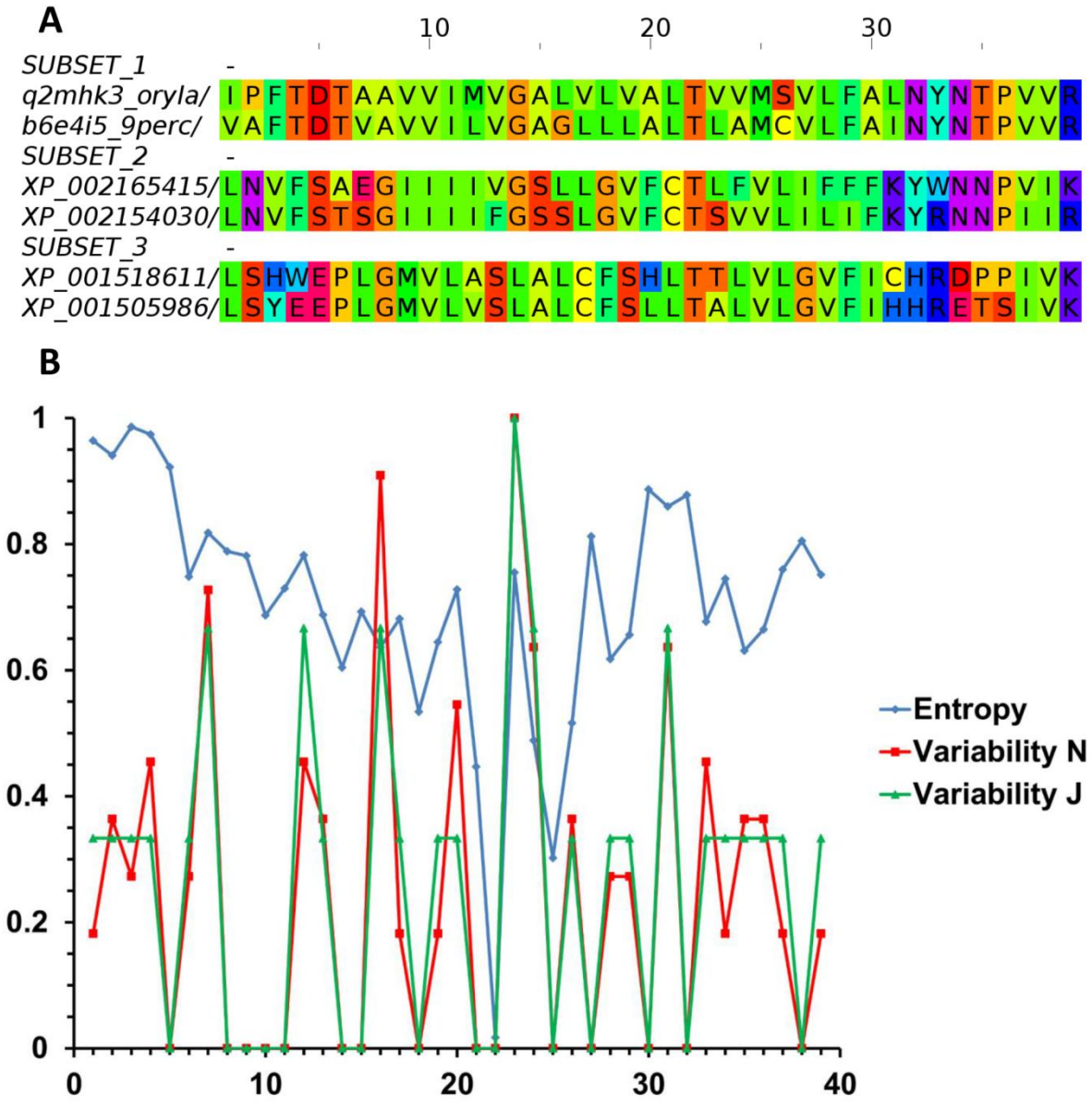

Figure A1. A. The three subsets of highly similar sequences generated at a similarity level of 70% (i.e. the pairs of sequences have a percentage identity of 70% or greater within each subset, but a percentage identity of less than 70% between subsets. The variability is taken from the differences within a subset (Vohra et al., 2013). B. The variation of entropy and variability along TM1 and its flanking regions. Variability N is based on the number of amino acid differences, regardless of the nature of the mutation while Variability J is based on the number of differences in the properties, as defined in the Jalview conservation score. Variability N is the form used in the thesis.