

HW 2: Introduction to homology and structural concepts

Zackery Field

February 3, 2014

Part 1. Interpreting protein multi-domain architectures and degrees of structural relatedness.

1. – Chinese scorpion toxin (P45697, PDB 1SN1)
– Drosomycin (P41964, PDB 1MYN).

RCSB and SCOP both give evidence that Chinese scorpion toxin and Drosomycin inducible antifungal protein have the same MDA. SCOP reveals that both proteins are in the same superfamily, [Scorpion toxin-like](#), although not in the same family. This implies that there is either structural or mechanistic similarity, but no evident sequence similarity (from [Protein superfamily wiki page](#)). RCSB supports the claim that SCOP makes that these proteins are in the same superfamily. Using the JFATCAT-flexible structural alignment algorithm in RCSB, a p-value of 2.12e-03 was calculated which implies significant structural similarity. Visually, the overlaid structures reveal that the major secondary structure motifs: a single α -helix and 3 strands of β sheet, almost entirely overlap. Furthermore, the entries in the Pfam database for these proteins reveals that these proteins are in the same clan, [Knottin_1](#). The shared clan, and structural similarity imply that these proteins share the same MDA.

2. – Chinese scorpion toxin (P45697, PDB 1SN1)
– Radish defensin (P69241, PDB 1AYJ)

The relatedness of Chinese scorpion toxin and radish defensin is very similar to the relatedness of Chinese scorpion toxin and Drosomycin. Drosomycin and radish defensin are in the same family, and have very similar structures. As with the above analysis, SCOP shows that they share the same superfamily, [Scorpion toxin-like](#). A JFATCAT-flexible structural analysis by RCSB reveals a p-value of 7.87e-05, implying even more similarity between 1SN1 and 1AYJ than 1SN1 and 1MYN. Pfam reveals that both of these proteins are in the same clan, [Knottin_1](#). The similarity in evidence between the 1st comparison and this lead to the conclusion that 1SN1 and 1AYJ also share the same MDA.

MDA:

3. – GenBank 37700354 (“putative LRR receptor-like protein kinase [Oryza sativa]”)
 - GenBank 568829967 (“probable receptor protein kinase TMK1-like [Citrus sinensis]”)

A search of PDB, Pfam, and SCOP reveal no relevant entries for these two receptor proteins. The analysis must rely on sequence comparison. A BLAST sequence alignment of these two proteins reveals a 39% sequence identity, with an E-value of 2e-80. Since there is no structural information, the MDA of these proteins is unknown. Each of these sequences is annotated as a receptor protein kinase, although the lack of experimental evidence does not independently support these annotations. It is likely that the standard gene annotation transfer protocol was applied to these sequences when it was discovered that they had high sequence similarity to a well studied receptor protein kinase. This is supported by the fact that Uni Prot displays that there is only transcript level evidence of sequence 37700354, and that the [sequence was annotated electronically](#). There is simply not enough evidence to imply that these proteins have the same MDA, although they are likely homologous.

4. – GenBank 6606080 (“human neutral sphingomyelinase”)
 - GenBank 3021480 (“mouse neutral sphingomyelinase”)

A sequence alignment of these two proteins using BLAST reveals that there is very little sequence similarity. The query returned an E-value of 0.70 and a query cover of only 4%. The cross referencing listing from a Uni Prot database revealed that there was not significant structural information to be found. There is a clear similarity in the function of these proteins, they are both sphingomyelinases, but the lack of structural similarity would lead me to conclude that they are presumed unrelated.

5. – UniProt MYD88_MOUSE
 - UniProt Q9FYH3_ARATH

A BLAST sequence alignment of these two proteins returns a E-value of 0.22, identity 31%. This alone does not imply that these proteins would be homologous, but the UniProt entry for Pfam reveal that they share a domain. These proteins share the [Toll-interleukin receptor](#) domain. Assuming that this is not a promiscuous domain, this shared domain would indicate partial homology.

6. – UniProt MYD88_MOUSE
 - UniProt APAF_HUMAN

Following the same procedure as above, a BLAST sequence alignment was done on these two proteins. The alignment returned an E-value of 3.2 and a query cover of only 12%. This lack of sequence alignment supports the conclusion that these proteins are not homologous. The Pfam architecture of each of these proteins is well-defined, and reveals that these proteins ([MYD88 Mouse](#), [APAF Human](#)) do not share a single

domain. This, in conjunction with the lack of sequence similarity implies that these proteins are definitely not homologous.

Part 2. Exploring the connection between structure and function.

1. 2YVI and 2NSN

Not available in SCOP. Instructed to skip.

2. 1AYJ:A and 2SN3:A

This is a comparison between a radish defensin (1AYJ:A) and a scorpion toxin (2SN3:A). According to SCOP, these proteins share the same fold (Knottin) and superfamily (scorpion toxin-like). Although, they do not share the same family. The radish defensin is in the plant defensins family, while the scorpion toxin is in the Long-chain scorpion toxins family. Since these proteins are in the same superfamily, SCOP is asserting that they are homologous. A comparison using the JFATCAT-flexible method returns a P-value of 6.38e-03 and a score of 63.2. Quantitatively, these sequences are structurally similar. A qualitative analysis of the pictured overlaid structures supports that the proteins are structurally similar. Each has an alpha helix that is connected by a single loop to the first of 3 β strands that form a single β sheet. The loop regions do not superimpose as well as the other major secondary structure elements. A sequence alignment of 1AYJ with 2SN3 using BLAST returns a sequence identity of 34% with an E-value of 0.06. The PDB database reveals that each of these proteins has an N-terminal signal region, and a single domain that makes up the remainder of the peptide. For 1AYJ this region contains a [Gamma thionin](#) domain, and this same region in the 2SN3 protein contains a [Toxin_3](#) domain. The sequence identity returned from a BLAST sequence comparison, and the quantitative and qualitative evidence of structural similarity largely support the SCOP classification. That is, that these proteins are homologous, but do are not in the same SCOP family.

3. 2EL9:A and 12AS:A

These proteins are both aminoacyl-tRNA synthetases. 2EL9 corresponds to a histidine synthetase and 12AS corresponds to an Asparagine synthetase. Not only are these proteins placed in the same fold (Class II aaRS and biotin synthetases) and superfamily (Class II aaRS and biotin synthetases), SCOP also places them in the same family (Class II aminoacyl-tRNA synthetase (aaRS)-like, catalytic domain). This classification implies that these proteins have a [recent common ancestor](#). A comparison using the JFATCAT-flexible method returns a P-value of 6.90-05 and a score of

338.37. Quantitatively, these sequences are structurally similar. A qualitative analysis of the overlaid structures supports that the proteins are structurally similar. 2EL9 has a long loop region that connects a set of 3 α helices that encircle a set of 3 β sheets. This is the only major structural element that they do not share. Of the elements that they do share, there is a 5-strand β sheet surrounded by 4 or 5 α helices. A sequence alignment of 2EL9 with 12AS using BLAST returns a sequence identity of 38% with an E-value of 1. Since both of these proteins are aminoacyl-tRNA synthetases, but do not have similar enough structures and sequences that a recent common ancestor is likely to be found, I do not support the SCOP classification that they share the same family. Instead, I would classify them as being in the same superfamily, but separate families.

4. 1BT5:A and 2PBY:A

1BT5 is a β -Lactamase, and 2PBY is a hypothetical protein that is purportedly related to 1BT5. SCOP classifies them as being in the same fold and superfamily (both of which are, beta-lactamase/transpeptidase-like). Since 2PBY is a hypothetical protein it is not given a family classification. A comparison using the JFATCAT-flexible method returns a P-value of 3.76-06 and a score of 292.69, an indicator of structural similarity. Qualitatively, the two proteins in question have very similar structure. As with the protein comparisons before, entire secondary structural elements seem to overlap almost entirely. A sequence alignment of 1BT5 with 2PBY using BLAST returns a sequence identity of 38% with an E-value of 1.8. This sequence comparison implies that there is not much sequence similarity between these two proteins. It should be noted that the two major similar regions have a 200 aa gap between them. It is not independently proven that 2PBY has a lactamase function. Assuming that the automated function match by SCOP is correct, and taking the sequence and structural evidence from BLAST and RCSB into account, I support SCOP's assertion that 1BT5 and 2PBY share the same superfamily.

5. 1W37:A and 1WYI:A

1W37 is a gluconate aldolase, and 1WYI is triosephosphate isomerase. SCOP classifies these proteins as being in the same fold (TIM β/α barrel), but different superfamilies. Data from both a BLAST query, a JFATCAT-flexible structural comparison, and a functional analysis support the classification that these proteins are in different superfamilies. The Pfam entries for each of these proteins support the functions that are listed on SCOP. If these proteins were in the same superfamily, then it would be likely that they have related functionality. This is not the case with these two proteins, they have entirely different chemistries (aldolase vs isomerase). The JFATCAT-flexible structural comparison reveals that while the likelihood of a chance matching (p-value: 9.15e-06) is slim, there is not a high level of structural similarity (Score: 215.11). When the structures are superimposed it becomes clear that what

structural similarity that these proteins do have is restricted to the TIM barrel fold that they share. The high E-value and high % identity, 0.26 and 36% respectively, show that there is a decent amount of sequence similarity. However, there is a high chance that this sequence similarity would be seen elsewhere if more proteins were queried. All of this data is in agreement with SCOP's classification that these proteins are in the same fold, but are not in the same superfamily.

6. 1SBT:A and 1AB9:A

These two proteins are in entirely different folds. They have no structural or evolutionary relatedness. 1SBT is a chymotrypsin in the Trypsin-like serine proteases fold. 1AB9 is a Subtilisin in the Subtilisin-like fold. A BLAST sequence comparison returns that there is no significant sequence similarity found. A JFATCAT-flexible comparison done on the RCSB webserver returns a p-value of 0.982 and a score of 22.16. This structural comparison implies that there is no significant structural similarity. This is confirmed visually as the only secondary structure that is overlapping is a single turn of an α helix. These two pieces of evidence are enough to corroborate SCOP's classification that these two proteins are in entirely different folds.