Functional prediction of hypothetical proteins in human adenoviruses

Shane Dorden & Padmanabhan Mahadevan*

Department of Biology, University of Tampa, 401 W. Kennedy Blvd., Box 3F, Tampa, FL, 33606, USA; Padmanabhan Mahadevan - Email: pmahadevan@ut.edu; *Corresponding author

Received October 11, 2015; Accepted October 15, 2015; Published October 31, 2015

Abstract:
Assigning functional information to hypothetical proteins in virus genomes is crucial for gaining insight into their proteomes. Human adenoviruses are medium sized viruses that cause a range of diseases. Their genomes possess proteins with uncharacterized function known as hypothetical proteins. Using a wide range of protein function prediction servers, functional information was obtained about these hypothetical proteins. A comparison of functional information obtained from these servers revealed that some of them produced functional information, while others provided little functional information about these human adenovirus hypothetical proteins. The PFP, ESG, PSIPRED, 3d2GO, and ProtFun servers produced the most functional information regarding these hypothetical proteins.

Keywords: Hypothetical protein; human adenovirus; pathogen; protein function prediction; web server

Background:
Human adenoviruses (HAdVs) are double stranded DNA viruses that are around 35 kb in size [1]. These viruses cause a wide variety of diseases such as acute respiratory disease [2], keratoconjunctivitis [3], and gastroenteritis [4]. Therefore, HAdVs are important human pathogens. There are 7 species of human adenoviruses, species A-G which are further divided into different strains/types increasingly based on bioinformatics and genomics approaches due to the availability of whole genome sequences, whereas earlier, this was done based on serum neutralization and hemagglutination inhibition [5]. In recent years, the availability of whole genome sequences of various organisms has increased dramatically due to next generation sequencing methods. For example, there was a 21% annual increase in the number of virus nucleotide base-pairs in GenBank and an overall annual increase in all GenBank nucleotide base-pairs of 43.6% in 2014 [6]. Many of the proteins in sequenced genomes are annotated as “hypothetical proteins.” These are predicted proteins that do not have experimental evidence for their translation [7] nor do they have a characterized function [8]. In order to better understand the genomes to which these proteins belong, it will be extremely helpful to assign functions to these hypothetical proteins. Even with their relatively small genome size compared with prokaryotes and eukaryotes, HAdVs possess several hypothetical proteins that need to be functionally annotated.

A myriad of computational approaches to protein function prediction have been developed ranging from template based methods where a template with known function and structure is used to predict function of a query sequence [9], to text mining methods [10] to computational intelligence methods [11]. In this study, we used several well known protein function prediction servers to annotate HAdV hypothetical proteins. We found that some of these servers provided little to no information about the function of these HAdV hypothetical proteins, while others provided information that could potentially be used by wet bench biologists to further experimentally characterize these proteins. These results can serve as a guide to users, particularly wet bench biologists, as to which servers to use to predict the function of hypothetical proteins, particularly those belonging to viruses.

Methodology:
Twenty-eight hypothetical proteins across 11 HAdVs Table 1 (see supplementary material) were obtained from GenBank [6] by searching these genomes for the keyword “hypothetical”. Three additional proteins not explicitly annotated as
hypothetical (AAAT97486, AAAT97487, AAAT97489 from HAdV-4) were chosen as they are very likely hypothetical due to BLASTP hits to other hypothetical proteins. One of the 31 proteins, ADN06471 from HAdV-40/41, although annotated as hypothetical, is known to be expressed [12]. All thirty-one of these proteins were then submitted to several sequence-based protein function prediction servers. These were PFP [13] (Protein Function Prediction), ESG [13] (Extended Similarity Group), ARGOT2 [14], BAR+ [15], PSIPRED [16], ProtFun [17], and dcGO [18]. The hypothetical proteins were also submitted to the fold recognition server Phyre2 [19] in order to determine the fold of these proteins. Protein domain prediction was performed using the protein families database server Pfam [20] and the SMART server [21] (Simple Modular Architecture Research Tool). The homology modeling server SWISS-MODEL [22] and the MuFold protein structure prediction server [23] were used to predict the structures of the hypothetical proteins. Successfully predicted structures were then submitted to the structure-based server 3d2GO [24]. Tables were then constructed for all servers’ predictions for function of each individual protein, protein domain predictions, and fold predictions.

Results:
The average length of the 31 hypothetical proteins from 11 different human adenovirus genomes was 124 amino acids, with the high being 224 and a low of 58 (Table 1). The PFP server predicted functions for all 31 hypothetical proteins, some of which with high confidence, such as betal-adrenergic receptor activity at 92% confidence for protein ACN88103 and purine nucleotide binding at 100% for protein AAW65500 (Table 2 (see supplementary material)). The ESG server was not as successful as the PFP server, but still managed to predict functions for 26 of the 31 possible hypothetical proteins. For instance, GTPase activity and GTP binding at 99% confidence was predicted as the function of AGF90820, and lyase activity and aldehyde-lyase activity at 89% confidence was predicted for ACN88103 as shown in Table 2.

ARGOT2 was only capable of predicting the function of 7 hypothetical proteins, such as hydrolase activity at 100% confidence for protein AEG46441 and transferase activity at 85% confidence for protein AAAT97487 (Table 3 (see supplementary material)). Additionally as shown in Table III, BAR+ was unable to predict a function for any of the hypothetical proteins. Similarly, the dcGO server was unable to predict a function for any of the hypothetical proteins (table not shown). The PSIPRED server predicted functions for all 31 hypothetical proteins such as GTP binding at 94% probability for AFH58045 and oxidoreductase activity at 99% probability for protein AAAT97539 (Table 4 (see supplementary material)). The fold recognition server Phyre2 identified potential folds in 8 of the 31 hypothetical proteins as shown in Table 4. These folds include: pyruvate kinase C-terminal domain-like at 17.70% confidence for AFH58048 and barrel-sandwich hybrid at 25.10% confidence for protein AAW65505. The ProtFun server predicted functions for 24 of the 31 proteins, along with categorical information concerning gene ontology and whether the protein was an enzyme or not (Table 5 (see supplementary material)). Protein AAAT97531 was predicted to play a role in the cell envelope with 53% probability, be an enzyme with 46% probability, and finally, be a structural protein with 27% probability. Additionally, protein AFH58048 was predicted to play a role in transport and binding with 74% probability, be a non-enzyme with 82% probability, and finally, be a growth factor with 7% probability as shown in Table 5. The homology modeling server SWISS-MODEL did not produce a structure output for any of the 31 hypothetical proteins for use with the 3d2GO server. However, the structure-based 3d2GO server predicted a function for 22 of the 31 hypothetical proteins from proposed structures of these proteins, provided by MuFold (Table 6 (see supplementary material)). For example, 3d2GO predicted oxidoreductase activity at 29% confidence as a function for AAW33184 and transport at 61% confidence for protein AAW65506. The protein family server Pfam found no domains for any of the hypothetical proteins Table 7 (see supplementary material). In contrast, the protein domain prediction server SMART produced results for 25 of the 31 hypothetical proteins, with the majority containing low complexity regions as shown in Table 7.

Discussion:
The PFP server predicted some form of “binding” for 25 of the 31 function predictions, and had an average prediction confidence of 81% (Table 2). Additionally, the ESG server made function predictions for 26 of the 31 proteins, averaging 50% confidence. ESG did not predict a function for all proteins as PFP did, but it provided more complete functional information, albeit with average to low confidence. For example, for protein AAAT97533, 4-hydroxy-tetrahydrodipicolinate reductase, oxidoreductase activity, oxidoreductase activity, acting on CH or CH2 groups, NAD or NADP as acceptor, NADP binding, NAD binding, and NADPH binding was predicted at 32% confidence (Table 2). Also, for protein ADN06471 N-acetyltransferase activity, transferase activity, transferase activity, transferring acyl groups, transferase activity, and transferring acyl groups other than amino-acyl groups was predicted at 53% confidence.

ARGOT2 predicted only 7 functions, averaging 80% confidence (Table 3). The BAR+ and dcGO servers were both unable to predict a function for any of the proteins as shown in Table 3. PSIPRED was capable of predicting a function for all 31 proteins, averaging 91% confidence in the process (Table 4). The function of “structural constituent of ribosome” was predicted for 7 of the 31 proteins. Also, some form of “binding” was predicted for 16 of the 31 proteins and ranged from “calcium ion binding” to “actin binding”. While the PSIPRED predictions were rather vague, the confidence of the predictions remained high across all 31 hypothetical proteins. Additionally, the fold recognition server Phyre2 only identified 8 potential matching folds out of a possible 31 and had an average confidence of 16.60% which is the probability of the query sequence and template being homologous (Table 4). Moreover, since Phyre2 utilizes fold recognition, the information the server provided allows users to gain insight into the fold of that protein.

ProtFun provided a more thorough functional prediction for each protein that it could predict a function for. ProtFun managed to make 24 of the possible 31 hypothetical protein function predictions (Table 5). Not only did ProtFun predict functions for the 24 proteins, it also predicted whether the protein was an enzyme or nonenzyme, and its gene ontology (GO). Across the 26 predictions, function prediction confidence averaged at 29%, enzyme/nonenzyme prediction confidence averaged at 63%, and gene ontology prediction confidence averaged at 17%. SWISS-MODEL did not find templates for any
of the proteins and therefore, could not produce a structure to use as input to the 3d2GO server. However, MuFold predicted a structure for 22 of the 31 hypothetical proteins (Table 6). Furthermore, structure-based server 3d2GO utilized those predicted structures to predict a function for the 22 proteins as shown in Table 6. Average prediction confidence was 50% and the server was able to predict a function from all structures proposed by MuFold. The function for protein AAW33433 was predicted to be RNA binding, ribosome, ribonucleoprotein complex, structural molecule activity, intracellular, translation, rRNA binding and structural constituent of ribosome at 99% confidence, but aside from this thorough prediction, most other predictions were rather vague, such as “cytosol”, “cytoplasm”, and “membrane” as shown in Table 6. While Pfam and SMART are not strictly protein function prediction servers, we wanted to investigate whether they could provide pertinent domain information for the HAdV hypothetical proteins. Pfam also did not find any domains in these proteins. Further, while the SMART server did find matching regions for 26 of the 31 hypothetical proteins, the information provided from the server was very minimal as 23 of the 26 matches were “low complexity regions” and the other 3 were classified as “signal peptide regions” (Table 7).

Conclusions:
It is apparent from the results no single server produces the most complete functional determination of these “hard” HAdV hypothetical proteins. The servers that provided the most information were PFP, ESG, PSIPRED, 3d2GO, and ProtFun. The servers which provided very little or no functional information were ARGOT2, BAR+, and dcGO. We believe that the best option for functional prediction of hypothetical proteins is to use a multitude of servers and analyze the results produced. Furthermore, we agree with Radivojac et al. [25] that these servers need to be improved in order to better predict protein function.

Acknowledgement:
We wish to thank the Department of Biology at the University of Tampa for awarding Shane Dorden a summer research fellowship in order to facilitate completion of this work.

References:
[1] Davison AJ et al. J Gen Virol. 2003 84: 2895 [PMID: 14573794]
[2] Lai CY et al. PLoS one 2013 8: e53614 [PMID: 23326469]
[3] Huang G et al. PLoS One 2014 9: e110781 [PMID: 25343525]
[4] Lee JI et al. Microbiol Immunol. 2012 56: 523 [PMID: 22530970]
[5] Ghebremedhin B, Eur J Microbiol Immunol. 2014 4: 26 [PMID: 24678403]
[6] Benson DA et al. Nucleic Acids Res. 2015 43: D30 [PMID: 25414350]
[7] Ijaq J et al. Front Genet. 2015 6: 119 [PMID: 25873935]
[8] Galperin MY & Koonin EV, Nucleic Acids Res. 2004 32: 5452 [PMID: 15479782]
[9] Petrey D et al. Curr Opin Struct Biol. 2015 32: 33 [PMID: 25678152]
[10] Verspoor KM, Methods Mol Biol. 2014 1159: 95 [PMID: 24788263]
[11] Tiwari AK & Srivastava R, Int J Proteomics. 2014 845479 [PMID: 25574395]
[12] Dix I & Leppard KN, J Gen Virol. 1995 76: 1051 [PMID: 9049357]
[13] Khan IK et al. Bioinformatics 2015 31: 271 [PMID: 25273111]
[14] Falda M et al. BMC bioinformatics 2012 13: S14 [PMID: 22536960]
[15] Piovesan D et al. Nucleic Acids Res. 2011 39: W197 [PMID: 21622657]
[16] Buchan DWA et al. Nucleic Acids Res. 2013 41: W349 [PMID: 23748958]
[17] Jensen LJ et al. Bioinformatics 2003 19: 635 [PMID: 12651722]
[18] Fang H & Gough J, Nucleic Acids Res. 2013 41: D536 [PMID: 23161684]
[19] Kelley LA et al. Nat Protoc. 2015 10: 845 [PMID: 25950237]
[20] Finn RD et al. Nucleic Acids Res. 2014 42: D222 [PMID: 24288371]
[21] Letunic I et al. Nucleic Acids Res. 2015 43: D257 [PMID: 25300481]
[22] Biasini M et al. Nucleic Acids Res. 2014 42: W252 [PMID: 24782522]
[23] Zhang J et al. Proteins 2010 78: 1137 [PMID: 19927325]
[24] 3d2GO server: http://www.sbg.bio.ic.ac.uk/phyre/pfd/index.html
[25] Radivojac P et al. Nat Methods. 2013 10: 221 [PMID: 23353650]

Edited by P Kangueane

Citation: Dorden & Mahadevan, Bioinformation 11(10): 466-473 (2015)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.
### Table 1: List of human adenovirus hypothetical proteins, corresponding genomes, and protein length

| Protein GenBank Accession Number | HAdV (accession number)       | Length (aa) |
|----------------------------------|--------------------------------|-------------|
| AFH58036                         | HAdV-14 (JN032132)            | 102         |
| AFH58045                         |                                | 77          |
| AFH58048                         |                                | 68          |
| AFH58052                         |                                | 83          |
| AGF90820                         | HAdV-14 (JX892927)            | 59          |
| AGE46441                         |                                | 129         |
| AAW33161                         | HAdV-3 (AY599834)             | 91          |
| AAW33157                         |                                | 189         |
| AAW33158                         |                                | 173         |
| AAT97531                         | HAdV-7 (AY594255)             | 106         |
| AAT97533                         |                                | 114         |
| AAT97535                         |                                | 133         |
| AAT97539                         |                                | 58          |
| AAT97549                         |                                | 69          |
| AAW33433                         | HAdV-16 (AY601636)            | 95          |
| AAW33435                         |                                | 114         |
| AAQ10567                         | HAdV-1 (AF534906)             | 132         |
| AAW65500                         | HAdV-5 (AY601635)             | 115         |
| AAW65501                         |                                | 106         |
| AAW65502                         |                                | 112         |
| AAW65505                         |                                | 137         |
| AAW65506                         |                                | 121         |
| ACN88099                         | HAdV-6 (FJ349096)             | 215         |
| ACN88101                         |                                | 168         |
| ACN88103                         |                                | 176         |
| ACN88132                         |                                | 134         |
| AAT97486                         | HAdV-4 (AY594254)             | 189         |
| AAT97487                         |                                | 106         |
| AAT97489                         |                                | 224         |
| ACR78236                         | HAdV-22 (FJ619037)            | 130         |
| ADN06471                         | HAdV-41 (HM565136 & DQ315364)| 130         |

### Table 2: PFP and ESG function predictions with probability (%).

| Accession Number | HAdV   | PFP Function, Probability | ESG Function, Probability                                                                 |
|------------------|--------|---------------------------|-------------------------------------------------------------------------------------------|
| AFH58036         | HAdV-14| binding, 90%              | DNA binding, DNA-directed RNA polymerase activity, transferase activity, ribonucleoside binding, 68% |
| AFH58045         | HAdV-14| binding, 66%              | N/A                                                                                       |
| AFH58048         | HAdV-14| binding, 65%              | N/A                                                                                       |
| AFH58052         | HAdV-14| binding, 83%              | nucleotide binding, DNA binding, DNA topoisomerase activity, DNA topoisomerase type I activity, ATP binding, isomerase activity, metal ion binding, 7% GTPase activity, GTP binding, 99% transferase activity, 59% |
| AGF90820         | HAdV-14| binding, 64%              | N/A                                                                                       |
| AGE46441         | HAdV-14| binding, 74%              | N/A                                                                                       |
| AAW33161         | HAdV-3  | binding, 82%              | N/A                                                                                       |
| AAW33157         | HAdV-3  | myristoyltransferase activity, 78% | N/A                                                                                       |
| AAW33158         | HAdV-7  | binding, 86%              | nucleotide binding, 58% transferase activity, 70%                                         |
| AAT97531         | HAdV-7  | transition metal ion binding, 86% | 4-hydroxy-tetrahydrodipicolinate reductase, oxidoreductase activity, oxidoreductase activity, acting on CH or CH2 groups, NAD or NADP as acceptor, NADP binding, NAD binding, NADPH binding, 32% nucleotide binding, 52%  |
| AAT97533         | HAdV-16 | binding, 67%              | N/A                                                                                       |
| AAT97539         | HAdV-16 | binding, 74%              | N/A                                                                                       |
| AAT97549         | HAdV-16 | adenyl nucleotide binding, 83% | N/A                                                                                       |
| AAW33433         | HAdV-16 | adenyl nucleotide binding, 83% | N/A                                                                                       |
| AAW33435         | HAdV-16 | binding, 69%              | adenyl nucleotide binding, 83% | ligase activity, 79%                                                                 |
| AAQ10567         | HAdV-1   | adenyl nucleotide binding, 83% | metal ion binding, 20%                                                                     |
|                  | HAdV-1   | adenyl nucleotide binding, 83% | lyase activity, aldehyde-lyase activity, 61%                                                  |
| Accession Number | HAdV   | ARGOT2 Function, Confidence | BAR+ Function |
|------------------|--------|-----------------------------|---------------|
| AAW65500         | HAdV-5 (AY601635) | purine nucleotide binding, 100% | translation initiation factor activity, 32% |
| AAW65501         |        | oxidoreductase activity, acting on the CH-NH₂ group of donors, 68% | ATP binding, 16% |
| AAW65502         |        | binding, 82% | nucleotide binding, 44% |
| AAW65505         |        | interleukin-1 receptor antagonist activity, 94% | methyltransferase activity, S-adenosylmethionine-dependent methyltransferase activity, transferase activity, 58% |
| AAW65506         |        | beta1-adrenergic receptor activity, 92% | ATP binding, 28% |
| ACN88099         | HAdV-6 (FJ349096) | purine nucleotide binding, 98% | ATP binding, 28% |
| ACN88101         |        | adenyl nucleotide binding, 85% | metal ion binding, 23% |
| ACN88103         |        | beta1-adrenergic receptor activity, 92% | lyase activity, aldehyde-lyase activity, 89% |
| ACN88132         |        | purine nucleotide binding, 83% | transferase activity, 61% |
| AAT97486         | HAdV-4 (AY594254) | binding, 81% | metal ion binding, 51% |
| AAT97487         |        | cation binding, 90% | oxidoreductase activity, 26% |
| AAT97489         |        | binding, 63% | metal ion binding, 85% |
| ACR78236         | HAdV-22 (FJ619037) | lactate transporter activity, 85% | catalytic activity, 33% |
| ADN06471         | HAdV-41 (HM565136 & DQ315364) | binding, 77% | N-acetyltransferase activity, transferase activity, transferase activity, transferring acyl groups, transferase activity, transferring acyl groups other than amino-acetyl groups, 53% |

Table 3: ARGOT2 and BAR+ function predictions with confidence (%).
Table 4: PSIPRED function prediction with probability (%) and Phyre2 fold prediction where confidence (%) is the probability that the sequence and template are homologous.

| Accession Number | HAdV | PSIPRED Function, Probability | Phyre2 Fold | Confidence |
|------------------|------|-------------------------------|-------------|------------|
| AFH58036         | HAdV-14 (JN032132) | calcium ion binding, 93% | N/A | 8.10% |
| AFH58045         | HAdV-14 (JX892927) | GTP binding, 94% | Ferredoxin-like | 13.50% |
| AFH58048         | HAdV-14 (JX892927) | sodium ion transmembrane transporter activity, 98% | Pyruvate kinase C-terminal domain-like | 17.70% |
| AFH58052         | HAdV-14 (JX892927) | channel activity, 96% | N/A | 38.20% |
| AGF90820         | HAdV-14 (JX892927) | structural constituent of ribosome, 100% | N/A | 8.10% |
| AEG46441         | HAdV-14 (JX892927) | serine-type peptidase activity, 88% | N/A | 33.80% |
| AAW33161         | HAdV-3 (AY599834) | receptor binding, 93% | N/A | 46.40% |
| AAW33157         | HAdV-3 (AY599834) | oxidoreductase activity, 77% | N/A | 11.60% |
| AAW33158         | HAdV-3 (AY599834) | structural constituent of ribosome, 94% | N/A | 20% |
| AAAT97531        | HAdV-7 (AY594255) | GTP binding, 99% | Spectrin repeat-like | 16.20% |
| AAAT97533        | HAdV-7 (AY594255) | DNA binding, 91% | N/A | 55.50% |
| AAAT97535        | HAdV-7 (AY594255) | protein kinase activity, 80% | DHS-like NAD/FAD-binding domain | 11% |
| AAAT97539        | HAdV-7 (AY594255) | oxidoreductase activity, 99% | N/A | 26% |
| AAAT97549        | HAdV-7 (AY594255) | zinc ion binding, 98% | N/A | 31.20% |
| AAW33132         | HAdV-16 (AY601636) | ATP binding, 97% | N/A | 25.90% |
| AAW33135         | HAdV-16 (AY601636) | ATP binding, 99% | N/A | 18% |
| A AQ0567         | HAdV-1 (AF534906) | structural constituent of ribosome, 92% | N/A | 20.30% |
| AAW65500         | HAdV-5 (AY601635) | zinc ion binding, 79% | N/A | 10.20% |
| AAW65501         | HAdV-5 (AY601635) | structural constituent of ribosome, 99% | N/A | 10.10% |
| AAW65502         | HAdV-5 (AY601635) | structural constituent of ribosome, 97% | N/A | 11.20% |
| AAW65505         | HAdV-5 (AY601635) | structural constituent of ribosome, 97% | Barrel-sandwich hybrid | 25.10% |
| AAW65506         | HAdV-5 (AY601635) | receptor binding, 91% | N/A | 12.30% |
| ACN88099         | HAdV-6 (FJ349096) | transcription factor binding, 80% | N/A | 0% |
| ACN88101         | HAdV-6 (FJ349096) | receptor binding, 86% | N/A | 36.30% |
| ACN88103         | HAdV-6 (FJ349096) | receptor binding, 80% | N/A | 21.80% |
| ACN88132         | HAdV-6 (FJ349096) | structural constituent of ribosome, 98% | SOCS box-like | 14.30% |
| AAT97486         | HAdV-4 (AY594254) | actin binding, 74% | N/A | 24% |
| AAT97487         | HAdV-4 (AY594254) | receptor binding, 92% | Spectrin repeat-like | 13.30% |
| AAT97489         | HAdV-4 (AY594254) | peptidase activity, 89% | N/A | 11% |
| ACR79236         | HAdV-22 (FJ619037) | receptor binding, 80% | N/A | 33.30% |
| ADN06471         | HAdV-41 (HM565136 & DQ315364) | cytokine activity, 82% | Nop domain | 21.70% |

Table 5: ProtFun function with probability (%), enzyme/nonenzyme with probability (%), and gene ontology predictions with probability (%).

| Accession Number | HAdV | ProtFun Function, Probability | Enzyme/Nonenzyme, Probability | Gene Ontology, Probability |
|------------------|------|-----------------------------|-----------------------------|--------------------------|
| AFH58036         | HAdV-14 (JN032132) | Translation, 14% | Nonenzyme, 82% | Transcription regulation, 24% |
| AFH58045         | HAdV-14 (JX892927) | Energy metabolism, 31% | Nonenzyme, 77% | Immune response, 32% |
| AFH58048         | HAdV-14 (JX892927) | Transport and binding, 74% | Nonenzyme, 82% | Growth factor, 7% |
| AFH58052         | HAdV-14 (JX892927) | N/A | N/A | N/A |
| AGF90820         | HAdV-14 (JX892927) | Translation, 8% | Nonenzyme, 74% | Structural protein, 10% |
| AEG46441         | HAdV-14 (JX892927) | Amino acid biosynthesis, 21% | Nonenzyme, 71% | Growth factor, 6% |
| AAW33161         | HAdV-3 (AY599834) | N/A | N/A | N/A |
| AAW33157         | HAdV-3 (AY599834) | Translation, 30% | Enzyme, 37% | N/A |
| AAW33158         | HAdV-3 (AY599834) | Transport and binding, 49% | Enzyme, 42% | Growth factor, 3% |
| AAT97531         | HAdV-7 (AY594255) | Cell envelope, 53% | Enzyme, 46% | Structural protein, 27% |
| AAT97533         | HAdV-7 (AY594255) | Translation, 5% | Nonenzyme, 74% | Structural protein, 24% |
| AAT97535         | HAdV-7 (AY594255) | Translation, 30% | Nonenzyme, 74% | Structural protein, 5% |
| AAT97539         | HAdV-7 (AY594255) | Energy metabolism, 37% | Nonenzyme, 76% | Growth factor, 3% |
| AAW33433         | HAdV-16 (AY601636) | Energy metabolism, 22% | Nonenzyme, 77% | Structural protein, 21% |
| AAW33435         | HAdV-16 (AY601636) | Regulatory functions, 27% | Nonenzyme, 78% | Growth factor, 3% |
| A AQ0567         | H AdV-1 (AF534906) | N/A | N/A | N/A |
| AAW65500         | HAdV-5 (AY601635) | Energy metabolism, 13% | Enzyme, 47% | Structural protein, 20% |
| AAW65501         | HAdV-5 (AY601635) | Energy metabolism, 23% | Enzyme, 33% | Structural protein, 29% |
| AAW65502         | HAdV-5 (AY601635) | Translation, 20% | Enzyme, 29% | Structural protein, 31% |
| AAW65505         | HAdV-5 (AY601635) | N/A | N/A | N/A |
| AAW65506         | HAdV-5 (AY601635) | N/A | N/A | N/A |
| ACN88099         | HAdV-6 (FJ349096) | Regulatory functions, 19% | Nonenzyme, 73% | Transcription, 27% |
| ACN88101         | HAdV-6 (FJ349096) | N/A | N/A | N/A |
| Accession Number | HAdV                | 3dGo, Confidence                                      | MuFold                   |
|------------------|---------------------|-------------------------------------------------------|--------------------------|
| ACN88103         | HAdV-4 (AY594254)   | N/A                                                   | N/A                      |
| ACN88132         | Translation, 27%    | Nonenzyme, 82%                                       | Growth factor, 13%        |
| AAT97486         | Translation, 21%    | Nonenzyme, 72%                                       | Immune response, 9%       |
| AAT97487         | Cell envelope, 53%  | Enzyme, 37%                                          | Structural protein, 27%   |
| AAT97489         | Translation, 30%    | Nonenzyme, 76%                                       | Structural protein, 8%    |
| ACR78236         | HAdV-22 (FJ619037)  | Energy metabolism, 35%                               | Growth factor, 11%        |
| ADN06471         | HAdV-41 (HM565136 & DQ515364) | Energy metabolism, 30%                               | Signal transducer, 14%    |

Table 6: 3dGo function predictions and confidence (%) and whether MuFold predicted a structure for the protein.

| Accession Number | HAdV                | 3dGo, Confidence                                      | MuFold                   |
|------------------|---------------------|-------------------------------------------------------|--------------------------|
| AFH58036         | adenyl ribonucleotide binding, adenyl nucleotide binding, carbohydrate metabolic process, ribonucleotide binding, purine ribonucleotide binding, purine nucleotide binding, 12% | Predicted structure     |
| AFH58045         | metabolic process, 40% | Predicted structure                                  |
| AFH58048         | membrane, 34%        | Predicted structure                                  |
| AFH58052         | methyltransferase activity, 47% | Predicted structure                                  |
| AGF90820         | metal ion binding, 50% | Predicted structure                                  |
| AGE46441         | ion binding, 21%     | Predicted structure                                  |
| AAW33161         | cytoplasm, 25%       | Predicted structure                                  |
| AAW33157         | N/A                  | No structure predicted                               |
| AAW33158         | membrane, 61%        | Predicted structure                                  |
| AAT97531         | HAdV-7 (AY594255)    | N/A                                                   | No structure predicted   |
| AAT97533         | membrane, 98%        | Predicted structure                                  |
| AAT97535         | cytosol, 66%         | Predicted structure                                  |
| AAT97539         | protein binding, 12% | Predicted structure                                  |
| AAT97549         | ion binding, 47%     | Predicted structure                                  |
| AAW33433         | RNA binding, ribosome, ribonucleoprotein complex, structural molecule activity, intracellular, translation, rRNA binding, structural constituent of ribosome, 99% | Predicted structure     |
| AAW33435         | N/A                  | No structure predicted                               |
| AAQ10567         | oxidation reduction, 20% | Predicted structure                                  |
| AAW65500         | translation, intracellular, 100% | Predicted structure                                  |
| AAW65501         | N/A                  | No structure predicted                               |
| AAW65502         | N/A                  | No structure predicted                               |
| AAW65505         | transferase activity, transferring one-carbon groups, 26% | Predicted structure     |
| AAW65506         | transport, 61%       | Predicted structure                                  |
| ACN88099         | HAdV-6 (FJ349096)    | protein binding, 38%                                 | Predicted structure      |
| ACN88101         | N/A                  | No structure predicted                               |
| ACN88103         | metabolic process, 69% | Predicted structure                                  |
| ACN88132         | N/A                  | No structure predicted                               |
| Accession Number | HAdV     | Pfam                          | SMART                          |
|------------------|----------|-------------------------------|--------------------------------|
| AAT97486         | HAdV-4   | N/A                           | No domain found                |
|                  | (AY594254) |                              | Low complexity region          |
| AAT97487         | protein modification process, 59% |                              |                                |
| AAT97489         | N/A      |                               |                                |
| ACR78236         | HAdV-22  | hydrolase activity, acting on acid anhydrides, 14% |                                |
| (FJ619037)       |          |                               |                                |
| ADN06471         | HAdV-41  | intracellular, 96%            | Predicted structure            |
| (HM565136 & DQ315364) |    |                               |                                |

Table 7: Pfam and SMART domain predictions