ASSP; the Antibody Secondary Structure Profile search tool

Dimitrios Vlachakis[†], Alexandros Armaos[†], Kasampalidis I, Arianna Filntisi, Sophia Kossida[*]

Bioinformatics & Medical Informatics Team, Biomedical Research Foundation
Academy of Athens, Soranou Efessiou 4, Athens 11527, Greece
[*]skossida@bioacademy.gr

Abstract

Antibodies constitute the first line of defense against harmful invaders. In the post genomics era the sheer size of antibody related NGS information is a major bottleneck in the quest of understanding and tackling complex genetic diseases and immunological disorders. Bioinformatics is becoming hugely involved in the processing of this data with the development of new, more accurate and efficient algorithms. However, one of the major drawbacks of modern bioinformatics is the fact that protein similarity and blast searches are still based on primary amino acid sequence rather than structural data. Primary sequence searches are inadequate, as they fail to provide a realistic fingerprint for the query protein. Antibody function is much more related to its 3D structure and physicochemical profile rather than its primary amino acid sequence. After all, structure is much more conserved than sequence in nature. In this direction, a novel platform has been developed, which is capable of performing a customized hydropathy blast using traditional sequence blast filtering and an integrated fast similarity search algorithm that uses protein secondary structure information. The Antibody Secondary Structure Profile (ASSP) tool will use secondary structural information from the PDB database when available, whereas if the query antibody is not indexed in the RCSB PDB database, it will automatically determine the secondary elements of the given antibody by performing an "on the fly" secondary structure prediction. All query antibodies are then blasted against the RCSB PDB secondary elements database. Hits are scored, ranked and returned to the user via a well-organized and user friendly graphical interface.

In: Costas S. Iliopoulos, Alessio Langiu (eds.): Proceedings of the 2nd International Conference on Algorithms for Big Data, Palermo, Italy, 7-9 April 2014, published at http://ceur-ws.org/

[†] These authors have contributed equally to this study.

[*] Corresponding author: Sophia Kossida, Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Soranou Efessiou 4, Athens 11527, Greece
Tel: + 30 210 6597 199, Fax: +30 210 6597 545 E-mail: skossida@bioacademy.gr

1 Introduction

The hydrophobic effect is the tendency of non-polar substances to avoid contact with water. The hydropathy of an amino acid, which is derived from the physicochemical properties of its side chains, determines in part the orientation of its side chains in the three-dimensional protein structure. In particular, when a protein folds into a three-dimensional structure, the majority of the hydrophobic side-chains cluster together within the core of the protein. This removal of the hydrophobic side-chains out of contact with water generates sufficient free energy to maintain the folded structure of the protein. The determination of the hydrophobic or hydrophilic inclinations of a given amino acid side-chain has been approached in a number of ways. Measuring the partition coefficient of a given amino acid side-chain between water and a non-interacting, isotropic phase as well as calculating a transfer free energy from that coefficient is one such approach. Another way to calculate the hydropathy of a given side-chain is the tabulation of residue accessibilities from the atomic co-ordinates of twelve globular proteins, taking into consideration that the ensemble average of the actual locations of a side-chain should be a direct evaluation of its hydropathy. An additional approach combined those previously mentioned methods, resulting in the construction of a hydropathy scale, according to which each amino acid has been given a value reflective of its relative hydrophilicity and hydrophobicity [4, 17].

The twenty amino acids found in nature have been categorized in hydropathy classes based on the previously mentioned amino acid hydropathy index developed by Kyte and Doolittle (1982). Specifically, the amino acids with a hydropathy index equal to or more than 1.8 were defined as hydrophobic. The amino acids with a hydropathy index equal to or less than -3.3 were defined as hydrophilic, while the amino acids with a hydropathy index less than 1.8 and more than -3.3 were defined as neutral. Three classes were thus defined: the hydrophobic class (I, V, L, F, C, M, A, W), the neutral class (G, T, S, Y, P, H) and the hydrophilic class (D, N, E, Q, K, R). Tryptophan (W) was included in the hydrophobic class, its hydropathy index varying from -0.9 to 1.9, depending on the study. As a general observation, amino acids with large, nonpolar or largely nonpolar side-chains tend to be hydrophobic, while the least hydrophobic amino acids are the ones that are charged and largely polar, such as asparagine. Statistical analysis, in particular correspondence analysis (COA) and hierarchic classification (CAH), has been conducted according to those three hydropathy classes upon 2474 sequences of antibody variable regions, which were extracted from human productively rearranged sequences [17, 25, 8]. There is a significant variation among the hydrophobicities of the amino acids. Some are strongly hydrophobic, others are strongly hydrophilic, while others include both hydrophobic and hydrophilic parts and are called amphiphilic. For such amphiphilic molecules it is sometimes useful to define a hydrophobic moment, which is analogous to a dipole moment. For a single amino acid, the hydrophobic moment can be defined as a line that points from the Ca atom to the middle of the side-chain, and whose length is proportional to the hydrophobicity of the side-chain. The dipole moment of a protein, or a part of it, is obtained by summing the individual vectors (in magnitude and direction) corresponding to the amino acids the protein is composed of. For example, an α helix located on the surface of a protein will have one side of the helix exposed to solvent and the other side facing the interior of the protein. The amino acids that comprise the buried side of the α helix will, in general, be much more hydrophobic than those on the solvent-exposed side of the helix. This asymmetry results in the α helix having a large hydrophobic moment directed towards the center of the protein [21].

The three-dimensional structure of a protein is determined by the balance between a number of destabilizing and stabilizing forces, such as conformational entropy, electrostatic interactions, hydrogen bonds, van der Waals interactions and hydrophobic interactions. However, hydropathy is considered to be the most prominent driving force responsible for the folding of proteins. Protein folding occurs in the presence of water, the properties of which are dominated by its inclination to form hydrogen bonds. Polar compounds can share hydrogen bonds with water and, for this reason, are readily soluble. In contrast, when a hydrophobic nonpolar surface is introduced into an aqueous environment, it prevents hydrogen bonding from occurring, which forces the water molecules to adopt alternative arrangements that permit hydrogen bonding to other water molecules. This inflicted restriction on the alignment of the water molecules has an energetic cost and is the physical basis of the hydrophobic effect. It has been calculated that when a protein folds, 81% of the nonpolar side-chains (Ala, Val, Ile, Leu, Met, Phe, Trp, Cys), 70% of the peptide groups, 63% of the polar side chains (Asn, Gln, Ser, Thr, Tyr) and 54% of the charged side chains (Arg, Lys, His, Asp, Glu) are buried in the interior of the protein, out of contact with water [21, 23].

2 Description of ASSP

Hydropathy is a physicochemical property known to be well conserved among antibodies, which can be explained to a large extent by the significant contribution of the hydrophobic residues to the folding of antibodies. Numerous studies on proteins and antibodies have demonstrated that the information necessary to produce a given three dimensional protein structure can be encoded by many different amino acids. In contrast, it has been demonstrated that the periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. In fact, the choice between α-helical and β-sheet secondary structure is influenced by the sequence periodicity of polar and nonpolar amino acids. Even though amino acid residues may differ in their intrinsic preferences for one secondary structure versus another, these preferences can be overwhelmed by the drive to form amphiphilic structures capable of burying hydrophobic surface area. It can be observed that structural similarity among antibodies is reflected on the distribution of hydropathicity along their amino acid sequences, since the hydrophobicity patterns of residues match the periodicity of secondary structures.

Homologous antibodies and proteins within a antibody/protein family as well as proteins with related structures appear to have similarities in their hydropathy distributions, even when sequence similarities could not be detected [14, 2, 32, 24].

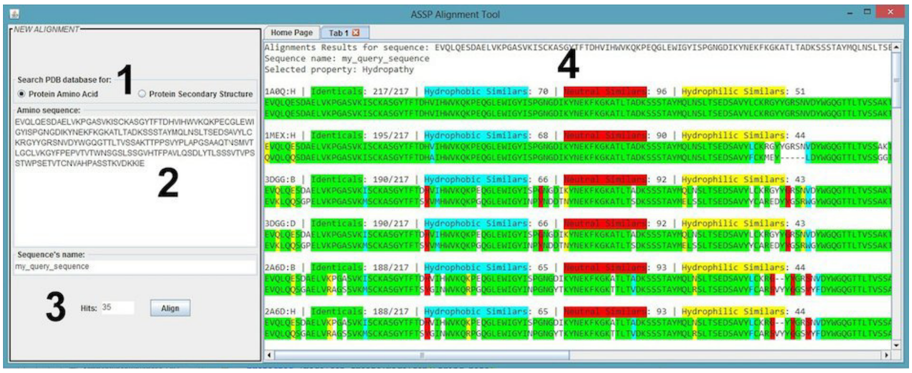
Since the hydropathy distribution along the antibody sequence has been recognized as a feature useful for the characterization of protein structure in the form of hydropathy profiles, a number of methods based on hydropathy have been developed in order to explain the folding and the structural features of antibodies. The realization that protein sequence contains hydropathy patterns led to the development of reduced amino acid alphabets based on hydropathy for the prediction of secondary structure.

Hydropathy has also been utilized for the detection of analogous and distantly related proteins and the classification of new protein sequence data. The use of hydropathy profile analysis has made possible the identification of more distantly related antibodies than could be done by sequence comparison. In addition, antibody sequence databases have been analyzed using hydropathy patterns with the goal of identifying new members of functional classes [17, 24, 7, 26, 5, 33, 19, 20, 6].

Many homologous proteins share very low primary sequence identity and similarity scores amongst them. The most characteristic example of such proteins is viral enzymes. Helicases, proteases and polymerases are just few of the many examples of protein families that are structurally conserved but may share no more than 10% sequence identity with each other. Consequently, looking for homologues of a certain viral enzyme or even for a suitable template structure during homology modelling using traditional amino acid based blast searches is futile. However, careful structural analysis of any of the above enzymes reveals that those proteins are actually highly conserved in their secondary, tertiary and quaternary structures. Moreover, all evolutionary protein relationships as well as protein function analysis should also be based on searches that utilize structural information. Overall, it has been established that homologous proteins are much more conserved in their structures than in their amino acid primary sequences [4, 17].

Herein, the ASSP tool takes advantage of the full RCSB PDB secondary structure database in order to perform blast-like searches in the secondary element level amongst proteins. To date, even though long studies have been conducted in many fields of structural biology and modern bioinformatics this problem not been yet satisfactorily addressed [15, 1, 22]. This is a fact that necessitates the need to the development of such a platform.

ASSPs main-window is a menu-driven interface as well as a tab step-by-step layout. Initially the user has an option regarding the query input type that will be used. ASSP will handle both primary amino acid sequence as well as DSSP-formatted secondary element protein sequence [13]. The user can follow two main routes for the ASSP run: Firstly, the user may input either raw primary amino acid sequence for a conventional blast search or opt for a quick secondary structure prediction of the amino acid sequence using the built-in STRAP module [9]. STRAP will perform a very fast, over the internet secondary structure prediction, which will eventually return the predicted secondary element composition of the query protein [9]. Eventually a DSSP compatible secondary structure determination code will have been obtained for the actual secondary structure similarity search [13]. Secondly, an existing DSSP compatible secondary structure determination code may be used as input from the user straightaway, which will then be automatically blasted against the secondary structural index database of the RCSB Protein Databank. If a secondary element antibody sequence description is used,



NEW ALIGNMENT

Search PDB database for: ● Protein Amino Acid ○ Protein Secondary Structure

Amino sequence:
EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPECGLEWI
GYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYLC
KRGYYGRSNVDYWGQGTTLTVSSAKTTPPSVYPLAPGSAAQTNSMVT
LGCLVKGYFPEPVTVTWNSGSLSSGVHTFPAVLQSDLYTLSSSVTVPS
STWPSETVTCNVAHPASSTKVDKKIE

Sequence's name: my_query_sequence

Hits: 35

```
Alignments Results for sequence: EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSE
Sequence name: my_query_sequence
Selected property: Hydropathy

1A0Q:H | Identicals: 217/217 | Hydrophobic Similars: 70 | Neutral Similars: 96 | Hydrophilic Similars: 51
EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYLCKRGYYGRSNVDYWGQGTTLTVSSAK
EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYLCKRGYYGRSNVDYWGQGTTLTVSSAK

1MEX:H | Identicals: 195/217 | Hydrophobic Similars: 68 | Neutral Similars: 90 | Hydrophilic Similars: 44
EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYLCKRGYYGRSNVDYWGQGTTLTVSSAK
QVQLQQSDAELVKPGASVKISCKASGYTFTDHAIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYFCKMEY-----LDYWGQGTTLTVSSGG

3DGG:B | Identicals: 190/217 | Hydrophobic Similars: 66 | Neutral Similars: 92 | Hydrophilic Similars: 43
EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYLCKRGYYGRSNVDYWGQGTTLTVSSAK
EVKLQQSGPELVKPGASVKMSCKASGYTFTSYVMHWVKQKPGQGLEWIGYINPYNDDTNYNEKFKGKATLTSDKSSSTAYMELSSLTSEDSAVYYCAREDYYGSRWGYWGQGTTLTVSSAK

3DGG:D | Identicals: 190/217 | Hydrophobic Similars: 66 | Neutral Similars: 92 | Hydrophilic Similars: 43
EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYLCKRGYYGRSNVDYWGQGTTLTVSSAK
EVKLQQSGPELVKPGASVKMSCKASGYTFTSYVMHWVKQKPGQGLEWIGYINPYNDDTNYNEKFKGKATLTSDKSSSTAYMELSSLTSEDSAVYYCAREDYYGSRWGYWGQGTTLTVSSAK

2A6D:B | Identicals: 188/217 | Hydrophobic Similars: 65 | Neutral Similars: 93 | Hydrophilic Similars: 44
EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYLCKRG--YYGRSNVDYWGQGTTLTVSSA
EVQLQQSGAELVRAGSSVKMSCKASGYTFTSYGINWVKQRPGQGLEWIGYINPGNGYTKYNEKFKGKTTLTVDKSSSTAYMQLRSLTSEDSAVYFCARSVYYYGSSYFDYWGQGTTLTVSSA

2A6D:H | Identicals: 188/217 | Hydrophobic Similars: 65 | Neutral Similars: 93 | Hydrophilic Similars: 44
EVQLQESDAELVKPGASVKISCKASGYTFTDHVIHWVKQKPEQGLEWIGYISPGNGDIKYNEKFKGKATLTADKSSSTAYMQLNSLTSEDSAVYLCKRG--YYGRSNVDYWGQGTTLTVSSA
EVQLQQSGAELVRAGSSVKMSCKASGYTFTSYGINWVKQRPGQGLEWIGYINPGNGYTKYNEKFKGKTTLTVDKSSSTAYMQLRSLTSEDSAVYFCARSVYYYGSSYFDYWGQGTTLTVSSA
```

Figure 1: The graphical user interface for the ASSP platform provides an intuitive pipeline for running similarity searches that eliminate the chance for human error, while at the same time providing a graphical, easy to comprehent results output window. The graphical user interface of the ASSP platform, consists of the following main windows: 1. The blast type selection option. 2. The input query window. 3. Sequence name and number of hits selections 4. The result-output area. The alignmed regions are color-coded in accordance to the IMGT coloring scheme for hydropathy. The identical residues are colored green, the hydrophobic residues are colored blue, the neutral residues are colored red and the hydrophilic residues are colored yellow.

then ASSP will move swiftly to the actual similarity blast search.

The screening process of ASSP is broken down in two steps. First a conventional primary sequence-based blast is performed with a threshold value of 30% identity, when temporary file is created with all sequences sharing more than 30% identity with the query antibody sequence (using the blosum62 substitution matrix). Then the custom made hydropathy substitution matrix is engaged and the previously filtered entries are ranked according to their identity/similarity scores based on their hydropathy profiles.

The hydropathy matrix has been created using the antibody hydropathy index from IMGT [18]. Results in the form of alignments and similarity percentages are calculated, scored, ranked and returned to the user through the same graphical interface that has been specifically designed to simplify the task for the user and to eliminate the possibility of a user-inflicted error (see Figure 1). The output window is color-coded in accordance to the IMGT coloring scheme for hydropathy. The identical residues are colored green, the hydrophobic residues $(4,5$ to $-0,9)$ are colored blue, the neutral residues $(-0,4$ to $-3,2)$ are colored red and the hydrophilic residues $(-3,5$ to $-4,5)$ are colored yellow. The results are outputted and saved in easy to manipulate text-based text/ascii files for future analysis.

The secondary description code that ASSP has adopted is the same with the one DSSP has been using for many years now [13]. This was intentionally done for ease of use and backward compatibility issues. More specifically an eight-letter description code is used. Using just eight letters, instead of the traditional twenty amino acid letters, makes similarity searches ever more efficient and faster than ever before. The eight letter secondary element code comprises of the following letters: H for α helix conformation, B for residues in isolated beta-bridge, E for extended strands that participate in beta ladders, G for 3/10 helices, I for pi helices, T for hydrogen bonded turns and S for bends. C is used for the blank space in the DSSP secondary structure, which represents a loop or an irregular element. Other major suites, such as the PDBFINDER suite, also adopt this convention with unstructured protein regions [11]. Same WHATIF uses C, as many times leaving a blank may be confusing, misleading and inconvenient [31, 10, 12]. A batch execultion mode has also been prepared for the ASSP suite. A simple text file is required with as many sequences as the user wishes, each one stored in a different line. The ASSP algorithm will then automatically read that file line by line and execute the antibody similarity search for as many times as the lines of the input batch file. This comes quite handy for those who wish to perform secondary structure similarity searches on large databases of protein or peptide sequences [28, 30, 29, 3, 27, 16]. Finally, an extensive manual and use-case based examples for the use of ASSP, will pop-up through the Help button, using the operating systems HTML browser application.

3 Conclusions

In conclusion, the ASSP toolkit provides a novel, quick and reliable tool for in silico antibody similarity searches in one pipelined platform under a user friendly graphical user interface. We therefore, propose that our structural similarities application described here would yield results of great interest to many antibody-related scientific disciplines. The ASSP platform is distributed as freeware under a GNU license.

4 Availability

Availability: ASSP can be freely downloaded via our dedicated server system at `http://www.bioacademy.gr/bioinformatics/assp/index.html`
ASSP is an open source, cross platform application available freely to all users under a GNU license basis. The full package, including installation scripts, figures, a full description, a detailed manual, complete tutorials as hands-on use cases, software prerequisites and various examples can be downloaded at: `http://www.bioacademy.gr/bioinformatics/assp/`. Prior to download; check the provided information on the website about software prerequisites. Please email comments and bug reports at *dvlachakis@bioacademy.gr*.

Acknowledgements

This work was partially supported by:
1. The BIOEXPLORE research project. BIOEXPLORE research project falls under the Operational Program “Education and Lifelong Learning” and is co-financed by the European Social Fund (ESF) and National Resources.
2. European Union (European Social Fund - ESF) and Greek national funds

through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

The authors would like to thank Prof. Marie-Paule LeFranc and the IMGT institute for their constructive comments and support.

References

[1] C. Berbalk, C. S. Schwaiger, and P. Lackner. Accuracy analysis of multiple structure alignments. *Protein Sci.*, 18(10):2027–2035, Oct 2009.

[2] J. U. Bowie, J. F. Reidhaar-Olson, W. A. Lim, and R. T. Sauer. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science (New York, N.Y.)*, 247(4948):1306–1310, Mar. 1990.

[3] C. S. Carvalho, D. Vlachakis, G. Tsiliki, V. Megalooikonomou, and S. Kossida. Protein signatures using electrostatic molecular surfaces in harmonic space. *PeerJ*, 1:e185, 2013.

[4] C. Chothia. The nature of the accessible and buried surfaces in proteins. *Journal of Molecular Biology*, 105(1):1–12, July 1976.

[5] S. Y. Chung and S. Subbiah. A structural explanation for the twilight zone of protein sequence homology. *Structure*, 4(10):1123–1127, Oct. 1996.

[6] J. D. Clements and R. E. Martin. Identification of novel membrane proteins by searching for patterns in hydropathy profiles. *Eur. J. Biochem.*, 269(8):2101–2107, Apr 2002.

[7] D. Eisenberg, R. M. Weiss, and T. C. Terwilliger. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci U S A*, 81(1):140–144, Jan. 1984.

[8] D. M. Engelman, T. A. Steitz, and A. Goldman. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annual review of biophysics and biophysical chemistry*, 15:321–353, 1986.

[9] C. Gille. STRAP: Structure based sequences alignment program. `http://www.bioinformatics.org/strap/index2.html`.

[10] M. L. Hekkelman, T. A. Te Beek, S. R. Pettifer, D. Thorne, T. K. Attwood, and G. Vriend. WIWS: a protein structure bioinformatics Web service collection. *Nucleic Acids Res.*, 38(Web Server issue):W719–723, Jul 2010.

[11] R. Hooft, C. Sander, M. Scharf, and G. Vriend. The pdbfinder database: a summary of pdb, dssp and hssp information with added value. *Computer applications in the biosciences : CABIOS*, 12(6):525–529, 1996.

[12] R. W. W. Hooft, G. Vriend, C. Sander, and E. E. Abola. Errors in protein structures. *Nature*, 381(6580):272, May 1996.

[13] W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, Dec 1983.

[14] S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, and M. H. Hecht. Protein design by binary patterning of polar and nonpolar amino acids. *Science*, 262(5140):1680–1685, 1993.

[15] R. Kolodny, P. Koehl, and M. Levitt. Comprehensive evaluation of protein structure alignment methods: scoring by geometric measures. *J. Mol. Biol.*, 346(4):1173–1188, Mar 2005.

[16] E. Krissinel. Enhanced fold recognition using efficient short fragment clustering. *Journal of Molecular Biochemistry*, 1(2), 2012.

[17] J. Kyte and R. F. Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1):105–132, May 1982.

[18] M. P. Lefranc, V. Giudicelli, C. Ginestoux, J. Bodmer, W. Muller, R. Bontrop, M. Lemaitre, A. Malik, V. Barbie, and D. Chaume. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.*, 27(1):209–212, Jan 1999.

[19] J. S. Lolkema and D. J. Slotboom. Estimation of structural similarity of membrane proteins by hydropathy profile alignment. *Mol. Membr. Biol.*, 15(1):33–42, 1998.

[20] J. S. Lolkema and D. J. Slotboom. Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins. *FEMS Microbiol. Rev.*, 22(4):305–322, Oct 1998.

[21] B. W. Matthews. *Hydrophobic Interactions in Proteins.* John Wiley & Sons, Ltd, 2001.

[22] G. Mayr, F. S. Domingues, and P. Lackner. Comparative analysis of protein structure alignments. *BMC Struct. Biol.*, 7:50, 2007.

[23] C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala. Forces contributing to the conformational stability of proteins. *FASEB journal*, 10(1):75–83, Jan. 1996.

[24] J. Pánek, I. Eidhammer, and R. Aasland. A new method for identification of protein (sub)families in a set of proteins based on hydropathy distribution in proteins. *Proteins*, 58(4):923–934, 03 2005.

[25] C. Pommié, S. Levadoux, R. Sabatier, G. Lefranc, and M.-P. Lefranc. Imgt standardized criteria for statistical analysis of immunoglobulin v-region amino acid properties. *Journal of Molecular Recognition*, 17(1):17–32, 2004.

[26] R. B. Russell, M. A. Saqi, R. A. Sayle, P. A. Bates, and M. J. Sternberg. Recognition of analogous and homologous protein folds: analysis of sequence and structure conservation. *J Mol Biol*, 269(3):423–439, June 1997.

[27] D. Vlachakis, D. Tsagkrasoulis, V. Megalooikonomou, and S. Kossida. Introducing drugster: a comprehensive and fully integrated drug design, lead and structure optimization toolkit. *Bioinformatics*, 29(1):126–128, 2013.

[28] D. Vlachakis, D. Tsagkrasoulis, G. Tsiliki, and S. Kossida. The future of structural bioinformatics in the post-genomic era. *EMBnet. journal*, 18(1):pp–3, 2012.

[29] D. Vlachakis, S. C. Tsaniras, C. Feidakis, and S. Kossida. An in silico 3D study of the biglycan core protein, using homology modelling techniques. *Journal of Molecular Biochemistry*, 2(2), 2013.

[30] D. Vlachakis, G. Tsiliki, D. Tsagkrasoulis, C. S. Carvalho, V. Megalooikonomou, and S. Kossida. Speeding up the drug discovery process: structural similarity searches using molecular surfaces. *EMBnet.journal*, 18(1), 2012.

[31] G. Vriend. WHAT IF: a molecular modeling and drug design program. *J Mol Graph*, 8(1):52–56, Mar 1990.

[32] H. Xiong, B. L. Buckwalter, H. M. Shieh, and M. H. Hecht. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proceedings of the National Academy of Sciences*, 92(14):6349–6353, 1995.

[33] X. J. Yu and D. H. Walker. Sequence and characterization of an Ehrlichia chaffeensis gene encoding 314 amino acids highly homologous to the NAD A enzyme. *FEMS Microbiol. Lett.*, 154(1):53–58, Sep 1997.