Structural Biology of Moonlighting — Lessons from Antibodies

Andrew C.R. Martin*
Institute of Structural and Molecular Biology,
Division of Biosciences,
University College London,
Gower Street,
London WC1E 6BT

August 1, 2014

Keywords: protein structure; complementarity determining regions; evolution

Abstract

Protein moonlighting is the property of a number of proteins to have more than one function. However, the definition of moonlighting is somewhat imprecise with different interpretations of the phenomenon. True moonlighting occurs when an individual evolutionary protein domain has one well-accepted rôle and a secondary unrelated function.

The 'function' of a protein domain can be defined at different levels. For example, while the function of an antibody variable fragment (Fv) could be described as 'binding'; a more detailed definition would also specify the molecule to which the Fv region binds. Using this detailed definition, antibodies *as a family* are consummate moonlighters. However, individual antibodies do not moonlight — the multiple functions they exhibit (first binding a molecule and second triggering the immune response) are encoded in different domains and, in any case, are related in the sense that they are both part of what an antibody needs to do. Nonetheless, antibodies provide interesting lessons on the ability of proteins to evolve binding functions. Remarkably-similar antibody sequences can bind completely different antigens, suggesting that evolving the ability to bind a protein can result from very subtle sequence changes.

1 Introduction

The traditional dogma of biology has been that DNA is copied into RNA which encodes a protein with a single function. Over the last 10–15 years, this dogma has been challenged in a number of ways. First, the importance of RNA as a functional molecule rather than simply a messenger and involved in protein synthesis has been revealed. Many types of non-coding RNA (ncRNA) have been identified including snoRNAs, miRNAs, siRNAs and snRNAs. In mammals, snoRNAs are involved in regulation of the 'spliceosome', a molecular machine which, like the ribosome, contains proteins and RNA, and performs RNA splicing. More importantly, many ncRNAs are involved in the regulation of many thousands of genes. Trans-acting ncRNAs such as micro-RNAs (miRNAs) regulate gene expression through partial complementary to mRNA molecules, generally down-regulating gene expression. Cis-acting ncRNAs and riboswitches are encoded within the untranslated regions of protein coding genes and can bind molecules that control the rate at which proteins are synthesized [1–4].

*andrew@bioinf.org.uk — to whom correspondence should be addressed.

Second, the frequency and importance of alternative splicing has been revealed. When the human genome project was completed, preliminary analysis suggested there were approximately 20,000 genes, far fewer than expected for an organism of our perceived complexity. Indeed that number is now believed to be an over-estimate as protein products cannot be identified for a number of the predicted genes. However, for approximately one third of genes, more than one protein is encoded — alternative splicing means that, on average, every gene encodes around three different proteins.

Third, more and more cases are being discovered where proteins 'moonlight'. In other words, in addition to a primary well-known function, they have a secondary function generally unrelated to the primary function.

2 What do we mean by function?

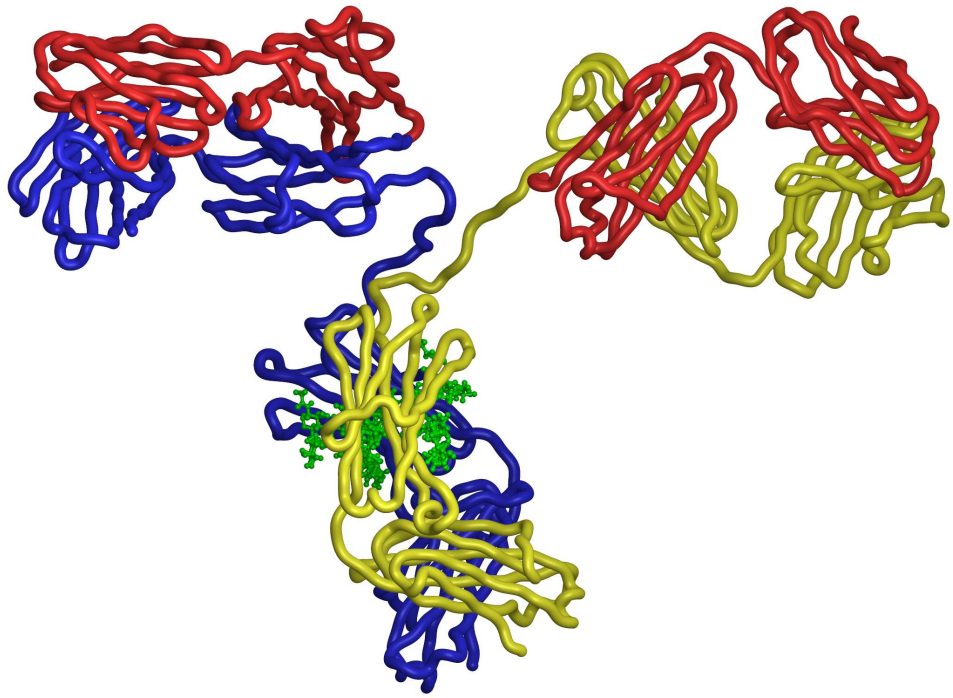

Figure 1: *The structure of a typical IgG antibody, PDB file 1igt rendered using QTree (*`http://www.bioinf.org.uk/software/qtree/`*). The antibody consists of two identical heavy chains (shown in blue and yellow) and two identical light chains (shown in red). Attached carbohydrates are shown in green. The Fc (crystallizable fragment) forms the stem of the Y-shape and is responsible for the effector functions of the antibody such as binding to Fcγ receptor or complement C1q, triggering the rest of the immune response. The two arms are known as antigen binding (Fab) fragments with the distal domains from light and heavy chains coming together to form the variable fragment (Fv) which is responsible for antigen binding. Within this, the actual antigen combining site consists of just six loops known as complementarity determining regions (CDRs) — three from the light chain and three from the heavy chain.*

Before elaborating on the field of protein moonlighting, it is necessary to understand what we mean by function. This may appear to be a simple concept, but actually hides huge layers of complexity. For example, one might describe the function of a protein as being that it is an enzyme and therefore acts as a catalyst. A more detailed description of the function of that protein might be that it catalyzes a particular type of reaction — for example hydrolysis. Further levels of detail can be applied indicating the type of bond it acts upon and more details of the substrate specificity. This hierarchical classification is encoded in the 'Enzyme Classification' (EC) numbering system (`http://www.chem.qmul.ac.uk/iubmb/enzyme/`). The first level of this hierarchy describes the general type of reaction (1 - Oxidoreductases; 2 - Transferases; 3 - Hydrolases; 4 - Lyases; 5

- Isomerases; 6 - Ligases) while three further levels of the hierarchy depend on the top level, but give more specific information. For example the EC number 3.4.11.4 describes hydrolases (EC 3) which act on peptide bonds (EC 3.4) cleaving off the N-terminal amino acid (EC 3.4.11) and are specific to tripeptides (EC 3.4.11.4).

Taking antibodies as another example, their overall function is to act as an 'adapter plug' between the huge variability of potential antigens (generally resulting from pathogens) and the constancy of the immune system. A typical IgG antibody is illustrated in Figure 1. Looking at the variable fragment (Fv) of the antibody, its function is to bind antigens, but more specifically, the function of a *given* antibody is to bind, with high affinity and specificity, to a particular antigen.

The categorization of function is further described by the Gene Ontology (GO, `http://www.geneontology.org`). GO defines three separate ontologies describing biological process (the rôle of this protein in the cell, tissue, organ or organism), cellular component (where the process happens) and molecular function (precisely what this protein is doing as part of that process).

Moonlighting is not trivial to define, and different people in the field use the term in different ways. A relatively well-accepted definition is that moonlighting occurs when a protein carries out two (or more) functions that are not merely different aspects of the same overall function. Stricter definitions could include multiple functions that have different EC numbers at the 3rd level of the hierarchy (differences at the 4th level — generally the specific substrate — would count as 'catalytic promiscuity' rather than true moonlighting), or that functions belong to different biological processes in GO.

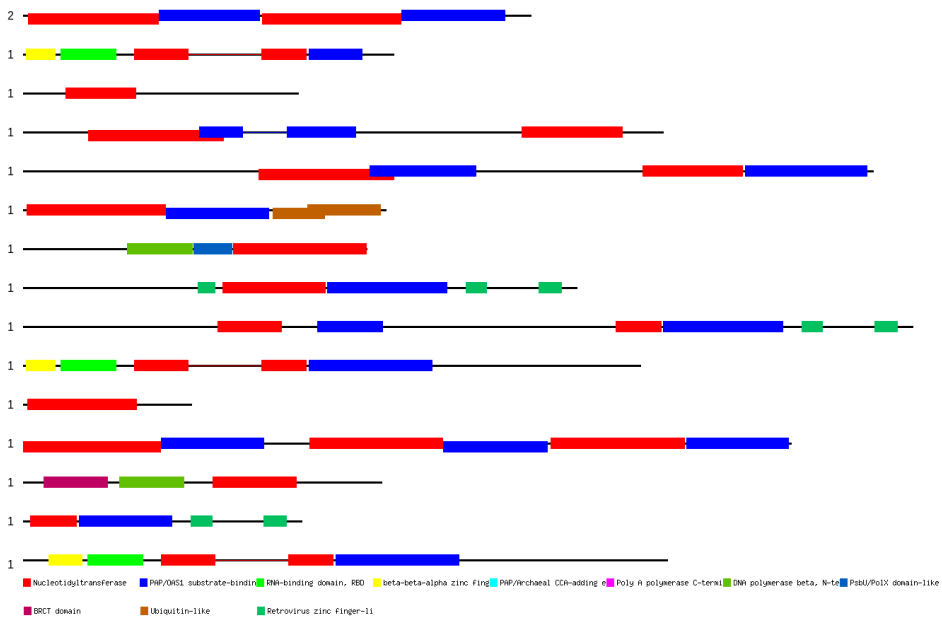

Figure 2: *An example from the Superfamily Database in which a nucleotidyl transferase domain (shown in red) is seen in human proteins to be coupled with many other types of domain.*

3 Moonlighting and structure

So how does moonlighting relate to protein structure? Proteins consist of one or more domains. Figure 1 also illustrates this. The antibody heavy chains each have four structural domains while the light chains have two. Structural domains are generally compact self-contained folding units. However, domains can also be viewed as 'units

of inheritance'. Two 'secondary databanks' (resources containing data derived from 'primary' data such as sequences and structures, typically grouping data into families) are available that classify protein domains. CATH [5] uses purely structural domains, while SCOP [6] also contains evolutionary domains; i.e. regions of sequence that are seen always to be inherited as a unit. Thus a SCOP evolutionary domain can include one or more structural domains. This concept is illustrated in Figure 2 which is taken from Julian Gough's Superfamily database (`http://supfam.cs.bris.ac.uk/SUPERFAMILY/cgi-bin/allcombs.cgi?sf=81301`). The figure shows occurrences of the nucleotidyl transferase domain — an evolutionary domain — in human proteins where it is partnered with several other domains including PAP/OAS1 binding domains, RNA-binding domains, zinc fingers, poly-A polymerase, DNA polymerase, PsbU/PolX, BRCT domains and ubiquitin.

As Figure 2 illustrates, it is common for a domain having a given function (nucleotidyl transferase) to partner with multiple different domains having different functions, generally as a result of gene fusion events. Consequently these protein chains will have multiple functions (which may or may not be aspects of some overall function), but cannot be considered to be examples of true moonlighting. Thus another strict definition of true moonlighting would be to require that multiple functions occur within a SCOP evolutionary domain.

4 Lessons from antibodies

Given our definitions of moonlighting above, do antibodies moonlight? It should be clear that they do not. Yes an antibody has two distinct functions — it must bind to an antigen (a function mediated by the Fv region) and it must bind to other proteins of the immune system (including the Fcγ receptor and the complement C1q molecule) to trigger an immune response. However, these are clearly different aspects of the same overall 'adapter plug' function.

Homology is the simple concept that two proteins (or more strictly the segments of genes that encode evolutionary domains within those proteins) are descended from a common ancestor. Of course we can never go back in time to find out if two proteins have truly descended from a common ancestor, but a sequence identity of more than 40% over at least 100 amino acids is generally taken as indicative of homology. It is the case that structure is better conserved than sequence and consequently two proteins can diverge quite considerably in sequence while maintaining a conserved structure. It is also the case that homologous families of proteins generally have related functions. Thus having a conserved structure and a related function can be used as an indicator of homology even when the signal available from the sequence information is too weak to be able to detect homology.

Since homologous proteins share related functions, one would expect the antibody Fv region in different antibodies (all of which easily pass the 40% over 100 amino acids threshold) to share a common function. Given our hierarchy of function, clearly Fv regions do indeed share the function of 'binding', but at a more detailed level the family (rather than an individual antibody) can be considered moonlighting because of the enormous range of antigens against which members of the family can bind. These antigens range from proteins and peptides through carbohydrates, lipids and small organic molecules known as haptens [7, 8]. Consequently, if one considers the shifts in antibody sequence that are needed radically to change the nature of the antigen to which the antibody binds, one obtains a picture of the types of changes that may be needed to evolve a secondary moonlighting binding function in other proteins.

To examine this concept, 'complete' antibodies (i.e. those for which both the light and heavy chain sequences are known) were extracted from the last public release of the Kabat sequence databank [9] using KabatMan [10]. Only antibodies for which the antigen is known were selected and the antigen information together with the sequences of the six complementarity determinig region (CDR) loops that form the antigen combining site were

a)

```
         L1           L2       L3         H1     H2                 H3
P65D6-3 RASQDISNYLN  YTSRLHS  QQGNTLPRT  SYGIN  YINPGNGYTKYNEKFKG  SHYYGGSYYFDY
        ||||||||||  |||||||  |||||      |||||  ||||||||||||||||       |||||||
ASWU1   RASQDISNYLN  YTSRLHS  QQGNT----  SYGIN  YINPGNGYTKYNEKFKG  EGA--GSYYFDY
```

b)

```
    L1          L2       L3         H1     H2                 H3
1G2 SASSSISYMH  GTSKLAS  HQRSSYP-T  SYWIE  EILPGSGSTNYNEKFKG  T-----DY
    |||||+||||   || |||   |||||| |  |||||  ||||||||||||||||         ||
E12 SASSSVSYMH  STSNLAS  QQRSSYPYT  SYWIE  EILPGSGSTNYNEKFKG  RARNYFDY
```

c)

```
     L1           L2       L3         H1     H2                 H3
D1.3 RASGNIHNYLA  YTTTLAD  QHFWSTPRT  GYGVN  MIWGDGNTDYNSALKS  ERDYRLDY
     ||||||||||  |  ||||  |||||||||  |||||  ||||||||||||||||  | || |||
AF14 RASGNIHNYLA  YAETLAD  QHFWSTPRT  GYGVN  MIWGDGNTDYNSALRS  ELDYTLDY
```

Figure 3: *Examples of antibodies having very similar CDRs but radically different antigens. a) Antibodies P65D6-3 which binds p-azophenylarsonate and ASWU1 which binds nucleolar particles U3 and U8; b) Antibodies 1G2 which binds cytochrome C and E12 which binds mesothelin; c) Antibodies D1.3 which binds hen egg lysozyme and AF14 which binds another antibody E5.2.*

obtained. The three light chain CDRs were concatenated with an 'X' character used as a boundary between each CDR to form a 'composite CDR sequence'; the three heavy chain CDRs were handled in the same way. An all-against-all comparison of the antibodies was then performed, aligning the light chain composite CDR sequences with one another and the heavy chain composite CDR sequences with one another. In performing the alignment, the scoring matrix was modified to give an X/X match a very high score and an X *vs.* anything else match a very low score thus forcing the alignments to align the X characters used as boundaries between the CDRs. Examples with >90% sequence identity in both light and heavy chain composite CDR sequences were identified and examined manually to identify examples where the antigens were radically different. Three such examples are presented here.

The first example consists of antibody P65D6-3 [11] which binds p-azophenylarsonate (a small organic molecule) and antibody ASWU1 [12] which binds nucleolar particles U3 and U8 (small nucleolar RNA and ribonucleoproteins or snoRNPs). As shown in Figure 3a, the CDRs of these antibodies differ by just three amino acids and six insertions over a total of 55/61 amino acids. This is less different than many antibodies that bind to the same antigen or, for example, antibodies 1G2 [13] and 5D9 [14] which bind mouse and pigeon cytochrome C respectively.

The second example is antibody 1G2 which binds cytochrome C and antibody E12 [15] which binds mesothelin. Cytochrome C and mesothelin are completely unrelated proteins having less than 20% sequence identity over the aligned regions. As shown in Figure 3b, the CDRs of these antibodies differ by just four amino acids (one of which is a conservative change) and six insertions over a total of 50/56 amino acids.

The third example is antibodies D1.3 [16] which binds hen egg lysozyme and AF14 which is an anti-idiotypic antibody binding the CDRs of another antibody, E5.2 [17]. Clearly hen egg white lysozyme and antibody E5.2 are not related proteins having less than 20% sequence identity when lysozyme is compared with either the light or heavy antibody Fv region sequences. As shown in Figure 3c, the CDRs of these antibodies differ by just four out of 56 amino acids. This is a rather interesting example because, while lysozyme and E5.2 are not related, E5.2 is itself an anti-idiotypic antibody which binds to D1.3. In other words AF14 binds E5.2 which binds D1.3 which binds lysozyme. Consequently E5.2 is somehow mimicking the surface of lysozyme.

The lesson from these examples is clear: antibodies can radically change their speci-

ficity by making minimal changes in the CDRs and thus only small changes may be needed to introduce new specific and high affinity binding ability into proteins allowing them to moonlight.

5 Conclusions

A key problem in understanding moonlighting is defining what we really mean by a moonlighting protein. It is clear that there is a hierarchy of definitions with varying strictness and that this needs to be considered in the context of evolution and protein structure. In some cases, proteins change conformation in order to moonlight. For example, in the presence of low iron levels, aconitase, an enzyme of the tri-carboxylic acid cycle, loses its iron-sulphur cluster and the concomitant conformational change allows it to bind to iron response elements in mRNA encoding proteins related to iron import [18]. In many other cases however, there is no evidence of any need for structural change to allow the moonlighting function(s) to occur. The lesson from antibodies in such cases is that only small changes are needed to the sequence on the surface of a protein for novel binding interactions to be generated, but in general little is known about the structural biology behind moonlighting functions, or indeed whether there are any trends to be discovered.

In the third antibody example of D1.3 and AF10, one can of course also look at D1.3 itself and realize that its CDRs bind to both lysozyme and antibody E5.2. Could we consider this to be a true moonlighting function? Probably not. Moonlighting functions, to be considered truly functional, must be under selective pressure; in other words there must be a selective pressure to gain and maintain the moonlighting function. Thus binding alone is not enough to be considered a moonlighting function. Indeed it is a distinct disadvantage for host proteins to bind to invading pathogens. The ability of the host ICAM-1 protein to bind rhinovirus [19], or of CD4 to bind HIV [20], cannot be considered a moonlighting function. On the other hand, if the pathogen exploits an existing protein to act as an adhesin or a virulence factor in order to allow infection, that is a clear case of moonlighting.

References

1. Nudler, E. and Mironov, A. S. (2004) The riboswitch control of bacterial metabolism. Trends Biochem Sci, **29**, 11–17

2. Tucker, B. J. and Breaker, R. R. (2005) Riboswitches as versatile gene control elements. Curr Opin Struct Biol, **15**, 342–348

3. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet, **20**, 44–50

4. Batey, R. T. (2006) Structures of regulatory elements in mRNAs. Curr Opin Struct Biol, **16**, 299–306

5. Sillitoe, I., Cuff, A. L., Dessailly, B. H., Dawson, N. L., Furnham, N., Lee, D., Lees, J. G., Lewis, T. E., Studer, R. A., Rentzsch, R., Yeats, C., Thornton, J. M., and Orengo, C. A. (2013) New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic Acids Res, **41**, D490–D498

6. Andreeva, A., Howorth, D., Chothia, C., Kulesha, E., and Murzin, A. G. (2014) SCOP2 prototype: a new approach to protein structure mining. Nucleic Acids Res, **42**, D310–D314

7. MacCallum, R. M., Martin, A. C., and Thornton, J. M. (1996) Antibody-antigen interactions: Contact analysis and binding site topography. J Mol Biol, **262**, 732–745

8. Collis, A. V. J., Brouwer, A. P., and Martin, A. C. R. (2003) Analysis of the antigen combining site: Correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. J Mol Biol, **325**, 337–354

9. Johnson, G. and Wu, T. T. (2001) Kabat Database and its applications: Future directions. Nucleic Acids Res, **29**, 205–206

10. Martin, A. C. (1996) Accessing the Kabat antibody sequence database by computer. Proteins, **25**, 130–133

11. Parhami-Seren, B., Wysocki, L., and Margolies, M. N. (1989) The amino acid residues at the VH-D-JH junctions affect the affinity of anti-p-azophenylarsonate antibodies. J Immunol, **143**, 4090–4097

12. Monestier, M., Losman, M. J., Novick, K. E., and Aris, J. P. (1994) Molecular analysis of mercury-induced antinucleolar antibodies in H-2S mice. J Immunol, **152**, 667–675

13. Wang, D. N., Chen, H. T., Liao, J., Akolkar, P. N., Sikder, S. K., Gruezo, F., and Kabat, E. A. (1990) Two families of monoclonal antibodies to alpha(1—6)dextran, VH19.1.2 and VH9.14.7, show distinct patterns of J kappa and JH minigene usage and amino acid substitutions in CDR3. J Immunol, **145**, 3002–3010

14. Minnerath, J. M., Wakem, L. P., Comfort, L. L., Sherman, F., and Jemmerson, R. (1995) The BALB/c mouse B-cell response to pigeon cytochrome c initiates as a heteroclitic response specific for the self antigen mouse cytochrome c. Proc Natl Acad Sci U S A, **92**, 12379–12383

15. Chowdhury, P. S., Chang, K., and Pastan, I. (1997) Isolation of anti-mesothelin antibodies from a phage display library. Mol Immunol, **34**, 9–20

16. Verhoeyen, M., Milstein, C., and Winter, G. (1988) Reshaping human antibodies: Grafting an antilysozyme activity. Science, **239**, 1534–1536

17. Wang, D., Liao, J., Mitra, D., Akolkar, P. N., Gruezo, F., and Kabat, E. A. (1991) The repertoire of antibodies to a single antigenic determinant. Mol Immunol, **28**, 1387–1397

18. Philpott, C. C., Klausner, R. D., and Rouault, T. A. (1994) The bifunctional iron-responsive element binding protein/cytosolic aconitase: the role of active-site residues in ligand binding and regulation. Proc Natl Acad Sci U S A, **91**, 7321–7325

19. Tomassini, J. E., Graham, D., DeWitt, C. M., Lineberger, D. W., Rodkey, J. A., and Colonno, R. J. (1989) cDNA cloning reveals that the major group rhinovirus receptor on HeLa cells is intercellular adhesion molecule 1. Proc Natl Acad Sci U S A, **86**, 4907–4911

20. McDougal, J. S., Nicholson, J. K., Cross, G. D., Cort, S. P., Kennedy, M. S., and Mawle, A. C. (1986) Binding of the human retrovirus HTLV-III/LAV/ARV/HIV to the CD4 (T4) molecule: Conformation dependence, epitope mapping, antibody inhibition, and potential for idiotypic mimicry. J Immunol, **137**, 2937–2944