Review

The adenomatous polyposis coli protein 30 years on

James Abbott a, Inke S. Näthke b, *

a Division of Computational Biology & D’Arcy Thompson Unit, University of Dundee, Dow Street, Dundee, DD2 1EH, United Kingdom
b Division of Molecular Cell and Developmental Biology, University of Dundee, Dow Street, Dundee DD2 1EH, United Kingdom

ABSTRACT

Mutations in the gene encoding the Adenomatous polyposis coli protein (APC) were discovered as driver mutations in colorectal cancers almost 30 years ago. Since then, the importance of APC in normal tissue homeostasis has been confirmed in a plethora of other (model) organisms spanning a large evolutionary space. APC is a multifunctional protein, with roles as a key scaffold protein in complexes involved in diverse signalling pathways, most prominently the Wnt signalling pathway. APC is also a cytoskeletal regulator with direct and indirect links to and impacts on all three major cytoskeletal networks. Correspondingly, a wide range of APC binding partners have been identified. Mutations in APC are extremely strongly associated with colorectal cancers, particularly those that result in the production of truncated proteins and the loss of significant regions from the remaining protein. Understanding the complement of its role in health and disease requires knowing the relationship between and regulation of its diverse functions and interactions. This in turn requires understanding its structural and biochemical features. Here we set out to provide a brief overview of the roles and function of APC and then explore its conservation and structure using the extensive sequence data, which is now available, and spans a broad range of taxonomy. This revealed conservation of APC across taxonomy and new relationships between different APC protein families.

1. Introduction

Mutations in the gene encoding the Adenomatous polyposis coli protein (APC) were discovered as driver mutations in colorectal cancers almost 30 years ago [1–4] initially in mouse models and also patients [5]. Since then, the importance of APC in normal tissue homeostasis has been confirmed in a plethora of other (model) organisms including C. elegans, Drosophila, rats, zebra fish, pigs, and more [6–9].

In this review we aim to increase our understanding of APC and its ability to contribute to many different cellular processes, by first outlining some of the key functions attributed to APC and then summarising results from interrogating and comparing sequences of currently known APC proteins.

1.1. Main functions of APC

A key role for APC is its ability to act as a scaffold protein in complexes involved in diverse signalling pathways. Most prominent is the Wnt signalling pathway. Here, APC is a crucial player in assembling the protein complex that phosphorylates β-catenin targeting it for degradation in the absence of Wnt signals. In the presence of Wnt signals, this complex is inactivated, β-catenin can accumulate and direct transcriptional changes that are associated with proliferative, less differentiated states [10]. In addition, APC is a cytoskeletal regulator with direct and indirect links to and impacts on all three major cytoskeletal networks [11].

Both the function of APC in Wnt signalling and in cytoskeletal regulation and the resulting contribution to the behaviour of cells and tissues have been extensively summarised and reviewed and highlight the complexity of the interactions of the APC protein and their outputs for cellular function [11]. Studies interrogating the APC interactions directly, further support the diversity and complexity of its links to many different cellular processes [12]. More recently, the idea that APC itself can undergo phase separation adds another layer of complexity to its regulation and could be a means for APC to respond to local intracellular conditions as has been shown in other disordered proteins [13,14]. It is thus not surprising that mutations in this one gene can have such profound effects, particularly on the lining of the intestinal tract, the most dynamic tissue in the body.

Additional complexity is introduced by the fact that two related but...
distinct APC proteins, APC and APC2, have been described [15]. The relationship between them is not entirely clear. Mutations in cancers, particularly colorectal cancers, are extremely common only in the former although both seem able to support Wnt signalling and there have been some reports about APC2 also able to affect the Wnt pathway [16–18]. The majority of research has focussed on APC with comparatively little information available about APC2. One exception is in Drosophila where both of the expressed APC proteins have been investigated. However, how each of these relate to APC and APC2 is not entirely clear (see below).

1.2. APC in the digestive tract

The human intestinal tract is estimated to shed 20–50 million cells per minute leading to the renewal of the entire lining every five days [19]. Normal homeostasis relies on stem cells that produce transit amplifying cells to generate the complement of cell types that constitute the intestinal epithelium. Most prominent are absorptive enterocytes that are protected by mucus-secreting Goblet cells, enteroendocrine and Paneth cells. The latter are found in intestinal crypts, where they provide constant production of new cells from stem cells, differentiation into lineages, and shedding from the tissue layer have to be well balanced for normal tissue function. In addition, the system can respond to injury and rapidly replace damaged areas to maintain the crucial barrier function that prevents entry of pathogens from the lumen of the intestinal tract. Directed migration of cells from crypts towards the lumen is an integral feature of these processes [23]. Directed migration is also important for the rapid closure of any areas denuded of the epithelial layer in response to injury and inflammation [24]. Integrating key signalling pathways allows careful tuning of this system to create required responses in a locally and temporally regulated manner. The ability of APC to interact with many different proteins and thus contribute to and coordinate many different signalling pathways, makes it a key integrator of such external signals. It plays an important role in coordinating the cellular responses required for normal homeostasis.

Mutations in APC in cancer most commonly produce truncated APC protein of varying lengths. Such truncated APC proteins lack many of its interaction sites, reducing its ability to integrate incoming signals. It is thus not surprising that loss of APC has profound effects on intestinal epithelial organisation, including loss of differentiation and changes in tissue shape [25–27] to initiate tumours. Losing fully functional APC impacts on all the processes required for normal homeostasis of a rapidly renewing epithelium: it activates Wnt signalling – promoting proliferation at the expense of differentiation, it reduces mitotic fidelity – increasing genetic instability, it compromises directed cell migration – increasing the residence time of APC-mutant cells in crypts, and it changes the direction of detachment of cells from the basal layer – reducing sloughing off into the gut lumen [28]. The latter two consequences directly bestow a competitive advantage over wild type cells by increasing the probability of APC-mutant cells to remain in the tissue while wild type cells are removed.

1.3. APC binding partners and structural features

Unsurprisingly, given the many processes APC impacts, many binding partners have been identified. The APC protein in humans contains 2843 amino acids (some alternatively spliced forms have been reported but here we will focus on the canonical isoform). Many APC interactions and their regulation have been described individually (11; 28) (Fig. 1 A) and binding sites for many of its binding partners have been mapped [29]. Those relating to Wnt signalling and some microtubule interactions are summarised in Fig. 1 A. The complexity of APC interactions is consistent with the idea that individual interactions affect each other [30]. In other words, there are likely many combinations of interactions that can occur simultaneously and some that are mutually exclusive. These will depend on the subcellular and molecular conditions and context. Relatively little is known about the combinations that do and do not occur. Two thirds of the APC protein is easily degraded and was predicted to be disordered by early biochemical experiments and resulting models [31]. Much has been learned about disordered protein domains in recent years and the tools to analyse them have become increasingly sophisticated. Intrinsically disordered domains, including some in APC, have been proposed to be highly dynamic and to provide interchangeable interaction sites for different partners to create a “cloud of bound conformations” [32]. Intrinsically disordered regions have also been suggested to act as localised sensors to different intracellular conditions [14,33]. One response can be for such domains to form liquid phase separated structures, as has been shown for APC [14,
This adds an additional layer of regulation of the functions APC can perform.

Currently, how APC’s diverse interactions are coordinated and regulated remains largely unknown. Similarly, how disease-associated mutations in APC affect the balance of all the responses it can integrate remains under investigated. The prevalence of mutations in the APC gene places this question at the heart of understanding this important player in cell behaviour. Answering this question will help to develop full and accurate predictions of the effectiveness of potential therapeutic or prognostic tools that aim to target such mutations. Using the molecular information available to identify common and different themes in the diverse organism where APC has been identified and studied is an important first step. To that end we interrogated and compared sequences of currently known APC proteins.

1.4. APC sequence analysis

1.4.1. Data resources and tools

The emergence of new tools to interrogate proteins computationally and our increased ability to predict structure - function relationships provide an opportunity to revisit the APC protein and its properties, which have remained somewhat enigmatic. To that end, we collated information currently available about the molecular details of APC across taxonomies, identified robust structural features, and compared and contrasted APC and APC2 to increase our understanding of its properties.

OrthoDB (https://www.ezlab.org/orthodb.html; [34]) provides curated sets of predicted orthologs from a wide variety of organisms, determined through a best-reciprocal blast and clustering approach. InterPro (www.ebi.ac.uk/interpro; [35]) produces protein signatures...
using a range of databases of protein domains and sites, enabling families of proteins to be identified based upon the presence of a number of representatives of a signature. Mappings of InterPro signatures to members of the UniProt database [36] provides an alternative approach to identifying related proteins. A single InterPro family (IPR026818) represents all APC sequences spanning a broad range of taxonomy from primitive phyla such as Cnidaria and Porifera throughout the Chordata, including Aves (birds), Mammalia and Actinopteri (ray-finned fish). The family incorporates two subfamilies containing representatives of APC (IPR026836), and APC2 (IPR026837), but also includes sequences that do not fall into either of these families (subsequently referred to as APC-like), including APC-related protein 1 (APR1) expressed in C. elegans and the two APC proteins expressed in Drosophila (Fig. 2).

1.4.2. Sequence analysis and comparison

To gain an indication of the conservation of APC, a set of metazoan APC orthologs were downloaded from OrthoDB 10.1 [34]. These were refined according to their annotations to remove APC2 sequences, which are included in the same ortholog set as APC, those annotated as low quality, and those which were not the canonical isoforms. The sequences were subsequently filtered to remove sequences shorter than 2750 AA, to remove partial length sequences, and to remove duplicates from the same organism. Multiple sequence alignment was carried out using ClustalW 2.1 [37] with default parameters. The alignment was then edited using Jalview 2.1.1.2.4 [38] and outlying sequences, based upon a principal component analysis, removed. Shenkin diversity scores were determined using AAccess 1.1 [39] and disorder predictions carried out using JRonn [40,41] within Jalview. Features annotated in the Uniprot record for the Human APC sequence (P25054) and InterPro domain mappings were projected onto the alignment positions along with the Shenkin diversity and JRonn disorder predictions (Fig. 1 C and 1D).

All sequences belonging to the InterPro Adenomatous polyposis coli (APC) family (IPR026818), which includes APC, APC2 and APC-like, were downloaded from InterPro release 89.0 [35]. Those sequences belonging to the APC family (IPR026836) were filtered to remove partial length sequences shorter than 2,000 AA, while sequences in the APC2 family (IPR026837) were similarly filtered with a 1,000 AA cut-off. Redundancy was removed from the group by removing identical sequences from the same organisms. The sequence set includes multiple isoforms, and sometimes duplicated sequences from different assemblies or sequencing projects, so the sequences were refined using a custom Python script. The UniProt record for each sequence was parsed to identify records in DNA sequence databases from which the protein sequence was derived, and both the genomic locus and assembly version identified. In cases where multiple assemblies were present, sequences from a single assembly were retained, with assemblies present in the Ensembl [42] database preferred, as was the longest protein sequence associated with each genomic locus. In addition to removing redundancy from the dataset, this process also allowed the identification of instances where genes were duplicated within an assembly, since the remaining protein sequences may be associated with multiple genomic loci within the same assembly. This resulted in a set of 1253 sequences covering 910 species.

A multiple sequence alignment was carried out as previously described, then phylogenetically informative regions were selected from the alignment using BMGE [43] with a Blosum62 similarity matrix, a phylogenetic tree was then constructed using RaxML-NJ 1.1 [44] using a JTT + G model with 100 bootstrap iterations, an MRE-based bootstrap convergence metric and Felsenstein branch support metric. The resulting tree was visualised and annotated using the Interactive Tree of Life [45].

The three-dimensional structures of the AlphaFill [46] models for Human APC and APC2 were visualised using Jalview and ChimeraX [47], with the intrinsically disordered regions downstream of the Armadillo (ARM) repeats (residues beyond 8000 AA) hidden for clarity.

2. Findings

- Phylogenetic analysis shows a clear separation between the members of the InterPro families APC, APC2 and APC-like (Fig. 2; Interactive versions of trees available online at https://itol.embl.de/shared/jca). APC and APC2 are found exclusively in the Chordata, while APC-like sequences are found across the remaining taxa, including Insecta, Amphibia, Hydracaro and Trematoda.

- Multiple sequence alignment of 158 Chordata APC sequences shows an extremely high degree of conservation. Shenkin divergence scores across the alignment demonstrate extensive conservation across the length of the protein, including disordered regions (Mean Shenkin Divergence score: 8.66, SD 4.24) (For reference: complete conservation would be indicated by a score of 4) (Figs. 1 and 2). Such extreme conservation across the entire sequence of APC suggests that all residues and their relative position to each other is crucial for the full complement of its functions and their coordination. Binding domains coincide with more conserved regions (Fig. 1 B/C), and also tend to occur in regions with lower predicted disorder, even in the highly disordered region associated with β-catenin binding (Fig. 1 D). However, in general, there appears an association between known binding domains and increased sequence conservation, and decreased disorder.

- The APC2 family is also highly conserved, but to a lesser degree than APC (Mean Shenkin Divergence score: 11.25, SD 6.65) (Fig. 2).

- The proteins in the APC-like family are much more divergent than those in the APC or APC2 families (Fig. 2).

- Most Chordata species have a single copy of the APC and APC2 genes (Supplemental Fig. 1). A small number of species carry two copies of the APC and/or APC2 genes, primarily the teleost fish i.e., rainbow trout, Atlantic salmon (Supplementary Table 1, and Supplementary Figure 3), and also Aves and Amphibia. The duplications within the teleost fish are likely a consequence of a genome duplication, which occurred ~90 million years ago [48]. The duplicated copies of APC within these species form separate sub-groups within the phylogeny (Supplementary Figure 1), which have independently diverged, while still retaining a high degree of conservation. Duplicated APC2 sequences, show a similar pattern. Most species with duplications have both APC and APC2 duplicated, although some have only one duplication (Supplementary Figure 2).

- Chordata have either APC, APC2 or both, but never an APC-Like gene. 62.1% of chordate organisms carry both APC and APC2, while 25% have only APC, and 12.9% have only APC2 (Supplementary Figure 1).

- AlphaFill models are available for both Human APC (first 1,400AA) and APC2 (full length). These are high-confidence models for N-terminal coils and the armadillo repeats. While the length of the disordered regions is included in the model for APC2, these are only partially included in the model for APC. These disordered regions, unsurprisingly, are regions with low confidence in the predicted models and lack any evident predicted structure. Disordered regions have been proposed to be highly dynamic and to provide interchangeable interaction sites for different partners by creating a “cloud of bound conformations” [32] (Fig. 3). One response of intrinsically disordered regions (IDR) can be formation of liquid phase separated structures [14,53]. Indeed, APC undergoes liquid phase separation, and this involves a number of different domains including the first half of the second ARM repeat and also consensus motifs in the 20R3 and 20R5 repeats [13,32,49].

- Armadillo repeats are inconsistently annotated. UniProt records for the Human APC (P25054) and APC2 (O95996) sequences are annotated with 7 and 6 armadillo (ARM) repeats respectively, while InterPro reports 7 repeats in each protein, including an 'APC Repeat' (IPR041257) corresponding to the second armadillo repeat. Comparison of the AlphaFill structures indicates that this region is
extremely similar between APC and APC2 (Fig. 3; Supplementary files 1 and 2), with no obvious difference in repeat number. The discrepancy appears to be a result of differing interpretations of what forms an armadillo repeat, with the main differences in the terminal repeats. It is notable that the annotated repeat units differ between APC and APC2, with each APC2 repeat unit starting about halfway through the corresponding APC repeat. Structural comparison suggests the presence of 8 armadillo repeats in both APC and APC2 (Table 1 and Fig. 3). The second of these (ARM2) relates to the APC Self-Association Domain (ASAD), which has been demonstrated to mediate self-association and is important in the assembly of the β-catenin destruction complex [50]

Fig. 3. AlphaFill structural model of the N-terminal region (residues 1–800) of APC (A) and APC2 (B). Helices annotated within the Uniprot records are highlighted in red. The labelled residues are experimentally identified phosphorylated serine residues in APC, and the conserved S744 residue in APC2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Annotated positions of armadillo repeats in UniProt and InterPro and proposed ‘consensus’ positions.

<table>
<thead>
<tr>
<th>ARM Repeat</th>
<th>APC (P25054)</th>
<th>APC2 (O95996)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UniProt</td>
<td>InterPro</td>
</tr>
<tr>
<td>ARM2 (ASAD)</td>
<td>453–495</td>
<td>393–466</td>
</tr>
<tr>
<td>ARM6</td>
<td>639–683</td>
<td>600–644</td>
</tr>
</tbody>
</table>
• N-terminal coils and armadillo repeats are the most prominent structural features (Fig. 3). Comparing the relative position of these structural features in APC and APC2, suggests that the N-terminal coils may be oriented differently relative to armadillo repeats in APC and APC2. The predicted aligned error in AlphaFold implies that, while the N-terminal coils and armadillo repeats are likely correctly predicted, their relative positioning is uncertain. Given the established importance of the armadillo repeats for many APC interactions, it is possible that these interactions vary between and are differently regulated in APC2 and APC, and result in different access to the armadillo repeats where many interacting partners bind (Figs. 1 and 3).

• AlphaFill models are representative of solved structures. Superposition of experimentally determined structures of portions of the APC protein with the predicted models using ChimeraX’s MatchMaker tool suggests these structural regions have been predicted extremely well by AlphaFold. PDB:SIZA represents most of the armadillo repeat region (P25054:407–751), which when superimposed has an RMSD of 0.566Å. Similarly, superposition of three alpha-helices from the annotated N-terminal coils (PDB:1MSI; P25054:126–250) results in an RMSD of 0.896Å.

• The ability of the disordered regions to create flexible clouds around the armadillo repeats could be affected by binding of proteins to the disordered regions. This may explain why truncated forms of APC, lacking much of the disordered regions, are more active. For instance, truncated APC appears to be more active in stimulating ASEF [51] suggesting that a regulatory feature provided by the disordered region is lacking in truncated forms of APC.

• Phosphorylated serine residues may play a structural role. A loop between ARM 6 and 7 in Human APC has two phosphosites (SER744 and SER748) that have been experimentally confirmed [52] (Fig. 3). These may be important for positioning the disordered regions beyond the armadillo repeats. One of these sites is conserved in APC2 (SER710) where it is similarly located in a loop within the final armadillo repeat (Fig. 3).

3. Conclusion

The extreme degree of conservation even across the extensive disordered regions confirms the importance of APC existing as a single protein. Dividing it into smaller proteins that work independently to perform the same functions individually would remove the ability to coordinate them. This is consistent with the idea that the ability to integrate and coordinate many different pathways and processes is a key feature supporting the multifunctionality of APC and can also explain why mutations leading to its truncation have such deleterious consequence in the highly dynamic epithelium of the intestinal tract.

This high conservation across the entire APC sequence underscores the importance of investigating the complement of APC interactions and potential functions, even when targeting only a single interaction (for instance by deletion mutations). The ability or many of the annotated domains of APC to interact with several different binding partners further illustrates the complex interplay of APC interactions and their regulation. The high content of intrinsically disordered domains provides an extremely dynamic means to create a plethora of conformations to support many different combinations of interactions for APC. Intrinsically disordered domains can respond to different subcellular local conditions to create different conformations, which in turn creates possibilities to regulate individual or combinations of interactions spatially.

The presence of three distinct APC protein families suggests that different organisms have evolved these proteins to be optimised for their needs. Of note is that the APC proteins in Drosophila and C. elegans for instance are members of the APC-like protein family, which may contribute to differences in how they interact with specific binding partners, how they are regulated and the specific details of how they contribute to distinct functions. This also may need to be considered when relating either of the two APC-like proteins in Drosophila to APC and APC2 directly.

Possible differences between the conservation and structure of APC and APC2 suggest there might be differences in their interactions and regulation, particularly in the N-terminal region where the position of coils relative to the armadillo repeats is predicted to vary (Fig. 3). Again, this may be one reason why mutations in APC not APC2 are common in cancer.

One surprising finding was that the APC-like protein in Hydra vulgaris (UniProt:T2MG20) does not have annotated armadillo repeats; however, the InterPro classifications (https://www.ebi.ac.uk/interpro/protein/UniProt:T2MG20/) identify six armadillo repeats at the C-terminus of the protein. This is contrary to other APC proteins, where these are typically found in close proximity to the N-terminus. Only a single armadillo ‘repeat’ is annotated in the C.elegans APR-1 sequence, between residues 314–358 in the 1,188AA protein. In neither case are the 15AA/20AA/SAMP domains, which are implicated in binding β-catenin and Axin, annotated in the database records. These findings may be a shortcoming of annotation and warrant further investigation.

3.1. The future

We anticipate (and hope) that our analysis and results will help to guide future work to understand and elucidate mechanisms for APC functions, their role in normal tissue function, and the consequence of mutations in disease Comparing the conservation of APC binding partners, particularly their APC-binding domains, may also shed light on contributions of APC to different signalling pathways across organisms.

Declaration of Competing Interest

None.

Acknowledgements

We are immensely grateful to S. MacGowan, M. Tsenkov, and G. Barton (Computational Biology, University of Dundee) for immensely helpful discussions.

Code availability

The code used in the generation of results and figures is available from https://github.com/bartongroup/30_years_of_APC.

Supplemental material

Supplemental Fig. 1: UpSet Plot of distribution of species with APC, APC2, APC-like or combinations of these.

Supplemental Fig. 2: Version of Fig. 2 that plots branch lengths to indicate degree of conservation evident across the range of taxonomy. Taxonomy of APC protein family members.

Supplemental Fig. 3: Phylogenetic tree showing APC and APC2 species within the Actinopeti, which include a number of species with duplications of APC and/or APC2. Duplicated proteins are linked with arrows that are colour coded according to the identity of the duplicated sequences.

Supplemental Table 1: Taxonomic details of species included in phylogenetic analysis including copy number of each class of APC, and the Uniprot accessions of the sequences.

Supplemental File 1: ChimeraX file of Human APC structure shown in Fig. 3 A.

Supplemental File 2: ChimeraX file of Human APC2 structure shown in Fig. 3 B.

Supplemental files 1 and 2 can be opened using the UCSF ChimeraX molecular visualisation software available from https://www.cgl.ucsf.edu/chimerax/.
Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.semcdb.2023.04.004.

References