

## Domain organization of activation-induced cytidine deaminase

### To the editor:

A recent article by Ta *et al.*<sup>1</sup> describes an experimental analysis of activation-induced cytidine deaminase (AID) mutants presented in the context of a postulated domain structure. The domain organization of AID is extrapolated from the homology model of the mRNA-editing enzyme APOBEC-1 (ref. 2), which shares substantial sequence homology with AID. The APOBEC-1 structure was in turn modeled using *Escherichia coli* cytidine deaminase (ECCD) as a template. Based on these studies, Ta *et al.* proposed four protein domains in AID that are equivalent to the ECCD domains: an N-terminal helical domain (residues 1–20), a cytidine deaminase active site domain (residues 20–110), a linker (residues 110–148) and a cytidine deaminase pseudoactive site domain (residues 148–198).

Since the homology modeling of APOBEC-1 in 1998, there has been considerable progress in both protein structure prediction methods and the number of experimental protein structures that can be used as a template in comparative modeling. We have analyzed the human AID sequence with the 3DJury method<sup>3</sup>, which predicts protein structures based on the consensus between various independent online structure prediction servers. Structure prediction approaches used by the servers include both threading methods (in which the query sequence is matched with the statistical models representing three-dimensional structures of the given structural family) and methods in which remote sequence similarity is detected by comparison of the query sequence with the database of sequence profiles. According to the 3DJury results, the structure of human AID can be modeled on the recently published

structure of yeast cytosine deaminase<sup>4</sup> (yeast CD; Protein Data Base accession number, 1UAQ). This structure was found to be the best template by five independent threading methods, including the frequently used 3DPSSM server<sup>5</sup>, which assigned a 95% significance level. Methods based on sequence similarity (BLAST, PSI-BLAST and HMM searches) did not indicate any protein with known structure sharing substantial sequence similarity with human AID. According to the consensus of the threading methods, computed by 3DJury, residues 10–151 of yeast CD are aligned with the stretch of the AID sequence between residues 1 and 167. Amino acids belonging to the active center of yeast CD are conserved in the alignment.

Yeast CD and ECCD belong to the same structural family and share 22% sequence identity. The main structural difference between these two proteins is the presence of a pseudoactive site domain in ECCD, which is absent from yeast CD. This domain of ECCD is structurally similar to the active site domain, which indicates a gene duplication event during evolution. Another difference between yeast CD and ECCD is the lack of a short, N-terminal helical domain in yeast CD. The geometry of conserved residues within the active sites of both enzymes is very similar. However, in the ECCD dimer, the contact between the active site and residues belonging to the pseudoactive site domain of the second molecule is necessary. The yeast CD dimer contains two independent active sites. We postulate therefore that residues 1–167 of AID form the globular domain with a structure similar to that of yeast CD. This domain contains the cytidine deaminase active site. In the AID dimers or multimers studied by Ta *et al.*, the active sites are most probably independent, similar to those of yeast CD.

In addition to the evidence provided by applying the threading methods, the AID sequence is too short to accommodate all three domains of ECCD. The length of the ECCD pseudoactive site domain, however, is 104 amino acids. According to Ta *et al.*, the pseudoactive site domain in AID is 50 amino acids long, too short to form the globular structure resembling that of the ECCD pseudoactive site domain. Also, the N-terminal domain postulated by Ta *et al.* is half the size of the N-terminal helical domain of ECCD. Our analysis indicates that the AID monomer contains a single globular domain with an independent cytidine deaminase active site. We were not able to predict the structure of the C-terminal 31 amino acids.

Because of the essential functions of AID in class-switch recombination and somatic hypermutation, the pioneering efforts of Ta *et al.* will certainly be followed by other work that attempts to functionally dissect the AID protein by mutational studies. These studies could be aided considerably by the comparative modeling of protein structures. We believe that the results presented above will help to select a proper template structure for the homology modeling of the AID protein.

Jolanta Zaim & Andrzej M Kierzek\*

*Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawlinskiego 5a, 02-106 Warsaw, Poland.*  
e-mail: andrzejk@ibb.waw.pl

1. Ta, V.-T. *et al.* *Nat. Immunol.* **4**, 843–848 (2003).
2. Navaratnam, N. *et al.* *J. Mol. Biol.* **275**, 695–714 (1998).
3. Ginalski, K., Elofsson, A., Fischer, D. & Rychlewski, L. *Bioinformatics* **19**, 1015–1018 (2003).
4. Ko, T.P. *et al.* *J. Biol. Chem.* **278**, 19111–19117 (2003).
5. Kelley, L.A., MacCallum, R.M. & Sternberg M.J.E. *J. Mol. Biol.* **299**, 499–520 (2000).