



## Delta-Interacting Protein A and the Origin of Hepatitis Delta Antigen

Manyuan Long *et al.*

*Science* **276**, 824 (1997);

DOI: 10.1126/science.276.5313.824

---

*This copy is for your personal, non-commercial use only.*

---

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of July 24, 2012 ):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/276/5313/824.full.html>

This article **cites 4 articles**, 4 of which can be accessed free:

<http://www.sciencemag.org/content/276/5313/824.full.html#ref-list-1>

This article has been **cited by 1** articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/276/5313/824.full.html#related-urls>

This article appears in the following **subject collections**:

Microbiology

<http://www.sciencemag.org/cgi/collection/microbio>

highly selective neurons; across the population, selectivity for what and for where varied along a continuum from nonselective neurons to highly selective neurons.

13. To eliminate any optimistic bias in the classification, we performed the discriminant analysis with cross-validation; that is, the distribution of means for each class was computed on half the data, chosen randomly, and these means were used to classify the objects or locations in the other half of the data.
14. Because the activity of individual neurons is "noisy," they rarely perform as well as the animal as a whole. By pooling the activity of multiple neurons, it is possible to reduce noise, and a neural classification rate equal to behavioral performance can thus be achieved [for details, see E. K. Miller, L. Li, R. Desimone, *J. Neurosci.* **13**, 1460 (1993)].
15. Their relatively low classification rate for objects on the basis of where delay activity (28.3%) was not due to the generally poor ability of these neurons at

classifying objects; their mean classification rate for objects on the basis of what delay activity was significantly greater (32.3%,  $P < 0.001$ ). The lower rate could have been due, in principle, to an effect of the nonmatching object, which was not included in the classification and thus could add "noise." However, the mean classification rate of a discriminant analysis of the where delay activity that attempted to classify the nonmatching stimulus (24.8%) was not different from chance ( $P = 0.680$ ), indicating no effect of the nonmatch on the where delay activity. Thus, the effect of object on the where delay activity appears to be due to the match object alone.

16. See J. M. Fuster, *Memory in the Cerebral Cortex* (MIT Press, Cambridge, MA, 1995).
17. J. Bullier, J. D. Schall, A. Morel, *Behav. Brain Res.* **76**, 89 (1996).
18. V. P. Ferrera, K. K. Rudolph, J. H. R. Maunsell, *J. Neurosci.* **14**, 6171 (1994); A. B. Sereno and J. H. R. Maun-

sell, *Invest. Ophthalmol. Visual Sci.* **36**, 692 (1995).

19. A neural network model also suggests that the extent to which what and where are segregated or integrated in the prefrontal cortex depends on task demands [T. S. Braver and J. D. Cohen, *Proc. Cognit. Neurosci. Soc.* **2**, 95 (1995)].
20. G. H. Recanzone, M. M. Merzenich, W. M. Jenkins, A. G. Kamil, H. R. Dinse, *J. Neurophysiol.* **67**, 1057 (1992).
21. Animal experiments were conducted in accordance with the MIT Committee on Animal Care. This work was supported by the Pew Charitable Trusts and the McKnight Foundation. We thank M. Histed for help in animal training and W. Asaad, R. Desimone, Y. Munakata, J. Mazer, R. O'Reilly, and M. Wickerski for valuable comments. S.C.R. was supported by a fellowship from the McDonnell-Pew Foundation and the Markey Foundation.

9 December 1996; accepted 5 March 1997

## TECHNICAL COMMENTS

### Delta-Interacting Protein A and the Origin of Hepatitis Delta Antigen

Robert Brazas and Don Ganem (1) propose that the cellular protein, delta-interacting protein A (DIPA), interacts with hepatitis delta antigen (HDAg), affecting hepatitis delta virus (HDV) replication. Although their work provides useful information about the biology of HDV, the main conclusion, that DIPA is the cellular homolog of HDAg, is not supported by their data.

We have examined the statistical significance of the match between HDAg and DIPA protein sequences by Monte Carlo simulation. In their comparison between HDAg and DIPA protein sequences, Brazas and Ganem reported an identity of 24% and a similarity of 56%, using the GES scale, which considers hydrophobicity when determining the distance matrix for substitutions (2). We compared HDAg with 10,000 randomized DIPA sequences, using the GAP program with the same parameters as Brazas and Ganem (1) (a gap weight of 3.0 and gap length weight of 0.1). The probability distributions for identity match and for similarity values that are determined using the GES scale (1) show that the match between HDAg and DIPA is not significant (Fig. 1A): The probability for an identity match greater than or equal to 24% is 13.2% and the probability for a similarity match greater than or equal to 56% is 14.1%. This does not support the proposed common ancestral relationship between HDAg and DIPA.

Furthermore, it is inappropriate to use the GES scale to determine homologous relationships between protein sequences, because convergent evolution could easily affect the hydrophobicity of a protein se-

quence, a relatively simple chemical property. However, the match between HDAg and DIPA is also not significant, with the use of the PAM-250 matrix (data not shown).

While various matrices may give different similarity measurements, the identity remains the same given a particular alignment. However, the identity match is a result of a biased amino acid composition. A Monte Carlo simulation comparing HDAg to 10,000 random sequences that

have the average amino acid compositions of an overall protein with the same length of the DIPA protein sequence shows that the observed similarity is again not significant ( $P = 18.2\%$ ), but the identity match would have been significant (Fig. 1B). Thus the reported "match" is biased by the amino acid compositions of HDAg and DIPA. [We used the amino acid composition derived from the exon database developed from GenBank release 90, where redundant sequences are deleted by a similarity criterion of 20%. For detailed procedures, see (3)].

The three amino acid compositions are listed (Table 1). Both HDAg and DIPA have similarly biased amino acid compositions with overrepresented residues like Glu, Gly, and Arg and underrepresented His, Thr, and Tyr. This will lead to elevated identity matching between the simulated random sequences and HDAg.

We conducted a test of the effect of amino acid composition on the identical

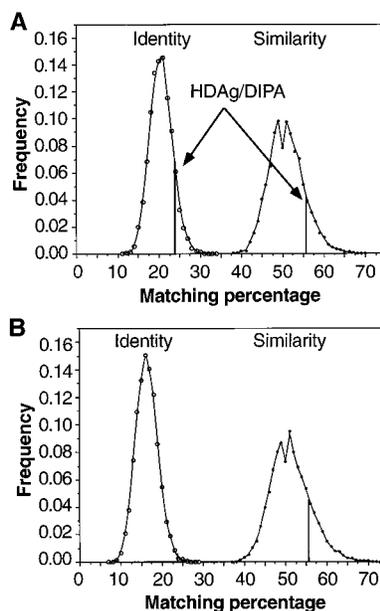


Fig. 1. Probability distributions for (A) DIPA sequence randomized and (B) randomization based on average amino acid composition.

Table 1. Amino acid compositions of HDAg, DIPA, and the exon database.

Residues	HDAg	DIPA	Database
Ala	0.0421	0.0990	0.0683
Cys	0.0047	0.0297	0.0202
Asp	0.0561	0.0446	0.0538
Glu	0.1168	0.1287	0.0648
Phe	0.0280	0.0099	0.0427
Gly	0.1262	0.1238	0.0612
His	0.0047	0.0050	0.0235
Ile	0.0327	0.0050	0.0548
Lys	0.1028	0.0198	0.0607
Leu	0.0794	0.1584	0.0886
Met	0.0093	0.0149	0.0245
Asn	0.0280	0.0198	0.0463
Pro	0.0841	0.0495	0.0529
Gln	0.0327	0.0644	0.0418
Arg	0.1262	0.1188	0.0532
Ser	0.0561	0.0495	0.0796
Thr	0.0234	0.0198	0.0574
Val	0.0327	0.0248	0.0633
Trp	0.0140	0.0149	0.0118
Tyr	0.0000	0.0000	0.0306

**Table 2.** Distribution of identical matches of amino acid compositions between HDAG and DIPA.  $\chi^2 = 6.6$  with 10 degrees of freedom;  $P = 0.76$ . The expected numbers were calculated as products of expected frequency and total identical matches (47). The expected frequency,  $f_{aa} = [N_{aa}(\text{HDAG}) + N_{aa}(\text{DIPA})] / \sum_{j=1}^{11} [N_{aa}(\text{HDAG}) + N_{aa}(\text{DIPA})]$ , where  $N_{aa}$  is the number of residues in HDAG and DIPA.

	Gly	Glu	Ala	Leu	Arg	Gln	Phe	Asp	Pro	Cys	Lys
Observed	7	10	3	6	10	2	1	1	5	1	1
Expected	7.1	7.0	4.0	6.7	7.0	2.7	1.1	2.9	3.8	1.0	3.6

matching of the residues given by Brazas and Ganem. With the use of the amino acid compositions of HDAG and DIPA, we calculated the expected number of identical matches and then compared that expectation with the observed matches. The 47 identical matches reported (1) can be explained as chance matches with a probability of 76% that occur as a consequence of the biased amino acid composition of HDAG and DIPA (Table 2).

In an accompanying Perspective, Hugh D. Robertson proposes that spliceosomal introns could have arisen from the self-replicative viroid (4). This proposal provides a specific candidate (viroid) as an ancestor of modern introns, in line with the exon theory of genes (5) that spliceosomal introns are descendants of self-splicing introns that existed in an RNA world. How-

ever, the relationship between HDAG and DIPA cannot yet be taken as evidence for this particular scenario.

**Manyuan Long  
Sandro J de Souza  
Walter Gilbert**

*The Biological Laboratories,  
Harvard University,  
16 Divinity Avenue,  
Cambridge, MA 02138, USA*

REFERENCES

1. R. Brazas and D. Ganem, *Science* **274**, 90 (1996).
2. D. M. Engelman, T. A. Steitz, A. Goldman, *Annu. Rev. Biophys. Chem.* **15**, 321 (1986).
3. Long *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 12495 (1995).
4. H. D. Robertson, *Science* **274**, 66 (1996).
5. W. Gilbert, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 901 (1987).

12 November 1996; accepted 7 February 1997

*Response:* Long *et al.* raise an important issue related to the determination of an evolutionary relationship between two proteins that share only limited amino acid identity. While we agree that the biased amino acid composition of L-HDAG and DIPA complicates the interpretation of their 24% identity (1), this compositional bias does not invalidate our inference that these two proteins are related.

The GAP program used by Long *et al.* is a global alignment algorithm that aligns two proteins; it uses a substitution matrix to provide a value for each aligned pair of amino acids, and in addition assesses a penalty for the introduction of a gap into the alignment [gap penalty score of  $-(3 + 0.1k)$  for a gap of length  $k$ ]. The program then determines the alignment that produces the maximal score when all the alignment values and gap penalties are added together. By using only percent identity and percent similarity individually, however, Long *et al.* do not include other important alignment characteristics (gap number, gap length, and the combined identity and similarity) that are inherent in each alignment score. This and other factors lead to a significant underestimate of the relatedness of the two proteins.

Like Long *et al.*, we have compared L-

HDAG to randomized (shuffled) DIPA sequences; such a pairwise comparison is justifiable because independent biological experiments pointed to a potential relationship (see below). However, we used the alignment scores rather than percent identity to determine the significance of the L-HDAG–DIPA alignment. We also used a more recent substitution matrix, the BLOSUM-62 matrix (2), which has been shown to outperform most other matrices (including PAM-250) in detecting significant protein similarities (3), together with the GAP alignment algorithm [gap penalty

score of  $-(9 + k)$  for a gap of length  $k$ ] (4) to align L-HDAG to 10,000 randomized DIPA sequences. Only 0.2% of the randomized DIPA–L-HDAG alignments had a score equal to or greater than the optimal L-HDAG–DIPA alignment (score = 64) (Fig. 1). (Although this statistical result would not be noteworthy in a database search, it is significant in light of the biological data favoring homology; see below.) This alignment contains 55 identical residue pairs (29.6% identity) compared to 47 for the original GAP alignment (5). Our analysis explicitly takes into account the issue of the contribution of the amino acid composition bias presented by Long *et al.*, yet points much more strongly than theirs to a relationship between L-HDAG and DIPA.

In addition to their shared compositional bias, which would be expected if indeed L-HDAG and DIPA are related, both proteins (i) are nearly identical in size, (ii) contain putative coiled coil domains, and most importantly, (iii) specifically oligomerize with each other in a manner that is dependent on the integrity of their coiled coil domains. [As we have shown (1), this interaction can have important functional consequences *in vivo*.] The computer-based alignment analyses do not take these facts into consideration, but biologists must do so in coming to a judgment about relatedness.

The assignment of an evolutionary relationship between two proteins of limited similarity is often difficult, and the compositional bias identified by Long *et al.* further complicates this risk for DIPA and HDAG. However, our analysis shows that when the totality of the evidence linking the proteins is considered, the confounding effect of compositional bias is not sufficient to negate the relationship between them that we proposed (1).

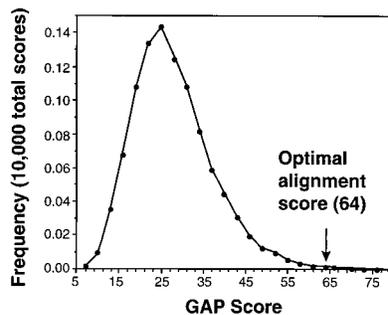
**Robert Brazas  
Don Ganem**

*Howard Hughes Medical Institute,  
Department of Microbiology,  
University of California,  
San Francisco, CA 94143–0414, USA*

REFERENCES AND NOTES

1. R. Brazas and D. Ganem, *Science* **274**, 90 (1996).
2. S. Henikoff and J. Henikoff, *Proc. Nat. Acad. Sci.* **89**, 10915 (1992).
3. S. Henikoff and J. Henikoff, *PROT.: Struct. Func. Gen.* **17**, 49 (1993).
4. W. Pearson, *Prot. Sci.* **4**, 1145 (1995).
5. R. Brazas and D. Ganem, data not shown.
6. We thank S. Altschul (National Center for Biotechnology Information) for helpful discussions and for assistance with the alignment analyses in Fig. 1.

12 November 1996; accepted 7 February 1997



**Fig. 1.** GAP alignment scores using BLOSUM-62 substitution matrix with DIPA sequence randomized.

Downloaded from www.sciencemag.org on July 24, 2012