Vol. 160, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 14, 1989 Pages 333-339

DOMAIN-SPECIFIC BIAS IN ARGININE/LYSINE USAGE BY PROTEIN TOXINS

Erwin London and Cindy L. Luongo

Department of Biochemistry, State University of New York, Stony Brook, New York 11794-5215

Received February 1, 1989

The content of lysine and arginine residues in a number of A-B type protein toxins has been examined. It is found that the A subunit, or its equivalent, often shows a strong bias in the type of basic amino acid residue used tending towards nearly exclusive use of either arginine or lysine rather than use of both, whereas the B subunit or its equivalent shows no such bias. Although arginine codons are GC-rich and lysine codons are AT-rich, the content of GC and AT in the genes coding for the toxins does not adequately explain this bias. Other explanations are discussed, including the possibility that the bias is linked to catalytic function or membrane interaction. Understanding this bias may yield valuable insights into toxin structure and function. Furthermore, identification of bias in sequences may be a useful tool for identifying new toxins and their domains.

Protein toxins of various types have been identified. One such type of toxin is the A-B class of toxins of medium molecular weight. These toxins have become targets of intensive research for a variety of reasons including the insights they give into membrane signalling and translocation processes, the interest in development of immunoglobulin-toxin hybrids (immunotoxins) as therapeutic agents, and the similarities between toxin and viral entry into cells. In these proteins the toxic enzymatic activity is generally carried by the A subunit, while the B subunit both carries the cell receptor binding activity and aids the penetration of the A subunit into the cytoplasm. The A-B toxins have a variety of enzymatic activities including the following: ADP-ribosylation of the modified His residue (diphthamide) of elongation factor 2 (1) in the case of diphtheria toxin (DT) and Pseudomonas exotoxin A (PE); ADP-ribosylation of the adenylate cyclase regulating G-proteins in the case of cholera toxin, the closely related E. coli heat-labile toxin, and pertussis toxin (2,3), and removal of the adenine base at position 4324 in 28s rRNA in the case of ricin, Shiga toxin, and presumably Shiga-like toxin of E. coli (4,5). Although different classes of A-B toxins appear to share only very little amino acid homology (6,7) and apparently catalyze quite different processes, all the toxins listed above cleave ribose 1-carbon to nitrogen bonds, and therefore their activities are related. Recently, we

0006-291X/89 $1.50

noted homology between two toxins, diphtheria toxin and Pseudomonas exotoxin A, and also briefly noted some peculiarities in the distribution of Arg and Lys residues in these two toxins (8). In this report, we have examined the sequences in a series of A-B toxins, and find that a very strong bias tending towards usage of only Arg or only Lys as the strongly basic residue is a general feature of the catalytic subunit.

RESULTS AND DISCUSSION

Figure 1 summarizes the structure of the mature forms of several protein toxins for which sequence data is available. Their subunit or domain structure is also illustrated. The Arg and Lys content in the various domains of these toxins was calculated from their sequences in order to evaluate the distribution of these two residues (Table I). In addition to the number of Lys (n_K) and Arg (n_R) residues in each domain, we have computed a Arg/Lys bias ratio (b_{RK}) defined by:

$$b_{RK} = (n_R - n_K)/n_R \qquad \text{if } n_R \leq n_K$$
$$b_{RK} = (n_R - n_K)/n_K \qquad \text{if } n_R \geq n_K$$

A b_{RK} value of 5 or greater indicates a strong bias towards Arg usage relative to Lys (> 85% Arg), and a value of -5 or less indicates a strong

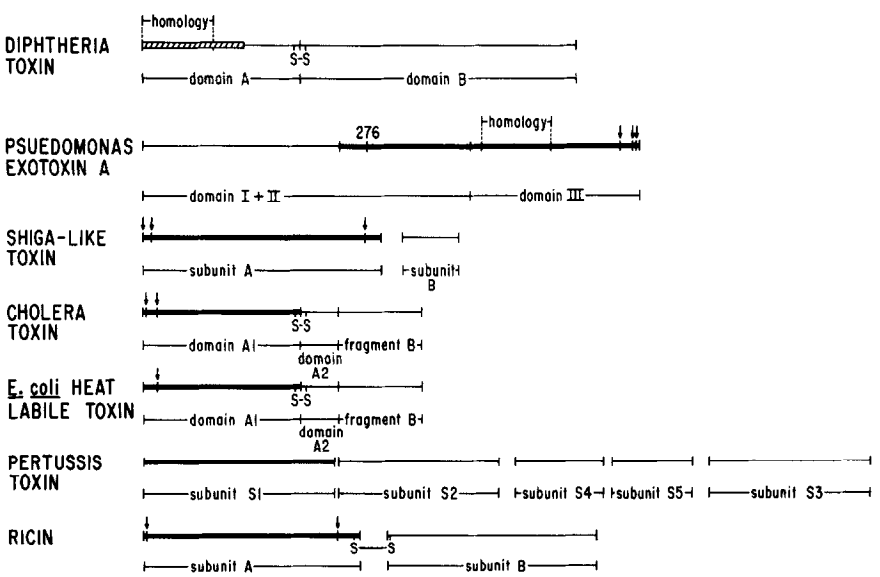

Fig. 1. Structure of Protein Toxins. Hatched line shows domain with bias towards Lys use. Filled heavy line shows domains with bias towards Arg use. Arrows show location of Lys residues in domains with Arg bias. Sequences are represented with N-termini at left. The lengths of the lines are proportional to sequence lengths. Domains and/or subunits are indicated below each sequence line. The domain closest to the left is the catalytic domain in each case, except for PE, in which domain III is the catalytic domain (40). Disulfide bonds linking domains are also shown. Sequences labeled "homology" refer to regions in DT and PE homologous to one another (see reference 8).

bias towards Lys usage. This parameter was chosen specifically because it identifies sequences in which there is extreme Arg or Lys bias clearly.

Table I shows that in cholera toxin, E. coli heat-labile toxin, Shiga-like toxin, and ricin the catalytic domains have a strong bias towards Arg relative to Lys ($b_{RK} > 6$). One can also conclude that Shiga toxin shows Arg bias in its A subunit because the amino acid composition of its subunits

Table I: Summary of Domain-Specific Bias in Arg/Lys Usage in Protein Toxins

Toxin[a]	Domain[b]	n_{Arg}	n_{Lys}	b_{RK}[c,d]	n_{Ile}	n_{Val}	b_{IV}[e]	%GC	%GC adjusted
Diph-	homology	0	9	$-\infty$	2	8	-3.0	40.7	43.8
theria	A	7	16	-1.3	6	16	-1.67	43.5	46.1
	B	9	23	-1.6	27	30	-0.11	41.1	43.1
	1-125	0	14	$-\infty$*	4	11	-1.75	41.1	44.4
	126-535	10	25	-0.56	29	35	-0.21	42.3	44.1
	total	16	39	-1.4	33	46	-0.39	42.0	44.1
Exotoxin	homology	10	0	∞*	4	7	-0.75	69.7	68.5
A	domain I+II	24	11	1.2	15	25	-0.67	68.4	67.8
	domain III	19	3	5.3*	11	12	-0.091	71.6	70.7
	1-240	13	11	0.18	12	16	-0.33	65.0	64.4
	241-613	30	3	9*	14	21	-0.50	72.4	71.6
	total	43	14	2.1	26	37	-0.42	69.5	68.8
Shiga-	A	26	3	7.67*	17	22	-0.29	42.9	41.1
Like	B	2	5	-1.5	3	6	-1.0	39.1	41.1
	total	28	8	2.5	20	28	-0.4	42.2	40.6
Cholera	A1	15	2	6.5*	10	9	0.11	42.4	40.9
	A2	2	5	-1.5	3	2	0.5	29.0	31.3
	B	3	9	-2.0	11	4	1.75	34.0	36.1
	total	20	16	0.25	24	15	0.6	38.1	36.9
E. coli	A1	19	1	18.0*	13	9	0.44	40.2	38.2
Heat-	A2	4	2	1.0	4	2	1.0	36.2	34.3
Labile	B	4	9	-1.25	13	4	2.25	38.2	40.5
	total	27	12	1.25	30	15	1.0	39.1	37.4
Pertussis	S1	22	0	∞*	9	17	-0.89	62.4	61.2
	S2	12	6	1.0	11	12	-0.091	59.0	58.1
	S3	13	5	1.6	13	10	0.3	61.5	60.6
	S4	6	8	-0.33	2	12	-5.0	61.8	65.0
	S5	3	5	-0.67	3	3	0.0	60.3	62.4
	total	56	24	1.33	38	54	-0.42	61.0	60.2
Ricin	A	21	2	9.5*	23	15	0.53	41.9	40.4
	B	13	7	0.86	18	17	0.059	42.7	41.8
	total	34	9	2.8	41	32	0.28	42.3	41.1
Tetanus	L chain	14	33	-1.4	not determined			29.6	31.1
	H chain	24	74	-2.1				26.9	28.6
	total	38	107	-1.8				27.9	29.5

[a]Sequences from (3,6,9,12,18,34-39). Only the mature protein sequence is shown, leader sequences being omitted. The value n is the number of residues of the given amino acid. [b]Numbers in domain column represent amino acid sequence numbers. [c]b_{RK} is defined as $(n_R-n_K)/n_R$ if $n_R \leq n_K$ or $(n_R-n_K)/n_K$ if $n_R \geq n_K$. [d]Asterick denotes regions of Arg/Lys bias as defined in the text. [e]b_{IV} is defined as $(n_I-n_V)/n_I$ if $n_I \leq n_V$ or $(n_I-n_V)/n_V$ if $n_I \geq n_V$. [f]Adjusted % GC is GC content after bases at the first two positions of all codons for Arg and Lys are discounted. This adjustment corrects for the inherent distortion of GC content by Arg and Lys due to the fact that Arg codons tend to be C and G in these positions and Lys codons are always A in these positions.

is almost identical to that of Shiga-like toxin (9). In addition, figure 1 shows that in each of these toxins the few Lys residues present in the catalytic subunits are not randomly distributed, but rather are confined to the extreme N- or C-terminal 10% of the sequence. In contrast, Arg bias does not occur in the other domains of these toxins and is not present in the overall amino acid composition of the complete toxins (Table I).

Bias is also present in both diphtheria toxin and Pseudomonas exotoxin A, toxins which have an identical catalytic function (10) and some sequence homology (8,11). In the catalytic fragment A of DT the bias is towards Lys usage rather than Arg usage. This bias is somewhat limited, in the sense that it is found only the N-terminal two-thirds of fragment A (residues 1-125). However, within that region there is exclusive use of Lys over Arg. In contrast, the catalytic domain III of PE shows a strong Arg bias similar to that observed with the other toxins described above. The Arg bias in PE continues beyond domain III on its N-terminal side to residue 241. The few Lys residues within domain III of PE are all found within the C-terminal 10% of its sequence, again paralleling the behavior described for the other toxins.

It should be noted that not all large protein toxins show an Arg or Lys bias. We find no evidence for subunit-specific bias in the sequence of tetanus toxin (12,13) or the related botulinum neurotoxins (12) (at least as far as can be determined from chemically deduced amino acid compositions in the latter case (14)). These larger toxins have sometimes been compared to A-B type toxins in their structure (15), but the exact degree of similarity in structure or function is not yet clear. Anthrax toxin and adenylate cyclase toxin also do not show bias (16,17, S. Leppla, personal communication). These results may not be surprising if bias is somehow specifically linked only to the behavior of ribose-nitrogen bond attacking A-B toxins (i.e. those that act by ADP-ribosylation or by RNA base cleavage). On the other hand, the antimicrobial protein colicin E1 shows a distinct C-terminal region rich in Lys residues (18). As the list of ADP-ribosylating Clostridium toxins and RNA base cleaving toxins is growing rapidly (19-24), we may soon have a clearer picture of which kinds of toxins show bias and of how general the phenomenon of bias is.

The most important questions concerning the Arg/Lys bias found in these toxins involve why it exists and what its significance is. One cause of bias could arise from the fact that Arg and Lys residues are similar, and their replacement by one another is considered a very conservative substitution (25). Selective pressure to maintain a particular %GC composition in a gene to be consistant with the overall %GC for the organism would tend to select either Arg or Lys residues because Arg codons are GC-rich and Lys codons are

AT-rich. This is a common explanation of bias in amino acid composition (18). Indeed, we previously noted an apparent correlation of this type may occur in the case of the homologous region of DT, which shows Lys bias and a low %GC, and that of PE, which shows Arg bias and high %GC (8).

To test whether this is a general relationship, Arg/Lys bias and %GC were compared for all the toxins. Table I shows the %GC for each domain of each toxin In terms of total protein sequence there is at most only a poor correlation between extremes of %GC and fractional Arg usage. The correlation remains poor when fractional Arg usage and %GC are compared to one another only in the domains showing bias. Furthermore, there is no marked correlation seen within each individual toxin when the %GC and Arg/Lys bias of individual subunits or domains are examined. In other words, within a given toxin the domains showing bias are not markedly more or less GC-rich than domains that do not show bias.

A relative lack of influence of %GC upon Arg/Lys bias is in agreement with results of measurements on Ile/Val bias. Exchange of Ile and Val for each other is another very conservative substitution (25), and Ile codons are AT-rich. Therefore, one might expect some bias towards Ile in those toxin domains that have low %GC if this were a source of bias. However, Table I shows there is no marked Ile or Val bias in any of the toxins or their individual domains. Overall, it seems %GC alone cannot explain Arg/Lys bias, although it is still possible %GC has some influence when combined with other factors in some cases.

A different explanation of bias is that it is somehow related to the enzymatic functions of the toxins. When the fact that the toxins showing bias all cleave ribose 1-carbon to nitrogen bonds is combined with the observation that bias always involves the catalytic domain of a toxin it seems plausible that there is a relationship of bias to function. However, a simple mechanism for a direct connection between bias and catalytic function is not apparent. Self modification could be a factor but most of the toxins do not attack Arg or Lys residues. Furthermore, although PE and DT catalyze the same reaction the former shows Arg bias and the latter Lys bias. It is hard to see how in this case both types of bias could be necessary for the same catalytic function at the same time. On the other hand, many of these toxins modify G proteins, and a clue to the functional significance of bias may come from the report that cationic regions on receptors and mastoparan peptides may play a central role in interaction with, and activation of, G proteins (26).

Another explanation of Arg/Lys bias is that it is somehow a result of the interaction of toxins with membranes. The A domains of toxin must generally penetrate membranes and emerge on the cytoplasmic side of the membrane, and

complex conformational changes are thought to be involved in this process (2, 27-29). Therefore, toxins may face special structural constraints encountered neither by ordinary soluble proteins nor by ordinary membrane proteins, and this may result in bias. One plausible possibility is the difference in Lys and Arg ionization constants and the recently noted strong bonding properties of Arg relative to Lys (30,31) could play a role in this regard. Another possibility is that selective use of Lys or Arg help toxins avoid inactivating proteolysis or enhance proteolytic activation. Proteolytic events are thought to be important for processing of several toxins and of viral proteins that are involved in membrane penetration (2,32,33). In any case, we have no definitive explanation(s) of bias as yet.

The idea that Arg/Lys bias will provide a valuble clue to the nature of toxin structure and function is an exciting possibility worthy of further exploration. Irrespective of the origin of bias, identification of bias in sequences of proteins of unknown function may be a powerful way of identifying toxins, their catalytic domains, and domain boundaries.

ACKNOWLEDGMENTS

The authors would like to thank Paul M. Bingham, Bernard S. Dudock, Rolf Sternglanz, and Jian-Min Zhao for helpful discussions. This work was supported by N.I.H. grant GM 31986 and the Center for Biotechnology of the State University of New York at Stony Brook.

REFERENCES

1. Van Ness, B.G., Howard, J.B., and Bodley, J.W. (1980) J. Biol. Chem. 255, 10710-10716.
2. Gill, D.M. (1977) Adv. Cycl. Nucl. Res. 8, 85-118.
3. Spicer, E.K., and Noble, J.A. (1982) J. Biol. Chem. 257, 5716-5721.
4. Endo, Y., and Tsurugi, K. (1987) J. Biol. Chem. 262, 8128-8130.
5. Endo, Y., Tsurugi, K., Yutsudo, T., Takeda, Y., Ogasawara, T., and Igarashi, K. (1988) Eur. J. Biochem. 171, 45-50.
6. Locht, C., and Keith, J.M. (1986) Science, 232, 1257-1264.
7. Wozniak, D. J., Hsu, L.-Y., and Galloway, D. R. (1988) Proc. Natl. Acad. Sci. USA 85, 8880-8884
8. Zhao, J.-M. and London, E. (1988) Biochemistry 27, 3398-3403.
9. Calderwood, S.B., Auclair, F., Donohue-Rolfe, A., Keusch, G.T., and Mekalanos, J.J. (1987) Proc. Natl. Acad. Sci. USA 84, 4364-4368.
10. Iglewski, B.H., Liu, P.V., and Kabat, D. (1977) Infect. Immun. 15, 138-144.
11. Carroll, S.F., and Collier, R.J. (1987) J. Biol. Chem. 262, 8707-8711.
12. Eisel, U., Jarausch, W., Goretzki, K., Henshen, A., Engels, J., Weller, U., Hudel, M., Habermann, E., and Niemann, H. (1986) EMBO J. 5, 2495-2502.
13. Fairweather, N.F., and Lyness, V.A. (1986) Nuc. Acids Res. 14, 7809-7812.
14. Sathyamoorthy, V., and DasGupta, B.R. (1985) J. Biol. Chem. 260, 10461-10466.
15. Simpson, L.L. (1986) Ann. Rev. Pharmacol. Toxicol. 26, 427-453.
16. Glaser, P., Ladant, D., Sezer, O., Pichot, F., Ullman, A., amd Danchin, A (1988) Mol. Microbiol. 2, 19-30
17. Weklos, S. L., Lowe, J. R., Eden-McCutchan, F., Vodkin, M., Leppla, S. H., and Schmidt, J. J. (1988) Gene 69, 287-300

18. Yamada, M., Ebina, Y., Miyata, T., Nakazawa, T., and Nakazawa, A. (1982) Proc. Nat. Acad. Sci. USA 79, 2827-2831.
19. Aktories, K., Barmann, M., Ohishi, I., Tsuyama, S., Jakobs, K.H., and Habermann, E. (1986) Nature 322, 390-392.
20. Florin, I., and Thelestam, M. (1986) Biochem. Biophys. Res. Commun., 139, 64-70.
21. Popoff, M.R., and Boquet, P. (1988) Biochem. Biophys. Res. Commun. 152, 1361-1368.
22. Roberts, W.K., and Selitrennikoff, C.P. (1986) Biosci. Reports 6, 19-29.
23. Vanderkerckhove, J., Schering, B., Barmann, M., and Aktories, K. (1987) FEBS Lett. 225, 48-52.
24. Narumiya, S., Sekine, A., and Fujiwara, M. (1988) J. Biol. Chem. 263, 17255-17257
25. Feng, D.F., Johnson, M.S., and Doolittle, R.F. (1985) J. Mol. Evol. 21, 112-125.
26. Higashijima, T., Uzu, S., Nakajima, T., and Ross, E. M. (1988) J. Biol. Chem. 263, 6491-6494 27. Blewitt, M.G., Chung, L.A., and London, E. (1985) Biochemistry 24, 5458-5464.
28. Farahbakhsh, Z.T., Baldwin, R.L., and Wisnieski, B.J. (1987) J. Biol. Chem. 262, 2256-2261.
29. Neville, Jr. D.M., and Hudson, T.H. (1986) Ann. Rev. Biochem. 55, 195-224.
30. Wigley, D.B., Hart, K.W., and Holbrook, J.J. (1987) Biochem. Biophys. Res. Commun. 149, 927-929.
31. Dallas, W.S., and Falkow, S. (1980) Nature 288, 499-501.
32. Cieplak, W., and Eidels, L. (1987) UCLA Symp. Mol. Cell. Biol. 45, 19-26.
33. White, J., Kielian, M., and Helenius, A. (1983) 16, 151-195.
34. Janin, J., Miller, S. and Chothia, C. (1984) J. Mol. Biol. 204, 155-164.
35. Gray, G.L., Smith, D.H., Baldridge, J.S., Harkins, R.N., Vasil, M.L., Chen, E.Y., and Heyneker, H.L. (1984) Proc. Natl. Acad. Sci. USA 81, 2645-2649.
36. Greenfield, L., Bjorn, M.J., Horn, G., Fong, D., Buck, D.A., Collier, R.J., and Kaplan, D.A. (1983) Proc. Natl. Acad. Sci USA 80, 6853-6857.
37. Halling, K.C., Halling, A.C., Murray, E.E., Ladin, B.F., Houston, L.L., and Weaver, R.F. (1985) Nuc. Acids Res. 13, 8019-8032.
38. Mekalanos, J.J., Swartz, D.J., Pearson, G.D.N.,Harford, N., Groyne, F., and de Wilde, M. (1983) Nature 306, 551-557.
39. Nicosia, A., Perugini, M., Franzini, C., Casagli, M.C., Borri, M.G., Antoni, G., Almoni, M., Neri, P., Ratti, G., and Rappuoli, R. (1986) Proc. Natl. Acad. Sci. USA 83, 4631-4635.
40. Allured, V.S., Collier, R.J., Carroll, S.F., and McKay, D.B. (1986) Proc. Natl. Acad. Sci. USA 83, 1320-1324.