

Prediction of chain turns in globular proteins on a hydrophobic basis

George D. Rose

Department of Chemistry, University of Delaware, Newark, Delaware 19711

Peptide chain turns are those parts of a globular protein where the backbone changes its direction. A simple, non-empirical hypothesis is presented to account for both the turns and the hydrophobic core of the protein using only the amino acid sequence.

KAUZMANN's generalisation¹ that globular proteins have a hydrophobic core forms the basis of our ideas about protein tertiary structure. In this article that generalisation is extended and used to understand the hydrophobic basis of peptide chain turns. Subsequent publications will deal with other aspects of secondary and tertiary structure.

Peptide chain turns are those parts of a globular protein where the backbone is folded back on itself; they are in effect changes in the overall direction of the polypeptide chain. Different investigators have used the concept of a turn in qualitatively differing ways. Turns in the general sense, as used here, were first described by Kuntz², Lewis *et al.*³ and Crawford *et al.*⁴, but significant elements of this concept are seen in earlier work of Dickerson *et al.*⁵, Blake *et al.*⁶, Venkatachalam⁷ and Kendrew⁸.

Although the concept of a chain turn seems intuitively obvious on visual examination of a physical model of a protein, the lack of an objective structural definition has resulted in variant catalogues of turns compiled by different investigators for the same protein. For this reason, Rose and Seltzer⁹ devised an algorithm to identify turns from X-ray-elucidated coordinates. This algorithm treats the polymer chain as a curve in space and computes a discrete radius of curvature for that curve. A turn corresponds to a locus where the chain direction vector is changing rapidly and the value of the radius of curvature is at a local minimum. Although turns are shown to correspond to local minima in the radius of curvature, the correspondence is not 1 : 1; that is, a small fraction of the minima is not associated with turns, and additional analysis was developed to identify these loci. α -Carbon coordinates are the only information required by the algorithm, and the existence of hydrogen bonds at turn loci is irrelevant to the geometrical nature of the procedure. In this sense, the algorithm provides an objective criterion for the recognition of turns as strictly structural components in proteins.

Rose and Wetlaufer¹⁰ later showed that the number of turns in a globular protein is a linear function of the molecular weight, and this relationship was interpreted to mean that turns are determined by linearly local sequences of amino acids along the polypeptide chain.

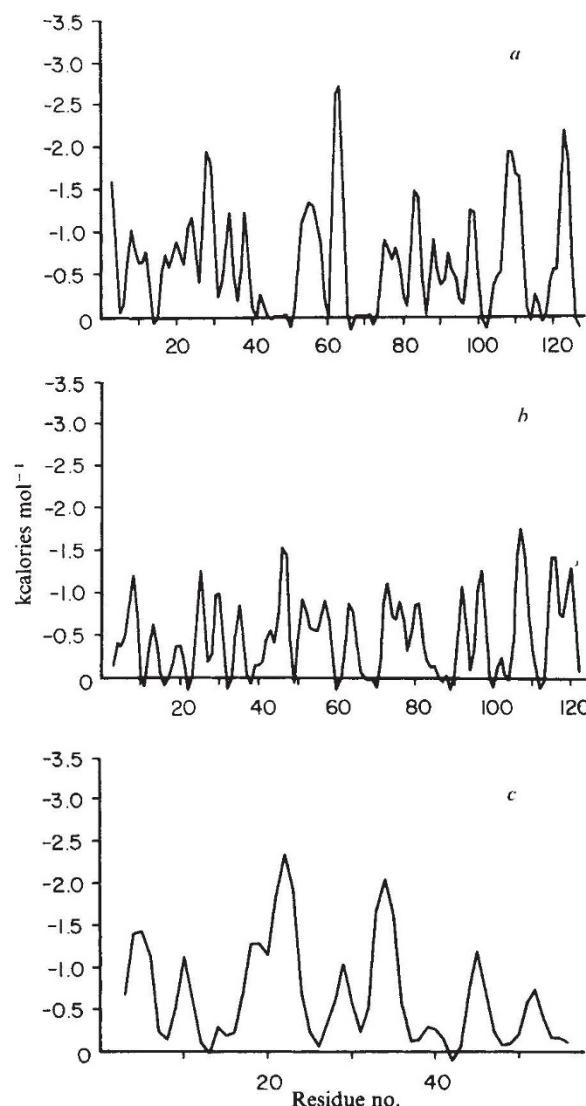
Hypothesis for turns

It is hypothesised here that turns occur at those sites in the polypeptide chain where the hydrophobicity is at a local minimum. A correspondence between turns and minima in hydrophobicity is not too surprising, for it has been observed repeatedly since Kuntz's work² that the sequence composition at turn sites is comparatively polar. Such an observation is implicit in the published 'predictive' schemes for finding turns^{11,12} and β bends⁷ (which are a subset of the turns). These predictive strategies usually proceed by seeking a

correlation between local collections of certain amino acids and a structural preference for turns, and in these schemes the observed turn-forming residues are all non-hydrophobic. Such an approach is almost entirely empirical, and prediction is about 70% successful at best¹³. In contrast, the hypothesis that turns correspond to minima in hydrophobicity has a chemical rationale and better predictive capability.

A method for locating the turns specified by the turns hypothesis is presented here. Linearly local clusters of large hydrophobic residues can be identified by scanning along a

Fig. 1 Smoothed hydrophobicity profiles for lysozyme (*a*), ribonuclease (*b*) and pancreatic trypsin inhibitor (*c*). A hydrophobicity profile plots the Nozaki and Tanford free energy of transfer against the residue number. Local minima on this curve correspond to the peptide chain turns.



protein's amino acid sequence. These clusters consist of from one to a few nearly consecutive residues in the sequence. Large hydrophobic residues are amino acids with hydrophobic side chains larger than alanine; they include tryptophan, phenylalanine, tyrosine, leucine, isoleucine, valine and methionine. In this approximate partitioning, adjacent clusters are separated from each other by a section of chain that is devoid of large hydrophobic groups. It is apparent from the three-dimensional structure of X-ray-

elucidated proteins that groups of clusters are brought into proximity, thereby kinking the intervening chain. These kinks will occur at the comparatively non-hydrophobic sites along the chain, and they become the peptide chain turns; while the interactions of proximate hydrophobic clusters establish the apparent hydrophobic core of the protein. Using this strategy for partitioning the chain, it is possible to locate most of the turns merely by looking at an amino acid sequence.

Table 1 Comparison of turns

Lysozyme	Measured	Predicted	Levitt-Greer	Measured	Predicted	Levitt-Greer	Measured	Predicted	Levitt-Greer
1.	5	5.5		4.	19.4*		22.	110.0*	113
2.		(10.4)		5. 26	25.8	26-27	23. 116	117.9	116-119
3.	15	14.4		6. 29, 33	31†		24. 120-124	121.2	
4.	18-19	18.2	18-22	7. 38	37.6	38-39	25.	(127.7)	
5.	21	21.5		8. 41-42	42.2		26. 135-136	134.2	
6.	24-26	25.8		9. 47-48	48.5		27. 142	142.8	137-141
7.		(31.4)		10.	55.2*	56			
8.	35-36	36.0	36	Flavodoxin			Subtilisin		
9.	38	40.7	41	1. 9	8.9	8-9	1. 11	9.3	11
10.	48	45.0, 50.0	48	2.	(12)		2. 2	13.3*	
11.	55-56§			3.	(16)		3. 19-20	18.4	
12.	62	59.6	62-64	4.	(21)		4. 26-28	23.6	24-25
13.	67	65.9		5. 26-27	26.8		5. 32	33.0	32-34
14.	70	72.1	67-71	6.	31.4		6. 40	38.5	38, 40-42
15.	78	76.9‡		7. 36	36.5	35-37	7. 45-46	47.1	46
16.		80.5‡		8.	41.2*		8. 53	54.6	51-53
17.	86	86.0	86-87	9. 44-45†			9. 57-58	57	
18.		90.2*		10. 47	46.1	45	10. 61-62	61.0	61-62
19.		(95.6)		11. 58	58.0	58-59	11. 61-62	(65.1)	
20.	102	101.8	101-106	12. 64-65	63.4		12. 72	70.3	
21.	109-110	113.8‡		13. (67.3)			13. 79	77.1	75-79
22.	117-118	117.3		14. (71.8)			14. 84		87
23.		126.8*	126	15. 75-76	75.0	76	15. 87	87.2	
Ribonuclease				16. 79	79.6		16. 95-96	93.5	
1.	3-4	5.0		17. 90	86.8, 92.5	89-90	17. 99	98.7	98-99
2.		(10.7)		18. (97.5)			18. 103-104	101.2	
3.	14	16.1		19. (101.4)			19. (109.7)		
4.	18¶		17-18	20. 106-107	107.5		20. 117	117.4	
5.	22	22.2	22	21. 112	112.9	112-113	21. 125	124.0	126
6.	24-25	27.3		22. 119	119.8	119-120	22. 120	129.2	129-130
7.	33-34	32.3		23. 123	125.9	122-123	23. 132-133	131.0	
8.	36	37.7	37-40	24. (133.1)		137	24. (136.7)		
9.	38	40.1					25. (140.7)		
10.	42	43.6		Staphylococcus nuclease			26. 145-146	145.3	146
11.	50-51	48.9		1. 5-6	3.7	3-7	27. 155	155.8	155-158
12.		(54.2)		2. 9.9*			28. 161	161.8	160-162
13.	58-59	60.2	59-60	3. 15	16.7		29. 169	168.9	
14.	62-63†		64	4. 20	20.6		30. 172	172.6	172-173
15.	67	67.7	66-67	5. 28-29	29.6	28-29	31. 178.1*		
16.	70	69.5		6. 38-40	41.8	42	32. 183	183.0	182-183
17.	77	74.7	77	7. 45	45.0	44-46	33. 188-189	185.2	188-189
18.	80	78.2		8. 48	48.0	48-53	34. 194	194.0	193-195
19.	88	86.8, 89.0	88	9. 53-57	51.7		35. 201	201.3	
20.	93	94.3	92-93	10. (56.9)			36. 206.9*		
21.		99.8*		11. (63.3)			37. 211	211.4	211
22.	104	103.6		12. 70	68.8		38. 215-216	215.3	
23.	108-109†			13. 73.4			39. 218-220	219.6	
24.	113-114	112.2	112-114	14. 79	79.0	78-80	40. (224.7)		
25.		117.7*		15. 81-82	82.0		41. (229.8)		
26.		121.5*		16. 85		84-85	42. 240	237.9	239-240
Trypsin inhibitor				17. 87.3‡			43. 242-243	243.7	
1.	5	7.8	8-9	18. 90-91			44. (247.9)		
2.	13-14	12.7	12-13	19. 95	95.7	96	45. 252	253.5	
3.	16	15.3		20. 100	100.8		46. 258	259.0	257-262
				21. 107	105.9		47. 261-262	265.0	
							48.	271.5*	270-274

Measured turns were found using the algorithm of Rose and Seltzer⁹ that identifies turns from structural criteria only. Predicted turns are the local minima in hydrophobicity. Levitt-Greer turns¹⁸ are the output of another algorithm that identifies reverse turns from structural criteria. Taking the measured turns as a standard, there are three ways the predicted turns can fail to be in agreement. (1) Overprediction: predicted turn does not correspond to any measured turn. (2) Underprediction: measured turn does not correspond to any predicted turn. (3) Alternate prediction: measured turn and predicted turn parse the chain differently. In cases of disagreement between measured and predicted turns, an α -carbon bent-wire physical model was used to help clarify the discrepancy. When visual inspection of the model affords a plausible explanation of the discrepancy, the turn is labelled with an asterisk, dagger or double dagger. These cases are discussed further in the text.

* Plausible overprediction: physical model reveals a believable alternate parsing that places an additional turn at this site.

† Plausible underprediction: physical model reveals a locally distorted helix or β strand in which the two pieces at either side of the local perturbation are almost colinear.

‡ Alternate prediction: measured turns and predicted turns both appear acceptable on strictly structural grounds in the physical model. Alternate parsings occur frequently around β_{10} helical turns.

§ Turn not at solvent accessible surface of the molecule. Hydrophobic minimum fails to predict this turn.

¶ Algorithm was applied to the atomic coordinates for RNase-S.

To refine this analysis, a measure of the hydrophobicity of the amino acid side chains is needed, and this is taken to be the Nozaki and Tanford¹⁴ free energy of transfer from water to an organic solvent. With this information, a graph of residue number against hydrophobicity can be constructed easily for any protein. This protein hydrophobicity profile is then smoothed by taking a five-point moving average in place of the individual values of hydrophobicity. Any smoothing technique should suffice here; I have used least squares fitting to a quadratic polynomial according to a method of Savitzky and Golay¹⁵.

The smoothed hydrophobicity profile for three proteins is shown in Fig. 1. Use of the free energy of transfer allows each point in the sequence to be associated with its appropriate hydrophobic energy.

The hydrophobicity profile can now be differentiated to find local minima. A complete set of local minima for six proteins is listed in Table 1, including the three proteins of Fig. 1. This set of six was used to correspond to the examples in the report⁹ giving a structural basis for the recognition of turns. However, a set of approximately two dozen proteins was analysed and the results were validated by the use of physical models. The prediction works equally well in all cases.

It must be acknowledged that one or more local minima in hydrophobicity are always found to be associated with helical secondary structure. Of course, the peptide chain does change direction at the endpoints of a helix and local minima would be expected there. However, additional minima may occur between the ends of helices, and in these cases it might seem that the proposed solution overpredicts turns. Because a helix is comprised entirely of turns, the structural effect of a local minimum in hydrophobicity is masked when embedded within residues that promote a helix. For this reason, turns in Table 1 that are interior to a helix are given in parentheses. In a pragmatic sense, this means that practical prediction of the apparent turns also requires a knowledge of the helical residues. This is to say, apparent turns correspond to the set of local minima in hydrophobicity minus any minima that are at the interior of helices.

Empirical correlations usually avoid the problem of ostensible overprediction by first dealing with the helices, and then removing the predicted helical residues from further consideration before the prediction of turns. The relationship between hydrophobic minima and helices will be dealt with in another publication in which a method for predicting the helices will be presented. (In the six proteins in Table 1, 'in-helix' minima represent 14% of the total turns predicted.)

Identification of turns

A standard for comparison is needed to judge the merit of the set of predicted turns. Table 1 compares the predicted set and the results of the algorithm⁹ which locates turns on a strictly structural basis. Although both the Rose-Seltzer algorithm and the minima in hydrophobicity give rise to discrete points as the loci of turns, it is not sterically possible for the polypeptide backbone to execute a turn within a single residue. For example, the subset of turns known as β bends extends over four consecutive residues, and very gentle turns involving up to eight residues have been observed². In general, a physical turn occupies an interval $I = [a, b]$ where a is the first residue involved in the turn, b the last residue, and $(b-a) \approx 4$. The turn locations cited in Table 1 should be viewed as internal to such an interval. For this reason, a difference of two or three residues between measured and predicted values in Table 1 need not be critical. It is reassuring that the number of hydrophobic minima compares favourably with the number of measured turns, excluding minima within the helix.

As mentioned above, there is no certified list of turns for

a protein. Indeed, the absence of such a list prompted development of the Rose-Seltzer analysis. In any analysis, turns are ultimately a property of the protein's three-dimensional geometry, and they must be apparent on visual inspection of a physical model. It should be emphasised that of the order of 15% of the turns are ambiguous in any protein model, and this limitation constrains the optimum correspondence that can be expected between any two methods for locating turns from strictly structural criteria. Ambiguity arises when the backbone experiences frequent changes of direction within a small number of residues. For example, the S turn in lysozyme from residues 17–22 can be construed as having either two turns (at residues 17 and 22) or three (at residues 17, 20 and 22); in the absence of additional information, either alternative is acceptable. Ambiguity also arises when there is more than one way to parse the backbone. Again in lysozyme, residues 109–110 can be viewed as a turn, but deferring the turn location to residue 114 is equally acceptable on structural grounds. In this latter example, placing the turn at residue 114 seems preferable because residue 110 is within a 3_{10} helix, but this preference is due to chemical, not structural criteria.

In addition to the comparison made in Table 1, the predicted turns were compared with α -carbon bent-wire backbone models^{16,17}, and where the algorithmic measure and the hydrophobic minimum are not in accord, a model was used to clarify the discrepancy. This visual comparison usually suggested an acceptable alternative way to parse the chain at these loci. Moreover, visual inspection showed that the turns prediction hypothesis is consistently more sensitive to the secondary structure of the protein than the structural algorithm. Sensitivity to secondary structure is demonstrated by resolving instances of multiple parsings in favour of a choice that avoids situating the turn at the interior of some other evident secondary structure, as, for example, in the case of residue 114 in lysozyme. Although the Rose Seltzer algorithm deliberately excludes secondary structure as an analytical criterion, the sensitivity to secondary structure shown by the turns hypothesis indicates that some of the hydrogen bonding is an implicit consequence of the local chain hydrophobicity.

Using the structural algorithm as a standard, 117 turns are identified for the six proteins in Table 1, and there are 26 discrepancies with non-helical minima. In the worst case, this amounts to 78% agreement between measured and predicted turns at non-helical loci. However, visual inspection of bent-wire models reveals that most of these discrepancies are due to a plausible alternate parsing of the chain. Because plausible alternate parsings inevitably arise from the inherent ambiguity in locating turns on strictly structural grounds, 22 of the discrepancies in Table 1 should not necessarily be counted in error. Thus, in the best case, the agreement between measured and predicted turns is 97% at non-helical loci.

From the foregoing discussion, it seems that local minima in hydrophobicity give rise to a set of turns that is as authentic as the set identified on structural grounds from X-ray coordinates. The evidence strongly supports the proposed hypothesis for turns.

Comparison with other algorithms

As mentioned above, the literature contains various definitions of turns, and these discrepancies complicate possible comparisons. Most often, investigators have used β bends as a paradigm for recognising the sharply defined reversals in chain direction, known as reverse turns; consequently these procedures can fail to identify the more gentle turns. Such a constraint was deliberately not built into the Rose-Seltzer algorithm, and it has no obvious relationship to the hypothesis for turns. For this reason, our structural algorithm ought to identify a set of turns that includes reverse-turns as a proper subset, and the turns hypothesis

ought to predict a set of turns that also includes any reverse turns.

In the following paragraphs, some comparisons are made between the results of the turns prediction hypothesis and an alternate algorithm for finding reverse turns from X-ray coordinates. In addition, the turns hypothesis was applied to the amino acid sequence of adenyl kinase in order retroactively to participate in an ingenious comparison sponsored by Schultz¹¹, who withheld the X-ray structure of this molecule to provide an unbiased test of the accuracy of procedures for prediction. Any comparison of the turns hypothesis with a method that reveals only the reverse turns is necessarily incomplete, however, because reverse turns are a proper subset of total turns and so possible overprediction cannot be detected. In compensation, however, reverse turns are, by definition, more sharply defined than gentle turns, and the inherent ambiguity that engenders alternate parsings is greatly reduced for this category of turns.

Levitt and Greer¹⁸ published an algorithm for identification of secondary structure from X-ray coordinates. Their criterion for turns was meant to identify only the reverse turns which should be a subset of the turns given by the Rose-Seltzer algorithm. The Levitt-Greer reverse turns for the six proteins used in this study are listed in Table 1.

Of the 70 reverse turns identified by the Levitt-Greer algorithm for the six proteins in Table 1, five of these (or 7%) are not found at all by the Rose-Seltzer algorithm. However, four of the five are final turns within a few residues of the C terminus of the chain where the two structural algorithms disagree due to chain-end effects. With these end effects set aside, only one of the Levitt-Greer reverse turns is missed completely by the Rose-Seltzer algorithm. In addition, three of the turns (4%) found by the Levitt-Greer algorithm are situated between three and four residues away from the corresponding turn identified by the Rose-Seltzer algorithm; these three cases correspond to alternate parsings of the chain.

Taking the Levitt-Greer algorithm as an alternate standard with which to judge the merit of the set of predicted turns, there are at worst six discrepancies (91% agreement) with the set of measured turns. Of these, two turns identified by the structural algorithm are missed completely by the turns prediction hypothesis; one turn is missed because the structural algorithm was applied to RNase-S; and there is one instance each of plausible overprediction, plausible underprediction and alternate prediction as described in Table 1. This amounts to a best case agreement of 97% between turns identified by the Levitt-Greer algorithm and the corresponding turns predicted by the turns hypothesis. Overprediction by the turns hypothesis cannot be scored in this comparison of course because the Levitt-Greer algorithm identifies only a subset of the total turns.

Comparison between the results of the hypothesis for turns and the literature on empirical correlations is also beset by a lack of congruency in the definitions of turns. For example, Schultz¹¹ withheld the X-ray-elucidated details of adenyl kinase until various groups had an opportunity to record their secondary structure predictions for that molecule. Comparisons of both the predicted and the experimentally determined results were then collected and published.

The X-ray data for adenyl kinase were derived from a 3-Å electron density map and, consistent with resolution at this level, the only bends identified by Schultz are those where the backbone changes in overall direction by more than 120°. Schultz found ten bends that satisfy this restriction. These are extracted from the bar graph in Fig. 1 of his paper¹¹ and shown here in Table 2.

In comparison, the Levitt-Greer algorithm to identify bends from X-ray coordinates defines as reverse turns those loci where the backbone changes in overall direction by

more than 90°, as measured by their α angle. Thus the Levitt-Greer algorithm ought to find a superset of bends with respect to the more restrictive criterion used by Schultz. Of the ten bends identified by Schultz, the Levitt-Greer algorithm finds eight of these (80% agreement) and four additional reverse turns as well.

Results of the turns prediction hypothesis are also shown in Table 2. All ten of the Schultz bends are successfully predicted by the hypothesis (100% agreement). Of the four additional reverse turns in adenyl kinase identified by the Levitt-Greer algorithm, three of these are also predicted by the turns hypothesis, while a turn corresponding to the fourth is predicted to occur three residues away. This amounts to a worst case agreement of 92% between the Levitt-Greer reverse turns and the turns prediction hypothesis, and a best case agreement of 100% if the turn at residues 66–67 is construed to be an alternate parsing. Once again, the predicted turns are a superset of the measured turns, and there is no way to score for possible overprediction.

Table 2 Comparison of measured and predicted turns for adenyl kinase

Schulz	Levitt-Greer	Predicted
16–23	16–19	16.6, 20.0, 22.3
29–32	32	30.4
		35.6
38–42		39.6
48–52		49.6
		54
62–65	62–64	63.4
	66–67	70.4
84–88	84–86	87.2
		93.4
	96–98	97.2
109–112	110–112	107.1
		111.6
		122.4
134–137	137–140	136.3
		143.3
	158	157.8
166–169	167–168	166.5
		171.6
175–178	177–178	176.5
	193	191.3

Turns in column 1 are taken from Fig. 1 of the comparison by Schultz¹¹. Turns in column 2 are taken from Table 10 of ref. 18. Turns in column 3 are local minima in hydrophobicity as specified by the turns prediction hypothesis, but excluding those local minima that are interior to a helix.

It should be noted that the turns prediction hypothesis is quite insensitive to any uncertainties in the free energy of transfer used to measure the chain hydrophobicity. Indeed, prediction is almost as successful when the side chains are designated as either nonpolar or polar and assigned $\Delta G_{\text{xf}}^{\circ}$ values of -1 and 0, respectively.

Persistence of local structure

Local maxima in hydrophobicity are also evident in Fig. 1, and the loci corresponding to these peaks appear to fold relative to each other to establish the tertiary structure and the hydrophobic core of the protein. Examination of distance plots¹⁹ confirms that the tertiary interactions (which appear as off-diagonal contacts in a distance plot) in the protein correspond to peaks in the hydrophobic profile and, conversely, that every peak is utilised at least once in a tertiary interaction. As mentioned earlier, interactions between hydrophobic maxima are presumed to be responsible for the occurrence of chain turns at the intervening sites where hydrophobicity is at a local minimum.

I have argued here that local minima in hydrophobicity provide a sufficient chemical basis for understanding the occurrence of peptide chain turns. The fact that these

short, linearly local sequences of amino acids are still discernible as turns in the three-dimensional structure has important implications for protein folding. In an earlier article, a structural segment was defined as a continuous sequence of linear chain neighbours bounded by consecutive chain turns⁹. I have shown here that linearly local interactions suffice to partition the polypeptide chain into its structural segments and turns, and that these locally defined moieties persist in the product of the folding process. The persistence of local structure is the basis of our model of protein folding²⁰.

I thank Professors D. Dennis, R. H. Winters and K. E. Van Holde for useful discussion, Dr Richard Feldmann for X-ray coordinates, and Professors Walter Kauzmann and Judith Voet for critical reading of the manuscript. This work was carried out in the laboratory of Professor Donald Wetlaufer, and was supported by US PHS grant RO1 GM-23713-01 and by a University of Delaware Provost's grant.

Received 8 November 1977; accepted 23 January 1978.

1. Kauzmann, W. *Adv. Protein Chem.* **14**, 1 (1959).
2. Kuntz, I. D. J. *Am. chem. Soc.* **94**, 4009–4012 (1972).
3. Lewis, P. N., Momany, F. A. & Scheraga, H. A. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2293–2297 (1971).
4. Crawford, J. L., Lipscomb, W. N. & Schellman, C. G. *Proc. natn. Acad. Sci. U.S.A.* **70**, 538–542 (1973).
5. Dickerson, R. E. *et al.* *J. biol. Chem.* **246**, 1511–1535 (1971).
6. Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. *Proc. R. Soc., Lond. B* **167**, 365 (1967).
7. Venkatachalam, C. M. *Biopolymers* **6**, 1425–1436 (1968).
8. Kendrew, J. C. *Sci. Am.* (December 1961).
9. Rose, G. & Seltzer, J. *J. molec. Biol.* **113**, 153–164 (1977).
10. Rose, G. & Wetlaufer, D. *Nature* **268**, 769–770 (1977).
11. Schultz, G. E. *et al.* *Nature* **250**, 140–142 (1974).
12. Matthews, B. W. *Biochim. biophys. Acta* **405**, 442–451 (1975).
13. Chou, P. Y. & Fasman, G. D. *Trends biol. Sci.* **2**, 128–131 (1977).
14. Nozaki, T. & Tanford, C. *J. biol. Chem.* **246**, 2211–2217 (1971).
15. Savitzky, A. & Golay, M. *Analyt. Chem.* **36**, 1627–1639 (1974).
16. Rubin, B. & Richardson, J. S. *Biopolymers* **11**, 2381–2385 (1972).
17. Goldberg, D., Saliterman, S., Wetlaufer, D. B., Rose, G. D. & Hopkins, T. E. *Biopolymers* **14**, 633–640 (1975).
18. Levitt, M. & Greer, J. *J. molec. Biol.* **114**, 181–293 (1977).
19. Rossmann, M. G. & Liljas, A. *J. molec. Biol.* **85**, 177–181 (1974).
20. Rose, G. D., Winters, R. H. & Wetlaufer, D. B. *FEBS Letts* **63**, 10–16 (1976).

DRB-induced premature termination of late adenovirus transcription

Nigel W. Fraser, Pravinkumar B. Sehgal & James E. Darnell

The Rockefeller University, New York, New York 10021

Exposure of HeLa cells to 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) late in the course of adenovirus 2 infection results in the inhibition of virus-specific RNA synthesis from all parts of the previously identified, very long (~ 28 kilobases) late transcriptional unit, except for the first ~ 400–800 nucleotides. It is suggested that DRB acts not as an inhibitor of RNA chain initiation, but rather causes 'premature' chain termination close to the promoter. A practical aspect of these findings may be that RNA sequences near promoter sites that are responsible for mRNA formation can be isolated from DRB-treated cells.

THE initiation sites for RNA synthesis (promoters) in adenovirus 2 (Ad 2) DNA during the infection of HeLa cells were the first promoters responsible for mRNA production to be identified in DNA-containing animal viruses or animal cells^{1–4}. As such, the DNA sequences in the region of the promoters, and the RNA transcribed from these regions are of considerable interest. A useful test for promoter proximity of nucleotide sequences in bacteria derives from the action of inhibitors like the rifamycins that allow chain completion but block chain initiation⁵. Thus, after rifamycin treatment, promoter proximal synthesis ceases first and promoter distal synthesis last. The nucleoside analogue 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) was previously thought to act in eukaryotic cells by inhibiting transcriptional initiation of HnRNA, the possible precursor to mRNA^{6–9}, and a test of this hypothesis was made possible based on the location of the promoter site for the majority of late adenovirus RNA synthesis. If DRB only stopped chain initiation, cessation of virus-specific nascent chain labelling would first occur near 16 (the major 'promoter region' late in Ad2 infection⁴), and then inhibition of labelling would gradually progress to the distal terminus of the transcribed region near 100 on the physical map. (Conventional description¹⁰ of the Ad2 genome places 0 at the left-hand end and 100 at the right-hand end of the DNA strand which is transcribed towards the right; each unit is 0.35 kilobases, 1% of the genome, the total length of Ad2 DNA

being 35 kilobases. Restriction endonuclease fragments of the DNA are designated by italics; for example, *Sma*I 11.6–18.2.) The results indicated that DRB did, in fact, inhibit Ad2 RNA synthesis near the origin of the transcription but, surprisingly, that RNA chains which possibly represent the first 400–800 nucleotides of the transcription unit continued to be labelled.

Late Ad2 RNA synthesis in DRB-treated cells

In a way similar to its effect on cellular HnRNA synthesis, DRB strongly inhibited both late Ad2 nuclear RNA synthesis (90–95%) and the appearance of label in Ad2-specific mRNA (> 95%) (Table 1). Additional experiments demonstrated that a concentration of 75 μM DRB exerted maximal inhibition of Ad2 RNA synthesis (data not shown), as was also true for cellular mRNA synthesis^{11,12}.

The time course of inhibition of Ad2 RNA synthesis by DRB was then determined; an inhibition of chain elongation should inhibit incorporation immediately, whereas inhibition of chain initiation would inhibit incorporation more gradually because of the completion of already initiated chains. Ad2-infected HeLa cells were exposed to DRB for various periods and then labelled for 2 min with ³H-uridine. A pulse of this duration should be much shorter than the anticipated synthesis time

Table 1 Inhibition of late Ad2 RNA synthesis by DRB

	Nuclear RNA c.p.m.	% of control	Polysomal mRNA c.p.m.	% of control
Control	14,667	—	937	—
DRB	732	5.0	15	1.6

One sample of HeLa cells (5×10^7 per sample) infected with Ad2 was exposed to DRB (75 μM) 16 h after infection, and one served as control. After 30 min, 1 mCi of ³H-uridine (New England Nuclear, 26 Ci mmol⁻¹) was added to each sample and incubation continued for a further 35 min. Nuclear RNA and EDTA-released polysomal mRNA were prepared as described^{1,2,22}. Ad2-specific incorporation was measured by hybridisation to nitrocellulose filters containing 2 μg each of Ad2 DNA^{1,2}. Incorporation of ³H in the presence of DRB has been corrected for an inhibition of ³H-uridine transport into cells by a factor of 1.23 (ref. 11).