solution.[13] However, since at least 25–45% of the histamine cation is in the trans conformation, this conformer could still be the reactive species. Furthermore, the recent crystal-structure study of *dl*-bromopheniramine maleate, a potent antihistamine, has also revealed a trans conformation about the C–C bond of the dimethylaminoethyl group.[14] Therefore, the hypothesis that the trans conformation is essential to antihistamine activity appears very reasonable. In addition, the dimethylamino group is more basic than the NH_2 group in histamine, allowing the antihistamine to displace or to compete favorably with histamine for a receptor site. The biological response caused by histamine is then related to the imidazole group, not the dimethylaminoethyl chain. The bulky groups found in most antihistamines are therefore selected to prevent the response triggered by the imidazole ring. Other structural studies of antihistamines and related compounds are planned to further explore these hypotheses.

The biological activity of the cephalosporins[15] was related to the planarity of the β-lactam nitrogen atom. Although I contains a planar tertiary amine nitrogen atom, the biological activity of the drug is probably not related to this function. The conformational arguments presented above, together with the fact that other potent antihistamines lack this group, suggest that a planar tertiary nitrogen is not essential for antihistamine activity. However, the present results together with the cephalosporin studies indicate that planarity may be a good indication of resonance interaction.

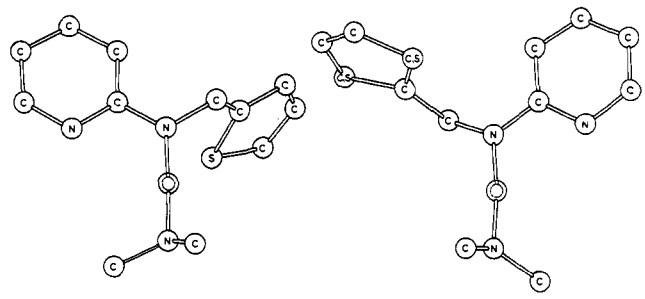

Figure 2. A view down the C–C bond of the dimethylaminoethyl side chain in the two molecules in the asymmetric unit. The disordered thiophene ring atoms are labeled CS.

Acknowledgment. We wish to thank Eli Lilly and Co. for the sample of Histadyl. Financial support was provided by the U. S. Public Health Services, National Institute of Allergies and Infectious Diseases, and the National Research Council of Canada. The research was carried out at the University of Waterloo, Waterloo, Ontario, Canada.

(13) A. F. Casy and R. R. Ison, *Chem. Commun.*, 1343 (1970).
(14) M. N. G. James and G. J. B. Williams, *J. Med. Chem.*, **14**, 670 (1971).
(15) R. M. Sweet and L. F. Dahl, *J. Amer. Chem. Soc.*, **92**, 5489 (1970).

Protein Folding

I. D. Kuntz

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94122.
Received September 17, 1971

Abstract: We have developed simple computer routines for locating regions in which the peptide backbone of globular proteins is folded back on itself. These programs have been used to locate "turns" in carboxypeptidase and α-chymotrypsin. The following generalizations can be made. (1) Turns occur on the surface of the protein molecules. (2) Turn segments are less hydrophobic than the protein as a whole. (3) Uninterrupted sequences of 3–8 hydrophilic residues are frequently associated with folding of the peptide chain.

Although more than ten protein crystal structures have been determined, the principles by which these molecules develop secondary and tertiary structure are not yet well understood. The considerations discussed to date have been rather general and nonrestrictive.[1,2] One approach to this problem is to focus on local structural features such as α helices and β-pleated sheets. Recently, attention has been drawn to the turning points of the peptide chain. "Hairpin" and 3–10 turns have been predicted theoretically[2,3] and identified as common features of globular protein structures.[4] We have developed computer procedures for determining the location of "hairpin" and other turns, given the protein coordinates. We report here some initial results on three characteristics of the turns in globular proteins. (1) Turns occur on the surface of the protein molecules. (2) The amino acid composition of these regions is considerably less hydrophobic than the composition of the proteins as a whole or the helical and β sections. (3) The occurrence of a turn or corner is very frequently associated with an uninterrupted sequence of 3–8 hydrophilic residues.

Methods and Results

Our observations are based on the following sources of data: (1) the atomic coordinates[5] and a Phillips–

(1) W. Kauzmann, *Advan. Protein Chem.*, **14**, 1 (1959).
(2) G. N. Ramachandran and V. Sasisekharan, *ibid.*, **23**, 284 (1968).
(3) C. M. Venkatachalam, *Biopolymers*, **6**, 1425 (1968).
(4) R. E. Dickerson, T. Takano, D. Eisenberg, O. Kallai, L. Samson, A. Cooper, and E. Margoliash, *J. Biol. Chem.*, **246**, 1511 (1971).

(5) J. J. Birktoft, B. W. Matthews, and D. M. Blow, *Biochem. Biophys. Res. Commun.*, **36**, 131 (1969).

Kendrew model of α-chymotrypsin;[6] (2) carboxypeptidase atomic coordinates;[7] (3) published analysis of the cytochrome *c* structure;[4] (4) stereoviews of lysozyme, ribonuclease, and myoglobin,[8] combined with the static accessibility data of Lee and Richards.[9]

We must begin with a precise and, hopefully, useful definition of a "turn." One feels, intuitively, that a turn is defined by the angles formed with the vectors which link α-carbon atoms in sequential peptide residues. The simplest form of this type of definition would involve the angle, θ, formed by the two vectors connecting three successive α-carbons, specifically, the vectors $\vec{\mathbf{R}}_{i,i+1}$ and $\vec{\mathbf{R}}_{i+1,i+2}$ linking the atoms C_i to C_{i+1} and C_{i+1} to C_{i+2}. Interestingly, this simple idea is not very useful because steric factors limit the approximate range of θ[10] to 30–100°. Thus *at least* four residues are required to generate a 180° turn. Two versions of the hairpin turn, for example, involve four residues with the angle made by $\vec{\mathbf{R}}_{i,i+1}$ and $\vec{\mathbf{R}}_{i+2,i+3}$ 180° and the separation of the first and fourth carbon atoms of 4.8–4.9 Å[11] Computations of the appropriate angles and separations in chymotrypsin and carboxypeptidase yield no angles >170°, and many high-angle turns have separations considerably different from 4.8 Å. Accordingly we adopted the arbitrary definitions for "hairpin" and "near hairpin" turns, "loops," and "corners" given in Table I. Our major results are not very sensitive to the precise limits in these definitions. Residues meeting the various definitions for carboxypeptidase and chymotrypsin are given in the first column of Table IIA. One should remember that the experimental uncertainties are not accurately known and vary from one part of the molecule to another. We assume errors of ±20° in angles and ±0.3 Å in separations. Turns for myoglobin, ribonuclease, and lysozyme are given in Table IIB. These latter entries were obtained from stereoviews[8] and cannot be accurately classified as to type. Location of the turns is also more uncertain than those in Table IIA.

Table I. Definitions of Turns

Type of turn	No. of residues	θ_n,[a] deg	S_n,[a] Å
Hairpin (H)	4	$\theta_4 \geq 135$	$S_4 \leq 5.5$
Near hairpin (NH)	4	$135 > \theta_4 \geq 120$ or $\theta_4 \geq 135$	$S_4 \leq 5.5$ or $5.5 < S_4 \leq 6.5$
Corners (C)	4	$135 > \theta_4 \geq 90$ or $120 > \theta_4 \geq 90$	$5.5 < S_4 \leq 6.5$ or $S_4 \leq 5.5$
Loops (L)	≥5	$\theta_n \geq 120°$	$S_n \leq 6.5$

[a] θ_n is the angle formed by the vector linking α-carbons i and $i+1$ and the vector linking atoms $i+n-2$ and $i+n-1$; n is the number of residues in the turn; S_n is the separation of the first and nth carbon atoms. These definitions permit overlapping turns. Loops will often include other turns. We only consider loops which have $S_n < S_{n-i}$ for all turns within the loop.

To test the observation that turns occur at the molecular surface we must have some working definition of "surface." A qualitative definition is "the boundary separating protein and solvent." The best attempt to reduce this to quantitative terms is the "static accessibility" calculated by Lee and Richards.[9] Their results are used in Table IIB. For carboxypeptidase and chymotrypsin we have used a less accurate but simple method which calculates the number of α-carbon atoms which lie outside (*i.e.*, further from the molecular center of gravity) of a given α-carbon but within a cone of ±20° of the test residue, the apex of the cone being located at the molecular center of gravity. Approximate geometric arguments indicate that five–ten "exterior" residues are required to "block out" solvent. In Table IIA we have listed all those residues with three or less such exterior residues as being on the molecular surface in the sense that they are very likely to be accessible to solvent. Our criterion with the static accessibility data was that at least one hydrogen bonding atom (O, N, S) per residue should have static accessibility of 10 or more.

Although these two definitions of "surface" are arbitrary, they are both on the conservative side in that some residues which we count as interior actually could have substantial exposure to the solvent. For example, given the small size of hydrogen atoms, a static accessibility of less than 5 and perhaps as low as 1 would permit full hydrogen bond formation. Using these two tests only 8 of the 130 turns tabulated in Table II might occur on the "inside" of the protein molecules (Table III).

The amino acid composition of turns, helices, and β structures is summarized in Table III. The low hydrophobicity of turns is apparent. Although it is premature to comment at length on the distribution of individual amino acids, we find that those residues commonly considered as "helix breakers" are more likely to appear in turns than in the protein as a whole. Thus Ser, Thr, Gly, Pro, Gln, and Asn concentrations are enhanced about 20% in turns. Conversely, the larger hydrophobic groups such as Val, Leu, Ile, and Phe are depleted by 30–50%. Similar conclusions have been reached by Lewis, *et al.*[12]

Given the compositional data, it is not surprising that there is a respectable correlation between the occurrence of turns and the strings of hydrophilic[13] residues that are a common feature of the amino acid sequences. We have listed in Table II all such sequences of three or more hydrophilic[13] residues. The correlation is summarized in Table III.

Discussion

Several comments can be made. First, the number of turns (10–50 per molecule), the number of residues involved in turns (~45% protein), and the fraction of the protein surface taken up by turns (~60%) are all quite high. As a rough check we note that the number of 180° turns to be expected for a chain connecting points within a cube is $N^{1/3} - 1$, where N is the number of points, if one assumes that all turns occur in two opposite faces of the cube. This model works surprisingly well for the two proteins on which detailed calculations have been made. The fact that the three

(6) Built in the laboratory of Professor W. Kauzmann.
(7) Courtesy of Professor W. Lipscomb.
(8) R. E. Dickerson and I. Geis, "Stereo Supplement to the Structure and Action of Proteins," Harper & Row, New York, N. Y., 1969.
(9) B. Lee and F. M. Richards, *J. Mol. Biol.*, **55**, 379 (1971).
(10) $\theta \equiv \arccos((\vec{\mathbf{R}}_{i,i+1} \cdot \vec{\mathbf{R}}_{i+1,i+2})/(|\vec{\mathbf{R}}_{i,i+1}| \cdot |\vec{\mathbf{R}}_{i+1,i+2}|))$.
(11) Ref 2, p 340.
(12) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Proc. Nat. Acad. Sci. U. S.*, **65**, 2293 (1971).
(13) This term is somewhat misleading in that we have included Gly, Pro, and Ala (see Table II, footnote *c*).

Table II. Turn Sequences

A. From Coordinates

Residues	Type of turn[a]	Residues on surface[b]	Predicted[c,d]	Residues	Type of turn[a]	Residues on surface[b]	Predicted[c,d]
		Carboxypeptidase		278–281	C	278–281	
1–8	L	1–3, 5, 6	1–6	282–285	C	282, 284–285	282–285
3–6	N	3, 5, 6	1–6	283–286	N	284–285	282–285
4–7	N	5, 6	1–6				
8–11	C	8, 10–11	8–11			α-Chymotrypsin	
13–17	L	13, 17	11–14	4–7	N	5–7	*e*
29–32	C	30–31	27–31	16–19	C	17–19	18–22
30–33	H	30–31	27–31	23–26	H	23–25	24–26
41–46	L	42–46	39–46	25–28	C	25	24–26
41–44	H	42–44	39–46	27–30	C		
56–59	H	56–58	53–61	34–40	L	35–38	34–38
67–70	N		69–72	35–38	C	35–38	34–38
69–72	C		69–72	48–51	H	48–51	48–50
70–73	C	73	69–72	55–58	C	55–58	54–57
89–92	H	91–92	87–93	61–64	C	61–64	61–64
90–93	C	91–93	87–93	67–70	C	69	68–70
108–111	N	110		72–75	C	73–75	72–79
110–113	N	113	111–115	75–78	C	75–78	72–79
119–124	L	120–123	119–124	91–94	C	91–94	90–96
123–126	C	123, 125	119–124	95–98	H	95–98	90–96
132–139	L	134–138	133–136[e]	96–99	N	96–98	
142–145	C		142–146	99–102	N	100	100–102
143–146	N		142–146	108–111	N	109–111	109–113
144–147	C		142–146	114–118	L	114–117	115–117
148–151	H	149–151	148–150	115–118	H	115–117	115–117
150–153	N	150–153	152–160	125–128	N	125–128	124–129
153–156	C	153–154	152–160	131–134	N	131–134	131–135
159–162	N	159	152–160	169–173	L	170–171, 173	169–171
162–165	N	163	162–173	172–175	N	173–175	173–175
169–172	H	169–172	162–173	173–176	C	173–175	173–175
171–176	L	171,172, 174, 175	162–173	177–180	C	178–179	177–179
206–209	N	209	202–212	184–188	L	186–187	183–187
213–216	H	214–215	214–218	190–194	L	192–193	193–198
232–235	H	232–235	232–242	191–194	N	192–193	193–198
242–245	N	245		194–197	N	195–196	193–198
243–246	C	245–246		202–207	L	203–206	202–206
244–247	H	245–246		202–205	N	203–205	202–206
250–253	C		248–254	203–207	N	203–206	202–206
259–262	C	260–262	258–262	217–220	H	217–219	216–219
263–266	C	263–265	264–266	221–224	C	222–224	221–226
273–276	N	274–276	272–278	230–233	C	230, 233	
275–278	H	275–277	272–278	231–234	H	233–234	
277–280	H	277, 279–280	272–278	232–235	C	233–234	

B. From Stereoviews

Residues	No. of turns[f]	Residues on surface[g]	Predicted[c,h]	Residues	No. of turns[f]	Residues on surface[g]	Predicted[c,h]
		Myoglobin		61–72	3	61–63, 65–68, 70, 72	65–74[e]
3–7	1	4, 6	3–6	74–77	1	74–77	*e*
18–21	1	18–20	18–20	79–87	2	81, 85–87	85–87[e]
36–39	1	36, 38	34–39	99–104	2	100–102	100–104
41–47	2	41–42, 44–45, 47		115–120	2	116–119	116–119
50–54	1	50–52, 54	50–55	124–127	1	125–127	*e*
78–81	1	78–81	77–85				
84–87	1	84–85, 87	77–85			Ribonuclease A	
95–98	1	95–96, 98	90–98	3–6	1	3	1–7
119–122	1	119–122	116–122	15–18	1	15–18	14–25
147–150	1	147–150	143–148, 150–153	21–25	1	21–24	14–25
				32–39	2	32–34, 37–39	31–34, 36–39
		Lysozyme		47–53	2	49, 53	48–50
4–7	1	5–7	*e*	59–62	1	59–62	59–62
13–17	1	13–14, 16	13–16	65–72	2	66–71	66–71
19–23	2	19–23	18–24	75–81	2	76–78, 80	73–78
34–42	3	34–35, 37, 39, 41	39–54	91–94	1	91–92, 94	90–94
46–50	1	46–48	39–54	101–105	2	101, 103–104	101–106
54–57	1			112–115	1	113, 115	111–115
60–63	1	61–63	59–61				

[a] Based on computer processing of coordinates according to definitions given in Table I. [b] Based on calculations given in text. [c] Uninterrupted sequences of three or more of any combination: Gly, Pro, Asp, Glu, Arg, Lys, His, Asn, Gln, Ser, Thr, Tyr. Ala is included if adjacent to one of the above. [d] Predicted but not found sequences: carboxypeptidase 75–79, 83–85, 127–131, 184–188, 197–200, 223–226, 288–293, 302–304; chymotrypsin, 43–45,[e] 138–140,[e] 144–146, 149–154, 156–159, 164–167,[e] 228–230, 239–241, 243–245. [e] Turns possibly altered by Cys. [f] Based on stereoviews, hence very approximate. [g] Using the data of ref 9; see text. [h] Predicted but not found: myoglobin, 22–27, 57–60, 62–65, 108–110, 124–130, 132–144, 139–141; lysozyme, 31–35,[e] 89–91, 95–97,[e] 112–114; ribonuclease, 9–12, 85–90,[e] 96–100, 121–123.

Table III. Summary of Results

Protein[a]	Helix + β[b]	Hairpin + near hairpin[c]	Total non-turns	Total turns[c]	Total protein	No. of turns with no residues at surface
	A. Occurrence of Structural Features					
CBP	168	82	166	141	307	
CTN	98	60	127	109	236	
MYO	121	*d*	108	45	153	
LYSO	58	*d*	59	68	127	
RIBO	65	*d*	59	65	124	
	B. Fraction of Residues on Surface[e]					
CBP	0.38	0.75	0.39	0.56	0.47	6
CTN	0.44	0.84	0.60	0.65	0.62	1
MYO	0.45	*d*	0.35	0.75	0.47	0
LYSO	0.36	*d*	0.38	0.71	0.56	1
RIBO	0.57	*d*	0.36	0.60	0.48	0
	C. Fraction of Hydrophobic Residues[f]					
CBP	0.47	0.27	0.43	0.27	0.36	
CTN	0.51	0.33	0.50	0.31	0.41	
MYO	0.39	*d*	0.47	0.27	0.41	
LYSO	0.43	*d*	0.50	0.31	0.40	
RIBO	0.43	*d*	0.46	0.22	0.33	
CYT C		0.24[g]				

Protein[a]	No. of turns calculated[h]	Found[i]	Predicted[j] and found	Found but not predicted	Predicted but not found
	D. Prediction of Turns				
CBP	45	45	40	5	8
CTN	37	38	32	6	9[k]
MYO	28	(11)	9	2	7
LYSO	25	(21)	14	7[l]	4[l]
RIBO	24	(16)	14	2	4

[a] CBP, carboxypeptidase, CTN, chymotrypsin, MYO, myoglobin, LYSO, lysozyme, RIBO, ribonuclease A, CYT *c*, cytochrome *c*. [b] Assignments taken from Dickerson and Geis.[8] [c] Calculated as described in text. Some residues appear both in columns 2 and 5. [d] Data not available. [e] Calculated as described in text. Different criteria were used for the first two proteins and for the last three. [f] Hydrophobic residues were taken as: Ala, Val, Leu, Ile, Phe, and Trp. [g] From Dickerson, *et al.*[4] [h] Calculation from $n^{2/3} - 1$ as discussed in text. [i] From calculations and stereoviews as described. [j] At least 50% overlap between "predicted" and "found" residues with the remaining residues directly contiguous to "predicted" regions. [k] Five of these in center region (138–167) perhaps altered by sequence break at 147–148. [l] All but one involve Cys.

proteins studied with the stereoviews have fewer turns than predicted could reflect the approximate nature of the model or that we were unable to generate a complete set of turns or both.

The location of some turns at the protein surface is expected but is not a geometrical necessity. The more interesting observation is that there are so very few interior turns. This is equivalent to noting that linear structures such as helices and strands of pleated sheets run from one side of the molecule to the other without major changes of direction. In fact, of the eight turns listed in Table III as possibly being internal, not more than five actually appear buried from the inspection of the stereoviews. Three of these are at or near active sites.

The compositional differences between turns and nonturn regions presumably reflect the well-known tendencies for hydrophilic groups to be near the solvent and hydrophobic groups to be "buried." This type of thermodynamic mechanism may actually provide the driving force for turn generation since a large number of the turns observed are "anchored" at the ends by hydrophobic residues. It may be that such composition *gradients* are the important factors. Any such mechanism is, of course, a probabilistic rather than a deterministic one.

The predictability of turns is similar to, but not as accurate as, the methods for helix prediction.[14] A major source of error in our predictions is proximity to disulfide bonds (Tables II and III). In the same way, the relatively poor results for myoglobin might suggest that the heme has a pronounced influence on folding. If so, we would predict that apomyglobin would have a lower helicity and more turns than myoglobin does. The existence of a simple compositional rule suggests that a local, sequence-related, chemical signal greatly increases the chances of turn formation along a particular stretch of peptide chain. Exact location of a turn is presumably based on the complete three-dimensional environment of the chain, including S–S bonds, interfold interactions, solvent effects, and, in the crystals, protein–protein interactions.

As a final point, a preliminary analysis of intramolecular hydrogen bonding in carboxypeptidase suggests that both backbone and side-chain hydrogen bonds are disrupted in turn regions. Less than 50% of the possible donor and acceptor sites are filled intramolecularly compared with more than 80% for helical regions. Since unfilled proton-donor or proton-acceptor sites are energetically unfavorable, we presume that these sites are actually occupied by water molecules, suggesting that solvent participation is an important influence in turn generation or stabilization. It is interesting, but perhaps coincidental, that the number of ureas required to denature globular proteins appears of order one to two urea molecules per turn.[15]

In summary, helices and pleated sheets have been considered important structural units because of their distinctive geometric features, particular amino acid compositions, and unique hydrogen bonding patterns. The present evidence suggests that turns also should be treated as discrete structural elements with characteristic properties.

Attention should be called to a paper by Lewis, *et al.*,[12] published after submission of this manuscript. These authors arrived at similar conclusions about the composition of turns.

Acknowledgment. Helpful discussions with Professor Walter Kauzmann and his research group and Professors V. Sasisekharan and Donald Wetlaufer are gratefully acknowledged. I wish to thank the National Institutes of Health for partial support of this work (GM 19269).

(14) For example, M. Schiffer and A. B. Edmundson, *Biophys. J.*, 7, 121 (1967), and ref 2.

(15) For bovine serum albumin (~600 residues), we estimate the number of turns as $N^{2/3} - 1$ or ~70. Approximately 50–70 ureas are thought to be required to half-denature BSA.[16]

(16) J. A. Gordon and J. Warren, *J. Biol. Chem.*, **243**, 5663 (1968), and I. D. Kuntz and T. S. Brassfield, *Arch. Biochem. Biophys.*, **142**, 660 (1971).