Request # 18652545 **NOV 28, 2005**

Ariel To: 64.40.17.85/140.163.217.217

Memorial Sloan-Kettering Cancer Center
Medical Library Nathan Cummings Center (METRO #146)
1275 York Avenue
New York, NY 10021

DOCLINE: Journal Copy Epayment

Title:	Nature.
Title Abbrev:	Nature
Citation:	1979 Nov 1;282(5734):109-11
Article:	Antiparallel and parallel beta-strands differ in a
Author:	Lifson S; Sander C
NLM Unique ID:	0410462 Verify: PubMed
PubMed UI:	503185
ISSN:	0028-0836 (Print) 1476-4687 (Electronic)
Publisher:	Nature Publishing Group, Basingstoke :
Copyright:	Copyright Compliance Law
Authorization:	A Artale
Need By:	DEC 01, 2005
Maximum Cost:	**$25.00**
Patron Name:	Gangi-Dino, Rita
Referral Reason:	Lacking
Library Groups:	BQSIMB,EFTS,RESOURCE,METRO
Phone:	1.212.639-7441
Fax:	1.646.422-2316
Email:	ill@mskcc.org
Comments:	**Please, we prefer as PDF or Ariel. Thank you! EFTS, METRO, BQSI.**
Routing Reason:	Routed to PAUPIT in Serial Routing - cell 1
Received:	Nov 28, 2005 (01:58 PM EST)
Lender:	University of Pittsburgh/ Pittsburgh/ PA USA (PAUPIT)

Bill to: NYUMSK

Memorial Sloan-Kettering Cancer Center
Medical Library Nathan Cummings Center
1275 York Avenue
New York, NY 10021

and the third letter of the codon (C or U). This also indicates, in accord with biochemical observations, that the codon recognition properties of Q are not significantly different from those of G (ref. 1 and S. Noguchi and S.N., unpublished). The cyclopentene ring is puckered; C(3″)—C(2″)=C(1″)—C(5″) is coplanar but C(4″) is displaced by 0.5 Å from this plane on the same side as the NH_2^+ group, possibly due to the existence of a diol {O(4″)H and O(5″)H} in the *cis* configuration. This puckered form is stabilised by the equatorial orientation of the large substituent ($NH_2^+—CH_2—\cdots$).

The atomic coordinates of yeast $tRNA^{Phe}$ (ref. 11) were used for the model-building study on the spatial arrangement of the modified nucleoside Q in the anticodon loop of tRNA. The first letter of the anticodon, Gm(34), was replaced by Q with the conformation found in the present study on the crystal of pQ. The second letter of the anticodon, adenine(35), was replaced by uracil, but the third letter, adenine(36), was retained. This sequence QUA corresponds to that in the anticodon of the $tRNA^{Tyr}$ of *E. coli*. Figure 3 shows a stereoscopic view of the three-dimensional structure involving U(33)—Q(34)—U(35)—A(36). It may be seen that Q is now incorporated into the first position of the anticodon without any steric repulsion from other parts of the anticodon loop. The cyclopentenediol group of Q is exposed out of the anticodon, in contrast to the short group of 2-thiouridine derivatives, which is more or less buried in the anticodon loop[13,14].

In addition to the intra-unit hydrogen bond with O(6), the NH_2^+ group of Q may also be involved in a hydrogen bonding interaction with ribose O(2′) of the preceding U(33); the distance between the imino nitrogen and ribose oxygen is as short as 3 Å and N—H···O is nearly linear. The side group of Q(34) is thus fixed by these two hydrogen bonds, and the *cis*-diol group is orientated out to the open space, away from the base-pairing moiety of the anticodon.

The biochemical function of Q in protein biosynthesis is not yet understood. Q has a wobbling property not significantly different from that of G. Extensive studies on the properties of $tRNA^{Tyr}$ containing G in place of Q, isolated from Q-deficient mutants of *E. coli* have not revealed any differences between Q and G with respect to binding of $tRNA^{Tyr}$ to ribosomes or aminoacylation (S. Noguchi and S. N., unpublished). Two species of rabbit reticulocyte $tRNA^{His}$ containing either Q or G were used with equal efficiency in haemoglobin biosynthesis[19]. These observations are consistent with the conformation of Q found in the present study by X-ray diffraction analysis. This hypermodified nucleoside Q is widely found in tRNAs of various organisms and thus is expected to be essential for some biological process other than protein biosynthesis. Q in *Drosophila* tRNA is suggested to have a role in cell differentiation or expression of a particular enzymatic activity[5,20]. In tumour cells, under-modified tRNA containing guanosine in place of Q appears specifically[21].

In conclusion, the conformational characteristics of the modified nucleoside Q appear to bring the cyclopentenediol group out of the anticodon loop, where it does not interfere with recognition of the codon of mRNA. This bulky side chain with its unique structure, charge and reactive group constitutes a suitable site for specific interaction with protein or other components. Thus the exposed cyclopentene ring (and orientation of the *cis*-diol group) of Q is possibly important for some biochemical function of tRNA other than protein biosynthesis.

Received 18 June; accepted 9 August 1979.

1. Harada, F. & Nishimura, S. *Biochemistry* **11**, 301–308 (1972).
2. Kasai, H. *et al. Biochemistry* **14**, 4198–4208 (1975).
3. Kasai, H. *et al. J. Am. chem. Soc.* **98**, 5044–5046 (1976).
4. Kasai, H., Kuchino, Y., Nihei, K. & Nishimura, S. *Nucleic Acids Res.* **2**, 1931–1939 (1975).
5. White, B. N., Tener, G. M., Holden, J. & Suzuki, D. T. *J. molec. Biol.* **74**, 635–651 (1973).
6. Ohgi, T., Goto, T., Kasai, H. & Nishimura, S. *Tetrahedron Lett.* 367–370 (1976).
7. McCloskey, J. A. & Nishimura, S. *Accts Chem. Res.* **10**, 403–410 (1977).
8. Kim, S.-H. *et al. Science* **185**, 435–440 (1974).
9. Robertus, J. D. *et al. Nature* **250**, 546–551 (1974).
10. Stout, C. D. *et al. Nucleic Acids Res.* **3**, 1111–1123 (1976).
11. Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. & Kim, S.-H. *J. molec. Biol.* **123**, 607–630 (1978).
12. Parthasarathy, R., Ohrt, J. M. & Chheda, G. B. *Biochemistry* **16**, 4999–5008 (1977).
13. Hillen, W., Egert, E., Lindner, H. J. & Gassen, H. G. *FEBS Lett.* **94**, 361–364 (1978).
14. Hillen, W., Egert, E., Lindner, H. J. & Gassen, H. G. *Biochemistry* **17**, 5314–5320 (1978).
15. Kasai, H., Kuchino, Y., Nihei, K. & Nishimura, S. *Nucleic Acids Res.* **2**, 1931–1939 (1975).
16. Germain, G., Main, P. & Woolfson, M. M. *Acta crystallogr.* **A27**, 368–376 (1971).
17. Ohgi, T., Kondo, T. & Goto, T. *J. Am. chem. Soc.* **101**, 3629–3633 (1979).
18. Voet, D. & Rich, A. *Progr. Nucleic Acid Res. molec. Biol.* **10**, 183–265 (1970).
19. McNamara, A. L. & Smith, D. W. *J. biol. Chem.* **253**, 5964–5970 (1978).
20. Jacobson, K. B. *Nucleic Acids Res.* **5**, 2391–2404 (1978).
21. Okada, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 4247–4251 (1978).

Antiparallel and parallel β-strands differ in amino acid residue preferences

Shneior Lifson & Christian Sander

Chemical Physics Department, Weizmann Institute of Science, Rehovot, Israel
and
Institut für Theoretische Physik, Heidelberg, FRG

A β-strand is a particular type of extended sequence of amino acid residues, an element of secondary structure of proteins. β-sheets are an assembly of strands, often bringing together parts of the protein which are separated along the backbone. As such, β-sheets are an element of tertiary structure. Parallel (β_P) and antiparallel (β_A) arrangements of strands in a sheet differ in the hydrogen bond pattern between strands, as shown schematically in Fig. 1, and in the type of chain connectivity they allow: short reverse turn connections for β_A and longer crossover connections for β_P (refs 1–3). Most present secondary structure prediction methods (for reviews refs 4–6) use a four-state distinction of secondary structure: α-helix, β-strand or extended, reverse turn, and 'random coil' (everything else). With a data base of 30–40 different protein structures, the conformational preferences for all amino acid residues in these four states seem to have converged[7]. However, the steadily increasing data base of structurally known proteins makes a refinement of the four-state description feasible. Although more refined classifications of conformational states based on finer subdivisions of (φ, ψ)-space have been made[8,9], we prefer making distinctions based on structural environment. Using a novel definition of β-sheet structure in terms of the tertiary structure juxtaposition of strands, we have analysed residue contacts in known β-sheets and report here secondary structure preferences for the 20 amino acids, separately for antiparallel and parallel arrangements of strands. The distinction between the two arrangements results in strikingly different and sharpened sets of preference parameters, including some of the largest values reported so far for any substructure. These results point the way towards a basic improvement of secondary structure predictions by further distinction of secondary structure elements according to tertiary structure environment. Beyond secondary structure prediction, the different preferences for β_A and β_P may aid in predicting the tertiary interaction between strands.

Our independent compilation of β_A- and β_P-residues in 30 proteins (a list of the proteins is given in ref. 10) defines β-sheets by their strand–strand interactions, rather than by single residue conformation, such as backbone (φ, ψ) angles. The protein coordinates are from ref. 11 (and updates as of August 1978). We identify two neighbouring antiparallel (or parallel) stretches of polypeptide chain (at least three residues each) as a pair of β_A (or β_P) strands (1) if pairs (i, j) of contacting residues have their side groups on the same side of the sheet (Fig. 1), and (2) if the C=O and N—H groups of the backbone are properly orientated[10]. Single strands are not counted. For regular stretches of

0028-0836/79/440109—03$01.00

Table 1 Amino acid residues in antiparallel and parallel β-strands

	Residue contact count			Frequency (%)				Conformational preference†							
	β_A	β_P	All β	β_A	β_P	All β	Global*	β_A	β_P	β_P/β_A	All β:	Here	Ref. 13	Ref. 14	Ref. 12
Gly	81	38	119	5.1	7.2	5.7	9.2	0.56	0.79	1.41		0.61	0.75	0.66	0.92
Pro	29	8	37	1.8	1.5	1.8	4.4	0.42	0.35	0.83		0.40	0.55	0.84	0.64
Asp	39	14	53	2.5	2.7	2.5	5.3	0.47	0.50	1.08		0.48	0.54	0.64	0.72
Glu	47	15	62	3.0	2.9	2.9	4.8	0.62	0.59	0.96		0.61	0.37	0.61	0.75
Ala	119	44	163	7.6	8.4	7.8	8.4	0.90	1.00	1.11		0.92	0.83	0.79	0.90
Asn	41	12	53	2.6	2.3	2.5	4.2	0.62	0.54	0.88		0.60	0.89	0.66	0.76
Gln	63	5	68	4.0	1.0	3.2	3.4	1.18	0.28	0.24		0.95	1.10	1.13	0.80
Ser	123	33	156	7.8	6.3	7.4	9.0	0.87	0.70	0.80		0.82	0.75	0.84	0.95
Thr	139	21	160	8.8	4.0	7.6	6.8	1.30	0.59	0.45		1.12	1.19	1.14	1.21
Lys	83	22	105	5.3	4.2	5.0	7.1	0.74	0.59	0.79		0.70	0.74	0.72	0.77
Arg	45	10	55	2.9	1.9	2.6	2.8	1.02	0.68	0.67		0.93	0.93	1.04	0.99
His	44	5	49	2.8	1.0	2.3	2.5	1.12	0.38	0.34		0.93	0.87	0.78	1.08
Val	188	108	296	11.9	20.5	14.1	7.8	1.53	2.63	1.72		1.81	1.70	1.97	1.49
Ile	112	63	175	7.1	12.0	8.3	4.6	1.54	2.60	1.69		1.81	1.60	1.95	1.45
Met	24	11	35	1.5	2.1	1.7	1.4	1.09	1.49	1.37		1.19	1.05	1.26	0.97
Cys	45	11	56	2.9	2.1	2.7	2.3	1.24	0.91	0.73		1.16	1.19	1.55	0.74
Leu	141	53	194	8.9	10.1	9.2	7.1	1.26	1.42	1.13		1.30	1.30	1.26	1.02
Phe	68	24	92	4.3	4.6	4.4	3.5	1.23	1.30	1.06		1.25	1.38	1.30	1.32
Tyr	98	21	119	6.2	4.0	5.7	3.7	1.68	1.08	0.64		1.53	1.47	1.49	1.25
Trp	47	8	55	3.0	1.5	2.6	1.7	1.75	0.89	0.51		1.54	1.37	0.90	1.14

The order of amino acids reflects their average structural similarity[17]. Amino acids which are observed to be more mutually exchangeable in evolution are nearer each other in this list.

* Numbers taken from ref. 11.

† The preference for $\beta_A(\beta_P)$ is a normalised frequency: the frequency in $\beta_A(\beta_P)$ divided by the global frequency. Note that the definitions of β-residues vary among different authors[7,13,15]. Furthermore, our preferences are for making β-sheet contacts, whereas those of the other authors are for occurrence in extended or β-strands.

β-strands (generally away from the edges of sheets) our identification of β-residues agrees with that of the crystallographers. As we define β-residues by their contact with neighbouring strands, we count residue contacts rather than residues in deriving the single residue frequencies. Thus, residues in strands at the edge of a sheet are counted once, residues in interior strands are counted twice, and residues making β_A contact on one side and β_P on the other are counted once for each type of contact.

The residue contact counts and residue contact frequencies are given in Table 1. β_A structures outweigh β_P by about three to one in the data base. The amino acid compositions of β_A and β_P are remarkably distinct. About one-third of all residue contacts in β_P are made by Val and Ile, the two non-polar side chains branched at C_β. Val (21%), Ile (12%), Leu (10%) and Ala (8%), with purely aliphatic side chains, together account for more than 50% of all contacts in β_P. In β_A six amino acids are needed to reach the 50% mark, including the polar side chains Thr and Ser, and the order of abundance (listed in decreasing order) is different: Val, Leu, Thr, Ser, Ala, Ile.

The 'conformational preference' parameters presented in Table 1 are defined as the ratios between the frequency of pair-contacts made by a residue in β-strands and the global frequency of occurrence of that residue anywhere. Values larger than 1.0 indicate preference of the particular residue for $\beta_A(\beta_P)$, values less than 1.0 indicate the opposite. Conformational preference parameters (derived from residue frequencies rather than from contact frequencies) have been used by many authors (for reviews see refs 4–6, 12) in the prediction of secondary structure (called 'potential', 'propensity', 'information measure'). The parameters given by Chou and Fasman[13], by Garnier *et al.*[14], and by Levitt[12] are included in Table 1 for comparison. There is overall agreement between the 'all β' (β_A and β_P together) preferences in the different studies. The differences, for example, for Pro, Glu or Trp, arise because the studies use not only different sets of proteins as their data base, but also different ways of identifying β-residues[7,13,15] and, in our case, different weights for interior and edge strands.

Comparison of conformational preferences between β_A (antiparallel) and β_P (parallel) shows: (1) the β_P structure is more selective: there are more extreme large and small values in β_P (standard deviation 0.66 for β_P but only 0.40 for β_A; high and low: 2.63 and 0.28 for β_P but only 1.75 and 0.42 for β_A), and, in β_P fewer residues are preferred (only 6, compared with 12 in β_A) and for these few the preference is stronger (except Tyr). In fact, the 20 amino acids can be divided into three groups (Table

Table 2 Conformational classification of β-sheet residues

		Conformational preference
Favourable in both β_A and β_P	Val Ile Met Leu Phe Tyr	>1.0 in β_A and β_P
Favourable in β_A but unfavourable in β_P	Gln Thr Arg His Cys Trp	>1.0 in β_A and <1.0 in β_P
Unfavourable in both β_A and β_P	Gly Pro Asp Glu Ala Asn Ser Lys	≤1.0 in β_A and β_P
Best sheet 'makers' in β_A	Trp > Tyr > Val > Ile	>1.5
Best sheet 'makers' in β_P	Val > Ile > Met	>1.5
Worst sheet 'breakers' in β_A	Pro < Asp < Gly < Glu < Asn	≤1/1.5
Worst sheet 'breakers' in β_P	Gln < Pro < His < Asp < Glu < Thr < Lys	≤1/1.5

2)—favourable in both β_P and β_A, favourable in β_A but not in β_P, and unfavourable in both β_A and β_P. (2) β_P 'abhors' polar and charged groups and favours hydrophobic side chains more strongly than β_A: only the purely hydrophobic residues Val, Ile, Leu, Met, Phe are significantly preferred (>1.2) in β_P, whereas in β_A also Thr, Cys, Tyr and Trp, all containing polar atoms, show values above 1.2. (3) Val and Ile are particularly favoured in β_P, with the second largest secondary structure enhancement values reported so far—2.6. Only the value for Pro at the second position in reverse turns[16] is larger—3.4. In β_A the preferences for Tyr and Trp (1.7) are exceptionally large. Note that both have a large non-polar ring in combination with a polar group. (4) The differences between β_A and β_P are particularly large for Gln, Thr, His and Trp (favourable in β_A and unfavourable in β_P); Gln is less favoured in β_P by about a factor of 4.

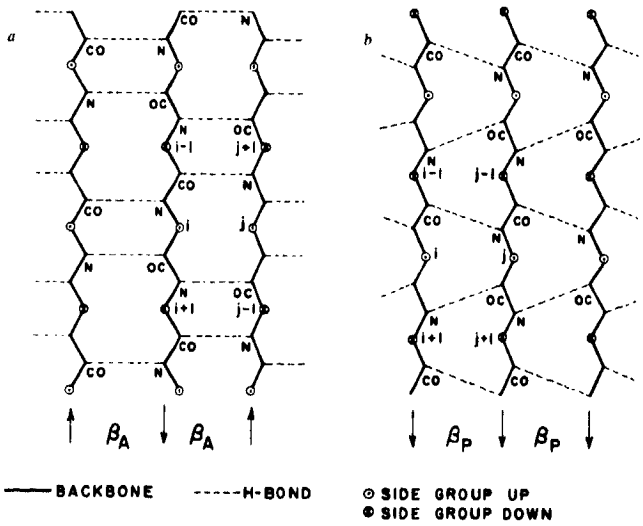

Fig. 1 A schematic drawing of antiparallel (*a*) and parallel (*b*) β-strands. The arrows indicate the N-terminal to C-terminal direction of the polypeptide backbone. Although the (ϕ, ψ) backbone angles are similar for β_A and β_P, the arrangement of the strands and the hydrogen-bonding patterns are fundamentally different. This difference is also reflected in different amino acid preferences.

The results presented here are potentially useful in two ways. On the one hand, they help to pose more specific questions, by focusing attention on those residues which have a special structural role. In this, residues which are both abundant and preferred are the most interesting. For example, one may ask: can Val and Ile serve as nucleation points for β_P-strands by restricting the backbone conformation with their side chains? Do they pack particularly well in the environment of the crossover connection between β_P-strands? Do they interact preferably with each other in forming clusters on the surface of β-sheets? The aim in investigating this type of question is to understand the physical basis of the statistical preferences.

On the other hand, the conformational preferences can be incorporated into structure prediction schemes, allowing distinct predictions for antiparallel and parallel β-stands. However, when this is done the limitations inherent in all statistical approaches to protein structure must be kept in mind: proteins are biologically evolved molecules with a particular structure tailored to a particular function. Only average properties of protein structures can be understood by statistical methods, but not the highly individual character of each protein.

Thus, the main conclusions regarding protein architecture are: (1) antiparallel and parallel β-strands are as distinct in their amino acid preferences as are α-helices and β-strands; (2) parallel β-strands impose much more severe constraints on their amino acid content than antiparallel β-strands or α-helices; (3) Val and Ile have an outstanding role in parallel strands. In a related paper[10], we analyse interstrand pair recognition of amino acid residues and its role in the tertiary structure assembly of strands into sheets.

As more protein structure data become available, further distinctions of secondary structure elements according to the type of tertiary contacts should be made. For example, one can distinguish different hydrogen-bonding positions in β-sheets, solvent-exposed and interior faces of sheets or helices, segments in tertiary contact with sheets compared with those in contact with helices. Such distinctions are likely to lead to more clearcut statistical preferences, and also serve as a starting point for predicting tertiary structure.

We thank John Moult, Barry Robson, Charlotte Schellmann and Joel Sussman for critical suggestions and Richard Feldman for supplying protein coordinates.

Received 25 May; accepted 30 August 1979.

1. Levitt, M. & Chothia, C. *Nature* **261**, 552-558 (1976).
2. Richardson, J. S. *Nature* **268**, 495-500 (1977).
3. Sternberg, M. J. E. & Thornton, J. M. *J. molec. Biol.* **115**, 1-17 (1977).
4. Chou, P. Y. & Fasman, F. D. *A. Rev. Biochem.* **47**, 251-276 (1978).
5. Sternberg, M. J. E. & Thornton, J. M. *Nature* **271**, 15-20 (1978).
6. Schulz, G. E. & Schirmer, R. H. *Principles of Protein Structure* (Springer, New York, 1979).
7. Robson, B. & Suzuki, E. *J. molec. Biol.* **107**, 327-356 (1976).
8. Robson, B. & Pain, R. H. *Biochem. J.* **141**, 869-882 (1974).
9. Burgess, A. W., Ponnuswamy, P. K. & Scheraga, H. A. *Israel J. Chem.* **12**, 239-286 (1975).
10. Lifson, S. & Sander, C. *J. molec. Biol.* (submitted).
11. Feldman, R. J. *AMSOM Atlas of Molecular Structures on Microfiche* (U.S. NIH, 1976).
12. Levitt, M. *Biochemistry* **17**, 4277-4285 (1978).
13. Chou, P. Y. & Fasman, G. D. *Biochemistry* **13**, 222-245 (1974).
14. Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120**, 97-120 (1978).
15. Levitt, M. & Greer, J. *J. molec. Biol.* **114**, 181-239 (1977).
16. Chou, P. Y. & Fasman, G. D. *J. molec. Biol.* **115**, 135-175 (1977).
17. Sander, C. & Schulz, G. E. *J. molec. Evolut.* (in the press).

Errata

The letter 'Local destabilisation of a DNA double helix by a T–T wobble pair', *Nature* **281**, 235, was printed with errors in the author line. This line should read:

A. G. Cornelis Haasnoot, Jeroen H. J. den Hartog, Jan F. M. de Rooij, Jacques H. van Boom & Cornelis Altona.

In the paper 'Isolation, cloning and sequence analysis of the cDNA for the α-subunit of the human chorionic gonadotropin', *Nature* **281**, 351, figures 1 and 2 should be transposed. The legends remain as published.

In the paper 'Unmasking of fetal determinants on adult bone marrow cells', *Nature* **281**, 484, the second author should be spelt B. M. Sheinberg.