Cotranslational protein folding—fact or fiction?

Charlotte M. Deane[1], Mingqiang Dong[2], Fabien P.E. Huard[2], Braddon K. Lance[2] and Graham R. Wood[2],*

[1]Department of Statistics, 1 South Park Road, Oxford University, Oxford OX1 3TG, UK and [2]Department of Statistics, Macquarie University, NSW 2109, Australia

ABSTRACT

Motivation: Experimentalists have amassed extensive evidence over the past four decades that proteins appear to fold during production by the ribosome. Protein structure prediction methods, however, do not incorporate this property of folding. A thorough study to find the fingerprint of such sequential folding is the first step towards using it in folding algorithms, so assisting structure prediction.

Results: We explore computationally the existence of evidence for cotranslational folding, based on large sets of experimentally determined structures in the PDB. Our perspective is that cotranslational folding is the norm, but that the effect is masked in most classes. We show that it is most evident in α/β proteins, confirming recent findings. We also find mild evidence that older proteins may fold cotranslationally. A tool is provided for determining, within a protein, where cotranslation is most evident.

Contact: gwood@efs.mq.edu.au

1 INTRODUCTION

It is well known that proteins are manufactured sequentially in the ribosome; whether they fold as they are manufactured is very much less well understood. In this article, we look for computational evidence of cotranslational folding in a large set of proteins and draw some broad conclusions.

Why is this question of importance? Accurate protein fold prediction remains one of the central scientific challenges of our time. To date, the mechanism of cotranslational folding, and in train, its structural consequences, has not been introduced into fold prediction. If it does occur and can be usefully incorporated, then it would provide a contribution to this challenging problem. Building cotranslational behaviour into algorithms may lead to greater efficiency and robustness of prediction. For example, it may provide a more direct energy path to the final fold, so be more efficient. It may also be the case that cotranslational algorithms have to search fewer possible paths, reflecting the known result that *in vivo* folding from the ribosome is far faster than in *vitro* folding from a fully extended denatured starting point (Baldwin, 1999).

Building on this first question, there are several further questions to be answered, listed now:

- If a protein folds as it emerges, does this sequential production influence the final fold?

- If it does influence the final fold, in what way (for example, asymmetry of the fold or perhaps progress to a local energy minimum)?

- If there is asymmetry, how is this seen (for example, in that secondary structures are more readily found at the N-terminus)?

- If progress is to a local energy minimum, how do we know this?

- Can incorporation of sequential folding into prediction algorithms be useful?

In addition to the central question of the existence of cotranslational folding, we address here aspects of the first three of these additional questions. An algorithm carrying out sequential folding will be described elsewhere.

The perspective emerging in this article is that while cotranslational folding is the norm, it is often masked by other activities. For instance the ribosome tunnel is known to provide physical constraints (Jenni and Ban, 2003; Nakatogawa and Ito, 2002); this exit tunnel for the protein can favour α helices (Ziv *et al.*, 2005). Chaperones were also observed interacting cotranslationally with nascent peptides (Srikakulam and Winkelmann, 2003; Ullers *et al.*, 2004). It also appears that for ancient proteins there may be more evidence of cotranslational folding.

We also consider that most proteins fold to the global energy minimum, but only because nature has selected proteins which can fold cotranslationally to the global energy minimum (Alexandrov, 1993).

The literature provides ample evidence for sequential folding. As early as 1967, Phillips revealed the structure of the hen egg-white lysozyme molecule and concluded 'the last 20 residues are folded around the globular unit built up by the first 40' (Phillips, 1967). Kolb provided an excellent summary of the experimental evidence that protein folding occurs during translation (Kolb, 2001). Alexandrov, on the other hand, put the arguments for and against cotranslational folding (Alexandrov, 1993). He provided early evidence, based on a small set of 170 proteins, that residues are in general closer to previously synthesized residues than those synthesized later. He also showed for this protein set that the N-terminus was more compact than the C-terminus. Elcock has carried out molecular simulations of cotranslational folding, examining the effects of the ribosome tunnel as well as the slow extrusion of amino acids (Elcock, 2006). More recent work (Taylor, 2006) considered topological accessibility (the ability of a protein to fold from a given residue as a starting point) and found

---

*To whom correspondence should be addressed.

evidence for cotranslation in α/β proteins and possibly ancient proteins. He postulated 'If these ancient proteins had to fold unassisted, it is possible that they had a bias to fold their amino segments first as they were synthesized'.

We discuss the data sources used and then detail two distance measures and a measure of 'previous contact', each used to detect evidence for cotranslational folding. All three measures consistently indicate such folding for α/β proteins.

2 METHODS

Simple HP-lattice protein models were folded cotranslationally in Huard *et al.* (2006). Such cotranslationally folded models favoured local contacts and sometimes produced a final fold that was not in the lowest possible energy state. These consequences of cotranslation were in accordance with theoretical (Morrissey *et al.*, 2004) and experimental (Baker, 1998; Baskakov *et al.*, 2001; Sohl *et al.*, 1998) findings. A prediction was also made that the N-terminus region would be more likely to be buried than the C-terminus region. Motivated by this we investigate evidence of cotranslational folding of real proteins in this article using two measures of end region burial and a measure of extent of 'previous contact'.

If a protein folds as it emerges from the ribosome, the N-terminus is expected to be more buried than the C-terminus, in the final fold. A number of studies, however, have shown that the terminal residues of proteins tend to be located on the surface (for example, Jacob and Unger, 2006). We overcome this phenomenon by 'snipping' off the ends and using distance measures which work with segments of the chain that are an appropriate number of residues from the extremes.

We look at a ratio of minimum distances of near-terminal segments to the centroid ($R_{\min}$) and a proportional distance ($P_{\min}$) from the N-terminus to the residue closest to the centroid. For a protein folded sequentially, we would expect a given residue to be more in contact with residues already folded. Envisage wrapping wool into a ball, constantly changing the direction; it is more likely that the wool you are currently wrapping is in contact with wool that has already been wrapped than wool that is yet to be wrapped. We develop a measure of such 'previous contact'.

We shortly present a detailed description of the measures.

2.1 Data

We use 'culledPDB' sets as presented on the Dunbrack Lab website (Wang and Dunbrack, 2003). We use three sets of proteins, extracted using the following criteria:

Set label	Number of PDB files	% Sequence identity cutoff	Resolution cutoff (Å)	R-factor cutoff
1	1122	20	1.6	0.25
2	3585	30	2.0	0.25
3	4298	30	2.2	1.0

2.2 Measures of closeness to the centroid

We represent a residue by a single point in space, the coordinates of the C_β atom of the protein (or C_α if the amino acid is glycine). We determine the centroid (the mean location of residues) C of the protein.

2.2.1 Ratio (R_{min}) of minimum distances of near-terminal segments to the centroid In this measure, the first 10 and last 10 residues are removed from consideration, to dampen any effect caused by the tendency for ends to be located on the surface. We then consider the neighbouring 10 residues at each end. We calculate the minimum distance of residues in the near-N-terminus segment to the centroid as

$$D_{\min}^N = \min\{d(R_i,C) : i = 11,\ldots,20\}$$

where $d(R_i,C)$ is the Euclidean distance between the ith residue R_i (i from 11 to 20) from the N-terminus and the centroid C of the protein. In the same manner, we determine $D_{\min}^C$, the minimum distance of a residue in the near-C-terminus segment to the centroid. We then form the ratio

$$R_{\min} = \frac{D_{\min}^C}{D_{\min}^N}$$

2.2.2 Proportion of length (P_{min}) until closest to the centroid We determine the residue i along the chain, measured from the N-terminus, which is closest to the centroid and define

$$P_{\min} = \frac{i}{n}$$

where n is the number of residues in the protein.

We will consider these measures, and the following one, for the main SCOP classes.

2.3 Measures of previous contact

We develop here a measure of previous contact and compare this value taken from the N-terminus with that taken from the C-terminus.

A previous contact with a residue at position $i \geq 7$ from the N-terminus is deemed to occur when a residue numbered from 1 to $i-6$ comes within $13\,\text{Å}$ of residue i. The five closest residues towards the N-terminus are eliminated from the pool of contact candidates, since such contacts are generally due to proximity rather than the folding process. We let A_i^N denote the actual number of such previous contacts and $P_i^N = i - 6$ be the potential number of such contacts. We form the ratio of actual contacts to potential contacts for each residue, from the seventh onwards, and compare the corresponding ratios formed from the C-terminus. Note that $P_i^C = n - i - 5$.

2.3.1 Sum of the ratios (SR) To retain sensitivity of the measure we choose to 'compare as we go', forming an actual to potential proportion from each end and immediately comparing them. A technical difficulty arises: if there are no actual contacts then division by zero would occur. We remedy this by grouping the residues until both ratios are non-zero. Formally, a group is defined as follows: we parse the chain simultaneously and symmetrically from the N- and C-termini. If both A_j^N and A_j^C are greater than zero, then a group is constituted. Else we keep parsing along the chain and sum the actual contacts until the sum of all A_j^N values and the sum of all A_j^C values are greater than zero. Our convention is that i indexes the groups and that the ith group contains J_i residues. We thus define (the average of) the sum of ratios as

$$SR = \frac{1}{I}\sum_{i=1}^{I} \left.\frac{\sum_{j=1}^{J_i} A_j^N}{\sum_{j=1}^{J_i} A_j^C}\right/\frac{\sum_{j=1}^{J_i} P_j^N}{\sum_{j=1}^{J_i} P_j^C} = \frac{1}{I}\sum_{i=1}^{I} \frac{\sum_{j=1}^{J_i} A_j^N}{\sum_{j=1}^{J_i} A_j^C}$$

where I is the number of groups (so index i runs from 1 to I).

2.3.2 Sum of the logarithmic ratios (SLR) We define a logarithmic version, mapping the positive values of *SR* onto the real line, so making visualization of the results easier, by taking (the average of) the sum of the log-transformed ratios,

$$\mathrm{SLR} = \frac{1}{I}\sum_{i=1}^{I}\log\left(\frac{\sum_{j=1}^{J_i} A_j^N}{\sum_{j=1}^{J_i} A_j^C}\right) = \frac{1}{I}\sum_{i=1}^{I}\left(\log\left(\sum_{j=1}^{J_i} A_j^N\right) - \log\left(\sum_{j=1}^{J_i} A_j^C\right)\right)$$

2.4 Statistical Analysis

If no cotranslation is evident in the proteins, $R_{\min}$ and $P_{\min}$ are expected to be centred on 1 and 0.5, respectively. Each of these measures is transformed to a binary variable, $R'_{\min}$ and $P'_{\min}$, respectively. The value '0' is assigned to $R'_{\min}$ when $R_{\min}$ is less than one, and the value '1' is assigned when $R_{\min}$ is greater than or equal to one. The value '0' is assigned to $P'_{\min}$ when $P_{\min}$ is greater than or equal to 0.5, and the value '1' is assigned to $P'_{\min}$ when $P_{\min}$ is less than to 0.5.

As described earlier, under cotranslation we expect the value of $R_{\min}$ to be greater than one. Furthermore, under cotranslation we expect the residue associated with $P_{\min}$ to be closer to the N-terminus. We evaluated the null hypothesis that each of $\mathrm{Prob}(R'_{\min}=0)=0.5$ and $\mathrm{Prob}(P'_{\min}=0)=0.5$, versus the alternative of being less than 0.5, by comparing the observed frequencies to a binomial distribution (with parameters *n* and $P=0.5$).

Results for each SCOP class were examined for evidence of cotranslation, as well as age categories within each SCOP class.

The relative age of protein structures calculated in Winstanley *et al.* (2005) were used to create a categorical variable with three levels, 'Old' comprising all proteins with a relative age of one, 'Middle' comprising all proteins with relative age in the interval [0.5,1) and 'Young' comprising all proteins with relative age in the interval [0,0.5).

Under the null hypothesis of no cotranslation, SLR is centred on zero. Under cotranslation, we expect this to be greater than zero. A one-sided *t*-test was used to assess whether the SLR mean for proteins within each SCOP class was greater than zero.

3 RESULTS

3.1 Structural evidence for cotranslation within the four major SCOP classes

The distributions of each of the measures $\log(R_{\min})$, $P_{\min}$ and SLR, stratified by major SCOP classes, are plotted in Figure 1. Each measure provides strong evidence for cotranslational folding in α/β proteins, as shown in Table 1. There is significant evidence for cotranslation in the α/β class using both the $R_{\min}$ measure ($P<0.0002$ for the small dataset and $P<1\times10^{-20}$ for the larger datasets), and the $P_{\min}$ measure ($P<0.0006$ for the small dataset, and $P<1\times10^{-9}$ for the larger datasets). There was also strong evidence that SLR is greater than zero in the α/β class ($P<1\times10^{-10}$ for each dataset). There was no structural evidence for cotranslational folding within any of the SCOP classes α, β or $\alpha+\beta$ ($P>0.1$ for all tests in Table 1; also see Fig. 1).

3.2 Evidence of cotranslational folding in SCOP-by-age classes

To establish the role of protein age in cotranslation, each of the SCOP classes was stratified by age, as shown in Table 2. We comment now on the results, first for $R_{\min}$ and $P_{\min}$ then
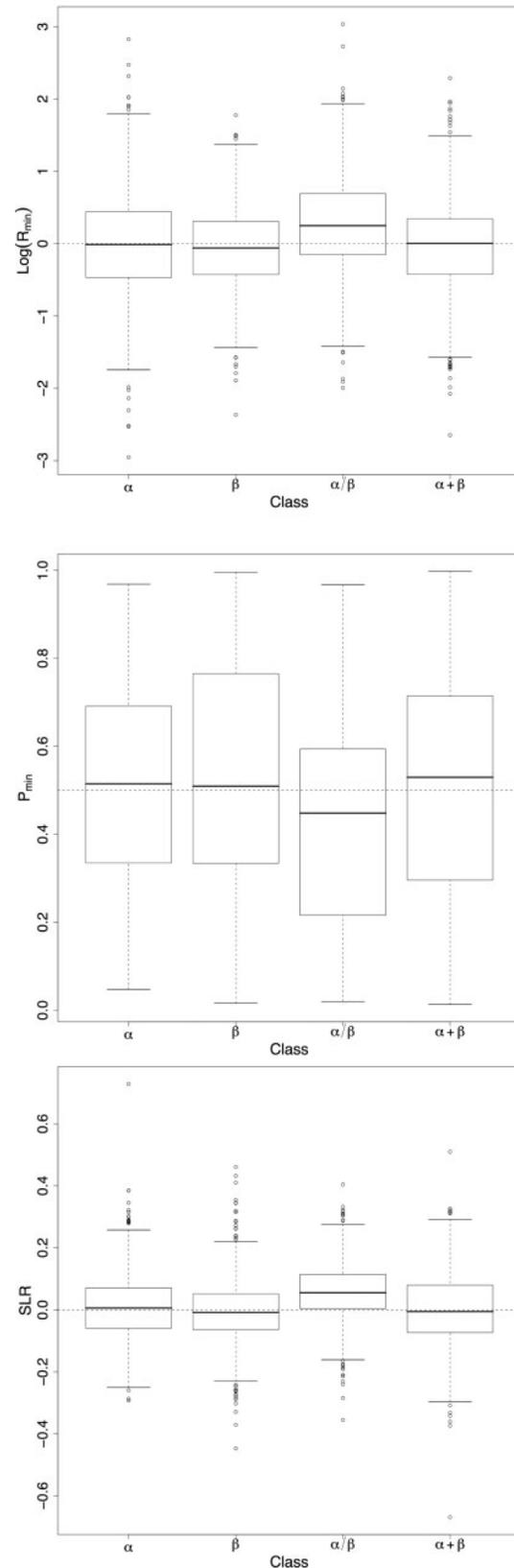
Fig. 1. Boxplots of $\log(R_{\min})$ (top), $P_{\min}$ (middle) and SLR (bottom), each for the four major SCOP classes.

Table 1. Evidence for cotranslational folding in major SCOP classes, using three measures of cotranslation

Class	Set	Size	$R_{\min}$	$P_{\min}$	SLR
α	20%, 1.6 Å	109	ns	ns	ns
	30%, 2.0 Å	397	ns	ns	ns
	30%, 2.2 Å	510	ns	ns	ns
β	20%, 1.6 Å	146	ns	ns	ns
	30%, 2.0 Å	572	ns	ns	ns
	30%, 2.2 Å	648	ns	ns	ns
α/β	20%, 1.6 Å	215	***	***	***
	30%, 2.0 Å	820	***	***	***
	30%, 2.2 Å	961	***	***	***
$\alpha+\beta$	20%, 1.6 Å	166	ns	ns	ns
	30%, 2.0 Å	620	ns	ns	ns
	30%, 2.2 Å	739	ns	ns	ns

Significance code: '***' <0.001; ns >0.1.

for SLR. The old α/β class shows very strong evidence of cotranslational folding, using both $R_{\min}$ and $P_{\min}$. There is some evidence for cotranslation in the α/β class of proteins of middle age using the $P_{\min}$ measure ($P \approx 0.09$ in the larger datasets) and marginal evidence from $R_{\min}$. Small sample size prevented calculation of statistical significance in the young α/β class for the smaller datasets, though the largest dataset afforded some cotranslation evidence.

There was no structural evidence for cotranslational folding within any of the age categories within the α, β or $\alpha+\beta$ class of proteins, with $P>0.1$ for all tests using $R_{\min}$ and $P_{\min}$.

Evidence for cotranslational folding from the SLR measure is summarized in Table 2 for each of the α, β, α/β and $\alpha+\beta$ classes stratified by age.

The α/β class has SLR values significantly greater than zero, for each dataset in the old and middle age classifications (Table 2, Fig. 2), suggesting cotranslational folding is the norm for this class. The α class has SLR values significantly greater than zero in the oldest age category for the larger two datasets (Table 2). The observed difference, however, is marginal (Fig. 3).

None of the remaining mean SLR values was significantly greater than zero in any of the age categories in any of the remaining SCOP classes (Table 2).

3.3 Proteins exhibiting evidence of cotranslational folding

Within the α/β class, $P_{\min}$ is bimodally distributed (Fig. 4). A cutoff that satisfactorily separates the populations is 0.3. This cutoff was used to separate the α/β folds into two groups, as shown in Table 3. The folds within the α/β class above and below this cutoff show a significant association ($P<1 \times 10^{-5}$ from Fisher's exact test using Monte Carlo simulation), with SCOP folds 108, 30 and 26 containing proteins with very strong evidence for cotranslational folding, and SCOP folds 52, 67 and 68 containing proteins that do not show strong structural evidence for cotranslation. (For the SCOP names of these folds, see Table 3.)

Table 2. Significance of the three estimates of cotranslation for each of the data subsets

Class	Age	Set	Size	$R_{\min}$	$P_{\min}$	SLR
α	Old	20%, 1.6 Å	50	ns	ns	ns
		30%, 2.0 Å	86	ns	ns	**
		30%, 2.2 Å	261	ns	ns	**
	Mid	20%, 1.6 Å	32	ns	ns	ns
		30%, 2.0 Å	106	ns	ns	ns
		30%, 2.2 Å	135	ns	ns	ns
	Young	20%, 1.6 Å	27	ns	ns	ns
		30%, 2.0 Å	205	ns	ns	ns
		30%, 2.2 Å	114	ns	ns	ns
β	Old	20%, 1.6 Å	80	ns	ns	ns
		30%, 2.0 Å	297	ns	ns	ns
		30%, 2.2 Å	334	ns	ns	ns
	Mid	20%, 1.6 Å	42	ns	ns	ns
		30%, 2.0 Å	171	ns	ns	ns
		30%, 2.2 Å	197	ns	ns	ns
	Young	20%, 1.6 Å	24	ns	ns	ns
		30%, 2.0 Å	104	ns	ns	ns
		30%, 2.2 Å	117	ns	ns	ns
α/β	Old	20%, 1.6 Å	205	***	***	***
		30%, 2.0 Å	775	***	***	***
		30%, 2.2 Å	908	***	***	***
	Mid	20%, 1.6 Å	9	*	**	**
		30%, 2.0 Å	39	ns	.	***
		30%, 2.2 Å	44	ns	.	***
	Young	20%, 1.6 Å	1	x	x	x
		30%, 2.0 Å	6	x	x	x
		30%, 2.2 Å	9	*	ns	***
$\alpha+\beta$	Old	20%, 1.6 Å	101	ns	ns	ns
		30%, 2.0 Å	399	ns	ns	ns
		30%, 2.2 Å	488	ns	ns	ns
	Mid	20%, 1.6 Å	46	ns	ns	ns
		30%, 2.0 Å	148	ns	ns	ns
		30%, 2.2 Å	169	ns	ns	ns
	Young	20%, 1.6 Å	19	ns	ns	ns
		30%, 2.0 Å	73	ns	ns	ns
		30%, 2.2 Å	82	ns	ns	ns

Significance codes: '***' <0.001; '**' <0.01; '*' <0.05; '·' <0.1; ns >0.1; 'x' indicates a total sample size less than nine, so no P-value was calculated.

3.4 Detecting cotranslational activity within a protein

The accumulating partial sums within the SLR measure can be used to detect regions within a protein showing evidence of cotranslational folding. An example of this cumulative SLR measure is shown for protein 1ejx, within the α/β class, in Figure 5. The cumulative measure shows a large increase over the first 100 residues. This indicates that the structural evidence for cotranslational folding in this protein resides in the first 100 residues.

3.5 Relationship between the measures and protein length

Plots in Figure 6 show the relationship between each of $R_{\min}$, $P_{\min}$, SLR and the number of residues in the protein,
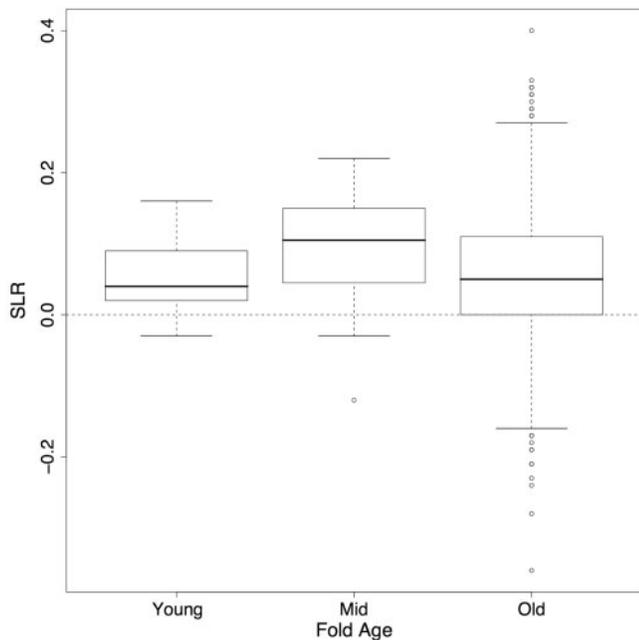
Fig. 2. α/β proteins have mean SLR values in each age category that are significantly greater than zero. The largest dataset (30%, 2.2 Å) was used to create this graphic.
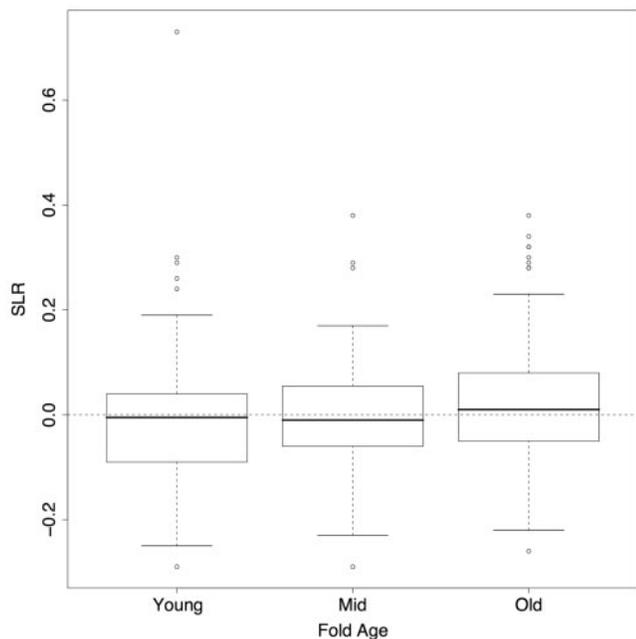

Fig. 3. α class proteins have mean SLR values significantly greater than zero for the 'old' age category. The largest dataset (30%, 2.2 Å) was again used to create this graphic.

summarized using a lowess smoother. There is no clear relationship between R_{min} and protein length. Each of P_{min} and SLR, however, surprisingly show decreasing evidence for cotranslationality as protein length increases. The waning of the measure with length may be due to a 'washing out' effect of
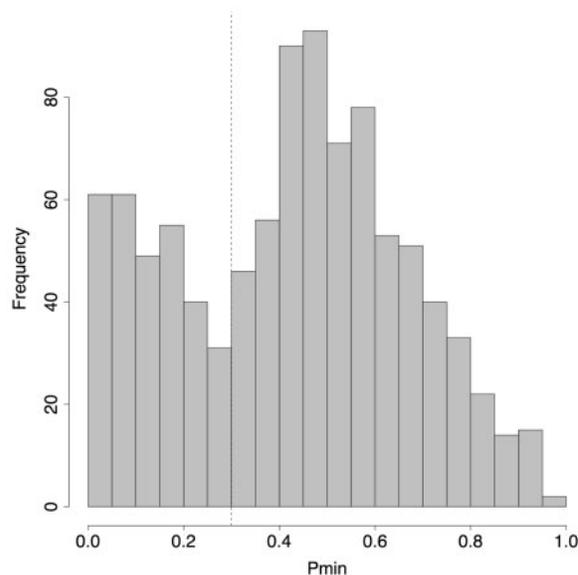

Fig. 4. Histogram of P_{min} for the α/β class, showing a bimodal distribution. The cutoff at 0.3 is marked by the dashed line. The largest dataset (30%, 2.2 Å) was again used here.

Table 3. α/β folds with 10 or more structures and their relationship to P_{min} classes

Fold name and number	$P_{min} > 0.3$	$P_{min} \leq 0.3$
Tryptophan synthase beta subunit-like PLP-dependent enzymes (108)	0	13
Cryptochrome/photolyase, N-terminal domain (30)	1	9
Methylglyoxal synthase-like (26)	6	18
FAD/NAD(P)-binding domain (3)	5	8
Alpha/beta-Hydrolases (93)	5	6
Thiamin diphosphate-binding fold (THDP-binding) (47)	16	18
Nucleoside hydrolase (94)	13	13
Indigoidine synthase A-like (55)	22	17
Tubulin nucleotide-binding domain-like (37)	41	29
IIA domain of mannose transporter, IIA-Man (72)	8	3
NAD(P)-binding Rossmann-fold domains (2)	58	20
TK C-terminal domain-like (66)	25	6
Arginase/deacetylase (56)	12	2
Hypothetical protein MT938 (MTH938) (69)	32	5
Ribosomal protein L13 (23)	41	6
Phosphotyrosine protein phosphatases I-like (61)	11	1
ClpP/crotonase (14)	13	1
Putative lysine decarboxylase (52)	24	1
Pyruvate kinase C-terminal domain-like (67)	23	0
Anticodon-binding domain-like (68)	17	0

Folds towards the top of the table exhibit greater evidence of cotranslation, while those towards the bottom of the table show little evidence.
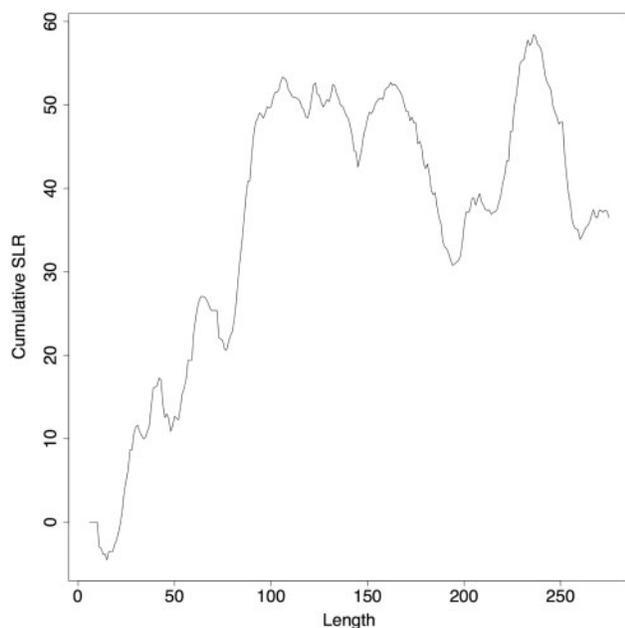
Fig. 5. The cumulative SLR score for protein 1ejx within the α/β class. The large SLR value is achieved within the first 100 residues.

length, namely that there is a bound to the amount of evidence to be found in each protein, and that it is therefore less easily seen in longer proteins.

4 DISCUSSION

This article summarizes the results of a first search of large sets of differing proteins for evidence of cotranslational folding. Three measures were used, a ratio of minimum distances of near N-terminus and near C-terminus segments to the centroid, the proportional distance from the N-terminus to the residue nearest to the centroid and a measure of 'previous contacts'. For all three measures, the SCOP α/β class stands out as that containing proteins exhibiting cotranslational folding (Table 1). There is also slight evidence (in the SCOP α class) that older proteins may evidence the results of sequential folding (Fig. 3).

Within the α/β class, three fold classes were found showing a strong propensity for cotranslational folding (Table 3). Finally, the 'previous contact' measure provides a tool for looking within a single protein for structural traces of cotranslational folding (Fig. 5).

Some terminal residues can be absent in PBD files due to high b-factor. In order to assess the effect that this might have, we computed R_{min}, P_{min} and SLR with five residues snipped from each terminus of the peptides. We found that the three measures were largely unaffected, and the conclusions were robust to removal of terminal residues. The correlations between variables calculated on snipped and unsnipped data were high ($r > 0.84$), except for R_{min}, which yielded a slightly weaker correlation ($r = 0.673$). This strongly suggests that missing end-residues in the PDB files have not affected these results.
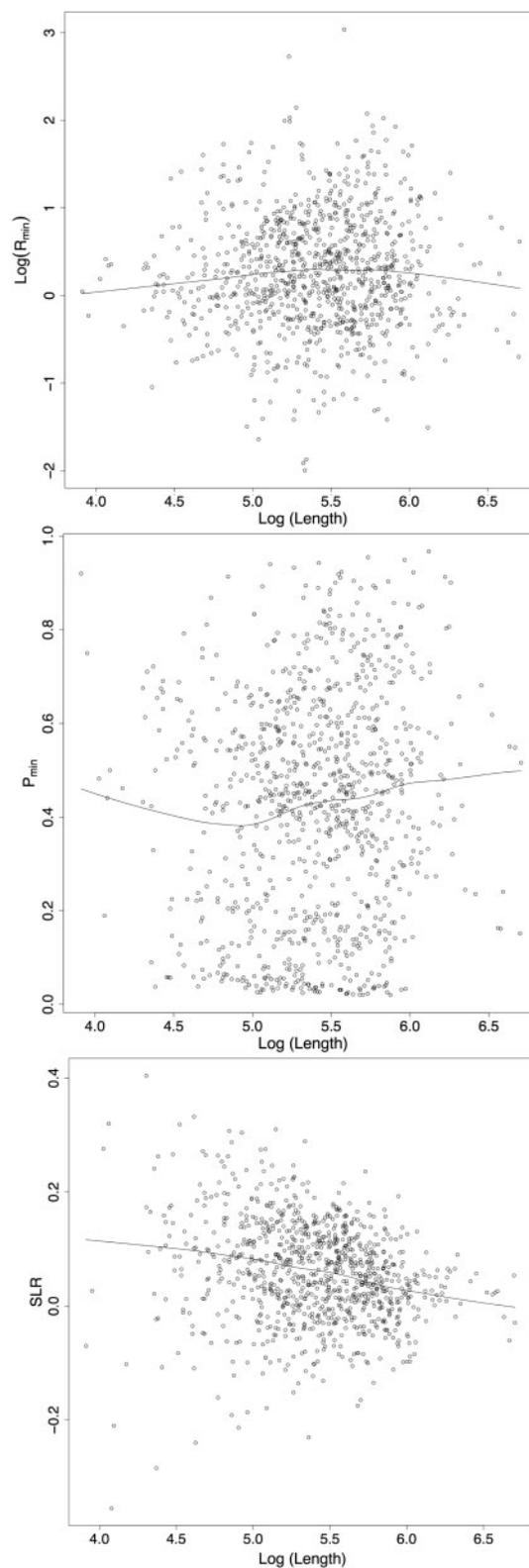
Fig. 6. Scatterplots of $\log(R_{min})$ (top), P_{min} (middle) and SLR (bottom) versus the log of protein length within the α/β SCOP class from the largest dataset (30%, 2.2 Å). The value of the lowess smoother is indicated by the line.

It was mentioned in the Introduction section that the prime purpose of this research was to aid structure prediction. It is therefore interesting to know whether evidence of cotranslational folding indicates that cotranslational structure prediction succeeds. Our sequential structure prediction algorithm, still in an early stage of development, has been used to assess whether the prediction for a polypeptide showing evidence of cotranslational folding is closer to the native state than for one which did not show evidence. We chose two proteins, each of class α/β and of length 138 residues, one (1nu0) showing strong evidence of cotranslation and the other (1m0d) showing little evidence. We performed 200 predictions of each structure and found that the one showing cotranslational evidence was considerably closer (using TMscore) to its native structure than that showing little evidence of cotranslation.

The results of this article are consistent with earlier findings of Alexandrov and Taylor (Alexandrov, 1993; Taylor, 2006). We caution, however, that the results provide evidence, not proof, of cotranslational folding; it is conceivable that the evidence is due to some other factor.

Conflict of Interest: none declared.

REFERENCES

Alexandrov,N. (1993) Structural argument for N-terminal initiation of protein folding. *Protein Sci.*, **2**, 1989–1991.

Baker,D. (1998) Metastable states and folding free energy barriers. *Nat. Struct. Biol.*, **5**, 1021–1024.

Baldwin,T.O. (1999) Protein folding in vivo: the importance of ribosomes. *Nat. Cell Biol.*, **1**, 154–155.

Baskakov,I.V. *et al.* (2001) Folding of prion protein to its native alpha-helical conformation is under kinetic control. *J. Biol. Chem.*, **276**, 19687–19690.

Elcock,A.H. (2006) Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput. Biol.*, **2**, 0824–0841.

Huard,F.P.E. *et al.* (2006) Modelling sequential protein folding under kinetic control. *Bioinformatics*, **22**, 203–210.

Jacob,E. and Unger,R. (2006) A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics*, **23**, 225–230.

Jenni,S. and Ban,N. (2003) The chemistry of protein synthesis and voyage through the ribosomal tunnel. *Curr. Opin. Struct. Biol.*, **13**, 212–219.

Kolb,V.A. (2001) Cotranslational protein folding. *Mol. Biol.*, **35**, 584–590.

Morrissey,M.P. *et al.* (2004) The role of cotranslation in protein folding: a lattice model study. *Polymer*, **45**, 557–571.

Nakatogawa,H. and Ito,K. (2002) The ribosomal exit tunnel functions as a discriminating gate. *Cell*, **106**, 629–636.

Phillips,D.C. (1967) The hen egg-white lysozyme molecule. *Proc. Natl Acad. Sci. USA*, **57**, 484–495.

Sohl,J.L. *et al.* (1998) Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature*, **392**, 817–819.

Srikakulam,R. and Winkelmann,D.A. (2003) Chaperone-mediated folding and assembly of myosin in striated muscle. *J. Cell. Sci.*, **117**, 641–652.

Taylor,W.R. (2006) Topological accessibility shows a distinct asymmetry in the fold of beta/alpha proteins. *FEBS Lett.*, **580**, 5263–5267.

Ullers,R.S. *et al.* (2004) SecB is a bona fide generalized chaperone in Escherichia coli. *Proc. Natl Acad. Sci. USA*, **101**, 7583–7588.

Wang,G. and Dunbrack,R.L. (2003) PISCES: recent improvements to a PDB sequence culling server. *Bioinformatics*, **33**, 94–98.

Winstanley,H.F. *et al.* (2005) How old is your fold? *Bioinformatics*, **21**, 449–458.

Ziv,G. *et al.* (2005) Ribosome exit tunnel can entropically stabilize alpha-helices. *Proc. Natl Acad. Sci. USA*, **102**, 18956–18961.