ELSEVIER

The family feud: do proteins with similar structures fold via the same pathway?

Arash Zarrine-Afsar[1], Stefan M Larson[3] and Alan R Davidson[1,2]

Theoretical and experimental studies of protein folding have suggested that the topology of the native state may be the most important factor determining the folding pathway of a protein, independent of its specific amino acid sequence. To test this concept, many experimental studies have been carried out with the aim of comparing the folding pathways of proteins that possess similar tertiary structures, but divergent sequences. Many of these studies focus on quantitative comparisons of folding transition state structures, as determined by Φ_f value analysis of folding kinetic data. In some of these studies, folding transition state structures are found to be highly conserved, whereas in others they are not. We conclude that folds displaying more conserved transition state structures may have the most restricted number of possible folding pathways and that folds displaying low transition state structural conservation possess many potential pathways for reaching the native state.

Addresses
[1] Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[2] Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[3] 110 Charles Street West, Toronto, Ontario M5S 1K9, Canada

Corresponding author: Davidson, Alan R (alan.davidson@utoronto.ca)

Current Opinion in Structural Biology 2005, **15**:42–49

This review comes from a themed issue on
Folding and binding
Edited by Gideon Schreiber and Luis Serrano

Available online 26th January 2005

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2005.01.011

Abbreviations

AcP	acylphosphatase
ADA2h	activation domain of procarboxypeptidase
Ig	immunoglobulin
SH3	Src homology 3
Spc	spectrin
TI I27	27th immunoglobulin domain from human cardiac titin
TNfnIII	third fibronectin type III domain of human tenascin
TS	transition state

Introduction

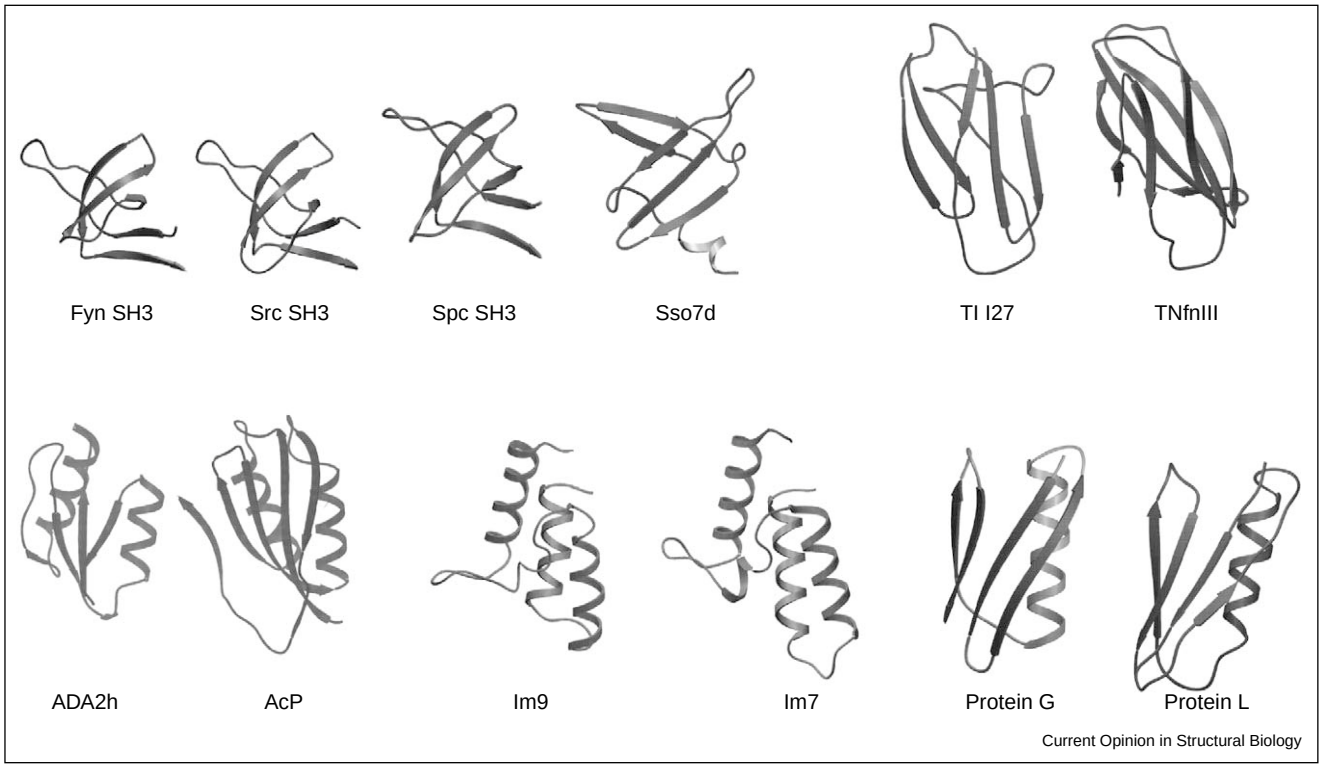
During the past five years, a considerable amount of research has been directed at answering the question of whether proteins with similar tertiary structures fold via the same pathway. Before we review some of this literature, it is worthwhile to consider why this question has gained such prominence in the field. The discovery of a good correlation between the folding rates of proteins and their contact order, which is a property of their native state topology [1], led to the intriguing hypothesis that the topology of a protein, not its specific amino acid sequence, might be the most important factor determining its folding pathway. If this hypothesis were true, then any two proteins with similar tertiary structures and topologies should fold via a similar pathway, even if their sequences are highly diverged. A consistent finding that proteins with similar structures fold through a similar pathway would impart a 'surprising simplicity' to protein folding [2], in that it would be possible to understand and predict the folding pathway of a protein largely from the topology of its native state. Thus, studies aimed at elucidating the folding pathways of proteins with similar structures but diverse sequences have become very important to the field of protein folding, because they provide a critical test of the relationship between native topology and folding pathways.

As a very thorough general review of folding studies on related proteins was published only a few years ago [3], in this review we have focused primarily on quantitative comparisons of transition state structures, a topic not previously dealt with in detail, and on recent developments in the field. Although a variety of theoretical studies provide useful insight into the issues discussed here [4–6], we have considered only experimental studies for the sake of brevity.

Proteins with similar topologies do share common transition state structures: a quantitative analysis

The majority of protein folding studies have been performed on small reversibly folding proteins that exhibit two-state behavior. For these proteins, the only folding 'intermediate' that can be investigated experimentally is the folding transition state (TS). The protein engineering method, developed by the Fersht laboratory [7], provides a means to identify interactions mediated by specific amino acid sidechains that stabilize the folding TS structure of a protein, even though this structure is impossible to observe by direct means due to its extremely short lifetime. This method involves measuring the folding kinetics and equilibrium thermodynamics of mutants containing amino acid substitutions located throughout the protein. Analysis of these data allows the calculation of the extent of structure formation surrounding a particular mutated sidechain in the folding TS; this is

Figure 1



The structures of proteins on which comparative Φ_f value analyses have been performed are shown. The PDB codes for these structures are shown in Table 1.

expressed as the Φ_f value. Φ_f values typically range from 0 to 1, where 1 represents complete structure formation in the TS and 0 represents a complete absence of native structure formation (for a recent review, see [8]).

Direct Φ_f value comparisons have now been made for aligned positions in seven pairs of structurally similar proteins (Figure 1). The amino acid identity between the tested proteins ranges from 75% down to 4% (Table 1). In the published studies, the similarities between the Φ_f values of these proteins were discussed mostly in qualitative terms. To add more precision to the current discussion pertaining to similarities and differences between TS structures, we felt it would be useful to

Table 1

Alignment information for proteins used in Φ_f value comparison studies.

Proteins compared	PDB code	Sequence identity (%)[a]	rmsd (Å)[b]	Residues aligned	References
Src SH3/Fyn SH3	1SHF/1FMK	73	1.23	58	[10,17,46]
Src SH3/Spc SH3	1FMK/1SHG	33	1.47	56	[9,10]
AcP/ADA2h	2ACY/1AYE	10	2.71	66	[13,14]
TNfnIII/TI I27	1TEN/1TIT	10	2.01	52	[11,12]
Im7/Im9	1CEI/1IMQ	72	2.14	72	[20••]
Protein G/Protein L	2IGD/2PTL	13	1.81	52	[26,27]
Sso7d/Src SH3	1SSO/1FMK	4	3.03	26	[29]
Suc1/Cks1	1SCE/1QB3	74	0.81	82	[30]

[a]Alignments were based on sequence similarity when the percent identity was higher than 30%, whereas the rest of the alignments were derived from structural comparison. In most cases, we utilized alignments included in the referenced publications. When alignments were not included in publications (e.g. AcP/ADA2h and protein G/protein L), we used structural alignments from the FSSP database [47].
[b]The rmsd between the two structures after structural alignment. This value was calculated using only residues that were aligned within 5 Å rmsd of each other. The next column indicates how many residues were aligned to calculate the rmsd value shown.