Methods 52 (2010) 38–50

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Review Article

What lessons can be learned from studying the folding of homologous proteins?

Adrian A. Nickson [1], Jane Clarke *

University of Cambridge, Department of Chemistry, MRC Centre for Protein Engineering, Lensfield Road, Cambridge CB2 1EW, UK

ARTICLE INFO

Article history:
Available online 4 June 2010

Keywords:
Protein folding
Φ-Value
Protein families
Homologous proteins

ABSTRACT

The studies of the folding of structurally related proteins have proved to be a very important tool for investigating protein folding. Here we review some of the insights that have been gained from such studies. Our highlighted studies show just how such an investigation should be designed and emphasise the importance of the synergy between experiment and theory. We also stress the importance of choosing the right system carefully, exploiting the excellent structural and sequence databases at our disposal.

1. Introduction

Studies of the folding of homologous proteins have been gaining in popularity since the first explicit studies over 10 years ago. The inspiration for such studies comes from structural and sequence family databases, such as SCOP [1] and Pfam [2]. Initially the questions asked were relatively straightforward: do all proteins with the same structure fold via the same pathway? Are residues important for folding highly conserved? How can differences in the kinetics of folding be explained? In Table 1 we list all folds where more than one protein has been subject to detailed analysis of the folding pathway. In some cases proteins are closely related, with significant sequence identity (such as the immunity proteins, ~60% identical); in others, only structure is conserved (for instance the immunoglobulin-like domains where proteins from different superfamilies have been compared); in other cases, in a twist to the tail, circular permutation has allowed the importance of chain connectivity to be investigated in proteins with identical sequence (such as the studies on the ribosomal protein S6).

Perhaps the most important insights into the folding field from studies of homologous proteins, have come from investigating *differences* between related proteins – proteins from the same family which have different folding mechanisms, proteins with different kinetic properties and those with completely different folding pathways.

The first importance of these studies is that they have allowed insight into some of the fundamental questions about protein folding mechanisms. These experimental studies have been particularly powerful when they have gone hand-in-hand with computational and theoretical work on folding. In the first part of this work we show how studies of the folding of families of proteins have been vital in developing our understanding of the relative importance of topology, sequence, entropy/enthalpy balance and secondary structure propensity in determining folding mechanisms.

In the second part we show how theoretical studies are adding to this work and highlight some very recent studies that show the power of this family approach to study protein folding: insights which would not have been possible from studies of individual proteins in isolation. We show that comparative studies have allowed rational design of folding pathways and altered kinetics.

Finally we address the design of studies of the folding of protein families.

2. Folding pathways and mechanisms

Although the 'protein folding problem' could theoretically be solved using a brute-force approach, where homologues to every conceivable protein sequence and structure are studied, a more subtle approach is to discover the pathways by which proteins attain their native state. This method also avoids the potential trap that proteins with very similar primary sequences can fold to different three dimensional structures [3,4].

As Valerie Daggett cogently observed: "To map [a] folding reaction we need to characterise all states along the way – native, transition, intermediate and denatured – as well as the mechanism of conversion between them" [5]. This feat has been achieved for a small number of proteins, by combining experimental studies on the free-energy maxima and minima with all-atom molecular dynamics simulations to add the remaining detail [6–8]. In addition, recent advances in NMR have made it possible to gain residue specific information about the structure of a polypeptide chain as it collapses towards the native state [9,10].

* Corresponding author.
E-mail address: jc162@cam.ac.uk (J. Clarke).
[1] Present address: Medimmune, Milstein Building, Granta Park, Cambridge CB1 6GH, UK.

1046-2023/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2010.06.003

Table 1
Protein folds where the folding of homologous proteins has been studied.

Class (fold)	Superfamily	Protein (species)	Method of investigation	PDB code	Experimental references	Comparative references[a]
All-α (Acyl-CoA binding protein-like)	Acyl-CoA binding protein	ACBP (Cow)	Φ-Value analysis	2ABD	[93,94]	[94]
		ACBP (Rat)	WT kinetics	2ABD[b]	[93]	
		ACBP (Yeast)	Φ-Value analysis	2ABD[b]	[94]	
All-α (Acyl carrier protein-like)	Colicin E immunity proteins	Im7 (*E. coli*)	Φ-Value analysis	1AYI	[28]	[29]
		Im9 (*E. coli*)	Φ-Value analysis	1IMQ	[29,30,95]	
All-α (Cytochrome *c*)	Cytochrome *c*	Cytochrome *c* (Horse)	Hydrogen exchange	1HRC	[96]	[97]
		Cytochrome *c*2 (*R. capsulatus*)	WT kinetics	1C2R	[98]	
		Cytochrome *c*551 (*P. aeruginosa*)	Minimal Φ-value analysis	2PAC	[99]	
		Cytochrome *c*552 (*H. thermophilus*)	WT kinetics	1AYG	[100]	
		Cytochrome *c*552 (*T. thermophilus*)	WT kinetics	1C52	[101]	
		Mitochondrial cytochrome *c* (Yeast)	WT kinetics	1YCC	[102]	
All-α (Four-helical up-and-down bundle)	Cytochromes	Cytochrome *b*562 (*E. coli*)	Hydrogen exchange	1APC	[103]	
	FKBP12-rapamycin-binding domain of FKBP-rapamycin-associated protein (FRAP)	FRB (Human)	WT kinetics	1AUE	[104]	
All-α (DNA/RNA-binding 3-helical bundle)	Homeodomain-like	DNA-binding domain of human telomeric protein hTRF1 (Human)	WT kinetics	1BA5	[47]	[47]
		En-Hd (Drosophila)	Φ-Value analysis	1ENH	[7,105]	
		c-Myb DNA-binding domain (Mouse)	Φ-Value analysis	1IDY	[47]	
		Rap1 (Human)	WT kinetics	1FEX	[47]	
All-α (Globin-like)	Globin-like	Leghemoglobin (Soybean)	Hydrogen exchange	1FSL	[106]	[106]
		Myoglobin (Sperm whale)	Hydrogen exchange	1A6M	[107]	
All-α (peripheral subunit-binding domain of 2-oxo acid dehydrogenase complex)	Peripheral subunit-binding domain of 2-oxo acid dehydrogenase complex	E3 binding domain of dihydrolipoamide acetyltransferase [E3BD] (*B. stearothermophilus*)	Φ-Value analysis	1EBD	[58]	[59]
		E3-binding domain of dihydrolipoamide succinyltransferase [BBL] (*E. coli*)	Φ-Value analysis	1BBL	[59]	
		POB (*P. aerophilium*)	Φ-Value analysis	1BBL[b]	[60]	
All-α (ROP-like)	ROP protein	ROP (*E. agglomerans*)	WT kinetics	1ROP[b]	[4]	[4]
		ROP (*E. coli*)	WT kinetics	1ROP	[4]	
		ROP (*P. vulgaris*)	WT kinetics	1ROP[b]	[4]	
All-α (spectrin repeat-like)	Spectrin repeat	Alpha chain R15 (Chicken)	Φ-Value analysis	1U5P	[46]	[46]
		Alpha chain R16 (Chicken)	Φ-Value analysis	1CUN	[81]	
		Alpha chain R17 (Chicken)	Φ-Value analysis	1CUN	[45]	
α/β (α/β knot)	α/β knot	YbeA (*E. coli*)	Φ-Value analysis	1NS5	[108]	[108]
		YibK (*H. influenzae*)	Φ-Value analysis	1J85	[109]	
α/β (Dihydrofolate reductase-like)	Dihydrofolate reductase-like	Dihydrofolate reductase (*E. coli*)	WT kinetics, Ligand binding	1RA9	[110]	[110]
		Dihyrofolate reductase (*L. casei*)	WT kinetics, Ligand binding	3DFR	[110]	
		Dihydrofolate reductase (Human)	WT kinetics, Ligand binding	1KMV	[110]	
α/β (flavodoxin-like)	CheY-like Flavoproteins	CheY (*E. coli*)	Φ-Value analysis	1EAY	[111]	[112]
		Apoflavodoxin (*A. vinelandii*)	WT kinetics	1YOB	[113]	
		Flavodoxin (*Anabaena pcc 7119*)	Φ-Value analysis	1FTG	[112]	
α/β (Phosphoglycerate kinase)	Phosphoglycerate kinase	Phosphoglycerate kinase (*B. stearothermophilus*)	Minimal Φ-value analysis	1PHP	[114]	
		Phosphoglycerate kinase (Yeast)	WT kinetics	3PGK	[115]	
α/β (RNase-H-like Motif)	RNase-H-like	RNase-H (*E.coli*)	Hydrogen exchange Limited mutagenesis	1F21	[168,169]	
		RNase-H (*T. thermophilus*)	Hydrogen exchange	1RIL	[170]	[170,171]

(*continued on next page*)