J Mol Model (2007) 13:587–594
DOI 10.1007/s00894-007-0185-8

ORIGINAL PAPER

Analysis of the differences in the folding mechanisms of c-type lysozymes based on contact maps constructed with interresidue average distances

Shunsuke Nakajima · Takeshi Kikuchi

Received: 28 January 2006 / Accepted: 6 February 2007 / Published online: 6 March 2007

Abstract A method for analyzing differences in the folding mechanisms of proteins in the same family is presented. Using only information from the amino acid sequences, contact maps derived from the interresidue average distances are employed. These maps, referred to as average distance maps (ADM), are applied to the folding of c-type lysozymes. The results reveal that the ADMs of these lysozymes reflect the differences in the detailed folding mechanisms. Further possible applications of the present method are also discussed.

Keywords Average distance map · Folding mechanism · Lysozyme family · Prediction of folding from sequence

Introduction

It is widely accepted that the folding mechanism of a protein is basically determined by its topology [1–3]. That is, members of the same family are assumed to fold basically through the same pathway. However, recent investigations [4, 5] revealed that some proteins in the same family fold through rather different pathways, although the basic features of the process is in common. Information on the folding mechanism of a protein must be coded in its amino acid sequence, but the differences of the folding pathways of proteins with the same fold cannot be detected by standard sequence analysis methods such as sequence alignment. Only the predicted location of a hydrophobic core, which is based on the prediction of buried surface, might be considered as a nucleation site in the early stage of folding events of a protein [6]. A method to predict the kinetic properties of folding has been attempted by using the native structure and native contacts of a protein [7–9]. Although this method has a high possibility of predicting the details of the folding kinetics based on the native higher order correlation of distances, it requires information of the native 3D structure; sequence information alone does not suffice. On the other hand, it has been revealed that protein folding rates exhibits correlation with protein hierarchical structures such as secondary sturcture and topology. A strong correlation between contact order, the parameter denoting the native topology, and the folding rates of proteins showing two-state folding has been found out by Baker and coworkers [1–3] and have been confirmed by several authors [10–15]. Calloni et al. [16] have made a comparison of the folding processes of proteins sharing the same folding topology and have shown the importance of hydrophobic content in folding. They also suggeted the possibility of the prediction of protein folding rates. Several proposals to predict the protein folding rates from amino acid properties have recently come out. Methods for the predictions of folding rates of proteins with two-state folding based on secondary structure content were presented by Gong et al. [17] and Prakash and Bhuyan [18], and based on amino acid rigidity and secondary structure propensities by Huang and Tian [19]. Thus, these works tried to predict protein folding rates from

S. Nakajima
Department of Chemistry and Bioscience,
College of Industrial Technology,
Kurashiki University of Science and the Arts,
Kurashiki, Okayama, Japan

T. Kikuchi (✉)
Department of Bioscience and Bioinformatics,
College of Information Science and Engineering,
Ritsumeikan University,
Nojihigashi, Kusatsu,
Shiga 525–8577, Japan
e-mail: tkikuchi@is.ritsumei.ac.jp

only sequence information, but all techniques predict just folding rates but not details of folding processes from amino acid sequences.

The aim of the present work is to perdict folding processes of proteins from only their sequence. We have reported [20, 21] that the differences in folding mechanisms of members of the lipid binding protein family and the globin family are reflected in the shapes of their average distance maps (ADMs) derived based on interresidue average distances in proteins. With the ADM method, it is possible to predict differences in folding pathways of proteins of a family from the amino acid sequence of a protein [22, 23].

The present study focuses on the folding mechanism of c-type lysozymes in the SCOP classification because many studies on the folding mechanisms of lysozymes have been performed in this decade. Interestingly, in spite of the 3D-structure and sequence similarity, the function of α-lactalbumin, a member of c-type lysozymes, is basically different from that of other members. The folding mechanisms of c-type lysozymes have been thoroughly studied as well as their sequences, 3D-structures and functions [24–29]. The folding pathways of three lysozymes are reported to be relatively similar, but it was also observed that the folding mechanisms are also slightly different among the three [24–27]. However, the relationships between the differences in the folding mechanisms of c-type lysozymes and their sequences have not been studied till now.

In this work, we construct ADMs of c-type lysozymes from their sequences and compare the folding mechanisms reported in the literature with the ADMs. The purpose of this work is to extract information on the differences in folding mechanisms of c-type lysozymes from their sequences via the ADMs.

Materials and method

Analysis of folding process of a protein based on the average distance map method

We employ the method in Ref. [22] in this paper, and briefly give a survey of the method. We refer this method as the average distance map (ADM) mehod. Using proteins with known structures, the average distances between Cα atoms of residues were calculated in each 'range' which is defined as the length of a loop made by a pair of the contacted residues, i.e., a range is defined as $1 \leqq k \leqq 8$ as the range M=1 where k=|i-j| and i and j are the residue numbers forming a contact in the sequence, and also $9 \leqq k \leqq 20$, $21 \leqq k \leqq 30$, $31 \leqq k \leqq 40$ and so on define respective ranges M=2, 3, 4 ···. For a protein with unknown 3D structure, a contact map is constructed by making a plot (i.e., defining a contact) on a map when the average distance of a pair of residues is less than a cutoff value determined in advance. A cutoff value is defined in each range so that the contact density of whole RDM of a protein is reproduced [22], where RDM (real distance map) stands for a contact map constructed based on the actual 3D structure. In this work, RDM is made based on the X-ray structure of a protein with the difinition of a contact as shorter interresidue Cα atomic distance than 15Å. When we regard ρ_{av} as the average values of contact density of the entire region of a map, the value of ρ_{av} in a RDM is roughly reproduced with the formula, $\rho_{av} = C/N$ [22] where N is the total number of residues and C is a constant. Cutoff distances for construction of an ADM for a protein are defined to reproduce a value of ρ_{av}. For the ADM construction, we prepare a different cutoff distance for a different range in contrast to the same cutoff distance used in construction of RDMs. Here, we make an assumption that the contact density in a fragment chain in a protein is also in invers proportion to the number of residues in this fragment. Then, we can compute a probability to form a contact between residues A and B in the range M which is roughly proportional to k=|i-j|. The number of residue pairs that make contact obeys the following equation.

$$P(M)_c = (D/M)P(M)_t$$

Here, $P(M)_c$ is the number of aminoacid pairs whose average distances in the range M is less than a cutoff distance and $P(M)_t$ is the total number of residue pairs in a given range, i.e., 210 pairs of residues minus the number of the pairs with statistically insufficient occurrence [22]. D is an adjustable parameter that is not dependent on M, i.e., constant for all M values in a given protein, and this value is chosen so that the overall average density ρ_{av} of the ADM is close to the value of $\rho_{av} = C/N$ by trial and error. Thus, we can define a kind of contact maps based on the statistics of the interresidue average distances. We call this map average distance map (ADM). Based on a contact map (ADM or RDM), a compact area on a map can be defined by the following procedure. The whole area of the map is divided into two parts by a line parallel to the abscissa at the ith residue or by a line parallel to the ordinate at the ith residue as illustrated in Fig. 1a and b, and the difference of contact density values between the triangle and the trapezoldal parts of the map, $\Delta\rho_i$, is calculated, i.e., $\Delta\rho_i = \rho_i - \rho_i^*$ where ρ_i and ρ_i^* denote the contact density of the triangle and trapezoldal parts respectively. The series of the values of difference of density, $\Delta\rho_i$, from residue 1 to residue N provides a scanning plot. We call the scanning plot produced by the division using the line parallel the ordinate as horizontal scanning, and the plot produced by the line parallel to the abscissa as vertical scanning. h of $\Delta\rho_i^h$ and v of $\Delta\rho_i^v$ denote the horizontal and vertical divisions of a map

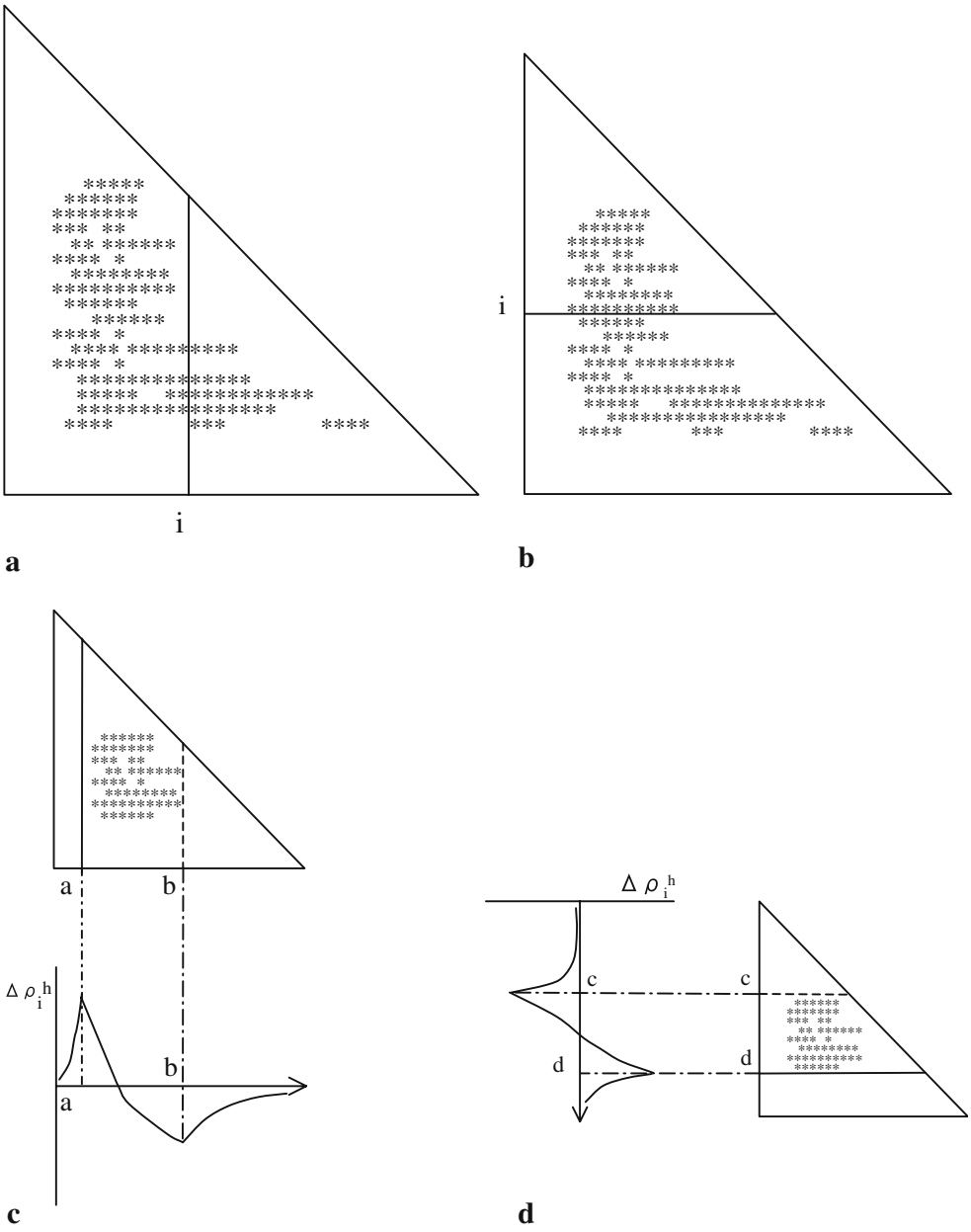

Fig. 1 (**a**) Schematic srawing of a contact map divided by a line parallel to the abcissa at th residue i. An asterisk on the map denotesa a contact. (**b**) Schematic srawing of a contact map divided by a line parallel to the ordinate at th residue i. An asterisk on the map denotesa a contact. (**c**) Schematic drawing of the horizontal scanning plot of the density difference ($\Delta\rho_i^h = \rho_i - \rho_i^*$) of the map with the hypothetical contact map. The maximum (peak) and minimum (valley) would be obtained in the horizontal scanning plot as the point of the largest change of contact density values. The peak and the valley appear at points *a* and *b*. The positions of the peak (solid line) and the valley (broken line) obtained in scanning plot are also indicated in the maps. The peak and the valley should appear at the boundary of the high and low density contact regions as schematically illustrated in these figure. (**d**) Schematic drawing the vertical scanning plot of the density difference ($\Delta\rho_i^v = \rho_i - \rho_i^*$) of the map with the hypothetical contact map. The maximum (peak) and minimum (valley) would be obtained in the horizontal scanning plot as the point of the largest change of contact density values. The peak and the valley appear at *c* and *d*. The positions of the peak (solid line) and the valley (broken line) obtained in the scanning plot are also indicated in the maps

respectively. The maximum (peak) and minimum (valley) would be obtained in the scanning plot as the point of a large change of contact density values. The horizontal scanning plot of $\Delta\rho_i^h$ from 1 to N of the residue number is schematically shown in Fig. 1c, and peak and valley appear at a and b in the figure at a large change of contact density values. The same situation is observed in the vertical scanning (Fig. 1d).

The locations of peaks and valleys are analyzed. The compact areas on a map can be defined by the positions of the peaks as shown in Fig. 2.

This figure shows the schematic drawing of a contact map having two compact areas near the diagonal, with the horizontal and vertical scanning plot, which is denoting the existence of two domains by the peaks at residues A and B in the horizontal scanning plot and residues C and D in the vertical scanning plot. Thus, regions A-C and B-D on the map is predicted as possible comapact regions or domains in the protein. The strength of the compactness of a region A-C can be measured by the η values defined by $\eta = \Delta\rho_A^h + \Delta\rho_C^v$ where the residue A shows a peak in the horizontal scanning and the residue C in the vertical scanning [22]. Thus, based on this procedure, we can make a prediction on location of domains and 3D structural nuclei in a protein from only its amino acid sequence. The region with the highest η value can be defined as a domain. The regions with high η values within a domain can be assigned as subdomains [22]. This procedure predicts also a folding process by predicting positions of subdomains formed during folding based of η values of subdomains. Hence the η value measures the strength of domain/subdomain in a protein, a high η value

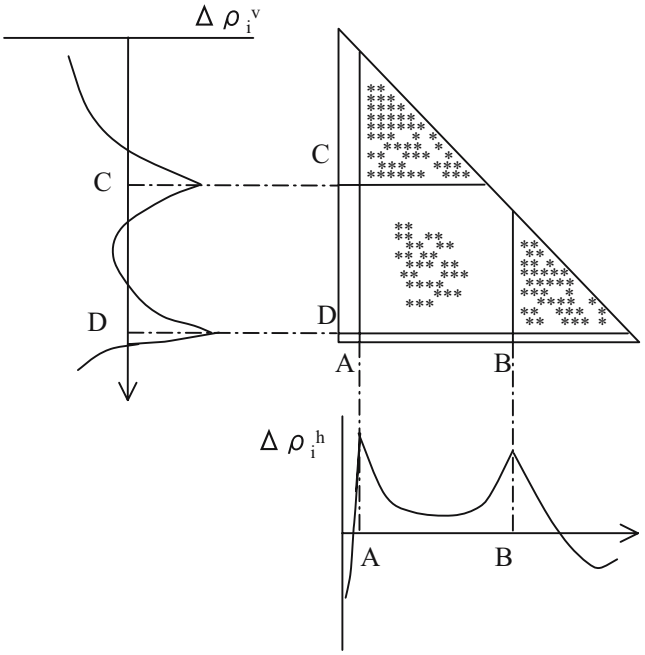

Fig. 2 Schematic drawing of a map with the two high density contact regions along the diagonal of the map with the horizontal and the vertical scanning plot of the density difference of the map. Two peaks would appear at positions denoted *A* and *B* in the horizontal scanning plot, and also two peaks would be at *C* and *D* in the vertical scanning plot

also denotes the existence of a stable compact region during folding. In the case that two subdomains with different η values in a protein, the subdomain with higher η value would fold first followed by the formation of the subdomain with lower η value. Two subdomains will fold simultaneously (i.e., without the formation of heterogeneous intermediates) if they have similar η values, i.e., if one η value is within the 85% of the other η value [20–23]. For example, the detailed procedure of the assignment of subdomain of hen egg white lysozyme from its ADM represented in Fig. 5a is as follows. The peaks obtained from the horizontal and vertical scanning plots listed in Table 1 with their respective values of η. The highest value of η (=0.354) is observed at the residue 1 of the horizontal and at the residue 34 of the vertical scanning plots. This observation is interpreted that the region 1–34 forms a subdomain. This region is presented as a triangle area enclosed by solid blue lines in Fig. 5a. However, the second highest η=0.329 value appears at the residue 1 of the horizontal and at the residue 127 of the vertical scanning plots. This value is within 85% of 0.354 and thus the subdomain 1–34 can be extended to 1–127 and this region is predicted as a domain, i.e., hen egg white lysozyme is predicted as one domain protein indicated by a region enclosed by red lines in Fig. 5a. On the other hand, the residue 28 of the horizontal and the residue 64 of the vertical scanning plots show the relatively small η value (=0.155), and this area includes a part of αB and β2–5, that is, this region corresponds to the β region. This observation implies that the β domain is predicted to form a relatively weak compact region during the folding. This region is presented as a triangle area enclosed by broken blue lines in Fig. 5a.

Based on these considerations, we analyze and predict the qualitative folding processes of c-type lysozymes and compare these results with the experimental data in the literature.

Proteins used in this work

The following c-type lysozymes were selected: hen-egg lysozyme (PDB code: 6LYZ), human lysozyme (PDB code: 1LZ1), equine lysozyme (PDB code: 1EQL) and bovine α-lactalbumin (PDB code: 1F6S). The 3D structures of these lysozymes are presented in Fig. 3a–d.

Basically, these proteins have common secondary structures, i.e., four α-helices, A, B, C, and D, and five β-strands, 1–5. The sequence positions of secondary structures in hen-egg lysozyme are helix: 5–15 (A), 25–35 (B), 80–84 (C), 89–96 (D); and β-strand: 1–3, 38–40, 42–46, 50–54, 57–60. The first and fifth strands, and the second, third and fourth strands form β-sheets respectively. Furthermore, two 3^{10} helices exist at 80–84 and 120–124. However, α-lactalbumin lacks the corresponding β-sheet formed by strands 1–3 and 57–60 in hen-egg lysozyme. The location of these secondary structures is illustrated in Fig. 4.

A folding process for each protein reported in the literature can be summarized briefly as follows. Hen-egg lysozyme folds with four helices and a C-terminal 3^{10} helix in a cooperative manner [24]. Folding of human lysozyme is accompanied by formation of helices A, B and C-terminal 3^{10} helix in the first few milliseconds, and these region interact with 109-Trp and 112-Trp [25]. Equine lysozyme initially folds at helices A, B and D via heterogeneous pathways [26]. In particular, high exchange protection of helices A and B in the NMR measurement was observed [26]. In bovine α-lactalbumin, the α helical regions form at an early stage of folding, and the highest exchange protection of the C helix in the molten globule state has been reported in the NMR study [27]. The role of the C helix in folding is controversial. It has been demonstrated that the C helix is not absolutely required for folding [28]. All lysozymes show a slow formation of β domain during folding [24–27].

We present sequence alignments in Fig. 4 and the mutual sequential homology of these lysozymes in Table 2. The homology of the sequences of these proteins is about 40–60% identical.

In general, the sequence of α-lactalbumin is less homologous than the other three proteins. However, it is difficult to extract differences in folding processes of these lysozymes from this sequence alignment.

Table 1 Position of residues with the peaks in the analysis of contact density differences of ADM for hen-egg white lysozyme for assignment of the location of subdomains

Location of residue	Δρ value in the horizontal scanning plot	Location of residue	Δρ value in the vertical scanning plot	η value	
1	0.255	34	0.099	0.354	m
1	0.255	58	0.054	0.309	
1	0.255	64	0.056	0.311	
1	0.255	80	0.025	0.280	
1	0.255	99	0.037	0.292	
1	0.255	111	0.033	0.288	
1	0.255	115	0.022	0.277	
1	0.255	120	−0.025	0.230	
1	0.255	124	0.039	0.294	
1	0.255	127	0.074	0.329	f
6	0.066	34	0.099	0.165	
6	0.066	58	0.054	0.120	
6	0.066	64	0.056	0.122	
6	0.066	80	0.025	0.091	
6	0.066	99	0.037	0.103	
6	0.066	111	0.033	0.09	
6	0.066	115	0.022	0.088	
6	0.066	120	−0.025	0.041	
6	0.066	124	0.039	0.105	
6	0.066	127	0.074	−0.140	
8	0.042	34	0.099	0.141	
8	0.042	58	0.054	0.096	
8	0.042	64	0.056	0.098	
8	0.042	80	0.025	0.067	
8	0.042	99	0.037	0.079	
8	0.042	111	0.033	0.075	
8	0.042	115	0.022	0.064	
8	0.042	120	−0.025	0.017	
8	0.042	124	0.039	0.081	
8	0.042	127	0.074	0.116	
28	0.099	34	0.099	0.198	
28	0.099	58	0.054	0.153	
28	0.099	64	0.056	0.155	b
28	0.099	80	0.025	0.124	
28	0.099	99	0.037	0.136	
28	0.099	111	0.033	0.132	
28	0.099	115	0.022	0.121	
28	0.099	120	−0.025	0.074	
28	0.099	124	0.039	0.138	
28	0.099	127	0.074	0.173	
55	0.145	58	0.054	0.199	
55	0.145	64	0.056	0.201	
55	0.145	80	0.025	0.170	
55	0.145	99	0.037	0.182	
55	0.145	111	0.033	0.178	
55	0.145	115	0.022	0.167	
55	0.145	120	−0.025	0.120	
55	0.145	124	0.039	0.184	
55	0.145	127	0.074	0.219	
75	0.160	80	0.025	0.185	
75	0.160	99	0.037	0.197	
75	0.160	111	0.033	0.193	
75	0.160	115	0.022	0.182	
75	0.160	120	−0.025	0.135	
75	0.160	124	0.039	0.199	
75	0.160	127	0.074	0.234	
88	0.129	99	0.037	0.166	
88	0.129	111	0.033	0.162	
88	0.129	115	0.022	0.151	
88	0.129	120	−0.025	0.104	
88	0.129	124	0.039	0.168	
88	0.129	127	0.074	0.203	
98	0.096	99	0.037	0.133	
98	0.096	111	0.033	0.129	
98	0.096	115	0.022	0.118	
98	0.096	120	−0.025	0.071	
98	0.096	124	0.039	0.135	
98	0.096	127	0.074	0.170	

m: maximum η value
f: second maximum η value
b: the region corresponding to the β domain

Results

The ADMs for the lysozymes treated in this work are shown in Fig. 5a–d. The analyses of the ADMs predict that these four lysozymes are all one-domain proteins, i.e., the compact regions in the sequences are 1–127 (0.329) in hen-egg lysozyme, 1–116 (0.339) in human lysozyme, 1–115 (0.312) in equine lysozyme, and 1–121 (0.493) in α-lactalbumin. The numbers in parentheses denote the η values.

The η values of the regions corresponding to central β-sheet of these proteins (These regions are denoted by regions encloed by blue broken lines in Fig. 5a–d) are not so large for all proteins. The η values of the these regions are: hen-egg 28–64 (0.155), human 23–65 (0.177), equine 25–65 (0.200), and α-lactalbumin 40–61 (0.171). These results suggest that each β domain is not so stable especially during folding.

Features of the ADM for hen-egg lysozyme in Fig. 5a are summarized as follows.

The highest η value (0.354) appears at 1–34 which corresponds to helices A and B; this region is regarded as a possible compact region but this region can be extended to 1–127 because its η value, 0.329, is within 85% of the region 1–34 (according to the method described in ref. [22]). Thus, based on the ADM appearance in hen-egg lysozyme, region 1–34 (helices A and B) folds first and then the whole structure of 1–127 is formed.

In the ADM of human lysozyme shown in Fig. 5b, regions 1–32 and 1–112 show similar η values, 0.357 and 0.350, and these regions are included in a domain 1–116

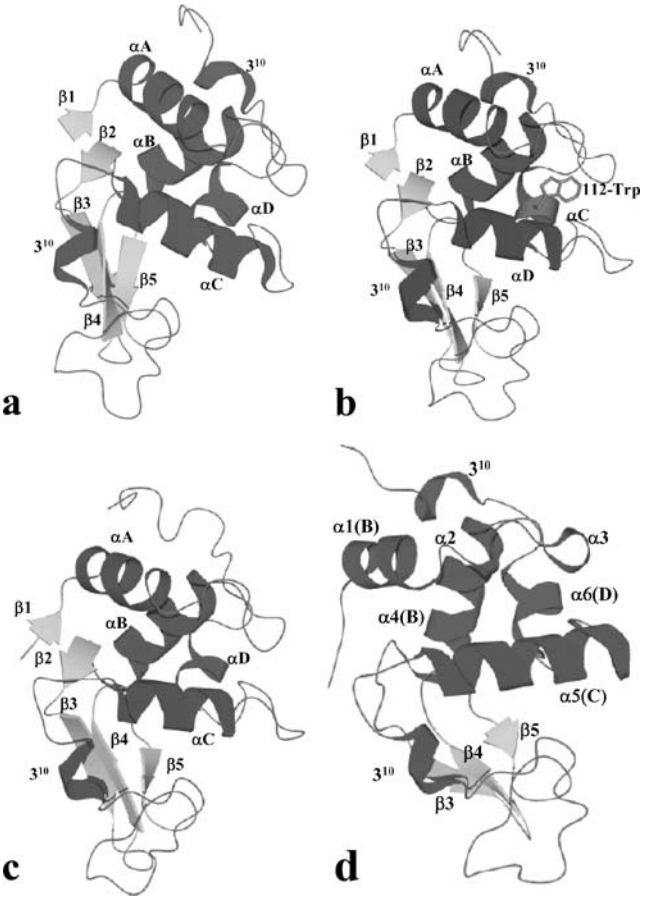

Fig. 3 3D structures of c-type lysozymes treated in this work. (**a**) hen-egg lysozyme, (**b**) human lysozyme, (**c**) equine lysozyme, (**d**) bovine α-lactalbumin. The location of the helices (Black) and β-strands (gray) is labeled in each structure

with an η value of 0.339 which is slightly lower than the η values region 1–32 and 1–112.

Hence the ADM of human lysozyme predicts that the protein starts to fold either at the site of the helices A and B, or at 1–112 including helices A and B, and then this folding unit extends to 1–116.

As shown in Fig. 5c, the ADM of equine lysozyme predicts a strong subdomain at 1–34 (helices A and B) with η=0.330, but the larger region 1–116 has a similar η value of 0.312.

The appearance of the ADM of equine lysozyme resembles that of human lysozyme, that is, the folding would proceed with the formation of the AB helices but immediately the folding region would extend to 1–116. However, the structural unit 1–112 as observed in human lysozyme does not appear in the ADM of equine lysozyme.

In the ADM for α-lactalbumin (Fig. 5d), the region 1–121, that is, almost the whole protein, is predicted as a strong folding unit with η=0.493.

Based on the ADM, the region of helices A and B is not predicted to be a compact region compared with the ADMs for other lysozymes. Regions 21–121, 26–121, 72–121 are predicted as possible subdomains with the relatively high value of η (0.366 for each). Region 72–121 corresponds to helices C, D and the C-terminal 3^{10} helix. This result suggests that the whole region folds in the initial stage of folding. A clearly identified folding unit in α-lactalbumin is not predicted, but the region 72–121 might be stable during the folding.

As a summary of the ADM prediction for the c-type lysozyme family members, all except α-lactalbumin generally fold with the AB helix region as the initial folding unit, and folding would proceed accompanied by formation of the α helix domain. For α-Lactalbumin, the whole region or the region of helices C and D might be a potential folding initiation site.

Discussion

We compare the ADM results and the reported experimental results for the folding of c-type lysozymes. The feature common to experiments and ADMs for all four members is

Fig. 4 Alignment of the sequences of hen-egg lysozyme, human lysozyme, equine lysozyme, and bovine α-lactalbumin[1)]. The location of α-helices is indicated by the black boxes, β-strands by the gray boxes and 3^{10} helices enclosed by rectangles

```
Hen Egg    KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNT-DGSTDYGILQIN
Human      KVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRSTDYGIFQIN
Equine     KVFSKCELAHKLKAQEMDGFGGYSLANWVCMAEYESNFNTRAFNGKNANGSSDYGLFQLN
Bovine[1)] EQLTKCEVFRELK--DLKGYGGVSLPEWVCTTFHTSGYDTQAIVQNN--DSTEYGLFQIN
           : : :**:   :*  :..: * ** :*:* :   *.::*:*   :    *::**::*:*

Hen Egg    SRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTD
Human      SRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQNRD
Equine     NKWWCKDNKRSSS-NACNIMCSKLLDENIDDDISCAKRVVRDPKGMSAWKAWVKHCKDKD
Bovine[1)] NKIWCKDDQNPHSSNICNISCDKFLDDDLTDDIMCVKKIL-DKVGINYWLAHKALCS-EK
           .: **:*.:  . : * *:: *. :*..::   : *.*::: *  *:  * *    *.  .

Hen Egg    VQAWIRGCRL
Human      VRQYVQGCGV
Equine     LSEYLASCNL
Bovine[1)] LDQWLCEK--
           :  ::
```

Table 2 Pairwise sequence identity (%) of four lysozymes

	4LYM	1LH1	2EQL	1HZF
4LYM	100			
1LH1	60	100		
2EQL	48	51	100	
1HZF	39	39	44	100

a slow folding of the central β sheet. The properties of the ADMs of the hen-egg, human and equine lysozymes are relatively similar in regard to helices A and B forming a folding unit.

A detailed discussion for each protein follows. The folding experiments in the literature [24] elucidated that the α domain and C-terminal 3^{10} helix fold first and cooperatively in hen-egg lysozyme. This result is reflected in the ADM prediction that the folding starts with helices A and B, but folding of the whole region of the protein follows immediately.

The experimental data for human lysozyme reveal that the initial event of the folding is the formation of the region of helices A, B and C-terminal 3^{10} helix, and this region interacts with 112-Trp and 109-Trp [25]. The ADM for the same protein predicts that the folding would initiate at helices A and B, but region 1–112 would also form a strong folding unit. This result strongly suggests the interaction of region A and B with 112-Trp. Thus, the ADM reflects the folding process in human lysozyme.

Whereas equine lysozyme was reported to fold through formation of helices A, B and D via a heterogeneous process [26], the ADM predicts that the folding starts at helices A and B and this region extends to 1-116, which can be interpreted as the interaction of helices A and B with helix D (see Fig. 5c).

Experiment suggests that the folding of α-lactalbumin starts with formation of α domain [27]. The amide protection measured by NMR is strongest at the C helix in both the molten globule and native state, but relatively small protection at the A and B helices was observed. In contrast, helices A and B showed the highest protection during the folding of equine lysozyme [26]. The ADM for α-lactalbumin shows that almost the whole region 1–121 exhibits strong compact propensity. However, the A and B helices are not predicted to be a folding unit in this protein, although the ADMs for equine lysozyme and others predict that helices A and B are a possible folding unit. In α-lactalbumin, helices C and D might be a possible stable unit based on the ADM prediction. This property of the ADM might correspond to the high protection of helix C in the molten globule state observed by NMR experiments.

Hence, we can confirm that the features of the positions of the compact regions predicted from ADMs reflect qualitative information on folding, although ADMs cannot predict details of folding such as cooperativity, heterogeneity and so on. In the case of c-type lysozyme family, except

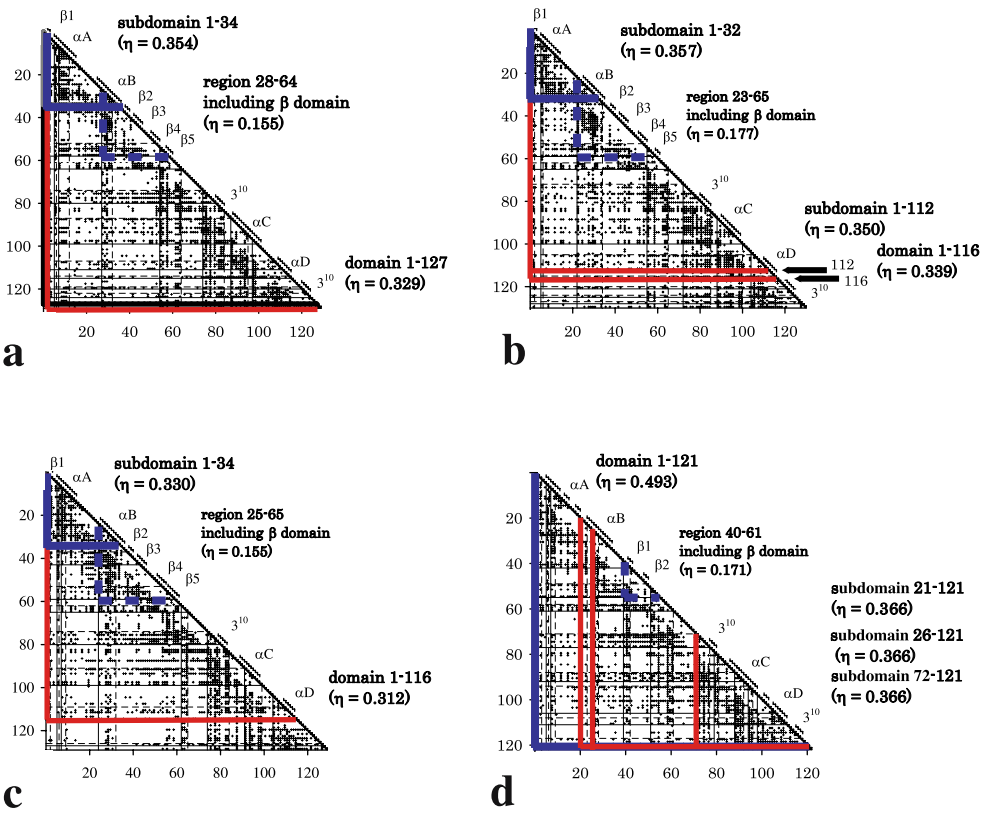

Fig. 5 ADMs for (**a**) hen-egg lysozyme, (**b**) human lysozyme, (**c**) equine lysozyme, and (**d**) bovine α-lactalbumin. The location of the α-helices and β-strands are labeled at the diagonal of each ADM. Regions with the highest η values are enclosed by broad blue lines. Subdomains with relatively high η values are enclosed by red lines. Regions with β strands are enclosed by broken blue lines. The η values of these regions are relatively small. The thin solid lines and the broken lines denote the location of the peaks and valleys obtained by the contact density analyses of ADMs. (see ref. [22])

for α-lactalbumin the common folding mechanism is that the α helices, mainly helices A and B, form at the early stage of folding. α-Lactalbumin also folds at α-helices, but the stable unit during folding might be the C helix. These essential differences in the folding mechanisms are seen in the ADMs.

Information on the folding mechanism of a protein is implicitly coded in its amino acid sequence and can be decoded by converting the sequence to ADM. In particular, the differences in folding processes of proteins in the same family can be detected by local differences in the ADMs. In the cases of the lipid binding proteins [20] and globin family proteins [21], to which our ADM method could be applied successfully, the differences in folding are relatively large. It should be noted that the overall appearance of ADMs of these proteins are similar reflecting the final results of respective folding processes, i.e., native structures or RDMs (This is also observed and discussed in Ref. [20]). The differences in folding processes are observed and predicted from the differences in the features of the locations of sundomains, i.e., differences in local appearances of ADMs (see also ref. [20]). We demonstrated in the present study that the relatively slight differences of folding pathways as seen in the c-type lysozymes can also be detected by ADMs. In other words, the ADM method is a simple method that predicts the folding mechanism of a protein from only its sequence, while the other published techniques to predict the folding mechanism usually require a native structure and detailed simulation. Thus, it is possible with our method to investigate evolutionary relationships in regard to the folding pathways as we partly presented in our recent paper [21].

References

1. Plaxco KW, Simons KT, Baker D (1998) J Mol Biol 277:985–994
2. Plaxco KW, Simons KT, Ruczinski I, Baker DK (2000) Biochemistry 39:11177–11183
3. Baker D (2000) Nature 405:39–42
4. Zarrine-Afsar A, Larson SM, Davidson AR (2005) Curr Opin Str Biol 15:42–49
5. Gunasekaran K, Eyles SJ, Hagler AT, Gierasch LM (2001) Curr Opin Str Biol 11:83–93
6. Nishimura C, Lietzow MA, Dyson HJ, Wright PE (2005) J Mol Biol 351:383–392
7. Portman JJ, Takada S, Wolynes PG (1998) Phys Rev Lett 81:5237–5240
8. Portman JJ, Takada S, Wolynes PG (2001) J Chem Phys 114:5069–5081
9. Portman JJ, Takada S, Wolynes PG (2001) J Chem Phys 114:5082–5096
10. Gromiha MM, Selvaraj S (2001) J Mol Biol 310:27–32
11. Makarov DE, Keller CA, Plaxo KW, Metiu H (2002) Proc Natl Acad Sci USA 99:3535–3539
12. Nölting B, Schalike W, Hampel P, Grundig F, Gantert S, Sips N, Bandlow W, Qi PX (2003) J Theor Biol 223:299–307
13. Weikl TR, Dill KA (2003) J Mol Biol 329:585–598
14. Jiang Z, Zhang L, Chen J, Xia A, Zhao D (2004) Polymer 45:609–621
15. Dixit PD, Weikl TR (2006) Proteins 64:193–197
16. Calloni G, Taddei N, Plaxco KW, Ramponi G, Stefani M, Chiti F (2003) J Mol Biol 330:577–591
17. Gong H, Isom DG, Srinivasan R, Rose GD (2003) J Mol Biol 327:1149–1154
18. Prabhu NP, Bhuyan AK (2006) Biochemistry 45:3805–3812
19. Huang JT, Tian J (2006) Proteins 63:551–554
20. Ichimaru T, Kikuchi T (2003) Proteins 51:515–530
21. Nakajima S, Álvarez-Salgado E, Kikuchi T, Arredondo-Peter R (2005) Proteins 61:500–506
22. Kikuchi T, Némethy G, Scheraga HA (1988) J Protein Chem 7:427–471
23. Kikuchi T (2002) In: Pandalai SG (ed) Recent research developments in protein engineering. Research Signpost, Kerala, India, pp 1–48
24. Radford SE, Dobson CM, Evans PA (1992) Nature 358:302–307
25. Hooke SD, Radford SE, Dobson CM (1994) Biochemistry 33:5867–5876
26. Morozova-Roche LA, Jones JA, Noppe W, Dobson CM (1999) J Mol Biol 289:1055–1073
27. Forge V, Wijesinha RT, Balbach J, Brew K, Robinson CV, Redfield C, Dobson CM (1999) J Mol Biol 288:673–688
28. Chowdhury FA, Fairman R, Bi Y, Rigotti DJ, Raleigh DP (2004) Biochemistry 43:9961–9967
29. Arai M, Ito K, Inobe T, Nakao M, Maki K, Kamagata K, Kihara H, Amemiya Y, Kuwajima K (2002) J Mol Biol 321:121–132