

A Novel Feature with Dynamic Time Warping and Least Squares Adjustment for Protein Structure Alignment[†]

HAN-WEN HSIAO^{1,*}, WEN-HUNG HSIAO², CHENG-KUANG HSU³
AND JEFFREY J. P. TSAI^{2,4}

¹Department of Computer Science and Information Engineering, Asia University, Taiwan

²Department of Biotechnology and Bioinformatics, Asia University, Taiwan

³Department of Health and Nutrition Biotechnology, Asia University, Taiwan

⁴Department of Computer Science, University of Illinois at Chicago, USA

ABSTRACT

Protein structure alignment is of importance in protein study. In general, such a task can be divided into two categories, *i.e.*, global and local structure alignments. In this paper, one-dimensional features are extracted first from the original protein structures. A hybrid approach combining dynamic time warping and least squares adjustment is proposed for global alignment of protein 3D structures in an iterative fashion, where dynamic time warping is responsible for coarse alignment of two structures and least squares adjustment handles the fine matching of amino acid residues. The residuals of matched pairs are utilized to calculate the weights to accelerate the convergence of coarse-to-fine matching. The preliminary results have demonstrated the effectiveness and efficiency of the proposed approach. However, there is still a room for improvement in terms of accuracy and memory usage.

Key words: protein structure alignment, dynamic time warping, least squares adjustment, coarse-to-fine matching.

1. INTRODUCTION

Comparing three-dimensional protein structures is one of the most important issues in structural proteomics and is helpful in solving the problems of protein folding, motif finding, drug design, etc. In general, the task of structure alignment is performed globally or locally. In global alignment, two protein structures are usually aligned by an affine transformation to calculate the root mean square deviation (RMSD) value of three-dimensional coordinates, while local alignment aims at matching segments with maximum local similarity between two structures. For example, Taylor and Orengo (1989) proposed an approach using double dynamic programming for the global alignment problem. Subsequently, their approach (Orengo & Taylor, 1993) was applied to the local alignment problem by using the torsion (ϕ and ψ) angles and solvent accessibility to accelerate the computation of local alignment. Evaluating the structural environment of a residue, however, is difficult. Later, Hiroike and Toh (2001) proposed a method to construct a structural environment that was robust against circular permutation. Akutsu and

[†] This work was partially supported under grant number NSC 93-2745-E-468-007-URD from National Science Council, ROC.

* Corresponding author. E-mail: hwhsiao@asia.edu.tw

Horimoto (2001) proposed a novel approach to multiple local structure alignment by integrating physicochemical characteristics and structural information of protein sequences to form a number of numeric profiles. These profiles were then recoded back to some alphabetic sequences to facilitate the task of local alignment. Lehtonen, Denessiouk, May, and Johnson (1999) developed a tool for the automatic identification of regions of local structural similarity in unrelated proteins having different folds, as well as for defining more global similarities that result from homologous protein structures. Zemla (2003) presented the LGA method for both local and global alignments in sequence dependent and sequence independent modes. It also took into account both local and global structure superposition and worked without a pre-assigned residue correspondence. Standley, Toh, and Nakamura (2004) proposed an alignment method based on maximizing the number of spatially equivalent residues and realigning structures using dynamic programming base on the proximity of residues in the superposition.

In this paper, a hybrid approach is presented to perform global alignment between two protein structures. For each protein, the feature along a three-dimensional structure is extracted first to form a numeric profile. The feature combines with the 3D coordinates to form a sequence of four-dimensional vectors, which is then aligned with that of the other protein by using dynamic time warping. The rough alignment is then refined by least squares adjustment. In the following section, the proposed approach is introduced in detail. Section 3 gives the experimental results and more discussions. Section 4 concludes the work with future improvement.

2. PROTEIN STRUCTURE ALIGNMENT

On the basis of coarse-to-fine matching strategy, the proposed approach consists of three major phases, *i.e.*, feature extraction from protein structures, dynamic time warping for coarse alignment of protein structures, and least squares adjustment for fine matching of amino acid residues. The entire framework is illustrated as a flowchart in Figure 1. Note that the extracted features as well as the original 3D coordinates of protein structures are utilized simultaneously for structure alignment. The advantage of using both original and extracted data is twofold. The extracted features are capable of representing the local characteristic that is more invariant to different coordinate systems, whereas the original 3D coordinates are considered for fine matching and accuracy calculation. The steps are given in more detail in the following.

2.1 Feature Extraction

Consider a structure of an amino acid sequence with a length of M , as illustrated in Figure 2. Each node represents the three-dimensional coordinates of C_{α} atom of an amino acid residue. A series of vectors $v_{i,i-1}$ can be obtained by

calculating the coordinate difference of any two adjacent residues i and $i-1$. This series of vectors can further be utilized to generate a new series of vectors p_i , where each of which is the cross product of two consecutive vectors $v_{i,i-1}$ and $v_{i+1,i}$. Instead of applying the cross product operation again to the series of vectors p_i , the volume pv_i of a parallelepiped formed by any triplet of vectors (p_{i-1}, p_i, p_{i+1}) is calculated as

$$pv_i = p_{i+1} \cdot (p_{i-1} \times p_i). \tag{1}$$

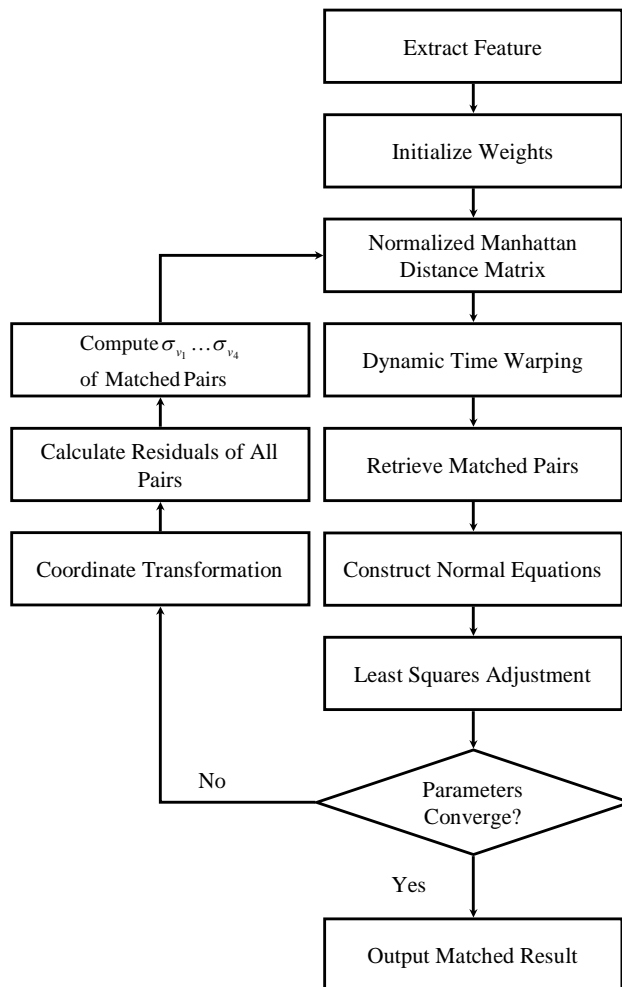


Figure 1. Flowchart of protein structure alignment.

The reason of extracting the parallelepiped volume is obvious. It simply reduces the original structure information in three-dimensional space to a profile in one dimension. Note that the order of calculating the parallelepiped volume for each residue should be consistent. Moreover, the value of the parallelepiped volume is signed, depending on the directions of three vectors. It is worthy of mention that the value of the parallelepiped volume becomes zero if five consecutive residues are aligned along a straight line. In the event that residues are aligned with equal intervals along a perfect helix, such as a spring, the volume profile then forms a function of just a constant for each residue. In any cases other than these two extreme examples, the parallelepiped volume is strongly capable of reflecting the characteristic of a local structure. Without normalizing the norm of all vectors, the value of the parallelepiped volume also indicates the extent of closeness among local residues.

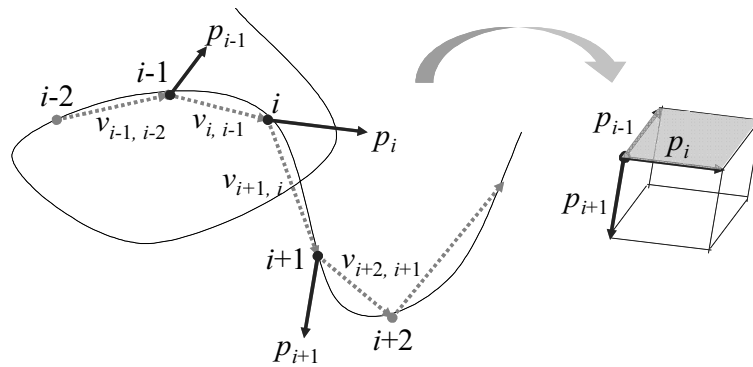


Figure 2. Extraction of parallelepiped volume.

After extracting the parallelepiped volume associated with each amino acid residue, a protein structure is represented by a series of four-dimensional vectors and expressed as $\mathbf{x}_i = [x_{1i} \ x_{2i} \ x_{3i} \ x_{4i}]^T$ and $i = 1, 2, \dots, M$, where the first three components are the 3D coordinates of the C_α atom of the i th amino acid residue and the fourth component is the corresponding parallelepiped volume pv_i . Since the parallelepiped volumes of four amino acid residues, two on each terminal, cannot be calculated, they are padded with pv_3 and pv_{M-2} , respectively. It is under an acceptable assumption that the curvature at each residue near the terminals does not alter dramatically.

2.2 Dynamic Time Warping

Once the profile of the parallelepiped volume along a protein sequence is calculated, the task of structure alignment is carried out mainly by dynamic time

warping (Rabiner, Rosenberg, & Levinson, 1978; Wang & Gasser, 1997). Two similar sequences of four-dimensional vectors are expected to be aligned through the introduction of expansion and contraction. Let two protein structures \mathbf{x} and \mathbf{y} with length of M and N be denoted as $\mathbf{x} = \{\mathbf{x}_1 \dots \mathbf{x}_i \dots \mathbf{x}_M\}$ and $\mathbf{y} = \{\mathbf{y}_1 \dots \mathbf{y}_i \dots \mathbf{y}_N\}$, respectively. An $M \times N$ matrix \mathbf{D} is then constructed, where the matrix element $d_{i,j}$ denotes the normalized Manhattan distance between \mathbf{x}_i and \mathbf{y}_j and will be described later. Let a warping path \mathbf{P} in the matrix \mathbf{D} be a set of contiguous elements $P_k = (i, j)$, where $k = 1, 2, \dots, K$ and $\min(M, N) \leq K \leq M + N - 1$. The relationship between P_k and P_{k-1} is defined as

$$P_k = (i, j) \rightarrow P_{k-1} = (p, q) \tag{2}$$

where $0 \leq (i-p) \leq 1$, $0 \leq (j-q) \leq 1$, $i \leq M$, $p \leq M$, $j \leq N$, and $q \leq N$. Therefore, the aim is to find an optimal warping path having the minimum accumulated distance between two structures by evaluating the recursive equation

$$D_{i,j} = d_{i,j} + \min(D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}) \tag{3}$$

2.3 Least Squares Adjustment

Although dynamic time warping is a very advantageous approach relevant to sequence matching problems and has been employed in a wide variety of areas, the outcome of protein structure alignment using its standard version may not be satisfactory. The reason partially lies in the ignorance of adjacency between residues. In other words, the optimal warping path does not take into account the fact that any little difference between two local structures that should be matched will result in a deviation from a perfectly matched path. It is because the algorithm always tries to find a path with the minimum distance (or difference). It turns out that the aligned structures may contain one-to-many matching. To remedy the potential problem mentioned above, the matched result is regarded as a rough alignment, which provides a number of one-to-one matched pairs of points for solving the transformation parameters.

In general, it suffices to transform protein structures by rotations and translations. Without loss of generality, suppose there are N_m pairs of one-to-one matched points extracted from the warping path \mathbf{P} . The transformation $\mathbf{z} = T(\mathbf{y})$ can be expressed as

$$\begin{bmatrix} z_{1s} \\ z_{2s} \\ z_{3s} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} y_{1s} \\ y_{2s} \\ y_{3s} \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}, \quad s = 1, 2, \dots, N_m. \tag{4}$$

The optimal transformation parameters can be solved by least squares adjustment with redundant observations, which are the matched pairs of points. The energy function is defined as

$$E = \sum_{s=1}^{N_m} \sum_{k=1}^4 v_{ks}^2 = \sum_{s=1}^{N_m} \sum_{k=1}^4 (x_{ks} - z_{ks})^2. \quad (5)$$

After minimization of the energy function, the normal equations are constructed in a matrix form as $\mathbf{AX} = \mathbf{B}$, where \mathbf{A} is the 12×12 design matrix, \mathbf{B} is the column vector of observation, and \mathbf{X} is the column vector of 12 parameters. Hence, the parameters solved at iteration t can be expressed as

$$\mathbf{X}_t = (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{B}). \quad (6)$$

The structure of the test protein needs to be transformed into the coordinate system of the reference protein by using the solved transformation parameters. This step provides an approximation between two coordinate systems for the following fine matching. The result of fine matching is evaluated by a root mean square deviation (RMSD) given as

$$RMSD = \sqrt{\frac{1}{N_m} \sum_{s=1}^{N_m} |\mathbf{x}_s - \mathbf{z}_s|^2}. \quad (7)$$

The convergence criterion for the least squares adjustment is then to examine if the relative change of the RMSD at the end of iteration t is within an insignificant tolerance ε , *i.e.*,

$$\left| \frac{RMSD_t - RMSD_{t-1}}{RMSD_t} \right| \leq \varepsilon \quad (8)$$

where the tolerance ε is given as 10^{-4} in this study.

2.4 Normalized Manhattan Distance Matrix

In the step of dynamic time warping for coarse alignment, searching the optimal path is based on the normalized Manhattan distance defined as

$$d_{i,j} = \sum_{k=1}^4 w_k |x_{ki} - z_{kj}| \quad (9)$$

where

$$w_k = \frac{1}{\sigma_{v_k}}; \quad \sigma_{v_k}^2 = \frac{1}{N_m - 1} \sum_{s=1}^{N_m} (x_{ks} - z_{ks})^2.$$

In other words, the Manhattan distance between any pair of \mathbf{x}_i and \mathbf{y}_j from two proteins` respectively is normalized by four weights $\mathbf{w} = [w_1, w_2, w_3, w_4]$ defined as

the inverse of the standard deviations of the matched pairs. At the beginning of least squares adjustment, the weights are initialized as $[0, 0, 0, 1]$, *i.e.*, the first three components representing the differences of 3D coordinates cannot be utilized due to different coordinate systems, while the fourth component is treated equally. As two coordinate systems get closer, the weights gradually reflect the importance of the matched pairs.

3. RESULTS AND DISCUSSION

A program has been implemented in MATLAB on a laptop computer equipped with a Pentium 1.6 GHz processor and 1.5GB RAM. All protein structures in this study are obtained from the Protein Data Bank (Berman et al., 2000). Twelve pairs of protein structures with various types of similarity measures (Orengo & Taylor, 1993; Shindyalov & Bourne, 1998; Chen, Zhou, & Tang, 2005), *i.e.*, global similarity, local similarity, and difficulty to be aligned, were used here for demonstration. Each pair of protein structures with PDB IDs, protein lengths (in parenthesis), sequence identity, the final RMSD value, the number of aligned residues N_m , and computation time are provided in Table 1. Furthermore, an additional pair of proteins with completely different structures is arbitrarily selected for comparison. These aligned results are also compared with those by the incremental combinatorial extension (CE) method (Shindyalov & Bourne, 1998). Figure 3 demonstrates four convergence profiles for different cases marked in gray in Table 1, respectively. In both of the globally and locally similar cases, the proposed approach rapidly converges to a stable state with a minimum RMSD. In the third case where two proteins are identified as less similar and difficult to be aligned, the alignment is not satisfactory and the final RMSD is larger. Two proteins in the fourth case are totally different and should not be aligned correctly. It is expected that the hybrid approach still tries to find a best match but the final RMSD value is obviously higher. A further discussion about these three cases (global, local, and difficult) is given below.

Table 1. Comparison of structure alignments for 13 protein pairs

Case	Test	Reference	Identity(%)	Time	Proposed		CE	
					RMSD(Å)	N_m	RMSD(Å)	N_m
Global	1DHFa (186)	3DFR (162)	27.8	7.48s	1.67	139	1.7	158
	1ATPe (350)	2CPKe (350)	100.0	15.19s	0.37	334	0.4	336
	1CDKa (350)	1CMKe (351)	100.0	24.20s	1.82	337	2.1	343
	1BPI (58)	1BUNb (61)	34.5	0.79s	1.54	52	1.5	55
Local	3ICB (75)	4CPV (109)	24.6	1.66s	3.32	66	3.4	65
	1PSM (38)	1LBD (282)	12.5	1.25s	3.40	29	0.6	24
	2ASR (142)	2BRD (247)	7.7	5.47s	2.89	84	4.3	117
	4ICB (76)	1CTDa (36)	29.0	0.56s	3.59	32	1.5	31
Difficult	1BGEb (175)	2GMFa (127)	12.1	4.35s	5.15	74	3.9	107
	1CEWi (108)	1MOLa (94)	17.3	2.09s	2.50	55	2.3	81
	1FXIa (96)	1UBQ (76)	9.4	1.18s	3.50	47	3.8	64
	2AZAa (129)	1PAZ (123)	11.9	3.80s	5.87	50	2.9	84
Different	1XWM (217)	1Q3B (262)	13.8	28.93s	7.53	77	5.3	80

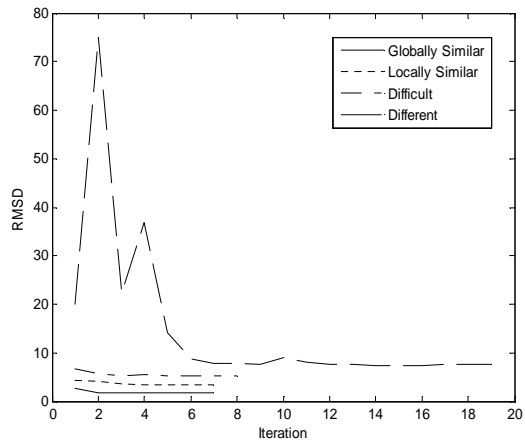


Figure 3. Convergence profiles for the four cases of “globally similar” (1DHFa/3DFR), “locally similar” (3ICB/4CPV), “difficult” (1BGEb/2GMFa), and “different” (1XWM/1Q3B), respectively.

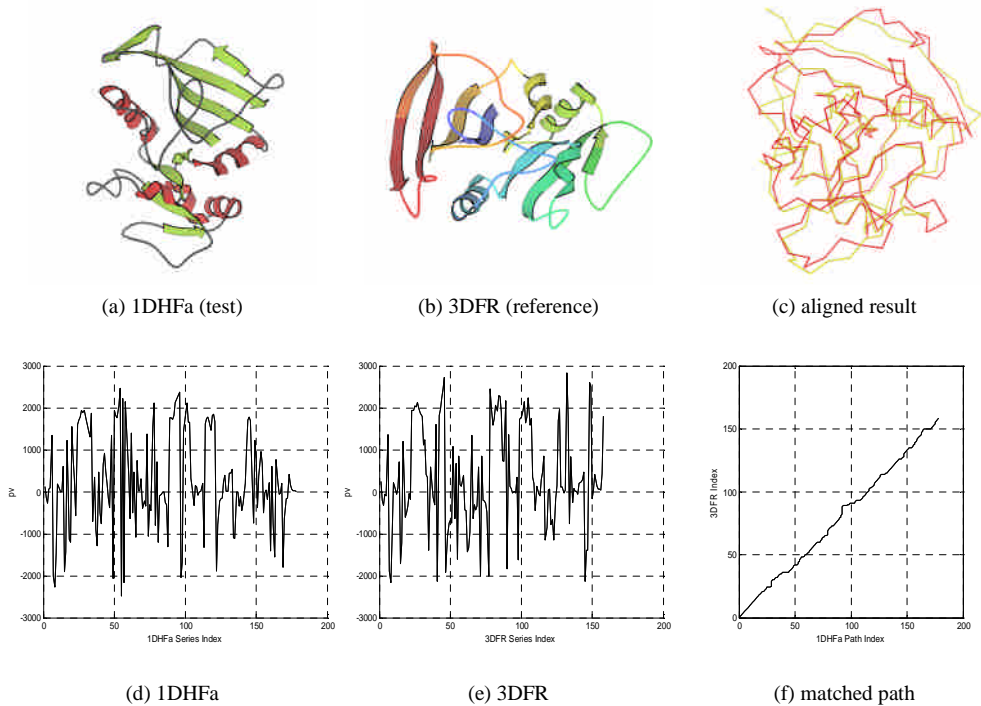


Figure 4. Two protein structures with global similarity. Protein structures drawn in darker and lighter gray in (c) represent the test and reference structures, respectively. (d) and (e) are the corresponding profiles of p_v values, and (f) is the matched path by dynamic time warping.

Two globally similar proteins 1DHFa and 3DFR are compared, as shown in Figure 4. The test protein in Figure 4(c) is transformed and superposed on the reference one for comparison. The profiles of parallelepiped volume ($p\nu$) are shown in Figures 4(d) and 4(e), respectively, where the x -axis indicates the residue index along the protein. As shown in Figure 4(f), the final matched path has been improved by dynamic time warping with least squares adjustment. The aligned result is rather perfect and comparable with those of the CE (Shindyalov & Bourne, 1998) and VAST data banks (Gibrat Madej, & Bryant, 1996). Two locally similar protein structures evaluated by Orengo and Taylor (1993) were selected for this study as well. The result of alignment in Figure 5 is satisfactory but not as accurate as that of the globally similar one.

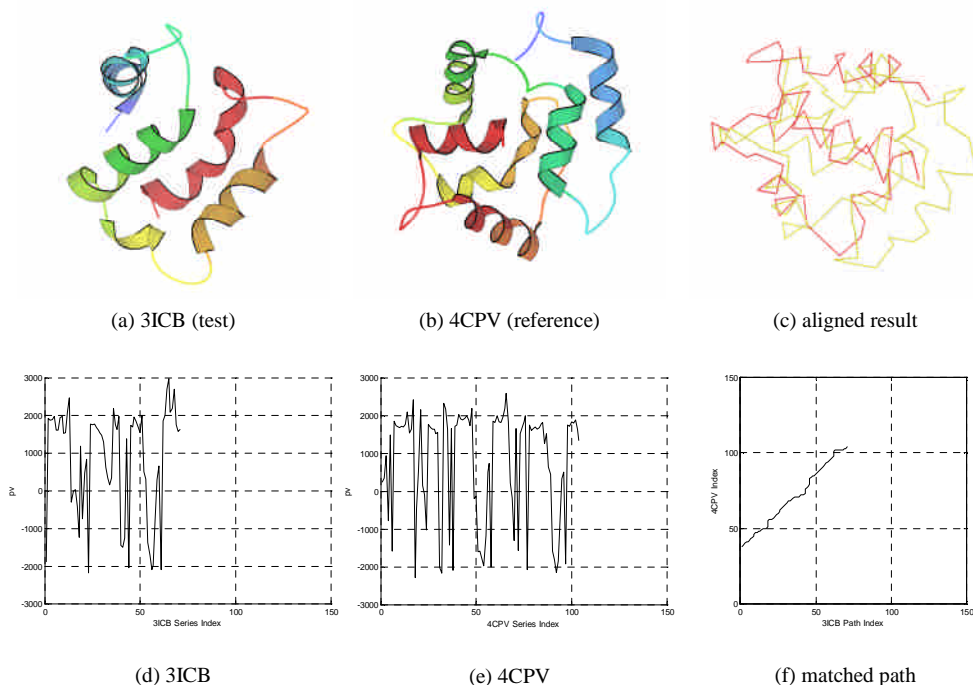


Figure 5. Two protein structures with local similarity. Protein structures drawn in darker and lighter gray in (c) represent the test and reference structures, respectively. (d) and (e) are the corresponding profiles of $p\nu$ values, and (f) is the matched path by dynamic time warping.

As mentioned previously and shown in Figure 6, two protein structures 1BGEb and 2GMFa reported in the literature (Fischer, Elofsson, Rice, & Eisenberg, 1996; Shindyalov & Bourne, 1998; Chen et al., 2005) are considered to be rarely similar and difficult for alignment. The ambiguous result shows that the alignment is a failure. After thorough examination of the experiment, two factors are identified. First, the proposed approach treats protein structures as rigid bodies and is not appropriate for the case of inexact alignment. A compromising solution for

coordinate transformation may not be an optimal one. Second, too many small pieces of matched structures are similar in terms of local features, not the actual coordinates. Accordingly, these matched points are not representative enough to put together for solving transformation parameters by using least squares adjustment.

More pairs of protein structures are provided in Figures 7, 8, and 9 for the cases of global, local, difficult alignments, respectively. Compared with those by the CE method listed in Table 1, the results of global alignment shown in Figure 7 demonstrate the effectiveness and efficiency of the proposed approach. In the cases of local alignment shown in Table 1 and Figure 8, the unsatisfactory results reveal that protein structures mainly consisting of α -helix structures will result in an ambiguous matching. Protein structures identified to be difficult for alignment are illustrated in Figure 9. The results show that these pairs are in fact aligned; however, the RMSD values are larger. The obvious reason is due to the lack of elastic matching for protein structures with less similarity.

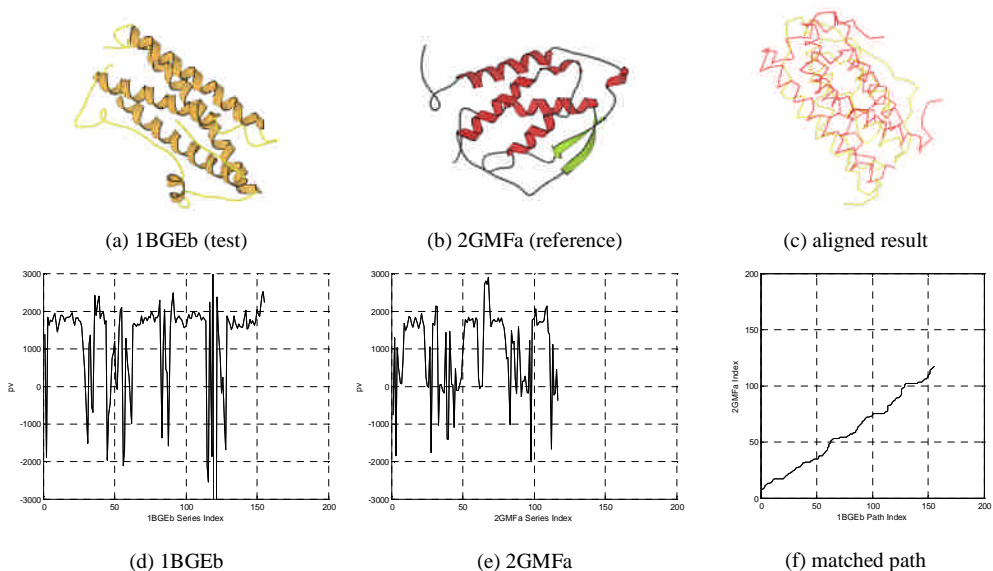
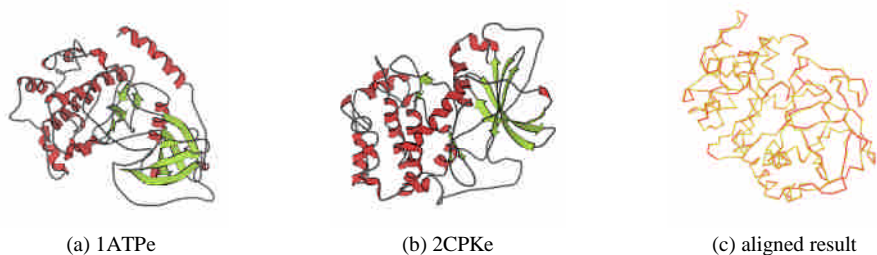


Figure 6. Two protein structures are difficult to be aligned. Protein structures drawn in darker and lighter gray in (c) represent the test and reference structures, respectively. (d) and (e) are the corresponding profiles of pv values, and (f) is the matched path by dynamic time warping.



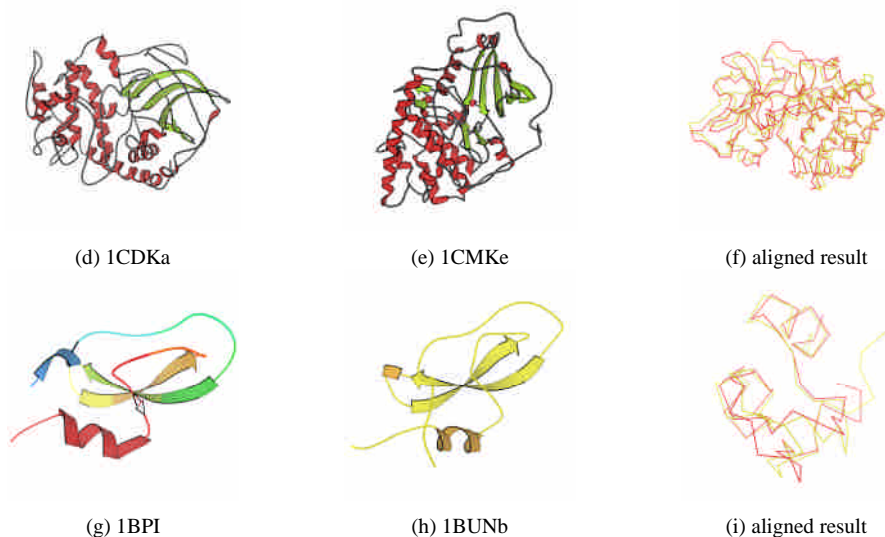


Figure 7. Three pairs of protein structures with global similarity. The first column indicates the test structures, while the second one shows the reference structures. Protein structures drawn in darker and lighter gray in the third column represent the test and reference structures, respectively.

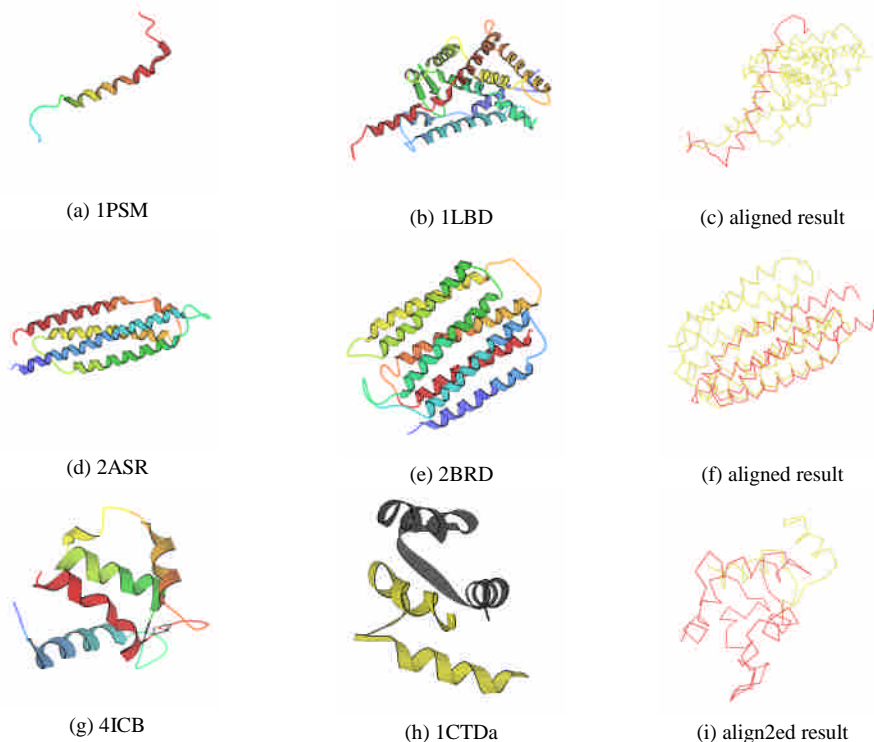


Figure 8. Three pairs of protein structures with local similarity.

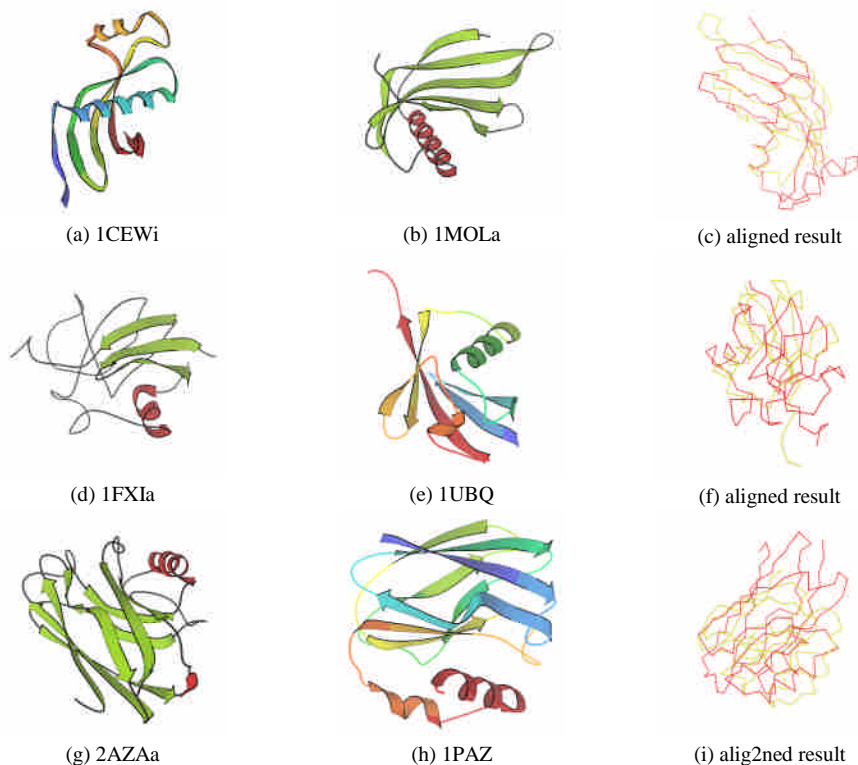


Figure 9. Three pairs of protein structures identified as difficult to be aligned.

4. CONCLUSION

In this paper, a hybrid approach to the problem of global alignment of protein structures has been presented. The representative feature of each amino acid residue is extracted by calculating the volume of a parallelepiped derived from the coordinates of five consecutive residues. Dynamic time warping is the key step for approximate alignment. Least squares adjustment provides a chance to refine the matched result, which then feeds back to the step of approximate alignment in an iterative fashion. It is expected that two similar structures can be optimally matched in only a few iterations. Thirteen pairs of protein structures with different measures of similarity are aligned for demonstration in this paper. The preliminary results are satisfactory. The proposed feature, in general, is representative, but cannot handle the ambiguous cases of too many α -helix structures. More works regarding accuracy evaluation instead of using the RMSD should be done soon. An automatic mechanism for filtering out dissimilar structures is necessary as well. However, there is still a room to improve the structure alignment in terms of accuracy and memory usage. Attention will be paid to the problems when locally similar structures in two proteins are located in different orders, which cannot be solved by

dynamic time warping. Elastic or non-rigid alignment of two structures should be addressed in the future work.

REFERENCES

- Akutsu, T., & Horimoto, K. (2001). Local multiple alignment of numerical sequences: Detection of subtle motifs from protein sequences and structures. *Proceedings of the 12th International Conference on Genome Informatics, 12*, 83-92. Tokyo, Japan.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242.
- Chen, L., Zhou, T., & Tang, Y. (2005). Protein structure alignment by deterministic annealing. *Bioinformatics*, 21(1), 51-62.
- Fischer, D., Elofsson, A., Rice, D., & Eisenberg, D. (1996). Assessing the performance of fold recognition methods by means of a comprehensive benchmark, *Proceedings of the Pacific Symposium on Biocomputing*, 300-318, Hawaii, USA.
- Gibrat, J. F., Madej, T., & Bryant, S. H. (1996). Surprising similarities in structure comparison. *Current Opinion in Structure Biology*, 6(3), 377-385.
- Hiroike, T., & Toh, H. (2001). A local structural alignment method that accommodates with circular permutation. *Chem-Bio Informatics Journal*, 1(3), 103-114.
- Lehtonen, J. V., Denessiouk, K. A., May, A. C. W., & Johnson, M. S. (1999). Finding local structural similarities among families of unrelated protein structures: a generic nonlinear alignment algorithm. *Proteins: Structure, Function, and Genetics*, 34(3), 341-355.
- Orengo, C. A., & Taylor, W. R. (1993). A local alignment method for protein structure motifs. *Journal of Molecular Biology*, 233(3), 488-497.
- Rabiner, L., Rosenberg, A., & Levinson, S. (1978). Considerations in dynamic time warping algorithms for discrete word recognition. *IEEE Transactions on Acoustics Speech and Signal Processing*, 26(6), 575-582.
- Shindyalov, I. N., & Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering*, 11(9), 739-747.
- Standley, D. M., Toh, H., & Nakamura, H. (2004). Detecting local structural similarity in proteins by maximizing number of equivalent residues. *Proteins: Structure, Function, and Bioinformatics*, 57(2), 381-391.
- Taylor, W. R., & Orengo, C. A. (1989). Protein structure alignment. *Journal of Molecular Biology*, 208(1), 1-22.
- Wang, K., & Gasser, T. (1997). Alignment of curves by dynamic time warping. *The Annals of Statistics*, 25(3), 1251-1276.
- Zemla, A. (2003). LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Research*, 31(13), 3370-3374.



Han-Wen Hsiao received a Ph.D. degree in civil engineering with a specialization in geoinformatics from the University of Illinois at Urbana-Champaign (UIUC) in 1999. During his graduate study in the US, he worked respectively for the Robotics Center and the Business Process Division of Construction Engineering Research Laboratories, Corps of Engineers, US Army. Later, he also participated in the UIUC Digital Library Initiative (DLI) Project for more than two years. After graduation, he then joined as a postdoctoral researcher in the information system research team of the Office of the National Science and Technology Program for Hazards Mitigation. He has been with Asia University (formerly Taichung Healthcare and Management University) since 2001. He is currently an Assistant Professor in the Department of Computer Science and Information Engineering, and affiliated with the Department of Biotechnology and Bioinformatics. His research interests include pattern recognition, data mining, image processing, bioinformatics, and geoinformatics.



Wen-Hung Hsiao has an undergraduate major in the field of applied mathematics at National Chung Hsing University, Taichung, Taiwan. He has worked as an assistant maintenance engineer for three years, and is currently pursuing his Master's degree in bioinformatics at Asia University. His research interests include pattern recognition and bioinformatics.



Cheng-Kuang Hsu received a Ph.D. degree in bioresource engineering from Oregon State University in 1995. He worked as a postdoctoral researcher at National Taiwan University for two years and as a research fellow with the Food Industry Research and Development Institute for three and half years. He is currently an Associate Professor and the head of the Department of Health and Nutrition Biotechnology, Asia University, Taiwan, ROC. His research interests include food engineering, food biotechnology and bioinformatics.



Jeffrey J. P. Tsai received a Ph.D. degree in computer science from Northwestern University, Evanston, Illinois. He is a Professor in the Department of Electrical Engineering and Computer Science at the University of Illinois at Chicago, where he is also the director of the Distributed Real-Time Intelligent Systems Laboratory. He co-authored *Knowledge-Based Software Development for Real-Time Distributed Systems* (World Scientific, 1993), *Distributed Real-Time Systems: Monitoring, Visualization, Debugging, and Analysis* (John Wiley & Sons, Inc., 1996), *Compositional Verification of Concurrent and Real-Time Systems* (Kluwer, 2001), coedited *Monitoring and Debugging Distributed Real-Time Systems* (IEEE/CS Press, 1995), and has published extensively in the areas of knowledge-based software engineering, software architecture, requirements engineering, formal methods, agent-based systems, and distributed real-time systems. Dr. Tsai was the recipient of a University Scholar Award from the University of Illinois in 1994 and was presented a Technical Achievement Award from the IEEE Computer Society in 1997. He is currently the Co-Editor-in-Chief of the *International Journal of Artificial Intelligence Tools*. He is also an Editor of the *Annals of Software Engineering*, the *International Journal of Software Engineering and Knowledge Engineering*, and the *International Journal of Systems Integration*. He is a fellow of the IEEE, the AAAS, and the SDPS.