Structural bioinformatics and data mining

DTA: dihedral transition analysis for characterization of the effects of large main-chain dihedral changes in proteins

Wataru Nishima¹, Guoying Qi², Steven Hayward²,³ and Akio Kitao⁴,⁵,*

¹Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Japan,
²School of Computing Sciences, ³School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK,
⁴Institute of Molecular and Cellular Biosciences, The University of Tokyo and ⁵JST, CREST, 1-1-1 Yayoi, Bunkyo, Tokyo 113-0032, Japan

Received on November 4, 2008; revised on December 25, 2008; accepted on January 10, 2009
Advance Access publication January 15, 2009
Associate Editor: Anna Tramontano

ABSTRACT

Motivation: The biological function of proteins is associated with a variety of motions, ranging from global domain motion to local motion of side chain. We propose a method, dihedral transition analysis (DTA), to identify significant dihedral angle changes between two distinct protein conformations and for characterization of the effect of these transitions on both local and global conformation.

Results: Applying DTA to a comprehensive and non-redundant dataset of 459 high-resolution pairs of protein structures, we found that a dihedral transition occurs in 82% of proteins. Multiple dihedral transitions are shown to occur cooperatively along the sequence, which allows us to separate a polypeptide chain into fragments with and without transitions, namely transition fragments (TFs) and stable fragments (SFs), respectively. By characterizing the magnitude of TF conformational change and the effect of the transition on the neighboring fragments, flip and hinge motions are identified as typical motions. DTA is also useful to detect protein motions, subtle in RMSD but significant in terms of dihedral angle changes, such as the peptide-plane flip, the side-chain flip and path-preserving motions. We conclude that DTA is a useful tool to extract potential functional motions, some of which might have been missed using conventional methods for protein motion analysis.

Availability: http://dynamics.iam.u-tokyo.ac.jp/DTA/
Contact: kitao@iam.u-tokyo.ac.jp
Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

The biological function of proteins, such as substrate binding, product release and allosteric behavior, and contraction, is often associated with significant structural changes (Berendsen and Hayward, 2000). By comparing protein structural differences in the Protein Data Bank (PDB), Gerstein et al. (1994) categorized structural change in domain proteins into two main types, hinge and shear. Kempner (1993) classified motions upon ligand or substrate binding into global (motion of two massive domain joined by a flexible loop) and local motions (a motion of a short segment of amino acids as flexible loop). Similarly, Karplus (2000) also discussed conformational change of specific proteins, categorizing motions into lid opening and closing motions, hinge-bending motions and relative motions of subunits. Interestingly, in most enzymes, induced fit motions are relatively small (RMSD across the whole protein ≤ 1.0 Å) (Gutteridge and Thornton, 2005).

To characterize the global motion of proteins, several approaches have been successful in describing significantly deformable regions in proteins and relatively rigid regions (typically called ‘domains’) (Flores and Gerstein, 2007; Flores et al., 2007; Hayward and Berendsen, 1998; Hayward et al., 1997; Hinsen, 1998; Lee et al., 2003; Maiovor and Abagyan, 1997; Shatsky et al., 2002; Wriggers and Schulten, 1997). However, these methods are not necessarily suitable to detect local motions, which may also play a key role in function. Since a set of dihedral angles is nearly sufficient to specify a protein’s structure, their variation can also be used to characterize structural change (Mao and McCammon, 1984). Korn and Rose (1994) systematically applied this idea to ~20 proteins and demonstrated that large dihedral changes often indicate local structural changes associated with ligand binding as well as the effects of crystal packing. Similarly more coarse-grained methods have used the dihedral angle of four consecutive Cα atoms (Cα torsion angle, or pseudo torsion angle) to characterize structural change in several proteins having hinge or shear motions (Flocco and Mowbray, 1995; Yan et al., 1999). Interestingly, a subtle change of main chain produced by a combination of changes in two consecutive Cα torsion angles can result in a larger two-state change of side-chain orientation (the so-called ‘backrub’ motion) (Davis et al., 2006). Consideration of Cα torsion angle changes are effective in such cases, however, very localized motions such as the peptide-plane flip (Hayward, 2001) cannot easily be detected using the pseudo torsion.

In this article, we propose a methodology, dihedral transition analysis (DTA), for the analysis of significant changes of main-chain dihedral angles and their effect in proteins. Since multiple dihedral transitions tend to occur in neighboring residues along the amino acid sequence as shown later, DTA is designed to identify consecutive peptide fragments with and without dihedral transitions, namely transition fragments (TFs) and stable fragments (SFs), respectively, and to characterize these fragments with three properties: fragment length, main-chain RMSD between the corresponding regions in the two structures and the non-locality
2 METHODS

2.1 The DTA procedure

The overall procedure is shown in Figure 1 (the details of each step are described later). Step 1: for a given protein structural pair, transitions of main-chain dihedral angles, φ, ψ and ω are assigned. Step 2: the polypeptide chain is divided into ‘TFs’ and ‘SFs’, defined as consecutive peptide fragments with and without dihedral transitions, respectively. Step 3: using a clustering algorithm, several SFs having relatively rigid structure (small RMSD) are hierarchically grouped into consecutive or disconnected regions called stable regions (SRs). Step 4: TFs, SFs and SRs are characterized using three properties: fragment length, RMSD and NLS.

2.2 Definition of dihedral transition

Dihedral transitions are defined by the changes of three main-chain dihedral angles, φ, ψ and ω, from three points of view: the (φ, ψ)-view defined by two consecutive main-chain dihedral angles φ(i) and ψ(i) (φ and ψ of the i-th residue), the (ψ, φ)-view defined by ψ(i) and φ(i+1) and the ω-view defined by ω(i). The (φ, ψ)-view shows the dihedral angle changes around the amino acid side-chain and the (ψ, φ)-view indicates the change around the peptide plane. In both of the (φ, ψ) and (ψ, φ)-views, dihedral transitions are defined considering both chemical properties of dihedral angle changes around each main-chain chemical bond and a correlated effect of φ and ψ angle changes. We regard that the dihedral transition has taken place when at least one of the following conditions is satisfied for the angle changes, Δφ and Δψ:

\[
\begin{align*}
|\Delta\phi| & \geq 120^\circ, \\
|\Delta\psi| & \geq 120^\circ, \\
|\Delta\phi + \Delta\psi| & \geq 120^\circ 
\end{align*}
\]

See Figure 2A and B for these conditions. A statistical analysis on Δφ, Δψ and Δφ + Δψ distributions showed local minima around ±120° (Supplementary Fig. S1). For visualization purposes, a cumulative (summing over residues) plot of Δφ + Δψ is useful (Supplementary Fig. S2). For the angle ω, a dihedral transition is defined to be taken place under the condition of cis–trans isomerization,

\[|\Delta\omega| \geq 90^\circ\]

The angles with dihedral transitions are termed ‘transition dihedrals’ in this article.

2.3 Flip motions: s-, p- and ω-flips

Among the regions with dihedral transitions, we further define the regions with a ‘side-chain flip’ (s-flip) in the (φ, ψ)-view and the ‘peptide-plane flip’ (p-flip) in the (ψ, φ)-view if |Δφ| ≥ 120° and |Δψ| ≥ 120°. Note that the ‘s-flip’ in this article means reorientation of side-chain direction due to a main-chain dihedral transition and not one associated with any side-chain χ-angle change (Word et al., 1999). An s-flip at φ(i) − ψ(i) followed by a p-flip at ψ(i) − ϕ(i+1) is termed an ‘sp-flip’ and a p-flip [ψ(i) − ϕ(i+1)] followed by an s-flip [ϕ(i+1) − ψ(i+1)] is called a ‘ps-flip’ in this article. Furthermore, the transition of angle ω[|Δω| ≥ 90°], which corresponds to cis–trans isomerization for a peptide bond, is called an ‘ω-flip’.

2.4 TF, SF and SR

We divide a polypeptide chain into ‘TFs’ and ‘SFs’, defined as consecutive peptide fragments with and without a dihedral transition, respectively (see Supplementary Material). After dihedral transitions are identified, TFs are selected as the fragments consisting of transition dihedrals within a specified gap length, \(p_{gap}\). In a TF, the number of consecutive φ and ψ angles between two consecutive transition dihedrals does not exceed \(p_{gap}\). Consecutive non-TF elements are considered to be SFs, comprising at least \(p_{gap}\) dihedral angles. The value of \(p_{gap}\) is variable in the database but it is set to be 10 unless stated otherwise. SFs are hierarchically clustered using a cutoff value for the RMSD (default 1.0 Å), to form ‘SRs’ which are considered to form a relatively rigid group of one or multiple SFs (see Supplementary Material).

2.5 Properties of TF, SF and SR: fragment length, RMSD and NLS

The fragment length (the number of residues involved) and the main-chain RMSD (\(N, C^\alpha\) and \(C\)) between the corresponding regions in the two structures are used to characterize TFs, SFs and SRs. The latter is used to quantify the flexibility of the fragments or the regions. For TFs, we further introduce a NLS, which is a quantity that is able to quantify the non-local effect of a dihedral transition on neighboring SFs (see Supplementary Material). If the NLS is relatively small, the effect of dihedral transition in the TF is considered to be localized.

2.6 Comprehensive and non-redundant dataset

A comprehensive and non-redundant dataset of 459 high-resolution protein structural pairs (≤2.0 Å), having identical sequences, were chosen from the PDB and used for the DTA. This dataset was derived from a dataset where...
conformational clustering was performed on sets of structures within a family and two representative structures with RMSD > 0.5 Å were selected. This non-redundant dataset consisting of 3120 pairs of protein structures from different protein families (Qi et al., 2005), was further filtered with the above sequence and resolution conditions. The resolution criterion (≤ 2.0 Å) is sufficient to eliminate spurious causes of apparent dihedral transitions. This dataset includes structural pairs from two distinct PDBs as well as two distinctive chains from the same PDB file. In all, 105,951 residues with 317,853 main-chain dihedral angles were examined by DTA. TFs next to the N- and C-termini and unresolved regions are excluded from the following analysis.

3 RESULTS

3.1 Properties of dihedral angle change

3.1.1 Statistics of dihedral transitions In this database the percentages of the dihedral transitions as defined in Section 2 in the (φ, ψ)-, (ψ, φ)- and α-views are 1.38%, 1.11% and 0.04%, respectively. The distributions of all values of Δφ(i) and Δψ(i) in the dataset in the (φ, ψ)- and (ψ, φ)-views are shown in Figure 2A and B, respectively. Three clusters are identified: in the (φ, ψ)-view (Fig. 2A), around a [Δφ(i) = Δψ(i) = 0°], b [Δφ(i) = 0°, Δψ(i) = ±180°] and c [Δφ(i) = ±180°, Δψ(i) = 0°] and in the (ψ, φ)-view (Fig. 2B) around a, b and d [Δφ(i) = ±180°, Δψ(i) = ±180° or Δφ(i) = ±180°, Δψ(i) = ±180°]. Note that clusters indicated by the same letter, b, c and d are connected through the periodic boundary. Comparing these two plots, the (ψ, φ)-view has a slightly more elongated pattern toward Δφ(i+1) = −Δψ(i) direction, originating from a cooperative local cancellation of the dihedral movement around the peptide plane (e.g. a p-flip, see Section 2) as shown by the higher correlation coefficients which are −0.08 between Δφ(i) and Δψ(i) and −0.23 between Δψ(i) and Δφ(i+1). Although the clusters around a and b [Δφ(i) = 0°, Δψ(i) = ±180°] are common in both views, the clusters around d and c are absent in the (φ, ψ) - and (ψ, φ)-views, respectively, indicating that the major patterns are the ‘ψ-only’ transition and the ‘p-flip’. The ψ-only transitions contribute to cluster b in both the (φ, ψ)- and (ψ, φ)-views. P-flips which are the reason for cluster d in the (ψ, φ)-view also contribute to cluster b and c in the (φ, ψ)-view. In comparison to the p-flip (percentage of occurrence, 0.29%) the occurrence of the s-flip is lower (0.09%), cluster d being sparsely populated in the (φ, ψ)-view.

For a given dihedral angle, φ, ψ or ω, the percentage involved with a transition are 1.72, 1.58 and 0.04%, respectively. For a given residue i, the percentage involved with a dihedral transition through ω(i−1), φ(i), ψ(i) or ω(i) is 1.90%. Given a transition, the percentage of cases where the transition involved a particular amino acid residue or a secondary structure type was calculated and is shown in Supplementary Figures S3A and B, respectively. In all types of transitions, GLY has the highest percentage, consistent with the fact that a φ > 0 value is more likely than for other amino acids. Although PRO has strong steric constraint to prevent it from occupying the φ > 0 area (Ho and Brassier, 2005; Mandel et al., 1977), its percentage in the ‘ψ-only’ transition is relatively high. Turn (T) and coil (C) tend to favor transitions much more than α-helix (H) and β-strand (E) (Supplementary Fig. S3). Typically, the p-flip in the β-turn causes a transition between type I and type II or type I′ and type II′ β-turns as pointed out by Gunasekaran et al. (1998).

3.1.2 The s-, p- and ω-flips The percentages for a given residue being involved with the s- and p-flips are 0.09% (91/104133) and 0.29% (304/104762), respectively. Since the p-flip involves two dihedral angles around two almost parallel axes, the Cα(i)−Cα(i−1) and N(i+1)−Cα(i) bonds, it can take place as relatively local conformational change (Hayward, 2001) and is also frequently observed in this work. However, the s-flip is much less likely as already shown. In the s-flip, the two rotation axes, the bonds N(i)−Cα(i) and Cα(i)−N(i+1), are not parallel and local cancellation may not be perfect even with considerable bond angle adjustments. Furthermore, protruding side chains should sterically suppress the

Fig. 2. Distribution of dihedral angle changes, Δφ, Δψ and Δω, sampled from the dataset containing 459 pairs of proteins (see text) and definition of dihedral transitions. (A and B) Ramachandran-type plot for dihedral changes for the (A) (φ, ψ)-view and (B) (ψ, φ)-view. Each gray dot represents Δφ and Δψ. Solid lines indicate the boundary of the transitions [Equations (1–3)]. Dashed lines show the boundary between the flips and other transitions (see text). (C) Probability distribution of |Δω| in the dataset. The vertical line at |Δω| = 90° shows the threshold for a dihedral transition in the ω-view.
As expected dihedral transitions tend to occur concomitantly with a high percentage, justifying the grouping of a polypeptide chain into TFs based on a gap length between transitions.

### 3.2 Statistics of TF, SF and SR

For the default value of $p_{\text{gap}} = 10$, the total number of TFs and SFs is 523 and 1287, respectively. The average number of TFs and SFs in one protein is 1.14 and 2.80, respectively. The distributions of fragment lengths, RMSDs and NLS for TF and SF are shown in Supplementary Figure S4. The average fragment length, RMSD and NLS for SFs are 3.9 residues (7.8 $\phi, \psi$ angles), 1.18 Å and 3.01 Å, respectively. The number of proteins with no dihedral transition is 83. The correlation of dihedral transitions along the polypeptide chain, the percentage of cases where the occurrence of a dihedral transition in residue $i$ is shown in Figure 3.

### 3.3 Typical motions found by TF ranking: hinge, flap, path-preserving motions

#### 3.3.1 Hinge motion

Table 2 shows the Top10 TFs ranked by NLS values in descending order. Interestingly, all the TFs can be considered to act as hinges. Among the Top10 TFs, six act as hinges in a global hinge bending motion connecting two SRs (e.g. Fig. 4A), the other four act as local hinges. Note TF #4 and #7 are parts of the same long flexible loop from the same protein pair (e.g. Fig. 4B). TF #8 is also involved in a local hinge motion of a flexible loop, and the remaining TF (#4) is a local hinge motion near the N-terminus.

In contrast to the classical view of hinge bending motions as high percentage, justifying the grouping of a polypeptide chain into TFs based on a gap length between transitions.

### Table 1. The number of s- and p-flip occurrences classified by amino acid type

<table>
<thead>
<tr>
<th>S-flip</th>
<th>P-flip</th>
<th>#(?–X)</th>
<th># (X–?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLY</td>
<td>PRO</td>
<td>38</td>
<td>GLY</td>
</tr>
<tr>
<td>SER</td>
<td>ALA</td>
<td>32</td>
<td>ASP</td>
</tr>
<tr>
<td>ALA</td>
<td>GLY</td>
<td>31</td>
<td>ASN</td>
</tr>
<tr>
<td>ASP</td>
<td>LYS</td>
<td>24</td>
<td>SER</td>
</tr>
<tr>
<td>GLU</td>
<td>ARG</td>
<td>21</td>
<td>ALA</td>
</tr>
<tr>
<td>THR</td>
<td>ASP</td>
<td>21</td>
<td>GLU</td>
</tr>
<tr>
<td>LEU</td>
<td>GLU</td>
<td>20</td>
<td>LYS</td>
</tr>
<tr>
<td>TYR</td>
<td>GLN</td>
<td>16</td>
<td>HIS</td>
</tr>
<tr>
<td>ARG</td>
<td>LEU</td>
<td>14</td>
<td>PHE</td>
</tr>
<tr>
<td>ASN</td>
<td>SER</td>
<td>14</td>
<td>ARG</td>
</tr>
<tr>
<td>GLN</td>
<td>THR</td>
<td>14</td>
<td>LEU</td>
</tr>
<tr>
<td>HIS</td>
<td>ASN</td>
<td>10</td>
<td>TYR</td>
</tr>
<tr>
<td>LYS</td>
<td>VAL</td>
<td>10</td>
<td>THR</td>
</tr>
<tr>
<td>MET</td>
<td>ILE</td>
<td>9</td>
<td>GLN</td>
</tr>
<tr>
<td>VAL</td>
<td>HIS</td>
<td>8</td>
<td>MET</td>
</tr>
<tr>
<td>CYS</td>
<td>TYR</td>
<td>7</td>
<td>VAL</td>
</tr>
<tr>
<td>ILE</td>
<td>MET</td>
<td>5</td>
<td>ILE</td>
</tr>
<tr>
<td>PHE</td>
<td>CYS</td>
<td>4</td>
<td>CYS</td>
</tr>
<tr>
<td>PRO</td>
<td>PHE</td>
<td>4</td>
<td>TRP</td>
</tr>
<tr>
<td>TRP</td>
<td>TRP</td>
<td>1</td>
<td>PRO</td>
</tr>
</tbody>
</table>

For p-flips the ‘?’ should be replaced by each amino acid type as it appears in the column. For example, there are 38 Pro–X flips where Pro and any amino acid X flank the flipping peptide plane. For p-flips the ‘?’ should be replaced by each amino acid type as it appears in the column. For example, there are 38 Pro–X flips where Pro and any amino acid X flank the flipping peptide plane.

3.1.3 Correlation of dihedral transitions

As already described, the total percentage of cases where a second dihedral transition separated by $m$ (the total number of $\phi$ and $\psi$ in between) from the first one occurred. Blue (hidden behind the cyan) and cyan show the percentage of independent $\phi$ and $\psi$ transition occurrences, respectively.

In contrast to the classical view of hinge bending motions as
### Table 2. Top 10 TFs ranked by NLS from the largest

<table>
<thead>
<tr>
<th>#</th>
<th>Protein</th>
<th>PDBID (CHAIN)</th>
<th>RMSD (Å)</th>
<th>NLS (Å)</th>
<th>Range (length)</th>
<th>#SR</th>
<th>Motion (DynDom assignment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Di-heme cytochrome c peroxidase</td>
<td>2c1u(A) 2c1v(B)</td>
<td>d d</td>
<td>5.14</td>
<td>30.56</td>
<td>227–236 (10)</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Dynactin 1</td>
<td>1txq(A) 2hkn(A)</td>
<td>m m</td>
<td>1.39</td>
<td>28.42</td>
<td>36–40 (5)</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Cytoplasmic protein NCKI</td>
<td>2ci8(A) 2ci9(A)</td>
<td>d m</td>
<td>1.19</td>
<td>26.28</td>
<td>335–338 (4)</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Ataxin-1</td>
<td>1oa8(B) 1oa8(C)</td>
<td>f f</td>
<td>2.16</td>
<td>25.68</td>
<td>568–577 (10)</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Forkhead box protein P2</td>
<td>2a07(F) 2a07(J)</td>
<td>m d</td>
<td>1.35</td>
<td>25.06</td>
<td>538–545 (8)</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Epidermal growth factor kinase</td>
<td>1i0c(A) 1i07(B)</td>
<td>m m</td>
<td>1.90</td>
<td>24.45</td>
<td>34–40 (7)</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Di-heme cytochrome c peroxidase</td>
<td>2c1u(A) 2c1v(B)</td>
<td>d d</td>
<td>7.91</td>
<td>23.80</td>
<td>244–257 (14)</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Phenol 2-monoxygenase</td>
<td>1poo(A) 1poo(D)</td>
<td>d d</td>
<td>0.54</td>
<td>21.57</td>
<td>210–211 (2)</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>Regulatory protein CII</td>
<td>1x6v(A) 1zs4(B)</td>
<td>h h</td>
<td>1.60</td>
<td>21.49</td>
<td>58–61 (4)</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>Bifunctional 3’-phosphoadenosine 5’-phosphosulfate synthetase 1</td>
<td>1isz(A) 1xsv(B)</td>
<td>d d</td>
<td>1.22</td>
<td>19.98</td>
<td>228–233 (6)</td>
<td>2</td>
</tr>
</tbody>
</table>

*Quaternary structure (m: monomer, d: dimer, t: trimer, h: hexamer). Some of the quaternary structures were predicted by the PQS server (Henrick and Thornton, 1998). Range: the range of TF in residue # and length in parenthesis. #SR: the number of SRs using a RMSD cutoff 1 Å per protein. Within the parenthesis in ‘motion’: hinge regions connecting 'dynamics domains’ assigned by DynDom (Hayward and Berendsen, 1998) are shown by residue number. ‘no domain’ shows the case where dynamics domains are not found by DynDom. Note the DynDom analysis was done for the whole protein chain.

Functional movements caused by the binding of a ligand (Gerstein et al., 1994), the formation of the quaternary structure underpins the hinge-like transitions seen in nine proteins here. Four global hinges motions (#2, 3, 5, 6) occur as typical 'domain swapping' upon oligomerization (Liu and Eisenberg, 2002) (Fig. 4A for #6), and five cases (#4,5,8,9,10) show the difference occurs between distinct chains in the same crystal. TF #1, 7, 8 located in long flexible loops are involved in allostery (Echalier et al., 2006) or protein–protein interactions (Enroth, 2003). Two cases (#5, 9) are associated with DNA binding. Since all of the Top 10 TFs are involved in protein–protein or protein–DNA interactions, the binding surfaces are relatively large; however, the range of significant effect on main-chain dihedrals is case dependent as the fragment lengths of the TFs vary from 2 to 14 residues.

#### 3.3.2 Flap motion

Table 3 shows the Top 10 TF rankings of localized motion (NLS ≤ 1.0 Å) in descending order according to RMSD (Table 3). These are the cases where the TFs cause large conformational change but their effect is localized, as on both sides the SFs do not move appreciably. Seven cases can be considered as ‘flap’ motions (#1 in Table 3 is shown as an example in Fig. 4C), where typically a fragment moves up and down relative to the main body of the protein. One of the well-known flap motions is that of the loop closure upon ligand binding in HIV-1 protease (Miller et al., 1989), which is not included in this Top 10 ranking. We also found the flap motion associated with ligand binding (#1) and closure upon ligand release (#2). To examine possible driving forces associated with the flap motion, we considered two factors: the TF is close to a ligand and/or it makes a crystal contact. Possible driving interactions are crystal contacts (four cases: #3, 5, 8, 9), both crystal contact and a small ligand (five cases: #1, 2, 4, 6, 7), and ligand only (one case: #10). As often seen in allostery, conformational change occurs at a distant region from a distant site of interaction (Cui and Karplus, 2008). We have not considered this here. Interestingly, these Top 10 proteins always comprise of single SR (with RMSD criterion 1.0 Å), indicating the existence of relative rigid protein ‘core’. These TFs are all located on the protein surface, indicating that they may play a role in compensating relatively localized interactions. Interestingly, the Top 10 TFs in Table 3 are relatively long as seven of them also occupy the Top 10 longest TF ranking (data not shown).

#### 3.3.3 Path-preserving motion

This motion preserves the main-chain path (therefore main-chain RMSD is small) and its effect is localized (NLS small). Their TFs have NLS ≤ 1.0 Å and they have the lowest RMSDs (see Supplementary Table S3 for the Top 10 ranking in descending order according to RMSD). Path-preserving motions are those TFs with RMSD ≤ 1.0 Å and NLS ≤ 1.0 Å. They comprise 24.3% of cases (127 out of 523 TFs, see Supplementary Fig. S4). Out of 127 cases, 94 contain a p-flip (75%), and both s- and p-flips were found in two cases. Therefore, the hydrogen-bonding pattern and side-chain packing can change considerably. A minimal example of a path-preserving motion is one p-flip (it appears in Top
Fig. 4. Example of TFs and their neighboring SFs. From the N-terminus side, 1st SF (blue)-1st TF (yellow)-2nd SF (green)-2nd TF (orange)-3rd SF (cyan), the other SFs (gray) and TFs (pink), and ligands (CPK). (A) and (B) Examples of hinge TFs. (A) Typical example of domain swapping. left: monomer, right: dimer (#6 in Table 2). (B) Hinge TF as a part of flexible loop [#1 (yellow) and #7 (orange) in Table 2]. (C) Examples of flap TF with distinct ligands (#1 in Table 3). (D) Examples of path-preserving TFs induced upon s-adenosylmethionine binding. (#7 in Table S3).

Table 3. Top 10 TFs ranked by RMSD from the largest within the group of data having relatively small NLS values (≤1.0 Å) (See Table 2 for details)

<table>
<thead>
<tr>
<th>#</th>
<th>Protein</th>
<th>PDBID (CHAIN)</th>
<th>RMSD (Å)</th>
<th>NLS (Å)</th>
<th>Range (length)</th>
<th>#SR</th>
<th>motion (DynDom assignment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Aspartate receptor</td>
<td>1vl(A)</td>
<td>d</td>
<td>4.59</td>
<td>0.69</td>
<td>76–87 (12)</td>
<td>Flap (no domain)</td>
</tr>
<tr>
<td>2.</td>
<td>Carbon monoxide dehydrogenase large chain</td>
<td>1n60(E)</td>
<td>h</td>
<td>4.40</td>
<td>0.69</td>
<td>165–177 (13)</td>
<td>Flap (no domain)</td>
</tr>
<tr>
<td>3.</td>
<td>Cell-cycle arrest protein</td>
<td>2i3s(A)</td>
<td>d</td>
<td>4.21</td>
<td>0.77</td>
<td>223–234 (12)</td>
<td>Flap (no domain)</td>
</tr>
<tr>
<td>4.</td>
<td>Purine nucleoside phosphorylase</td>
<td>1g2o(A)</td>
<td>t</td>
<td>3.65</td>
<td>0.91</td>
<td>60–71 (12)</td>
<td>Flap (no domain)</td>
</tr>
<tr>
<td>5.</td>
<td>Migration inhibitory factor</td>
<td>1hfo(B)</td>
<td>m</td>
<td>3.45</td>
<td>0.29</td>
<td>64–71 (8)</td>
<td>Flap (no domain)</td>
</tr>
<tr>
<td>6.</td>
<td>Sialic acid-binding IG-like lectin7</td>
<td>2hrl(A)</td>
<td>m</td>
<td>3.26</td>
<td>0.80</td>
<td>84–93 (10)</td>
<td>Deformation of helix (no domain)</td>
</tr>
<tr>
<td>7.</td>
<td>Hypothetical protein PH1933</td>
<td>1wwz(A)</td>
<td>d</td>
<td>3.25</td>
<td>0.49</td>
<td>28–39 (12)</td>
<td>Flap (no domain)</td>
</tr>
<tr>
<td>8.</td>
<td>Transcobalamin II</td>
<td>2bb6(B)</td>
<td>m</td>
<td>3.11</td>
<td>0.08</td>
<td>67–81 (15)</td>
<td>Deformation of long loop (no domain)</td>
</tr>
<tr>
<td>9.</td>
<td>Kynurenine aminotransferase</td>
<td>1yiz(A)</td>
<td>d</td>
<td>3.05</td>
<td>0.62</td>
<td>354–368 (15)</td>
<td>Deformation of long loop (no domain)</td>
</tr>
<tr>
<td>10.</td>
<td>Aminotransferase</td>
<td>1o69(A)</td>
<td>d</td>
<td>3.02</td>
<td>0.21</td>
<td>180–190 (11)</td>
<td>Flap (no domain)</td>
</tr>
</tbody>
</table>

10 of Supplementary Table S3 as #5 and #8). Commonly in the Top 10 are TFs that are relatively short (average 1.6 residues) and have low RMSD (average 0.40 Å). Although RMSD should be low by definition, path-preserving TF can consist of a few residues with a combination of multiple flips. Among the Top 10, three cases seem to be related to small ligand binding (#7 in Table S3 is shown in...
Fig. 3D), one case to protein–protein contacts, and five cases to crystal contacts. As far as we have checked in the literature, the authors of the original structures omitted to mention the functional relevance of this type of motion. One possible reason might be that this change in the main-chain is too subtle to notice.

4 DISCUSSION AND CONCLUSION

In this article, we carried out DTA, focusing on relatively large changes in main-chain dihedral angles. Although the probability of occurrence is relatively low (1.72%, 1.58% and 0.04% for \( \phi, \psi \) and \( \omega_1 \) respectively), multiple transitions tend to occur cooperatively along the sequence, which supports our splitting of a polypeptide chain into fragments with and without transitions. We also give statistics on dihedral transitions, which have not been systematically analyzed before. Table 1 shows important statistics for the s- and p-flips. Although the p-flip has been known to play an essential role in the function of the K⁺ pump (Obara et al., 2005), for example, it has not been investigated statistically. It is clear that DTA is a useful method for detecting potential functional motions. In this article, we found that the p-flip is frequently found in all kinds of TFs whereas the s-flip is mostly associated with hinge and flap movements.

As shown in Section 3, DTA was successful in extracting 'hinge', 'flap' and 'path-preserving' motions, many of which showed functional relevance. Among the Top 10 of the non-localized 'hinge' motions shown in Table 2, eight cases occur upon binding to large molecules, proteins and DNAs, and the other two cases are involved in allostery. Among the Top 10 of the localized 'flap' and other motions shown in Table 3, six cases are caused by binding to small compounds. ‘Path-preserving’ motions, which also occur upon ligand binding in three examples in Table S3, are found only after careful examination of dihedral changes as done in DTA. They cannot be detected by consideration of RMSD alone. It should be noted that ‘path-preserving’ motion occurs frequently (24% of TFs and 23% of the proteins in our database). We conclude therefore that DTA is a useful tool to identify potential functional motions, some of which might have been missed using conventional methods of protein conformational analysis.

Although DTA is not aimed at characterizing ‘hinges’ and ‘domains’ as the units of collective protein motion, TFs sometimes correspond to the ‘hinge’ of a domain motion assigned by the methods mentioned in Section 1 (Table 2). However, when the screw axis of domain motion almost coincides with one (or a few) main-chain dihedral angle(s), even a small change of \( \phi \) or \( \psi \) angles can produce a large domain motion. In principle such hinges cannot be detected by DTA. It should be noted therefore that small dihedral angle changes that would not qualify as TFs can also significantly affect protein structure. SRs detected by DTA often agree with ‘domains’, however, we also found some rare cases where SFs in domain proteins can contain the bending region involved with relatively small rotation of domain around the screw axis. It should be noted that SFs, are not necessarily ‘rigid’, because accumulated small changes of main-chain dihedrals, or even single small dihedral changes in the protein core, can produce a ‘soft’ deformation of the SF. The ‘classic’ view of protein motion characterized by the combination of ‘hinge’ and relatively rigid ‘domains’ is very useful and DTA provides a view of protein motion from a different perspective. To effectively characterize protein motion from different points of views, we propose to use DTA in combination with the other methods to detect ‘hinges’ and/or ‘domains’.


Conflict of Interest: none declared.

REFERENCES


