Optimized Torsion-Angle Normal Modes Reproduce Conformational Changes More Accurately Than Cartesian Modes

Jenelle K. Bray,* Dahlia R. Weiss, and Michael Levitt
Department of Structural Biology, Stanford Medical School, Stanford, California

ABSTRACT We present what to our knowledge is a new method of optimized torsion-angle normal-mode analysis, in which the normal modes move along curved paths in Cartesian space. We show that optimized torsion-angle normal modes reproduce protein conformational changes more accurately than Cartesian normal modes. We also show that orthogonalizing the displacement vectors from torsion-angle normal-mode analysis and projecting them as straight lines in Cartesian space does not lead to better performance than Cartesian normal modes. Clearly, protein motion is more naturally described by curved paths in Cartesian space.

INTRODUCTION

Normal-mode analysis can be a powerful tool for modeling conformational changes in proteins (1–5). Traditionally, normal-mode analysis is done in Cartesian space, where it is easy to formulate and solve the eigenvalue equations. However, the propagation of Cartesian modes can lead to the nonphysical deformation of bond length and angles because atoms move along straight lines and there is no way to describe rotation. This nonphysical deformation can be avoided by performing normal-mode analysis in torsion (dihedral) angle space (6–11). By fixing the bond lengths and angles, and performing the analysis in internal coordinates, all the predicted motions are physical and the number of degrees of freedom is reduced ~10-fold. This decrease in degrees of freedom leads to a substantial speed up in the calculation and the ability to treat larger systems.

Despite these apparent advantages, a systematic comparison of the lowest frequency displacement vectors arising from torsion-angle normal modes and those arising from Cartesian normal modes has never been carried out. Some studies that have used torsional modes have orthogonalized the linear displacement vectors (8,9,11,12), but we will show that orthogonalized torsional modes do no better than Cartesian modes in describing conformational changes. This has been shown to be the case for individual proteins (13), but not on an entire test set. Other studies (10,14) have derived a nonlinear second-order relationship between changes in torsion-angle space and curved motion in Cartesian space, which works better than the linear first-order treatment, but still only applies to small changes. The method we have developed changes the torsion angles along the normal modes to give the exact curved paths in Cartesian space; it is called optimized torsional normal-mode analysis.

The results on a set of 13 proteins with two conformational states show that with the same number of low-frequency modes, optimized torsional normal-mode analysis reproduces conformational changes more accurately than do either Cartesian or orthogonalized torsional modes. Optimized torsional normal-mode analysis is able to reproduce conformational changes by varying a small number of degrees of freedom. This leads to a dramatic reduction in dimensionality that is crucial if a conformational change is to be analyzed in detail.

COMPUTATIONAL METHODS

To compare the displacement vectors, we perform Cartesian and torsion-angle normal-mode analysis on a test set of proteins. The test set is made up of 13 pairs of proteins that have two conformations (A and B) with the same sequence, selected from the Protein DataBank (PDB). These proteins were chosen to represent a large range both in size (123–1079 amino acids) and in size of conformational change (root mean-square deviation, RMSD, from 1.7 to 27.6 Å). The test set also contains two complexes comprising more than one chain. See Table 1 for the protein names and the references to their structures.

To determine how well modes do at describing the conformational change from A to B, we use a metric called the fractional RMSD (fRMS), shown in Fig. 1. The fRMS is the RMS between the predicted conformation X reached by normal modes and the desired final conformation B, |RX|, normalized by the original RMS between A and B, |A−B|, i.e., fRMS = |RX|/|A−B|. An fRMS value of 0 represents the ideal description of A to B, whereas an fRMS of 1 means that the normal modes cannot predict a conformation X any closer to B than the original conformation A. We use the fRMS instead of the dot product defined as the cosine of the angle XAB. Although this dot product (termed dot in Fig. 1) has been used by most other studies (3), it is much less sensitive than the fRMS value. When conformations are defined in a linear, Euclidian space (like Cartesian coordinates), there is a simple relationship between fRMS and dot, with fRMS = √1 − dot². This shows that dot seems favorable (close to 1) for a wide range of fRMS values that look much less favorable. For example, an fRMS value of 0.5 (1/2), gives a dot product of 0.87, whereas an fRMS of 0.33 (1/3) gives a dot product of 0.94. In neither case is conformation X particularly close to conformation B: the initial RMSD has just been halved and reduced by two-thirds, respectively.

All-atom Cartesian normal modes (no hydrogen atoms) were calculated for both the A and the B conformations of each protein pair using the program NOMAD-Ref (15). A linear combination of the displacement...
TABLE 1  Fractional RMS for Cartesian, orthogonalized torsional, and optimized torsional modes

<table>
<thead>
<tr>
<th>Protein</th>
<th>PDB ID A</th>
<th>PDB ID B</th>
<th>No. of chains</th>
<th>RMS &lt;i&gt;fRMS&lt;/i&gt; (Å)</th>
<th>Min. num. modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ribose-binding protein</td>
<td>1BA2_A</td>
<td>2DRI</td>
<td>271</td>
<td>6.2</td>
<td>2 10</td>
</tr>
<tr>
<td>Calmodulin</td>
<td>1CLL</td>
<td>1CTR</td>
<td>144</td>
<td>14.8</td>
<td>2 1</td>
</tr>
<tr>
<td>LBO binding protein</td>
<td>2LBO</td>
<td>1LST</td>
<td>238</td>
<td>4.7</td>
<td>10 10</td>
</tr>
<tr>
<td>Ribonuclease III</td>
<td>1YZ9_AB</td>
<td>1YYO_AB</td>
<td>436</td>
<td>7.3</td>
<td>3</td>
</tr>
<tr>
<td>Diphtheria toxin</td>
<td>1DDT</td>
<td>1MDT_A</td>
<td>535</td>
<td>15.6</td>
<td>8</td>
</tr>
<tr>
<td>Lactoferrin</td>
<td>1LFI</td>
<td>1LFG</td>
<td>691</td>
<td>6.4</td>
<td>3 3</td>
</tr>
<tr>
<td>Aspate transcarbamoylase</td>
<td>8ATC</td>
<td>5AT1</td>
<td>912</td>
<td>4.9</td>
<td>10 10</td>
</tr>
<tr>
<td>Aspate aminotransferase</td>
<td>9AT_A</td>
<td>1AMA</td>
<td>401</td>
<td>1.7</td>
<td>4</td>
</tr>
<tr>
<td>Skeletal muscle Ca&lt;sup&gt;2+&lt;/sup&gt;-ATPase</td>
<td>1SU4_A</td>
<td>1IWO_A</td>
<td>994</td>
<td>1.4</td>
<td>8 10</td>
</tr>
<tr>
<td>5'-Nucleotidase</td>
<td>1HPU_D</td>
<td>1IOD_A</td>
<td>525</td>
<td>9.3</td>
<td>5</td>
</tr>
<tr>
<td>Scallop myosin II</td>
<td>1QVI_AYZ</td>
<td>1KK8_ABC</td>
<td>1079</td>
<td>27.6</td>
<td>9</td>
</tr>
<tr>
<td>T7 RNA polymerase</td>
<td>1MSW_D</td>
<td>1QLN_A</td>
<td>883</td>
<td>18.3</td>
<td>9</td>
</tr>
<tr>
<td>Nitrogen regulatory protein C</td>
<td>1DC8_A</td>
<td>1DC7_A</td>
<td>123</td>
<td>3.2</td>
<td>2 10</td>
</tr>
<tr>
<td>Average values</td>
<td></td>
<td></td>
<td></td>
<td>0.536 0.551 0.448 0.727 0.738 0.608</td>
<td>6.4 7.2</td>
</tr>
</tbody>
</table>

Fractional RMS (<i>fRMS</i>) values of Cartesian normal modes (<i>Cart.</i>), orthogonalized torsional normal modes (<i>Orth. tors.</i>), and optimized torsional modes (<i>Opt. tors.</i>). The <i>fRMS</i> values are calculated using the 10 lowest frequency modes. The A and B conformations are named so that the average value of the three methods of calculating normal modes describes A to B better than B to A. The lowest <i>fRMS</i> for each conformational change is highlighted. The minimum number of modes (<i>Min. num. modes</i>) is the number of optimized torsional normal modes required to give a value for <i>fRMS</i> that equals the value of 10 Cartesian normal modes; values ≤5 are highlighted. PDB file names for conformations A and B are as follows: Ribose-binding protein (1BA2 (16), 2DRI (17)), Calmodulin (1CLL (18), 1CTR (19)), LBO binding protein (2LBO (20), 1LST (20)), Ribonuclease III (1YZ9 (21), 1YYO (21)), Diphtheria toxin (1DDT (22), 1MDT_A (23)), Lactoferrin (1LFI (24), 1LFG (24)), Aspate transcarbamoylase (9AT (25), 1AMA (26)), Skeletal muscle Ca<sup>2+</sup>-ATPase (1SU4 (29), 1IWO (30)), 5'-Nucleotidase (1HPU (31), 1IOD (32)), Scallop myosin II (1QVI (33), 1KK8 (34)), T7 RNA polymerase (1MSW (35), 1QLN (36)), Nitrogen regulatory protein C (1DC8 (37), 1DC7 (37)).

RESULTS AND DISCUSSION

Although the torsion-angle normal modes are very different from Cartesian normal modes, the displacement vectors are very similar for the low-frequency motion. When these two different sets of displacement vectors are used to reproduce the experimentally observed conformational change using the 10 lowest frequency modes, the results shown in Table 1 are similar for the <i>fRMS</i> values of the Cartesian (<i>Cart.</i>) or orthogonalized torsion angle (<i>Orth. tors.</i>) displacement vectors. The average <i>fRMS</i> values for all the test cases for A to B are very similar for Cartesian modes and orthogonalized torsional modes (0.536 and 0.551, respectively). The same holds for the B-to-A test cases (0.727 and 0.738, respectively). For both the A to B and the B to A, the Cartesian modes are better than the orthogonalized torsional modes in half the cases (6 of 13 cases). Clearly, there is little
difference between the Cartesian displacement vectors and the orthogonalized torsional displacement vectors.

In Table 1, we show these $f_{RMS}$ values for all protein pairs in the test set for Cartesian modes, orthogonalized torsional modes, and optimized torsional modes calculated using the 10 lowest frequency modes. We chose 10 modes because this is the average number of modes for which the improvement in $f_{RMS}$ from additional modes is negligible. Optimized torsional modes describe conformational changes significantly better than Cartesian modes or orthogonalized torsional modes. For the A-to-B conformational change, the average $f_{RMS}$ for the entire test set is 0.448 for the optimized torsional modes, as compared to 0.536 for Cartesian modes and 0.551 for orthogonalized torsional modes. An $f_{RMS}$ of 0.431 corresponds to a dot product of 0.894, whereas an $f_{RMS}$ of 0.536 corresponds to a dot product of 0.844. For the A-to-B cases, the optimized torsional modes better describe the conformational change in 11 out of 13 pairs. For the B-to-A conformational changes, the difference in $f_{RMS}$ is even more striking. The optimized torsional modes give an average $f_{RMS}$ of 0.608 in comparison to $f_{RMS}$ values of 0.727 and 0.738 for Cartesian and orthogonalized torsional modes, respectively.

In some cases, most noticeably for both directions of calmodulin, optimized torsional modes do a markedly better job than the Cartesian or orthogonalized torsional modes. For the A-to-B conformational change of calmodulin, the $f_{RMS}$ of the optimized torsional modes is 0.194, whereas it is 0.561 and 0.580 for the Cartesian and orthogonalized torsional modes, respectively. An $f_{RMS}$ of 0.194 corresponds to a dot product of 0.981, whereas an $f_{RMS}$ of 0.561 corresponds to a dot product of 0.828. There is no case in the test set where the Cartesian or orthogonalized torsional modes perform significantly better than the optimized torsional modes. If we repeat the analysis with fewer optimized torsional modes we find that in many cases the dimensionality (number of normal modes) needed to reproduce the conformational change is significantly smaller for optimized torsional modes (see Table 1). With optimized torsional modes, the conformational changes in at least one direction (A to B or B to A) for five of 13 cases are as well described in three or fewer dimensions as in 10 dimensions for Cartesian modes. This is a dramatic reduction in the dimensionality of the problem! Based on these results, optimized torsional mode analysis is clearly the best method for modeling protein conformational changes.

CONCLUSION

Our finding that the Cartesian and orthogonalized torsional modes perform similarly, and significantly worse than the
optimized torsional modes, shows that protein motion is more naturally described by curved paths than by displacement vectors in Cartesian space. Optimized torsional normal mode analysis is a powerful, potentially new way to represent protein conformational changes.

We thank Russ Altman for helpful discussion.

This work was supported by National Institutes of Health grants GM041455 and GM072970. Computations were done on Stanford’s Bio-X2 computers (National Science Foundation award CNS-0619926). D.F.W. is partially supported by F32GM093580 from the National Institute of General Medical Sciences. Michael Levitt is the Robert W. and Vivian K. Cahill Professor of Cancer Research.

REFERENCES


