Exploring the range of protein flexibility, from a structural proteomics perspective
Mark Gerstein and Nathaniel Echols

Changes in protein conformation play a vital role in biochemical processes, from biopolymer synthesis to membrane transport. Initial systematizations of protein flexibility, in a database framework, concentrated on the movement of domains and linkers. Movements were described in terms of simple sliding and hinging mechanisms of individual secondary structural elements. Recently, the accelerated pace and sophistication of methods for structural characterization of proteins has allowed high-resolution studies of increasingly complex assemblies and conformational changes. New data emphasize a breadth of possible structural mechanisms, particularly the ability to drastically alter protein architecture and the native flexibility of many structures.

Addresses
Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Ave, New Haven, CT 06520, USA
e-mail: mark.gerstein@yale.edu

Introduction
Annotations in the Database of Macromolecular Movements (http://molmovdb.org) [1,2] currently include more than 240 distinct protein motions, the majority of which can be directly visualized from solved structures [3]. Domain motions of single subunits make up the largest subset, but an increasing number of molecular complexes, exhibiting large structural rearrangements have been solved. Initial attempts to classify motions used a convention of ‘shear’ versus ‘hinge’ movements [4] (based on the presence or absence of a maintained interface between moving parts) and typically focused on movements in single domains or large fragments. The repertoire of protein conformational changes has grown considerably, incorporating many cooperative movements of subunits and structural changes at a quaternary level, largely due to improvements in methods for structural characterization of large molecules. Efforts to computationally model the activity of macromolecular assemblages remain limited by time constraints, but recent studies have used simulation to investigate the global conformational changes of immense structures, such as the F1-ATPase [5], GroEL [6], and the 70S ribosome [7].

Here, we summarize several recent structural studies that illustrate the importance and diversity of protein motions, concentrating, primarily, on several groups of related protein structures or mechanisms. Although subtle conformational changes, down to the level of alternating sidechain rotamers, are often essential to protein function, our fundamental focus is on more global changes, involving significant movement of the protein backbone, and interactions between tertiary and quaternary elements. Furthermore, many of these changes might involve multiple distinct intermediate states or occur on time scales that are too large to permit conventional simulation. In particular, we have highlighted proteins that display considerable ‘plasticity’ or ‘fluidity’, in terms of changes in fold, interactions within the cell membrane, or movement in the native form.

An overview of several new motions, examined in the context of the database, is presented in Table 1 and Figure 1. With the exception of ATP sulfurylase, each structure listed has only a single chain that exhibits the described motion (although other subunits may be involved). Identical methods were used for gathering all of the presented statistics, but the nature of the structural changes varies widely, and most changes do not easily fall into one of the pre-existing categories; however, all of the structures shown have one or more flexible linker regions of multiple residues, from which much of the displacement is derived, and, although there are several cases of shearing helices, mobile interfaces are not usually maintained within a single chain.

Large-scale remodeling
T7 RNA polymerase
One of the most dramatic conformational changes that has been observed so far is seen in the elongation-phase structure of T7 RNA pol (Figure 1a). Although movement of some type is observed in RNA and DNA polymerases from a variety of organisms (for example, see [8] for a review of bacterial RNA polymerase structures), these typically involve flexible linkers between distinct rigid domains. The transition from initiation to elongation in the T7 polymerase requires refolding and massive translocation of the N-terminal domain, opening an exit tunnel for the seven-base mRNA strand, which would otherwise be blocked [9,10**,**11**]. The remainder of the protein undergoes comparatively little movement. The exact impetus for the rearrangement of the structure and...
promoter release is not yet clear, but a further intermediate conformation might be involved. The relocation of the upstream DNA and the apparent lack of movement by the specificity loop seem to prohibit any direct translation, but it is unclear to what extent, if any, the moving domains operate as rigid bodies or, indeed, whether they must partially unfold.

**Mad2**

On a smaller scale, the spindle checkpoint protein Mad2 undergoes a rearrangement of similar magnitude, involving the transposition of β strands (Figure 1b). Binding of the small peptide MBP1 disrupts the sheet and causes two of the strands to move to the opposite side (these are replaced by MBP1), while a smaller strand at the N-terminus dissociates with the sheet and, instead, adopts a helical conformation. Additional data indicate a similar transition on binding other proteins that are not related to MBP1 but are known to interact with Mad2 [12,13**]. Few examples exist of such rearrangement of a β sheet; the caspase inhibitor p35 [14] and serpin family [15] are the most similar proteins to Mad2, in this respect, but both involve cleavage and re-insertion of part of the peptide chain by another protein.

**Protein synthesis**

Various ribosome-binding proteins display considerable interdomain flexibility, observed by comparison of apo and ribosome-bound forms or their analogues. Ribosomal translase has been studied in both prokaryotic and eukaryotic hosts, and a recent pair of structures for the translocase has been studied in both prokaryotic and ribosome-bound forms or their analogues. Ribosomal transition on binding other proteins that are not related to MBP1 but are known to interact with Mad2 [12,13**]. Few examples exist of such rearrangement of a β sheet; the caspase inhibitor p35 [14] and serpin family [15] are the most similar proteins to Mad2, in this respect, but both involve cleavage and re-insertion of part of the peptide chain by another protein.

**Membrane proteins**

The improvement of techniques for structural characterization of membrane proteins has yielded several examples of structural changes in gating and transport, some involving considerable flexibility within the transmembrane region. An example of receptor functioning via conformational change was found in the structures of FccA [20], where ligand binding alters the conformation of extracellular loops, transmitting the signal to cytoplasmic proteins; but more complex motions are observed in several ionic transport proteins, where bending or shearing in the helical bundle is important.

**Potassium channels**

The role of large movements in ion channel gating has been demonstrated in several experimental studies [21,22], and MacKinnon and co-workers have recently investigated the specific structural elements that are involved in this movement, at atomic resolution. A comparison of the structurally related potassium channels, KcsA and MthK, in the closed and open forms, respectively, has illustrated a simple mechanism for gating, by bending of the inner helix at a conserved position [23*]. A more complicated model for the voltage-gated channel was proposed, based on separate structures of the channel and the voltage-sensing paddles, which are hypothesized to move up to 20 Å within the membrane and extend arginine residues almost to the solution on either side. This movement of the sensors would pull apart the outer helices to open the channel [24*].

**P-type ATPases**

One of the largest identified motions is exhibited by the Ca^{2+}Pase, switching from calcium-bound to calcium-free states, studied by crystallography and cryo-EM [25,26,27**]. The overall structure of the enzyme is

---

Table 1

Quantitative comparison of motions observed by comparison of recent protein structures.

<table>
<thead>
<tr>
<th>Structure Name</th>
<th>PDB IDs</th>
<th>Residues</th>
<th>RMSD of entire structure</th>
<th>RMSD of mobile domain(s)</th>
<th>Maximum Cα displacement (percentile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7 RNA polymerase</td>
<td>1qin, 1msw</td>
<td>883</td>
<td>18.1</td>
<td>36.9</td>
<td>75.6 Å (99%)</td>
</tr>
<tr>
<td>Mad2</td>
<td>1duj, 1k1q</td>
<td>197</td>
<td>10.0</td>
<td>21.4</td>
<td>37.1 Å (89.4%)</td>
</tr>
<tr>
<td>EF2</td>
<td>1m0v, 1m0u</td>
<td>842</td>
<td>13.8</td>
<td>38.0</td>
<td>70.7 Å (98%)</td>
</tr>
<tr>
<td>RF2</td>
<td>1gge, 1m16</td>
<td>362</td>
<td>17.1</td>
<td>29.0</td>
<td>56.7 Å (95%)</td>
</tr>
<tr>
<td>Ca^{2+} ATPase</td>
<td>1swo, 1elu</td>
<td>994</td>
<td>14.4</td>
<td>31.8</td>
<td>50.1 Å (93.5%)</td>
</tr>
<tr>
<td>ATP sulfurylase</td>
<td>12d2, 1m8p</td>
<td>572</td>
<td>4.1</td>
<td>7.3</td>
<td>14.6 Å (60.3%)</td>
</tr>
<tr>
<td>Anthrax oedema factor</td>
<td>1k8t, 1k93</td>
<td>507</td>
<td>10.0</td>
<td>17.4</td>
<td>31.2 Å (86.9%)</td>
</tr>
<tr>
<td>Acetyl-CoA synthase</td>
<td>1oa0 (CandD)</td>
<td>728</td>
<td>7.1</td>
<td>18.8</td>
<td>38.6 Å (91.5%)</td>
</tr>
</tbody>
</table>

*Residues* is the consensus length of the protein in the two PDB entries compared, accounting for truncations (only chain A was compared in ATP sulfurylase). RMSD of the mobile domains and maximum Cα displacement were both calculated with the non-moving parts of the protein fitted, using CNS [40] and the procedure described in [3], respectively. Percentile values are based on comparison to all other motions in the database. Graphical comparisons of the different states of each structure are shown in Figure 1.
Figure 1

Multiple conformations of proteins (structures not to scale). (a) T7 RNA polymerase, initiation and elongation states (PDB IDs 1qln and 1msw). (b) Spindle assembly protein, Mad2, with and without ligand MBP1 (PDB IDs 1duj and 1klq). (c) Ribosomal translocase EF-2, in the native state and with sordarin bound (PDB IDs 1n0v and 1n0u). (d) Ribosomal release factor 2, independent crystal structure and refitted to cryo-EM map (PDB IDs 1gqe and 1mi6). (e) Ca²⁺ ATPase, with and without calcium (PDB IDs 1iwo and 1eul). (f) ATP sulfurylase, showing half of the hexamer, in R and T states (PDB IDs 1i2d and 1m8p). (g) Anthrax toxin oedema factor, with and without bound calmodulin (not shown) (PDB IDs 1k8t and 1k93). (h) Acetyl-CoA synthase α subunits, both forms from a single tetramer, bound to Ni-Ni-[Fe₄-S₄] and Ni-Zn-[Fe₄-S₄] (chains D and C, PDB ID 1oao). All figures generated using PyMOL [41].
comprised of three independent and relatively rigid cytoplasmic domains, connected by flexible ‘stalks’ to the transmembrane domain, comprised of ten helices. Release of calcium involves a large rotation and translation of the cytoplasmic domains, accompanied by shearing of six of the transmembrane helices (Figure 1e). Although ATP hydrolysis and calcium transport involve several distinct steps, the transition from E1 to E2 states appears smooth and can be plausibly approximated without intermediates, unlike the more severe changes described above. Studies of the structurally homologous Na,K ATPase, using homology modeling and cryo-EM, have indicated a similar mechanism for this protein [28,29].

Ring complexes
Several known or suspected motions occur in complexes of identical subunits, arranged in a ring, whose motion is essentially cooperative. The best studied of these are GroEL, whose conformational cycle has been investigated by a host of biophysical techniques (including simulation), and aspartate transcarbamoylase. Recently, three new, completely unrelated structures of hexameric ATPases have been described, whose functionality depends on the conformation of the individual subunits.

ATP sulfurylase
ATP sulfurylase catalyzes the incorporation of inorganic sulfur, and is allosterically inhibited by a downstream intermediate in Penicillium. The hexamer consists of two stacked rings of three subunits; crystal structures of the R- and T-states differ, mainly, by the rotation of the C-terminal allosteric domain upon inhibitor binding, which slightly expands the volume of the overall hexamer (Figure 1f). A separate loop movement in the catalytic domain results in a more open active site [30,31].

VirB11
VirB11, an ATPase that is involved in secretion in Helicobacter pylori, forms a simple hexamer whose subunits adopt multiple conformations in the apo form. The N-terminal domain rotates away from the nucleotide-binding site, to varying degrees in each chain, but the structure is stabilized by the interaction of the C-terminal domains. ATP-analogue binding results in a stable configuration, with the N- and C-terminal domains closed; hydrolysis does not appear to be responsible for any further motion, because the ATP and ADP forms have identical conformations. The authors suggest a cycle of ATP binding, hydrolysis and release that occurs unevenly among groups of three subunits [32].

p97/VCP
p97/VCP is a multi-purpose enzyme, containing ‘AAA’ ATPase domains, that expands during hydrolysis, based on crystal and cryo-EM structures [33,34]. Disorder of the N-terminal domain in the cryo-EM maps indicates that the subunits of this structure also have considerable natural flexibility, even while assembled into the complete complex, but adopt a specific conformation during ATP hydrolysis.

Other structures
Several other structures that are not readily classified deserve mention here, particularly in the context of the inherently dynamic complexes described above. As in VirB11, evidence for multiple conformations is frequently found in single crystals, where two or more molecules in the asymmetric unit adopt different domain orientations. The two α subunits of the tetrameric Acetyl-CoA synthase (Figure 1g) are iron–sulfur binding proteins with three large domains, connected by hinges. In the crystal structure, the exchange of a nickel ion for a zinc results in a ‘closed’ configuration of the domains. Although the open form, with two nickel ions, appears to be the active state, the role of the conformational change is unclear [35]. The structure of ribose-5-phosphate isomerase A also has two conformations within a crystal, but without any change in bound heteroatoms. The mobile domains close around a cleft that is thought to be important for substrate binding, and the difference in orientation appears to be due to native flexibility of the tertiary structure [36].

Recent crystallographic studies, exploring the toxic activity of the anthrax bacterium are of particular relevance here. Structures of both the lethal factor and the oedema factor exhibit domain motions in different functional states. The lethal factor, a protease that attacks signaling pathway kinases in the host cell, displays several small re-orientations of elements throughout the structure (overall RMSD of 1.18) upon binding a target peptide [37]. Far greater flexibility is seen in the oedema factor, an adenyl cyclase, the activity and structure of which are altered by binding of calmodulin. The N-terminal helical domain is displaced and rotated to open a large cleft for calmodulin (Figure 1h). Alteration of the conformation of smaller segments results in activation of the cyclase [38]. This relatively large motion does not require any particular contortions of secondary or tertiary structure, but significant parts of the mobile domain are missing from the final model.

Conclusions
Theoretical studies of protein motion have traditionally focused on structures of single molecules, following a known transition (for example, domain closure in response to ligand binding [39]), or on detailed mechanistic and energetic analyses using simulation. Comparison of multiple structures is limited by available CPU power and by the diversity of tertiary arrangements; nevertheless, some trends might be seen by a proteomics approach. The degree of movement in many of the structures that have been examined is striking, particularly in light of the variety of mechanisms involved. Together with the
repeated observation of mobile domains in crystals, these studies indicate the importance of retaining a large degree of conformational freedom in folded proteins, and reinforce the importance of studying the mechanisms that enable structural malleability.

Supplementary material
Most of the structures discussed, for which 3D data are available, are listed online, at http://molmovdb.org/molmovdb/coch. These listings include additional images and animations.

Acknowledgements
We thank Thomas Steitz and Whitney Yin for discussions of T7 polymerase function, and Michael Barnett, Duncan Milburn and Vadim Alexandrov for assistance with the database. We acknowledge support from the NIH (5P50 GM062413) and apologize to the many authors whose work could not be included due to space constraints.

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:
* of special interest
** of outstanding interest

24. Comparison of crystal structures of structurally related MthK and KcsA channels indicates that bending of transmembrane helices controls gating.
26. Electrophysiological data, coupled with structural information, is used to elucidate a model involving large translations of the voltage-sensing domain within the membrane.
30. A high-resolution structure of the E1 state of Ca(2+)/+ATPase, demonstrating considerable reorientation of the cytoplasmic domains, accompanied by significant changes in the transmembrane region. Together with the T7 polymerase and haemaglutinin, this is one of the largest changes observed in a single polypeptide chain.
33. MacRae U, Segel IH, Fisher AJ: Crystal structure of ATP sulfurylase from Paracoccus chrysogenum: insights into the


Good example of domain movements in the absence of ligand/substrate; also demonstrates the role of ATP binding, rather than hydrolysis for stabilizing a closed structure.


Crystal structures show extension of a domain to allow calmodulin binding, which enables catalytic activity.

