Elastic network models capture the motions apparent within ensembles of RNA structures

MICHAEL T. ZIMMERMANN1,2,3 and ROBERT L. JERNIGAN1,2,4
1Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011, USA
2Bioinformatics and Computational Biology Interdepartmental Graduate Program, Iowa State University, Ames, Iowa 50011, USA

ABSTRACT
The role of structure and dynamics in mechanisms for RNA becomes increasingly important. Computational approaches using simple dynamics models have been successful at predicting the motions of proteins and are often applied to ribonucleoprotein complexes but have not been thoroughly tested for well-packed nucleic acid structures. In order to characterize a true set of motions, we investigate the apparent motions from 16 ensembles of experimentally determined RNA structures. These indicate a relatively limited set of motions that are captured by a small set of principal components (PCs). These limited motions closely resemble the motions computed from low frequency normal modes from elastic network models (ENMs), either at atomic or coarse-grained resolution. Various ENM model types, parameters, and structure representations are tested here against the experimental RNA structural ensembles, exposing differences between models for proteins and for folded RNAs. Differences in performance are seen, depending on the structure alignment algorithm used to generate PCs, modulating the apparent utility of ENMs but not significantly impacting their ability to generate functional motions. The loss of dynamical information upon coarse-graining is somewhat larger for RNAs than for globular proteins, indicating, perhaps, the lower cooperativity of the less densely packed RNA. However, the RNA structures show less sensitivity to the elastic network model parameters than do proteins. These findings further demonstrate the utility of ENMs and the appropriateness of their application to well-packed RNA-only structures, justifying their use for studying the dynamics of ribonucleoproteins, such as the ribosome and regulatory RNAs.

Keywords: RNA dynamics; normal mode analysis; elastic network model; coarse-graining; structure alignment; experimental dynamics

INTRODUCTION
Much effort has focused on elucidating the functional motions of proteins to understand the mechanisms by which their functions are conveyed, but similar structure-based studies of well-packed RNAs are rarer. The importance of structure-based insights from RNA has increased in recent years, for example, from the recently published ENCODE papers (Raney et al. 2011; Dunham et al. 2012) which emphasize the wide array of functional transcribed RNAs that are not translated into protein product. RNA molecules play diverse roles in physiology, and their de-regulation or functional disruption has been implicated in many diseases, including most genome-wide association studies, which often find significant risk loci in noncoding regions (Ward and Kellis 2012). While our knowledge of the roles RNAs play in physiology has increased at a rapid pace, comparatively little detail has been elucidated regarding the role of dynamics in their function, a critical aspect of understanding not only their normal behaviors but also their aberrant behaviors in complex diseases.

Defining the set of motions available to a biomolecule can be challenging but is necessary in order to test the accuracy of computational models. In this work, we will first define our method for extracting the experimental motions of various folded RNA structures. Next, the computational ENM method employed to simulate those motions and the breadth of parameterizations employed is explained. Finally, we show that, for a range of resolutions, the model accurately and consistently conveys motions apparent in the experimental motions.

© 2014 Zimmermann and Jernigan This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

792 RNA 20:792–804; Published by Cold Spring Harbor Laboratory Press for the RNA Society
In the Protein Databank (PDB) (Bernstein et al. 1977), many structures have been determined either under different conditions, for mutants, with different ligands bound or simply reported by different investigators. These static X-ray structures may not individually inform us about the dynamics of biomolecules, but collectively they constitute an ensemble of snapshots of a structure, or closely related structures, in its various conformations. These conformations are likely to be a meaningful sample of the inherent flexibility of the biomolecule. In order to better understand these functional motions, principal component analysis (PCA) can be employed to convert such an ensemble into a discrete set of vectors that capture its spatial variance. This set of vectors can be thought of as a set of “moves” that steer the structure between the observed conformations and can then be compared to in silico models of motion.

A computational method for representing the intrinsic flexibility of proteins that has achieved significant validation and widespread use for proteins is normal mode analysis (NMA) using elastic network models (ENMs) to obtain the motions of a structure at atomic, coarse-grained, or mixed resolution, based on a simplified force field (Tirion 1996; Doruker et al. 2002). ENMs, in particular the anisotropic network model (ANM), a type of ENM that provides the directions of motions in addition to relative mobility, have proven useful for obtaining the dominant functional motions of proteins in many studies (Bahar and Rader 2005; Ma 2005; Zheng and Brooks 2005; Sen et al. 2006; Tama and Brooks 2006; Jernigan and Kloczkowski 2007; Yang et al. 2007, 2008). It has been shown that the motions computed using ENMs correspond well to the principal components of molecular dynamics (MD) trajectories (Hayward and de Groot 2008), can aid in molecular structure determination (Wu and Ma 2004), refining of structural models (Delarue and Dumas 2004), and in flexible docking (May and Zacharias 2008; Gerek and Ozkan 2010).

In Yang et al. (2008), we found that the spatial variance seen in the superposition of many structures of the HIV-1 protease has a close relationship with the normal modes computed with ANM, and Bakan and Bahar (2011) investigated kinases with a similar approach. These studies have shown that the accessible conformations of a protein can be sampled by the dominant motions of one representative structure. A previous study has shown that these ENM models are capable of reproducing key aspects of nucleotide dynamics but that they may not be as accurate for loosely packed structures as they are for densely packed ones, such as globular proteins (Van Wynsberge and Cui 2005). In this work, we investigate the ability of normal modes from ANMs to capture dynamics apparent within 16 different ensembles of experimentally determined RNA structures. We test the utility and extent of applicability of ENMs to represent the motions of a variety of RNA structures.

How structure alignment is carried out can affect results, but despite the differences in algorithms, use of different methods is often not tested. Many different published algorithms exist, varying from superposition based on multiple sequence alignments, to aligning secondary structures, or most commonly minimizing the root mean square deviation (RMSD) of all points in the system, and even to algorithms that iterate to refine fits by down-weighting aligned pairs that are far apart spatially. Multiple structure alignment methods have also been devised (Taylor et al. 1994; Shatsky et al. 2004). Focusing on one example, an ensemble of tRNA structures, we compare the effect of alignment method chosen upon the overlap between normal modes from ANM models and the PCs extracted from the set of superimposed structures. Additionally, we investigate the extent of loss of dynamics information with various levels of coarse-graining and the effects of including the crystal environment.

We perform our study on a collection of 16 different groups of experimental RNA structures (described in Table 1), many of which are riboswitches, components of molecular motors, rRNA, and catalytic introns, and these range in size from 22 to 170 bases. Previous studies investigated the dynamics of ribonucleo-protein systems, including the ribosome, telomerase, and a partial reverse transcriptase structure, but only rarely has the model’s performance on RNA-
RESULTS

Sixteen diverse RNA groups

In this study, the variance present within each of the ensembles of the 16 RNA structures is analyzed and compared to the motions computed from a representative structure using normal mode analysis of elastic network models. A representative structure is sufficient for the dynamics simulations using ENM because these models are quite insensitive to the details of the structures, with the most dominant motions usually depending only upon the shapes of the structures. See Materials and Methods for a description of how the ensembles were generated and a description of the NMA method, and Table 1 for a brief characterization of each ensemble. See the Discussion section for more about riboswitch behavior. Briefly, at this time, we lack sufficient structural coverage to confidently predict base re-pairing events. The ability of ENMs to capture variation within the structural ensemble (where an ensemble is loosely defined by one set of base-pairings) is explored.

Figure 1A summarizes the agreement between computed models of motion for an arbitrarily chosen high resolution (not chosen to optimize the agreement) representative structure from each ensemble and the conformational variation within each respective ensemble. Each row represents one ensemble (see Table 1 for the identity of each ensemble ID), while columns correspond to results for different ANM models with different distance-dependent springs. Each comparison is between the first PC (most dominant conformational difference) and the first 20 motion vectors (normal modes) from the ANM. The agreement is better than 0.5 for all but three of the cases. Figure 1B shows the same comparison but for the first three PCs and groups of 5, 10, or 20 modes. Thus, Figure 1A corresponds to the upper right panel of Figure 1B. Unlike for globular proteins, there is relatively little sensitivity to the power dependency. While squared and 4th-power dependences exhibit the best performance, the decrease in performance for some ensembles, even for high powers of inverse distance, is relatively small.

The performance of ANM in capturing the first three PCs from an ensemble of structures is a highly specific measure. In Supplemental Figure S1, we show these overlaps as the percentage of total ensemble variance captured by the model, again varying the distance dependence and number of ANM modes considered. We find that, when 20 modes are considered with springs having a square power of distance dependence, the majority of our RNA structure groups have >50% of their variance explained. One third of them approach or

![Figure 1](https://example.com/figure1.png)

**FIGURE 1.** (A) Elastic network models with different inverse distance-dependent interaction springs capture well the fundamental motion of the first principal component apparent within most of the ensembles of RNA structures. Unlike in globular proteins, the performance of ANMs in capturing the respective ensemble’s first principal component does not depend strongly on the choice of distance dependence (see Materials and Methods). All comparisons are between the first PC of each ensemble (ensemble group numbers explained in Table 1) and the 20 most important normal modes of motion in the model using Equation 4. (B) Elastic network models with different inverse distance-dependent interaction springs capture well the fundamental motions apparent within most of the ensembles of RNA structures. Here, we compare each of the first three PCs to the k dominant normal modes of the ANM using cumulative overlap (Equation 4), which ranges from zero (orthogonal motions) to one (parallel motions fully capturing observation). See Table 1 for the identity of each of the members of the RNA groups 1–16. For each panel, the abscissa marks the distance dependence used in ANM (see Materials and Methods). Smaller values of the power dependence lead to a more cohesive system, while higher dependencies increasingly ignore long-range communication through the structure.
exceed the 75% level, indicating a strong agreement between the computed ANM motions and the variations seen within the experimental ensembles of structures.

**tRNA**

We have established above that ANM can be applied to various sets of folded RNAs to calculate their dynamics near the initial state, and from this point on, we focus on the tRNA ensemble in greater detail to characterize the behavior of the models across a wide range of representations (model granularity) and parameters. The variation seen in the ensemble of experimental tRNA structures is visualized using PCA in Figure 2 and Supplemental Figure S4. The majority of this variation is from the coupled motion between the acceptor and anticodon stems. A similar set of motions is apparent in the ANM and shown in Figure 3. Note that the model granularity used in Figure 3 is coarse-grained, with exaggerated motions of the acceptor arm. Coarse-grained models of this type do not restrain bond lengths to have their accepted highly restricted small ranges, sometimes rendering the absolute amplitudes of computed motions nonquantitative. The relative amplitudes still carry important directional information, but interpretation of the motions to have physical limits to the extent of deformation permitted is necessary, and some modifications of the ANM to better account for this have been explored (Stember and Wriggers 2009; Lin and Song 2010; Mendez and Bastolla 2010; Lu and Ma 2011).

In Table 2, we report the percentage of the variance captured by the first three PCs for the 75 tRNA structures using four different alignment algorithms and phosphate atom positions (further details are given in Supplemental Fig. S4). A more detailed structure model is also considered which includes all backbone and ribose sugar heavy atoms (visualized in Supplemental Fig. S5) in the 68 structures having these atom positions reported (sufficient resolution), computing a second ensemble and comparing to atomic ANM. We find that 59%–70% of ensemble variance is captured by only three PCs.

We immediately see that the mode 1 projection is very different from the PC1 projection. Mode 1 partitions the structures into those that have the acceptor arm in an “up” or “down” position coupled to a corresponding extension or compression of the anticodon stem–loop. Mode 2 involves the isolated movement of the anticodon stem–loop, while mode 3 is dominated by tangential movement of the acceptor arm relative to mode 1. Interestingly, in order to achieve the amount of flexing of the anticodon stem–loop that is observed in the PCs, the ANM modes require the counterbalance of a significantly exaggerated motion of the acceptor arm (mode 1). While the acceptor arm may appear to have exaggerated motions here, similar to the “tip effect” noted by the Ma group (Lu et al. 2006), nonetheless, its large motion does appear to meet the required deformation it undergoes when binding inside the ribosome.

**FIGURE 2.** PCs of the tRNA ensemble constructed from all common atoms, showing the physical limits of deformation. (A) Scatter plot of the structures in the space of the first three principal components (PCs). Each tRNA structure is plotted as a point in PC1-PC2, PC1-PC3, and PC2-PC3 space after being aligned using TM-align. Structure 1T3R is marked by a red star and is the central structure to which the other members of the ensemble were aligned. The structure closest to the origin in PC space is 2B9MW, which is marked with a blue star, and because of the proximity of 2B9MW to the origin, it is used for illustrating the PCs. (B–G) The first three PCs are shown as deformations to 2B9MW in two orthogonal views using PyMOL (32). We show the unaltered structure in yellow (ribose and bases) and orange (backbone trace) with the PC deformations of phosphate atoms only that are of significant amplitude as blue arrows, and the resulting backbone trace in red. PCs are shown as (B) +150 PC1, (C) −150 PC1, (D) +100 PC2, (E) −100 PC2, (F) +50 PC3, (G) −50 PC3. Magnitudes used for each mode are taken from the maximum deviation seen in A. From these images, we begin to understand the physical limits to each PC. For instance, the negative directions of PC1 and PC2 are quite restricted and correspond to compression of the minor groove of the anticodon arm and simultaneous stretching of the anticodon stem–loop. Neither of these actions can be sustained to a large extent by the structure. See text for more details. PC3 consists of a smaller motion (and accounts for a smaller contribution to the total variance) that also describes communication between the acceptor arm and anticodon stem–loop.
we select the subvector from each normal mode corresponding to the atoms selected for PC generation. We then re-orthogonalize the collection of subvectors using Gram-Schmidt renormalization (Apostol 1962), so that we have a basis set for the motion of the backbone in the context of the all atom model.

ANM using atomic details accurately captures the motions in the ensemble. Model performance for the heavy atom

Figure 2 also displays the structural ensemble mapped to PC space to show how the individual structures are distributed. We see that the structure to which others are aligned (1TRA) is not at the origin in PC space. Instead, 2B9M chain W is much closer. These two structures are only 1.6 Å RMSD apart from each other and possess highly similar mode shapes. For this reason, the deformations visualized in Figures 2 and 3 are considered, using the initial coordinates from 2B9MW. It is apparent that an asymmetry exists; the structures exhibit the closest sequence match at the anticodon stem–loop, whereas the positive PC1 direction exhibits simultaneous unwinding and bending, which is less physically constrained.

For the more detailed atomic models, we take all backbone heavy atoms and align the different structures with TM-align. Then, the covariance matrix is decomposed into principal components (PCs), and ANM models are constructed from all heavy atoms in the representative structure, ITRA. Backbone atoms are considered for the PCs to avoid complications arising from sequence variation and chemically modified bases. For comparing these PCs and the ANMs, we select the subvector from each normal mode corresponding to the atoms selected for PC generation. We then re-orthogonalize the collection of subvectors using Gram-Schmidt renormalization (Apostol 1962), so that we have a basis set for the motion of the backbone in the context of the all atom model.

The percentage of the total variance captured by each of the first three PCs for four different alignment methods with either a coarse-grained (phosphate atoms only) or all-heavy-atom backbone. The backbone selection is defined as the ribose sugar and backbone phosphate and oxygen atoms, where the bases themselves have been excluded because of their variable atom types (see Supplemental Fig. S5). The four alignment methods are abbreviated as follows: seqW, for sequence weighted—a sequence alignment is performed, and atoms aligned by sequence are given more weight in the structural alignment; minR, for minimum total RMSD; ENMa, mean-square fluctuation in the first two modes is computed, and the 14 points with the smallest motion are super-imposed; TM denotes TM-align. From this data, no alignment method is clearly better than another, but the seqW method is the least consistent in its structural alignments. The conciseness of the modes differs and TM-align performs similarly to ENMa. For the coarse-grained ensemble, TM marginally outperforms ENMa and minR, but for an ensemble at atomic detail ENMa, outperforms the others.

This alignment protocol may appear to be the best by conciseness of the PCs, but a large portion of the variance is captured by this PC because of a poor structural alignment. Some of the structures exhibit the closest sequence match at the anticodon stem–loop, leading to the larger part of the structure being under-weighted in the structure alignment. The first PC distinguishes these two groups—the bulk of the structure predominantly aligned or the anticodon stem–loop alone.

We align backbone atoms that meet the MSF criteria (see Materials and Methods).

### Table 2. Percent of variance in the tRNA ensemble for each alignment method

| Var | seqW | minR | ENMa | TM | seqW | minR | ENMa | TM |
|-----|------|------|------|----|------|------|------|----|
| PC1 | 33.2 | 41.0 | 40.7 | 43.0 | 69.2 | 29.9 | 38.3 | 35.0 |
| PC2 | 21.8 | 18.3 | 18.2 | 18.9 | 12.2 | 19.5 | 21.9 | 19.4 |
| PC3 | 12.5 | 8.2  | 9.6  | 7.5  | 8.3  | 9.5  | 9.3  | 8.0  |

The percentage of the total variance captured by each of the first three PCs for four different alignment methods with either a coarse-grained (phosphate atoms only) or all-heavy-atom backbone.
ANM model and atom-based PCs are evaluated for cutoff values up to 25 Å, but little change is seen either in cumulative overlap (CO; Equation 4) or root mean square inner product (RMSIP; Equation 5) above 8 Å. CO quantifies how well an individual PC is captured by a set of modes, while RMSIP considers a subspace defined by a set of PCs and the extent to which it is covered by a set of modes. We compare 3, 6, 10, or 20 modes to each of the first three PCs (CO) or the combination of up to three PCs (RMSIP) and find that considering 20 modes captures 75%–80% of the dominant PCs. See Supplemental Figures S4 and S6 for examples of overlaps of individual modes and PCs. Using fewer modes accounts for less of the ensemble variation; however, by using only three normal modes, ~55% of the space spanned by the first three PCs is covered.

One of the most remarkable findings from studies on proteins is the excellent performance of coarse-grained models for representing the motions. Some loss in the fidelity of computed motions occurs, but nonetheless, the essential aspects of these motions are observed with both the atomic and coarse-grained models. We have already demonstrated that the ANM motions of one representative structure capture the motions observed across a large set of experimental structures. Next, we will investigate the performance of coarse-grained representation of the tRNAs. Since RNA is a different material from protein, we test multiple structural representations of tRNA to determine the level of detail that is most informative for ANM motion computations. We consider the effect on elastic network model performance, as the resolution with which we treat the tRNA and its environment is changed. In order to compare the same experimental data to various models, we construct the PCs by using phosphorus atom coordinates from the backbone phosphate group (referred to hereafter as the phosphate atom). The models used vary in resolution from one point per base, all heavy atoms, including hydrogen atoms, or inclusion of increasing portions of the crystallographic environment (see Materials and Methods; Supplemental Fig. S5). Figure 4 shows model performance for distance-dependent springs, a one-point-per-base representation (only phosphate atoms), and CO using either six or 20 normal modes. Results for cutoff-based models and other representations are presented in Supplemental Figures S7 and S8.

The strong relationship between the first PCs and the normal modes points to the possibility of using the normal modes to align the structures. To test this hypothesis, we first calculate the mean-square fluctuation of each atom in ITRA from the two lowest frequency modes. We next select residues with the smallest mean-square fluctuation (Supplemental Fig. S10) and align the ensemble using least squares fit with only these atoms; this method is designated ENMa for “ENM-alignment.” The relationship between CO using the lowest frequency modes and all four alignment types is compared in Figure 5 where we find ENMa to be the most consistent (see further details in Supplemental Fig. S9).

To further investigate the amount of information lost upon coarse-graining of the system, we directly compare ANM motions built with heavy atom and phosphate only models. Supplemental Figure S11 depicts the correlations between computed temperature factors (the relative amplitude of motion for each atom) for atomic and phosphate-only models and metrics that describe the agreement of their anisotropy.
Anisotropy refers to how the atoms fluctuate to different degrees in each direction and is described by a $3 \times 3$ tensor. These tensors are compared using the modified real-space correlation coefficient and Kullback-Leibler distance, both of which are described by Zheng (2010). We find that the models exhibit best agreement with each other when the atomic cutoff is 7 Å or larger and the coarse-grained cases with a cutoff of 20 Å or greater. These cutoff values correspond to the best performance as judged by agreement with crystallographic B-factors (Fig. 6). Interestingly, the inclusion of hydrogen bonds does not significantly affect CO but does modestly improve temperature factor agreement (data not shown). At lower cutoff values more customarily used for atomic models (~5 Å), stable models are still generated and show interesting behaviors. Figure 7 describes both the overall fluctuation of the tRNA and the changes in interatomic distances from the ANM model. Bases experiencing the greatest fluctuations also experience the largest relative displacements. This is an important distinction as it confirms the internal movements in the structure and distinguishes the observed motions from the movement of relatively rigid domains. The ANM does not distinguish among interactions of different types and relies only on the proximity of atoms or coarse-grained points. Yet, bases that share hydrogen bonds between them (Watson-Crick or Hoogsteen) or are in base-stacking configurations are predicted by the model to be highly constrained. Thus, the ANM models that best reproduce experimentally determined flexibility at atomic or coarse-grained resolution are also the most consistent in their computed internal motions (see Supplemental Fig. S12 for a similar comparison using the experimental ensemble).

In this study, we consider the ability of ENMs to capture the spatial variance seen in a collection of 16 RNA ensembles and perform a more thorough analysis of tRNA dynamics. ENMs have become an increasingly popular method for determining the dominant motion of macromolecules from relatively small structures to the largest structures such as the ribosome. They have been confirmed previously to reproduce the principal components made from the covariance matrix of 164 superimposed HIV protease structures (Yang et al. 2008). Performing a similar analysis, we confirm that the low frequency ANM modes sample experimentally determined conformations of well-packed RNAs.

As many of the structures we consider here are riboswitches, it would be interesting to attempt to predict their transitions. While these are important challenges, the prediction of base re-pairing events is beyond the scope of this work. Further, determining whether computed motions point toward other known conformations that are far from the initial structure (as has been done in proteins) is more complicated here due to the more limited numbers of resolved structures. For many of the riboswitches in our data set, experimental structures exist for the highly base-paired sections but are lacking for the full structure in both conformations. Without having complete structures for both end points of the transition, this problem is poorly defined in structural terms.

While single-point-per-residue models are commonly used to represent proteins, such a level of coarse-graining may be too extreme for many applications in RNA. In nucleotide systems, multiple torsion angles are required for the backbones, as well as points corresponding to the base position and its orientation (Malathi and Yathindra 1981). While these denser models, as well as detailed models that take into account the crystal environment or closely packed water molecules, do not outperform the phosphate-only ANM model (data not shown), this model is limited in the resolution of questions it can address (like dynamics that alter base-pairing). In order to subvert issues caused by sequence variation and chemically modified bases, PCs were generated using the sugar-phosphate backbone atoms and compared to atomic ANM (Supplemental Fig. S5). Even at this level of detail, ANM can successfully recapitulate the important motions evident in the ensemble (Supplemental Fig. S6). Since inclusion of crystal environment has shown improved results for proteins, this may point to a characteristically different behavior for RNA.

We could have chosen the representative structure in such a way as to maximize the agreement between ENM and the structural ensemble, but in this initial study, a representative was chosen as the highest (best) resolution. A further
extension would be to also analyze the sensitivity of the agreement on initial structure choice and to attempt to determine the extent to which the structure is deformable in each normal mode by comparing to the observed conformations.

Combining all available information, we find that the most accurate and versatile model is an atomic level representation using a cutoff of 7 Å. This is a consistent model, achieving high agreement with coarse-grained models and concentrating the modes with highest overlap with atomic variations, and is capable of addressing more detailed questions. We visualize the mode-space of this model in Figure 3, finding that the computed motions are very similar to the PC subspaces (Fig. 2). The first mode of motion is noticeably dominated by a large motion of the acceptor arm which is coupled to motion of the anticodon stem–loop. This motion groups the structures into two groups—those with the acceptor arm flexed down or upward, relative to 2B9MW. The next two modes of motion refine the fit between structures by further modifying the anticodon stem–loop minor groove compression along with out-of-plane motion of the acceptor arm. After comparing many ANM models, we find that the regions of the tRNA that are constantly predicted incorrectly are the D and T arms. This region has relatively low experimental temperature factors, and cutoff-based ANM models tend to predict too much relative motion. This may be due to suppressed motion in the crystal due to packing but may also be caused by the stabilization of non-Watson-Crick (Hoogsteen) interactions, base-stacking, or ions that are not explicitly present in our coarse-grained models. Due to the high performance of all atom representations at a low cutoff, a more detailed hydrogen-bond ANM may exhibit further gains as has been indicated by recent studies of proteins (Stember and Wriggers 2009; Lin and Song 2010; Mendez and Bastolla 2010; Lu and Ma 2011; Seo and Kim 2012). These results may be indicative of the increased importance of long-range electrostatic interactions across RNA structures that play a more dominant role in fold determination and stabilization than in globular proteins. A surprising insensitivity to distance dependence is observed, contrary to what has been shown in globular proteins. This may be caused by the relative importance of long-range electrostatic interactions (often Mg$^{2+}$-mediated) that exist in RNAs.

In many ENM studies, researchers begin by animating the biomolecule using the normal modes, visualizing the effects on the structure of excitation of one or more modes. It is most common to begin with the lowest frequency mode and visually compare the first few—perhaps up to a dozen. Thus, it is not always as practically useful to show that 20 modes capture experimental ensembles when many researchers will only consider the first few. In Figure 5, we report the

**FIGURE 7.** Comparison between global motion, atom–atom displacements, and the relative rigidity of paired and stacked nucleotides. (A) The mean square fluctuation (MSF) of each nucleotide measures the overall displacement in the ANM model and is highly correlated with the median change in interatomic distance, taken across the structure. See Materials and Methods. (B) Median (upper triangle) and median absolute difference (lower triangle) of changes in internal distances $\Delta d_{ij}$ is shown, where each pairwise base interaction is summarized across all atoms. Color is blue to red on a relative scale. The secondary structure is shown by color (same as Supplemental Fig. S10) along the axes; acceptor arm in purple, D arm red, anticodon arm blue, variable loop orange, and T arm green. The anticodon stem–loop experiences the greatest motion, and the variation among atoms is high within the stem–loop. Comparing to the same metrics measured in the ensemble (see Supplemental Fig. S12), we find a strong correlation but the motions in our ANM model are more localized. (C) We use as a test statistic the sum of internal distance changes between nucleotides that have in-plane hydrogen bonds with each other; canonical Watson-Crick as well as Hoogsteen interactions (red line). Randomly sampling the same number of pairs and summing their $\Delta d_{ij}$, we find the hydrogen-bonded pairs exhibit significantly less deviation. (D) An even stronger relationship is seen when stacked nucleotides are compared.
alignment-dependent performance using only three or six modes and either one or three points per base. We find that the minimum total RMSD alignment can give the most easily interpreted mode space (the very lowest frequency modes are most meaningful for capturing PCs), but that the ENM-assisted alignment is the most consistent. A notable consideration for interpreting the results presented here is that the actual functional meaning of each mode is not changed upon alignment algorithm choice, but their apparent meaning is. That is, the modes are strongly related to the structural ensemble, but one must properly align the ensemble in order for the metric to reflect this agreement.

PCs generated using all four alignment procedures are similarly concise (capture a large portion of variance in a few PCs), though conciseness may, in fact, point to a poor structure alignment (see Table 2). In PC space, individual structures appear in different relative positions when different alignment algorithms are used. This underscores the importance of alignment algorithm choice for interpreting what PCs or computed motions may signify and the relative similarities of each structure within the ensemble. Aligning by using the atoms that move the least in the global mode of motion (ENMα) results in increasing the overlap of the space spanned by the first few normal modes and PCs, at the cost of some conciseness of the overlaps. This is a somewhat opposite approach from that of the recently published ALADYN method of Potestio and colleagues (Potestio et al. 2010) which optimally matches the largest motions from two protein structures. Refinement of this method could lead to further improvements and insight into the structure-function relationship. Numerous studies have shown that the lowest frequency modes are biologically relevant. ENMα structure alignment further emphasizes that the global modes are, indeed, biologically meaningful and may be useful for further exploring the conservation of functional motions in biomolecules.

The presence of bound protein can have a pronounced effect on conformation. As this study set out to determine the extent to which ENMs can accurately model RNA conformations in general, bound protein partners were not included in the models. Doing so is beyond the scope of this work but could be important for understanding the functional motions of physiological complexes. For tRNA, we find that structures far from the origin in PC space are almost always bound to proteins, but that numerous bound structures are close to the origin as well. Protein partners found near the PC1-PC2 origin include ribosomes (2B64, 2HGR, 2HGI, 3DEG) and amino acid charging enzymes (1SZ1). If we measure the extent of interactions by the number of amino acids within a close radius of the tRNA, we find that the extent of protein binding is also not clustered in PC space, nor do the most highly bound structures (more than 10 Cα atoms within 7 Å) represent outliers. Thus, neither the presence nor the extent of bound protein is observed to be a dominant factor in the sampled deviations. The distribution of structures exhibits a directional asymmetry seen in all cases, showing that the structure is easier to deform in one direction along a given PC than the other. Translating this asymmetry of motion (knowing the physical limits) to the modes would be a valuable improvement to future ANM studies.

**DISCUSSION**

The structure and dynamics of RNAs are important for investigations of the role of mutation in disease, particularly for the large number of noncoding variants. We find that the low frequency ANM modes reproduce experimentally determined structural variation for well-packed RNAs. Previously, this type of confirmatory analysis had only been performed thoroughly for proteins. This investigation is important in that it increases our confidence in using these simple dynamics models for nucleotide and ribonucleo-protein systems, confirming their ability to reproduce the variation seen in structural ensembles.

In this work, we have neglected the dynamical effects of any binding partners, or missing parts of structures, on the motions of each RNA structure. Previous studies have considered the change in dynamics upon tRNA and rRNA binding to protein partners. The tRNA-cognate synthase complex has been analyzed by us (Bahar and Jernigan 1998; Wang and Jernigan 2005) and by Luthey-Schulten’s group (Alexander et al. 2010). Yan et al. (2008) showed that the bacterial ribosome’s global dynamics are only modestly affected by the presence of protein, as the overall shape is largely defined by the rRNA. An important observation from the ENM models is that the motions of any part of a structure depend primarily on the shape of the entire structure (Doruker and Jernigan 2003; Lu and Ma 2005). The RNA group corresponding to the 23S pseudoknot is only a partial structure, ignoring the effects from the remainder of the rRNA and proteins, while most of the other structures are more complete. While a number of these have small missing parts of their structure (the riboswitches, for instance) or are known to bind other partners (like the tRNA), they are fairly complete or exist physiologically as a distinct entity. This is not the case for the 23S pseudoknot which is a relatively small piece of the whole 23S rRNA. For this reason, we are not surprised that the dynamics of this domain alone do not agree well with the dynamics sampled by this domain within the context of the ribosome, though the correspondence is still relatively high.

For proteins, by taking all residue pairs that interact, with spring strengths weighted by the inverse second power of their separation, there are consistent improvements in their dynamical behavior, but a less clear dependence has been observed here for RNA. This may point to a greater relative importance of longer range interactions in defining and stabilizing RNA structures. However, a comparison of the coarse-grained and atomic models directly shows their behaviors to be similar, emphasizing the dominant role of the overall
shape (Doruker and Jernigan 2003; Lu and Ma 2005) in determining the motions of RNA, just as for proteins. Models that weight local interactions more strongly yield a less accurate picture of RNA ensemble dynamics, pointing out the importance of the interplay between local and global motion that is not entirely captured in present models.

MATERIALS AND METHODS

Data set

Protein structure superpositioning has been used for many years for making structure comparisons that have led to well-established structural ontologies of the many protein folds, including CATH (Sillitoe et al. 2013), SCOP (Murzin et al. 1995), and Pfam (Punta et al. 2012). Because of the comparatively small number of available 3D atomic structures for RNA, fewer resources exist for their curation and comparison. The Rfam (Griffiths-Jones et al. 2003, 2005; Gardner et al. 2009, 2011), supported by the Wellcome Trust Sanger Institute in collaboration with Janelia Farm, is an important available resource that collects RNA sequences and available structures into families based on multiple sequence alignments and covariance models. To generate the ensembles studied here, we begin by filtering Rfam families. First, each family is divided into one or more subgroups based on sequence length; initially, all structures within a group may only differ by 12 bases in length. All groups having fewer than four members are discarded. For the remaining groups, 3D structures are downloaded and superimposed. Next, we consider the structural diversity within each group. If the largest difference between structures is too small (<1.5 Å RMSD) or too large (>8 Å RMSD), we would not expect the ENM to be as accurate. Structures falling into these two categories within a group are discarded and the group is discarded if fewer than four members remain. Any groups of structures remaining are retained for further analysis. In total, 16 ensembles of experimental structures have been found that are retained and analyzed here. Four structure alignment methods are compared, but the method used in determining if a structure will be retained in an ensemble is a modified multiple structure alignment procedure, based on the MUSTANG algorithm (Konagurthu et al. 2006), which is used to make all pairwise comparisons in a similar fashion to multiple sequence alignment. Unless stated otherwise, the modified MUSTANG was used in structural alignment of ensembles of RNA structures.

The tRNA ensemble corresponds to classifier RF00005 and, after removing structures with large extensions in the variable loop or missing residues (unresolved in the X-ray experiment), the group still contains 75 structures (Supplemental Table S1). Similar to the protein case (Yang et al. 2008), we capture conformations from wild type, antibiotic-bound, ribosome- or elongation factor-bound, cognate synthetase-bound, etc. The multiple sequence alignment, derived from structure matching, of the 75 structures is shown in Supplemental Figure S3, where the position of chemically modified bases is also marked. We consider principal components and ANM models built from phosphate atom coordinates but also more detailed models that include the ribose sugar and backbone atoms (see Supplemental Fig. S5). Only backbone atoms are considered for the more detailed analysis to avoid the complications that would arise from variable sequence and modified bases. Seven structures were not of sufficiently high resolution to resolve the backbone atoms and are thus excluded from the more detailed analysis (listed in Supplemental Table S2). Some of these remaining sequences have fewer than 76 nt. To permit retention of these structures in our data set, we consider only the 73 nt that best fit the 1TRA structure by permitting up to 2 nt on the N terminus and 1 nt on the C terminus to be missing in the reported structures.

Elastic network model

Coarse-grained protein structures are often represented by Cα atom coordinates with harmonic springs to connect spatially close residues, since it has been shown that the dynamics of such coarse-grained structures closely resemble that of the atomic structures (Sen et al. 2006). This has been demonstrated using an elastic network representation of the protein structure. The anisotropic network model (Atilgan et al. 2001) can be used to compute the directions of motions of all points within a structure. To generate an ANM model, we first construct a Laplacian (or Kirchhoff) matrix &F with using Equation 1, where \( r_c \) is a cutoff radius (typically 10–13 Å for proteins), \( d_{ij} \) is the distance between atoms \( i \) and \( j \), and \( \gamma \) is the spring constant, taken to be identical for all interactions between close atoms. We then compute a matrix of second derivatives of the potential energy—details are given in Atilgan et al. (2001)—the eigenvectors \( (Q_i) \) of which are called normal mode shapes, and the eigenvalues \( (\lambda_i) \) are the corresponding square frequencies \( (\omega_i^2) \). Only a few slowest modes are usually important contributors to the total motion, as these contributions decrease rapidly with increasing mode index corresponding to increasing frequency. For a given normal mode, \( i \), fluctuations of the structure \( (\Delta R_i) \) are computed with Equation 2 for normal mode \( Q_i \). Low-frequency normal modes represent the collective motions of the system and have been shown to be biologically relevant.

\[
\Gamma = \begin{cases} 
-\gamma & d_{ij} \leq r_c \\
0 & d_{ij} > r_c \\
-\sum_{k=1,N+1} \Gamma_{ik} & i = j 
\end{cases} 
\]  \hspace{1cm} (1)

\[
\Delta R_i = Q_i \cos(\omega_i t). \]  \hspace{1cm} (2)

We also generate atomic and coarse-grained models with different levels of detail in order to learn about how best to represent RNA to retain its atomic dynamics and also to learn about the amount of information lost upon coarse-graining. The crystal environment is also considered which Riccardi et al. showed can improve protein models (Riccardi et al. 2009). We also test an alternative to cutoff-based ANM by defining spring interactions using distance-dependent springs previously developed (Yang et al. 2009). In these models, all atoms in a structure are considered to be connected with springs weighted by \( y = d_{ij}^{-n} \), where \( x \) is the power dependence on the distance between points \( i \) and \( j \) in the structure. This is a highly cooperative model in which all points interact with each other, but differentially, with the close pairs interacting strongly and the distant pairs interacting only very weakly.

Structure alignment

Once the groups of structures are defined, they can be spatially superimposed. For each ensemble, all members are aligned to the highest resolution wild-type conformation using TM-align (Zhang
and Skolnick 2005). The TM-align algorithm was developed for comparing proteins, and the initial alignment used in the search for a global optimum employs a fitting of secondary structure elements. Our modification of the algorithm for RNA essentially treats each base as if it was a separate secondary structure element. In other words, there is no accounting for RNA secondary structure. The designation "wild type" used here is taken directly from the FDB file annotations. A more unbiased method would be to compute sequence alignments relative to a reference sequence, but the sensitivity of the results to the choice of representative structure is not the focus here.

To investigate the effect of the alignment algorithm choice, we focus on the tRNA ensemble, utilizing three established alignment methods, and propose an additional new ENM-assisted algorithm. The three established algorithms are a minimum total root mean square deviation sequence matching (minR) (Kendall 1989), sequence alignment followed by weighted RMSD superpositioning of the well-aligned atoms (seqW; align command in PyMol), and TM-align (Zhang and Skolnick 2005), which aligns secondary structure elements and employs a heuristic to refine the initial fit. After considering the performance of these three alignment types, a normal mode-based alignment protocol is considered where atoms with lowest mean square fluctuation in the global modes are aligned by minimum RMSD, which is here called ENM-assisted or ENMa.

**Principal component analysis (PCA)**

PCA is often used for dimensionality reduction on complex data where the idea is to rank-order the contributions of granular variables. The first few PCs capture a significant part of the ensemble variance, with the first PC giving the largest contribution, the second capturing the largest part of the remaining variance, and so forth. Often the first few PCs account for a majority of the variance in each ensemble. This rank-ordering of the PCs allows us to ignore most of them while retaining the most important information, thereby reducing the number of important degrees of freedom and simplifying the analysis.

We construct a matrix where each row holds all coordinates for a single structure. Columns are then variables, one for each structure coordinate. PCA is performed on this matrix using MATLAB 2010a.

**Structure representation**

ANM performance is tested across multiple structure representations at atomic and coarse-grained resolutions. These range from the simplest model which takes into account only the backbone phosphate atom positions, two points per base (P and C2'), three points (P, C2', and O4'), all heavy (nonhydrogen) atoms, all atoms including hydrogen atoms (with and without covalent bonds), including the X-ray resolved waters, and with symmetry-related intermolecular atoms within 5, 7, or 12 Å added into the structure. Another appropriate representation would be to use the C4' and either the N1 (pyrimidine) or N9 (purine) atom (Cao and Chen 2005, and references therein). Our first investigations of RNA dynamics with the ENM used this representation. However, we found that the stiffness within each strand was overly strong relative to the between-strand strength (data not shown). Often, base-paired helices would slide against each other or "break" open, displaying little tendency to remain structured. While these dynamics might be important predictions of unfolding events, they did not agree with the motions implied by the ensembles of crystal structures studied here. Thus, we utilized a representation that more accurately captures the packing density within the structures to avoid this unpairing problem. The C2 atoms capture density-linking base-pair interactions (canonical or Hoogsteen) better than do the N1 or N9 atoms. Further, we found the O4' atom to be more equidistant between a nucleotide's phosphate atom and its C2 carbon than is the C4' atom. See Supplemental Figure S5 for further explanation of these representations.

**Comparing PCs and modes**

Tama and Sanejouand (2001) defined a normalized overlap for comparing the ith PC ($P_i$) to the jth mode ($M_j$), and we use this here to compare the PCs from the experimental structures with the normal modes computed for a representative structure:

$$O_{ij} = \frac{|P_i \cdot M_j|}{\|P_i\| \|M_j\|}.$$  
(3)

The cumulative overlap between the first k normal modes and a given PC measures how well the first k modes together can capture the variance represented within a single PC (Equation 4). The overlap between the space spanned by the first l PCs and the first l low-frequency ENM modes was defined by the root mean square inner product in Leo-Macias et al. (2005) and is shown here in Equation 5.

$$CO(k) = \sqrt{\sum_{j=1}^{k} O_{ij}^2},$$  
(4)

$$RMSIP(J, I) = \sqrt{\frac{2}{I-1} \sum_{i=1}^{I} \sum_{j=1}^{J} (P_i \cdot M_j)^2}.$$  
(5)

The effect of the choice of representative structure was not explored but may be important for determining the performance. For each group, the highest resolution structure was arbitrarily chosen as the representative, but some other choices might represent the ensemble more accurately.

ANMs allow the efficient computation of changes in inter-atomic distances upon excitation of normal modes. These changes to internal distances can be computed directly from the system's Hessian, defined in Equation 1:

$$\Delta I_j = \langle (\Delta R_i - \Delta R_j)^2 \rangle = (3k_B T/\gamma) [\Gamma_+^{-1} + \Gamma_-^{-1} - 2\Gamma_0^{-1}].$$  
(6)

The interaction strength $k_B T/\gamma$ is an adjustable parameter and is typically set so that computed mean square fluctuations best match experimental crystallographic temperature factors. In summarizing the variation in $\Delta I_j$ between nucleotides, we use median and median absolute deviation (MAD), rather than mean and standard deviation, as the former is less sensitive to outliers.

**SUPPLEMENTAL MATERIAL**

Supplemental material is available for this article and includes the images of the superimposed structures and the animations of the structures along normal modes and also along the principal
components. These same files are available on our web site: http://ribosome.bb.iastate.edu/4papers/2014/Zimmermann_Jernigan/Index_of_Contents.html. In addition, on that web site are also included the coordinate files of the aligned ensembles (pdb formats) and the coordinate files used to generate the animations of the normal modes and the principal components.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the following NIH grants: R01GM073095 and R01GM72014. We thank Sumudu Leelananda for assistance with the manuscript.

Received July 12, 2013; accepted February 14, 2014.

REFERENCES

Alexander RW, Eargle J, Luthey-Schulten Z. 2010. Experimental and computational determination of tRNA dynamics. FEBS Lett 584: 376–386.

Apostol TM. 1962. Calculus of several variables with applications to probability and vector analysis, Vol. II, pp. 465–467. Blaisdell Publishing Co., New York.

Attilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80: 505–515.

Bahar I, Jernigan RL. 1998. Vibrational dynamics of transfer RNAs: comparison of the free and synthetase-bound forms. J Mol Biol 281: 871–884.

Bahar I, Rader AI. 2005. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 15: 586–592.

Bakan A, Bahar I. 2011. Computational generation inhibitor-bound conformers of p38 map kinase and comparison with experiments. Pac Symp Biocomput 2011: 181–192.

Berstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112: 535–542.

Cao S, Chen S-J. 2005. Predicting RNA folding thermodynamics with a reduced chain representation model. RNA 11: 1884–1897.

Delarue M, Dumas P. 2004. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. Proc Natl Acad Sci USA 101: 6957–6962.

Doruker P, Jernigan RL. 2003. Functional motions can be extracted from on-lattice construction of protein structures. Proteins 53: 174–181.

Doruker P, Jernigan RL, Bahar I. 2002. Dynamics of large proteins through hierarchical levels of coarse-grained structures. J Comput Chem 23: 119–127.

Dunham I, Kundaje A Aldred SF. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74.

Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen D, Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateeman A. 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res 33(Database issue): D121–D124.

Hayward S, de Groot BL. 2008. Normal modes and essential dynamics. Methods Mol Biol 443: 89–106.

Jernigan RL, Kloczkowski A. 2007. Packing regularities in biological structures relate to their dynamics. Methods Mol Biol 350: 251–276.

Kendall DG. 1989. A survey of the statistical theory of shape. Stat Sci 4: 87–120.

Konagarthu AS, Whisstock JC, Stuckey PJ, Lesk AM. 2006. MUSTANG: a multiple structural alignment algorithm. Proteins 645: 559–574.

Leo-Macias A, Lopez-Romero P, Lupsy D, Zerbino D, Ortiz AR. 2005. An analysis of core deformations in protein superfamilies. Biophys J 88: 1291–1299.

Lin TL, Song G. 2010. Generalized spring tensor models for protein fluctuation dynamics and conformation changes. BMC Struct Biol 10(Suppl 1): S3.

Lu M, Ma JP. 2005. The role of shape in determining molecular motions. Biophys J 89: 2395–2401.

Lu M, Ma JP. 2011. Normal mode analysis with molecular geometry restraints: bridging molecular mechanics and elastic models. Arch Biochem Biophys 508: 64–71.

Lu MY, Poon B, Ma JP. 2006. A new method for coarse-grained elastic normal-mode analysis. J Chem Theory Comput 2: 464–471.

Ma JP. 2005. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 13: 373–380.

Malathi R, Yahindra N. 1981. Virtual bond probe to study ordered and random coil conformations of nucleic acids. Int J Quant Chem 20: 241–257.

May A, Zacharias M. 2008. Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking. Proteins 70: 794–809.

Menendez R, Bastolla U. 2010. Torsional network model: normal modes in torsion angle space better correlate with conformation changes in proteins. Phys Rev Lett 104: 228103.

Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247: 536–540.

Potestio R, Alekseyev T, Pontiggia F, Cozzini S, Micheletti C. 2010. ALADYN: a web server for aligning proteins by matching their large-scale motion. Nucleic Acids Res 38(Web Server issue): W41–W45.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. 2012. The Pfam protein families database. Nucleic Acids Res 40(D1): D290–D301.

Raney BJ, Cline MS, Rosenbloom KR, Dreszer TR, Learned K, Barber GP, Meyer LR, Sloan CA, Malladi VS, Roskin KM, et al. 2011. ENCODE whole-genome data in the UCSC genome browser (2011 update). Nucleic Acids Res 39(Database issue): D871–D875.

Riccardi D, Cui Q, Phillips GN Jr. 2009. Application of elastic network models to proteins in the crystalline state. Biophys J 96: 464–475.

Sen TZ, Feng Y, Garcia JV, Kloczkowski A, Jernigan RL. 2006. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. J Chem Theory Comput 2: 696–704.

Seo S, Kim MK. 2012. KOSMOS: a universal morph server for nucleic acids, proteins and their complexes. Nucleic Acids Res 40(W1): W531–W536.

Shindyalov IN, Bourne PE. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 11: 739–747.

Sillitoe I, Cuff AL, Desaillly BH, Dawson NL, Furnham N, Lee D, Lees JG, Lewis TE, Studer RA, Rentzsch R, et al. 2013. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic Acids Res 41(D1): D490–D498.

Stemmer JW, Wriggers W. 2009. Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. J Chem Phys 131: 074112.

www.rnajournal.org 803
Tama F, Brooks CL. 2006. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu Rev Biophys Biomol Struct* 35: 115–133.

Tama F, Sanejouand YH. 2001. Conformational change of proteins arising from normal mode calculations. *Protein Eng* 14: 1–6.

Taylor WR, Flores TP, Orengo CA. 1994. Multiple protein structure alignment. *Protein Sci* 3: 1858–1870.

Tirion MM. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett* 77: 1905–1908.

Van Wynsberghe AW, Cui Q. 2005. Comparison of mode analyses at different resolutions applied to nucleic acid systems. *Biophys J* 89: 2939–2949.

Wang Y, Jernigan RL. 2005. Comparison of tRNA motions in the free and ribosomal bound structures. *Biophys J* 89: 3399–3409.

Ward LD, Kellis M. 2012. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* 30: 1095–1106.

Wu Y, Ma JP. 2004. Refinement of F-actin model against fiber diffraction data by long-range normal modes. *Biophys J* 86(Pt 1): 116–124.

Yang L, Song G, Carriquiry A, Jernigan RL. 2008. Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure* 16: 321–330.

Yang L, Song G, Jernigan RL. 2009. Protein elastic network models and the ranges of cooperativity. *Proc Natl Acad Sci* 106: 12347–12352.

Zhang Y, Skolnick J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33: 2302–2309.

Zheng W. 2010. Anharmonic normal mode analysis of elastic network model improves the modeling of atomic fluctuations in protein crystal structures. *Biophys J* 98: 3025–3034.

Zheng W, Brooks B. 2005. Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model. *J Mol Biol* 346: 745–759.