Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models

Sibsankar Kundu,* Julia S. Melton,† Dan C. Sorensen,‡ and George N. Phillips, Jr.*

*Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706; †Department of Biochemistry and Cell Biology, W.M. Keck Center for Computational Biology, and ‡Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005 USA

ABSTRACT The dynamic behavior of proteins in crystals is examined by comparing theory and experiments. The Gaussian network model (GNM) and a simplified version of the crystallographic translation libration screw (TLS) model are used to calculate mean square fluctuations of C_α atoms for a set of 113 proteins whose structures have been determined by x-ray crystallography. Correlation coefficients between the theoretical estimations and experiment are calculated and compared. The GNM method gives better correlation with experimental data than the rigid-body libration model and has the added benefit of being able to calculate correlations between the fluctuations of pairs of atoms. By incorporating the effect of neighboring molecules in the crystal the correlation is further improved.

INTRODUCTION

To understand a protein's function, one must know about both its structure and dynamics. X-ray crystallography can give good structural information by allowing the determination of the average position of atoms and the amplitudes of their displacements from these average positions. However, this classic analysis tells little about the ways the molecule moves. Many methods, such as molecular dynamic simulations, have been devised for modeling protein dynamics (MacKerell et al., 1998), but these often involve complicated and/or inaccurate potential functions and are computationally expensive. Some researchers have shown that simplified potentials, involving only a few parameters, can give results that are just as accurate as those of more complicated methods for many purposes (Tirion, 1996; Levitt et al., 1985; ben-Avraham and Tirion, 1998; Hinsen and Kneller, 1999; Higo and Umeyama, 1997).

The Gaussian network model (GNM) proposed by Bahar and colleagues (Bahar et al., 1997, 1998; Haliloglu et al., 1997; Haliloglu and Bahar, 1999) describe protein mobility in terms of the atoms' local packing density and exploits concepts developed in the theory of elastic networks (Eichinger, 1972; Kloczkowski and Mark, 1989). Tirion (1996) has shown that such single-parameter potentials can effectively model low-frequency modes of protein motion. Bahar et al. (1997) have further shown that for myoglobin the GNM gives measurable agreement with experimental crystallographic B-factors and furthermore can be used to calculate cross-correlations between motions of different atoms and compare them with NMR data (Haliloglu and Bahar, 1999). Although the GNM contains very little detail and is not amino acid specific, it gives remarkably reliable results for the C_α atoms with much less computation time than traditional dynamics simulations.

Crystallographic structure determination includes information about thermal and other fluctuations of the atoms in a crystal. Each atom can be assigned a Debye-Waller temperature factor or B-factor with the latter proportional to the mean square amplitude of the fluctuations. Although these factors have some limitations (Kuriyan et al., 1986) they represent a solid experimental source of information on the dynamics of proteins.

The translation libration screw (TLS) model (Schomaker and Trueblood, 1968; Sternberg et al., 1979; Kuriyan and Weis, 1991; Harata et al., 1999), developed by Schomaker and Trueblood, models a crystalline protein as an internally rigid body undergoing motion along translation, libration, and screw axes. Determining B-factors with the TLS model requires performing a six-parameter least-squares optimization of the observed and calculated diffraction patterns. In our study, we are interested in protein structure-based ab initio calculation of protein motion, so we modify the full TLS treatment to depend only on the molecular coordinates, calculating the square of the displacement of each C_α from the center of mass of the protein, corresponding to the lattice-independent libration component. For simplicity, we will refer to this simplified model as the libration model.

Although the GNM and libration models have both been shown to be capable of reproducing experimental B-factors for some test cases, no studies involving more than a few structures have been reported. Furthermore, debate continues as to which is more physically accurate and realistic and why (Haliloglu and Bahar, 1999).

In this context, we have completed a comparative study between librational and GNM methods in reproducing crystallographic B-factors with a set of 113 high-resolution (2.0 Å or better) proteins. We further modified the GNM calculations by incorporating the effects of neighboring atoms and molecules in the crystal lattice (Fig. 1). The

Submitted December 17, 2001, and accepted for publication April 17, 2002.

Address reprint requests to Dr. George N. Phillips, Jr., Department of Biochemistry, University of Wisconsin, 433 Babcock Drive, Madison, WI 53706. Tel.: 608-263-6142; Fax: 608-262-3453. E-mail: phillips@biochem.wisc.edu.

0006-3495/02/08/723/10 $2.00

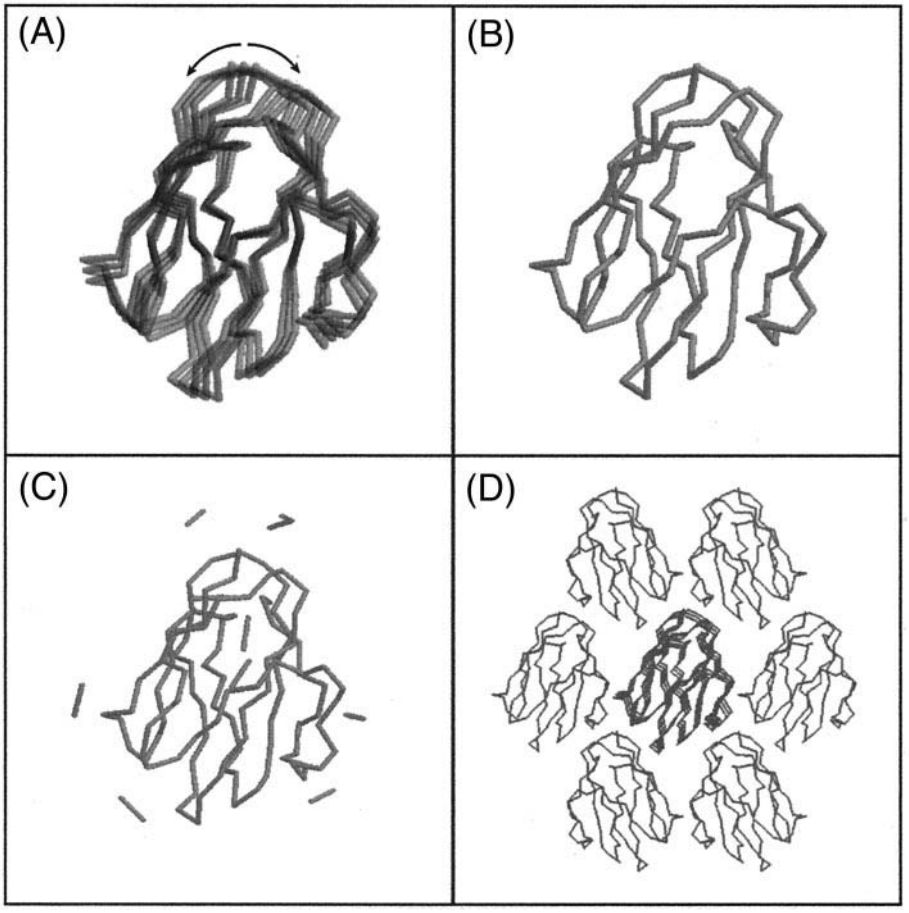

FIGURE 1 A pictorial diagram of the different models used in this work: (*A*) rigid body motion around the center of mass (RB); (*B*) GNM with an isolated protein molecule (GNM); (*C*) GNM with neighbor atoms within a certain distance equal to the spring length used for that calculation (GNM contact); and (*D*) GNM with all neighboring molecules (GNM neighbor).

GNM method has just two critical adjustable parameters, the maximum C_α-C_α distance for which the Hookean springs are attached and the associated force constant. The sensitivity of the calculation to this first parameter and analysis of the force constant are also explored.

MODEL AND METHODOLOGY

GNM background

The GNM describes a protein as an elastic network of α carbons attached by Hookean springs where the atoms fluctuate about their mean positions. The Kirchhoff or valency-adjacency matrix of such a structure is constructed using Eq. 1:

$$\Gamma = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{i,i\neq j} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (1)$$

where i and j are indices of α-carbons and r_c is the cutoff distance, normally ~7.0 Å. The close relationship of this matrix to the Hessian from classic normal mode analysis has been described (Atilgan et al., 2001).

A quantity proportional to the mean-square fluctuations of each atom and the cross-correlation fluctuations between different atoms are the diagonal and off-diagonal elements, respectively, of the pseudo inverse of the Kirchhoff matrix. This inverse can also be expressed as a sum of eigenvectors as in Eq. 2:

$$\Gamma^{-1} = \sum_{k=1}^{n-1} \lambda^{-1} q_k q_k^T, \quad (2)$$

where λ are the eigenvalues of Γ, arranged in descending order, with the smallest, zero-valued eigenvalue omitted. The q_k are the eigenvectors of Γ, and the superscript T indicates the transpose. For our symmetric positive semi-definite matrices the identical pseudo inverse can also be constructed using singular value decomposition

$$\Gamma^{-1} = V^T M_D^{-1} S,$$

where M_D^{-1} is a diagonal matrix of the reciprocals of M singular values and V and S are orthogonal and satisfy the usual singular value decomposition. Of course the terms with vanishing singular values must be omitted.

The variance/covariance matrix and B-factor of each atom can be calculated from the mean-square displacements by Eqs. 3 and 4:

$$\langle u_i u_j \rangle = (3k_B T/\gamma)[\Gamma^{-1}]_{ij} \quad (3)$$

$$B_i = 8\pi^2 \langle u_i u_j \rangle / 3, \quad (4)$$

where k_B is the Boltzmann constant, T is temperature, and γ is a constant scaling factor.

In the first part of the work we are interested in calculating the linear correlation coefficient between the experimental and calculated B-factors as given by

$$\rho = \frac{\sum_{j=1}^{n} (x_j - x)(y_j - y)}{\left[\sum_{j=1}^{n} (x_j - x)^2 \sum_{j=1}^{n} (y_j - y)^2\right]^{1/2}}, \quad (5)$$

where x_j is the experimental B-factor value of the jth C_α-atom, x is the mean value of the x_j values, y_j and y are the corresponding quantities for calculated B-factors, and n is the total number of C_α-atoms. This number measures only the relative rise and fall of the two curves (Eq. 5) and does not require that they be scaled properly. The correlation coefficient can range from -1 (perfect anti correlation) to $+1$ (perfect correlation). The absolute scale of the theoretical predictions depends on a spring constant, which can also be determined by comparing the experimental and theoretical curves (see below).

Method of calculation

Structures

For the comparative study between GNM and libration model to experimental B-factors, a set of 113 proteins from two different nonredundant sets were used. Researchers at Duke (Word et al., 1999) and Stanford (Singh and Brutlag, 1997) have each compiled a list of nonredundant high-resolution structures. We have combined these two lists, using only structures that were solved by x-ray diffraction, that are not oligomeric assemblies, and that have only one chain in the asymmetric unit, leaving 113 structures for comparisons of the two models. They are listed in Table 3.

All of the proteins examined in this study had a resolution better than or equal to 2.0 Å, with the exception of 1ACC, which had a resolution of 2.1 Å. All of the structural coordinates were obtained from the Brookhaven Protein Data Bank (PDB) (Bernstein et al., 1977). Some of the structures contained in the nonredundant lists were unavailable from the PDB, and the following substitutions were made: 1CYO for 3B5C (cytochrome B5) and 5PTP for 4PTP (b-trypsin). Because the distinction between homodimers and structures with two identical chains in the asymmetric unit was rather subjective, we have excluded such structures.

Consistent with past research, the α carbons alone were used to model the protein structures. Though many structures contained counterions or cofactors on the surface or in the active site, there exists no good systematic method for modeling these, and it must be done entirely ad hoc. The only cofactor we have modeled is the heme group, where the four bridging methylene carbons and the iron atom are all treated as C_α atoms. When calculating the center of mass, all atoms were given a mass of 12 atomic mass units.

Numerical calculations

We used Mathematica to calculate the GNM-based B-factors for each protein in a batch mode. The Kirchhoff matrix is formed first from Eq. 1. We invert the Kirchhoff matrix with the help of Eq. 2 and with each eigenvector contributing toward the B-factor. We ignore the eigenvector with value zero. Arranging the eigenvectors in the ascending order of their eigenvalues, we found that the first 30% of them are the major contributors toward the B-factor, but inclusion of all eigenvectors helps slightly (Fig. 2). For the anisotropic network model calculations we used Mathematica's Pseudo Inverse function, which uses a singular value decomposition formalism instead of eigen analysis.

GNM with contacts and neighbors

Usually GNM assumes springs between C_α atoms only within the same molecule. But these molecules are not isolated entities in crystals. Instead they reside in a lattice with neighbors. We included the neighbors in our calculation to incorporate the effect of environment on dynamic behavior. We first included only C_α within 7.0 Å of the concerned molecule but also considered all neighboring molecules surrounding the central molecule. We used the program CNS (Brunger et al., 1998) to identify the neighboring atoms and molecules.

Libration model

As previously mentioned, the TLS model is approximated by assuming mean square fluctuations are proportional to the square of the distance of each α carbon from the protein's centroid. The square of the distance, rather than the distance itself, is used to give the calculated B-factors the same units as those in the GNM calculations. Correlation coefficients were calculated by standard procedures.

Determination of k_BT/γ

To calculate B-factors from Eqs. 3 and 4 we determined the value of k_BT/γ by least-squares fitting to the observed B-factors. k_BT/γ' was also determined by least-squares fitting with a combined scale and offset parameter to allow a measure of rigid-body translation components.

RESULTS AND DISCUSSION

The comparison between theory and experiment was performed for several models, including libration only, GNM on all C_α atoms, GNM omitting from the correlation coefficient those atoms making crystal contacts, GNM including the central molecule and only neighboring atoms, and GNM including complete neighboring molecules in the lattice.

Because each eigenvector is weighted as the reciprocal of its respective eigenvalue, it is possible that only the small-eigenvalue terms contribute significantly to the total sum. Therefore, the sum in Eq. 2 was evaluated using the eigenvectors corresponding to the smallest 30% as well as using 100% of the eigenvalues. As described previously, a subset of the eigenvalues are most important in the computation of the inverse of the Kirchhoff matrix (Haliloglu et al., 1997), but for best results all nonzero eigenvalues/singular values should be included when computationally feasible.

The average correlation coefficient for the agreement between experimental B-factors and those calculated by the simple GNM procedure using $r_c = 7.3$ Å was 0.594 with all and 0.581 with 30% of the eigenvectors (Table 1). The best value for the cutoff for assigning C_α values to be connected, r_c, was also determined to be 7.3 Å by evaluating the GNM model with various values (Table 1). The individual protein correlations ranged from 0.000 (1DDT) to 0.831 (1FRD) (see Table 3). The average coefficient for the libration method was 0.515 (Table 2), with values ranging from -0.423 (1OSA) to 0.886 (1ARU) (Table 3). However, the GNM gave higher correlation coefficients than the libration method for 70 (62%) of the 113 structures.

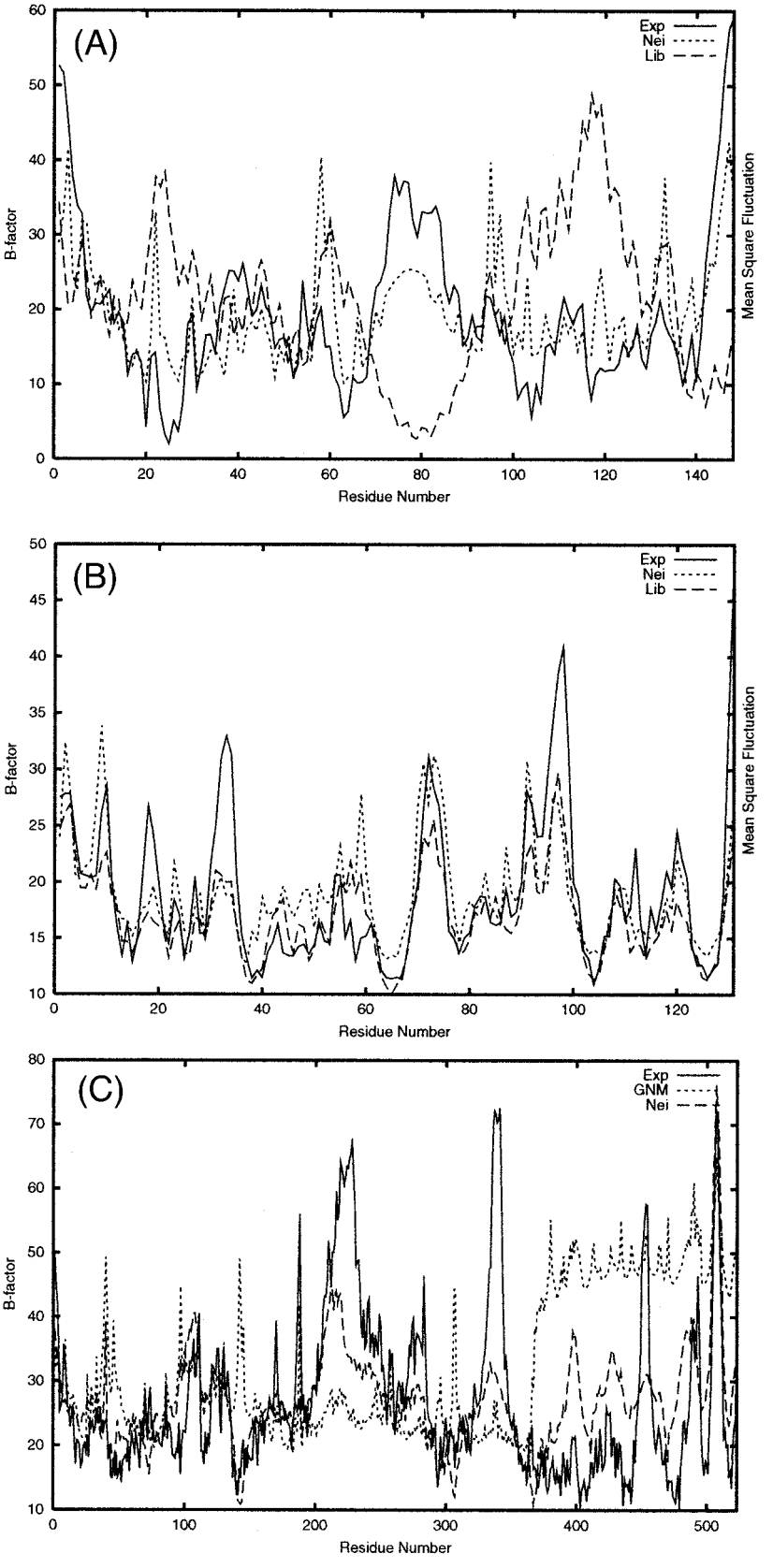

FIGURE 2 Plot of B-factor or mean square fluctuation with residue number. Exp is from the experimental data, Nei is that from the GNM model with neighbors, and Lib is from the libration model. (*A*) Calmodulin (1OSA). The libration model severely underestimates the mobility of the central helix and overestimates the mobilities of the end domains. The GNM model with neighbors does a much better job in this highly asymmetric molecule. (*B*) Lithostathine (1LIT). For this rather spherical protein, the GNM and librational models both predict the experimental B-factors reasonably well. (*C*) Diptheria toxin (1DDT). The last 200 amino acids of this protein form a distinct loosely connected domain that is predicted to be highly mobile in absence of crystal contacts (GNM) but pinned down in the crystal lattice (Nei).

TABLE 1 Average correlation coefficient with different spring length and using 30% and 100% of the eigenvalues in the GNM model

Maximum spring length, r_c	GNM	
	30%	100%
6.0 Å	0.520	0.525
6.5 Å	0.548	0.557
7.0 Å	0.572	0.582
7.1 Å	0.576	0.585
7.2 Å	0.579	0.590
7.3 Å	0.581	0.594
7.4 Å	0.579	0.592
7.5 Å	0.577	0.591
8.0 Å	0.565	0.583

There exists the question of whether the GNM or libration model might be more accurate for certain types of structures. There are 21 cases (for 7.3 Å) where the GNM and libration correlation coefficients differ by at least 0.2. In 18 of these, the GNM coefficient is better. Many of these structures are irregularly shaped (e.g., concave or dumbbell-shaped).

Indeed, one might expect that the libration method would work best on highly spherical structures. The libration theory implicitly assumes that atoms far away from the centroid are closer to the surface. It makes sense that this model would be most applicable when atoms that are the same distance from the centroid are also roughly the same distance from the surface.

In contrast, the theory underlying the GNM rests on local packing density as the determinant of thermal fluctuation. If a structure is not well packed with respect to its centroid, if it is concave, for example, there will be atoms quite near the centroid that are also near the surface. The GNM presumes, however, that tightly packed C_α atoms will fluctuate less than loosely packed C_α atoms. For irregularly shaped structures, the local C_α packing density becomes much more important in defining the mobility of the atoms, hence the overall superiority of the GNM models.

Fig. 2, *A* and *B*, shows the structures and experimental and calculated B-factors for calmodulin (1OSA) and lithostathine (1LIT). The GNM method gave much better results than the libration for calmodulin, which is shaped like a dumbbell. For litostathine, which is much more regular, though slightly oblong, the libration method gave a higher correlation coefficient, although both methods gave qualitatively similar results. A particularly interesting case is diphtheria toxin (1DDT) where an entire domain is tethered to the rest of the protein by only a single strand of polypeptide. Of course the simple GNM method, which does not include neighbors, severely overestimates the mobility of this domain as seen in the crystallographic B-factors (Fig. 2 *C*). This high degree of mobility is, however, likely for the protein in solution. (When contacts are included in the calculation, the domain is immobilized, and theory and experiment agree. (see below.) It should be noted, however, that although most of the structures where GNM is superior to libration are nonspherical, some of the structures where the two methods perform equally well are also shaped irregularly.

TABLE 2 Average correlation coefficient for different models with two different spring lengths

Model	7.0 Å	7.3 Å
Libration	0.515	
GNM	0.582	0.594
GNM omit	0.651	0.662
GNM contact	0.628	0.640
GNM neighbor	0.651	0.661

The standard spring length used by earlier workers is 7.0 Å, and 7.3 is the optimized spring length around 7.0 Å.

Additionally, the omission of cofactors from many of the structures may affect their behavior in the computations. A protein that has a cofactor in the active site may experience higher stability in that region than the models account for when the cofactor is omitted. Of the 113 structures in the list, 6 of them contained a heme group. For 5 of these, correlation coefficients were computed with and without the heme, and in all 5 cases, the values improved upon the inclusion of a subset of atoms from the heme in the structure. This suggests that the modeling of other structures can be improved with a systematic and reliable method of treating the cofactors.

Another issue to be noted is that the atomic coordinates used in these models comes from crystallographic structures. In some of these structures (mainly atoms near the surface), there is an interaction not only among atoms belonging to the same protein chain but also between proteins that are adjacent to one another inside the crystal. If one omits from the correlation coefficient calculation the C_α atoms within 7.0 Å of neighbors, the agreement between theory and experiment improves (Table 2).

By including these crystal packing effects the results are dramatically improved. Adding only neighboring atoms (GNM contact model) is not as effective as adding entire neighboring molecules (GNM neighbor model). By including neighboring molecules but taking only the central molecule for comparison the average correlation coefficient was improved from 0.594 in the simple GNM method to 0.661 in GNM with all neighboring molecules. Again a maximum spring length of 7.3 Å is better than 7.0 Å (Table 2). Attempts to use the anisotropic version of the GNM model (Atilgan et al., 2001; Doruker et al., 2000) failed to improve the results.

The correlation coefficient analysis measures the relative agreement between B-factors and GNM dynamics in terms of its positions of peaks and valleys in the functions, but does not include any measure of the overall scale of the motions. The factor k_BT/γ is essentially a force constant for

TABLE 3 Details of calculations for each protein

	Symbol	Protein	Cofactor	Molecular weight	R(Å)	% Sol	Libration	GNM	GNM omit	GNM contact	GNM neighbor	$\frac{k_BT}{\gamma}$	$\frac{k_BT}{\gamma'}$ Scale	$\frac{k_BT}{\gamma'}$ Offset
1	1aac	Amicyanin	Ca ion	11490	2.10	32.00	0.530	0.663	0.798	0.699	0.763	0.562	0.740	−0.995
2	1ads	Aldose reductase	NADP	35724	1.60	43.09	0.536	0.718	0.753	0.726	0.769	0.560	0.980	−3.257
3	1aky	Adenylate kinase	AP5 IMD	24036	1.63	46.00	0.588	0.706	0.752	0.745	0.767	1.390	2.715	−10.652
4	1amm	B-Drystallin		20966	1.20	35.0	0.151	0.720	0.604	0.720	0.584	0.291	0.448	−1.037
5	1arb	Achromobacter protease		27737	1.20	40.95	0.786	0.753	0.756	0.757	0.723	0.711	0.599	0.667
6	1aru	Peroxidase (+heme)	NAG Ca CN	35701	1.60	46.22	0.886	0.805	0.774	0.791	0.625	0.746	0.862	−0.788
7	1bkf	Fk506 binding protein	FK5	11754	1.60	39.40	0.271	0.428	0.523	0.591	0.662	1.110	1.549	−2.484
8	1bpi	BPTI (xtal form II)	PO4	6517	1.10	38.00	0.491	0.605	0.763	0.644	0.686	0.286	0.284	0.014
9	1cem	Cellulase cela		40308	1.65	40.51	0.533	0.666	0.681	0.673	0.678	0.635	0.721	−0.562
10	1cnr	Crambin		4736	1.05	29.36	0.430	0.633	0.829	0.693	0.747	0.248	0.359	−0.500
11	1cnv	Concanavalin B		33835	1.65	57.00	0.539	0.625	0.659	0.633	0.644	0.691	0.910	−1.397
12	1ctj	Cytochrome C6 (+heme)		9352	1.10	45.66	0.535	0.415	0.931	0.450	0.340	4.645	2.094	3.971
13	1cus	Cutinase		20723	1.25	41.65	0.769	0.766	0.813	0.791	0.807	1.021	0.873	0.863
14	1dad	Dethiobiotin synthase	ADP	24009	1.60	34.00	0.285	0.472	0.556	0.511	0.571	1.339	0.861	2.980
15	1ezm	Elastase	Ca Zn	33144	1.50	42.08	0.477	0.604	0.701	0.676	0.723	0.858	0.806	0.314
16	1fnc	Ferredoxin NADP + oxygen	A2P FDA SO	35335	1.70	49.00	0.474	0.574	0.615	0.600	0.652	0.884	1.355	−3.433
17	1fus	Ribonuclease F1	PCA	10874	1.30	35.49	0.598	0.593	0.607	0.641	0.614	0.714	0.633	0.434
18	1fxd	Ferredoxin II	Cs Fe-S cl	6262	1.70	34.97	0.557	0.533	0.700	0.576	0.594	0.791	0.595	1.164
19	1hfc	Fibroblast collagenase	HAP Ca Zn	18846	1.56	47.49	0.623	0.533	0.583	0.535	0.647	0.412	0.547	−0.976
20	1ifc	Intestinal f.a.b.p.		15125	1.19	35.54	0.342	0.602	0.672	0.724	0.688	1.445	2.066	−1.678
21	1igd	Protein G		6650	1.10	44.87	0.587	0.442	0.633	0.525	0.691	0.388	0.373	0.113
22	1iro	Rubredoxin	Fe(III)	6047	1.10	42.02	0.634	0.671	0.764	0.815	0.678	0.792	0.739	0.307
23	1jbc	Concanavalin A	Ca Mn	25599	1.20	46.96	0.680	0.677	0.731	0.676	0.686	0.639	0.633	0.034
24	1knb	Adenovirus type 5		21240	1.70	51.39	0.312	0.712	0.771	0.771	0.830	1.249	2.192	−6.430
25	1lam	Leucine aminopeptidase	MPD Zn CO3	52609	1.60	57.25	0.459	0.625	0.685	0.674	0.730	0.621	0.534	0.550
26	1lit	Lithostathine		16275	1.55	42.68	0.830	0.624	0.662	0.691	0.669	1.019	0.807	1.414
27	1mla	Malonyl-coenzyme A		32419	1.50	50.00	0.571	0.544	0.616	0.618	0.660	1.056	1.090	−0.203
28	1mrj	a-Trichosanthin	ADN	27144	1.60	42.57	0.273	0.490	0.512	0.512	0.451	0.734	1.168	−2.699
29	1nfp	*luxf* gene product	FMN MYR SO	26283	1.60	51.00	0.365	0.485	0.553	0.548	0.635	0.709	0.838	−1.021
30	1nif	Nitrate reductase	Cu	37018	1.60	42.65	0.683	0.604	0.815	0.826	0.818	1.016	0.832	1.218
31	1osa	Calmodulin	Ca	16672	1.68	48.86	−0.423	0.414	0.178	0.437	0.655	0.982	1.434	−3.256
32	1phb	Cytochrome P450(CAM)-heme	PFZ	46540	1.60	44.25	0.449	0.523	0.528	0.552	0.637	0.910	0.725	1.414
33	1php	3-Phosphoglycerate k	ADP Mg	42732	1.65	47.92	0.199	0.621	0.616	0.629	0.637	0.761	0.045	5.474
34	1plc	Plastocyanin	Cu	10486	1.33	35.42	0.495	0.477	0.684	0.563	0.501	0.704	0.578	0.607
35	1poa	Phospholipase A2	Ca	13144	1.50	32.64	0.381	0.678	0.705	0.677	0.616	0.740	1.371	−3.582
36	1ptf	His.-cont. phosphocarrier		9321	1.60	36.96	0.425	0.585	0.632	0.637	0.580	0.647	0.545	0.553
37	1ptx	Scorpion toxin II		7252	1.30	36.21	0.642	0.545	0.653	0.647	0.657	0.554	0.255	1.828
38	1ra9	Oxidoreductase	NADP	18001	1.55	46.77	0.419	0.602	0.660	0.655	0.650	0.773	0.895	−0.929
39	1rcf	Flavodoxin	FMN	18833	1.40	48.95	0.666	0.623	0.701	0.660	0.692	0.515	0.434	0.554
40	1rie	Cytochrome BC1-complex		14419	1.50	44.23	0.703	0.743	0.767	0.759	0.703	0.392	0.818	−2.877
41	1rro	Rat oncomodulin		12057	1.30	30.34	0.155	0.327	0.376	0.413	0.355	0.552	0.427	0.697
42	1smd	Human salivary amylase		55784	1.60	49.91	0.609	0.631	0.673	0.660	0.683	1.145	1.116	0.215

(*continued*)

TABLE 3 (*continued*)

	Symbol	Protein	Cofactor	Molecular weight	R(Å)	% Sol	Libration	GNM	GNM omit	GNM contact	GNM neighbor	$\frac{k_BT}{\gamma}$	$\frac{k_BT}{\gamma'}$ Scale	$\frac{k_BT}{\gamma'}$ Offset
43	1snc	staph. nuclease	Ca PTP	16812	1.65	43.46	0.01	0.687	0.770	0.676	0.685	1.121	1.385	−1.835
44	1whi	Ribosomal protein	L14	13346	1.50	37.93	0.497	0.302	0.593	0.598	0.570	0.799	0.617	1.017
45	1xic	D-xylose isomerase	Mn D-xylos	43201	1.60	55.87	0.461	0.389	0.448	0.550	0.766	0.444	0.825	−2.750
46	2ayh	. . . Glucano hydrolase	Ca	23916	1.60	40.28	0.682	0.754	0.725	0.740	0.658	0.580	0.771	−1.211
47	2cba	Carbonic anhydrase II	Zn	29116	1.54	43.06	0.854	0.741	0.762	0.750	0.732	0.639	0.727	−0.562
48	2cpl	Cyclophilin		18013	1.63	56.07	0.605	0.556	0.597	0.585	0.634	1.274	0.743	3.608
49	2ctc	Carboxypeptidase . . .	LOF Zn	34485	1.40	41.80	0.657	0.653	0.690	0.681	0.704	0.734	1.285	−3.735
50	2end	Endonuclease V		16079	1.45	36.49	0.490	0.722	0.747	0.739	0.687	0.601	0.516	0.685
51	2erl	Mating pheromone Er-1	EOH	4417	1.00	19.40	0.747	0.731	0.808	0.755	0.728	0.922	1.220	−1.443
52	2hft	Human tissue factor	SO4	24673	1.69	48.51	0.801	0.783	0.820	0.826	0.723	1.172	1.071	0.778
53	2ihl	Lysozyme	Na	14366	1.40	47.40	0.605	0.727	0.777	0.763	0.722	0.578	0.638	−0.416
54	2mcm	Macromomycin	Ca MPD	10751	1.50	45.39	0.597	0.820	0.858	0.839	0.800	0.873	0.880	−0.038
55	2mhr	Myohemerythrin	AZI EEO SO	13778	1.70	46.33	0.480	0.496	0.507	0.545	0.594	1.000	0.534	3.310
56	2phy	Photoactive yellow pigment	HC4	13874	1.40	35.25	0.608	0.515	0.557	0.619	0.539	0.655	0.898	−1.354
57	2rhe	Bence-Jones protein		11834	1.60	52.11	0.379	0.363	0.747	0.428	0.446	0.653	0.813	−0.900
58	2rn2	Ribonuclease H		17597	1.48	36.21	0.690	0.744	0.776	0.737	0.694	0.939	0.755	1.337
59	3b5c	Cytochrome B5 (+heme)		10635	1.50	41.12	0.471	0.458	0.430	0.390	0.520	0.688	0.438	1.688
60	3chy	Che Y	SO4	13966	1.66	41.03	0.617	0.753	0.798	0.781	0.826	0.722	0.911	−1.190
61	3ebx	Erabutoxin b	SO4	6869	1.40	32.71	0.330	0.578	0.844	0.739	0.704	0.865	0.720	0.739
62	3grs	Glutathione reductase	FAD PO4	51572	1.54	53.58	0.566	0.533	0.691	0.684	0.703	0.818	0.910	−0.671
63	3lzm	Lysozyme		18636	1.70	56.20	0.410	0.596	0.703	0.346	0.349	1.096	0.465	3.693
64	3pte	. . . Carboxypeptidase . . .		37393	1.60	47.74	0.532	0.818	0.853	0.833	0.845	0.493	0.651	−1.021
65	4fgf	Basic fibroblast growth factor	SEO SO4	16408	1.60	33.10	0.375	0.270	0.312	0.268	0.286	0.948	0.747	1.343
66	4ptp	b-Trypsin	Ca MIS	23306	1.34	47.12	0.590	0.353	0.378	0.416	0.358	0.767	0.336	2.606
67	5p21	c-H-Ras p21 protein	GNP Mg	18854	1.35	39.01	0.448	0.499	0.610	0.564	0.638	0.967	1.218	−1.578
68	7rsa	Ribonuclease A	TBU DOD	13690	1.26	43.34	0.640	0.637	0.607	0.670	0.641	0.670	0.512	0.949
69	8abp	1-Arabinose b.p.	GLA GLB	33193	1.49	47.73	0.401	0.822	0.852	0.842	0.861	0.792	1.069	−2.206
70	1ahc	a-Momorcharin		27369	2.00	49.52	0.461	0.690	0.710	0.708	0.722	1.111	1.483	−2.674
71	1amp	Aminopeptidase	Zn	31408	1.80	52.83	0.664	0.560	0.553	0.557	0.563	0.725	0.712	0.085
72	1ars	Asp aminotransferase	PLP	43575	1.80	60.12	0.511	0.411	0.735	0.590	0.763	1.539	1.773	−1.651
73	1cdg	Cyclodextrine glycosyltransferase	CA MAL	74518	2.00	58.67	0.583	0.620	0.660	0.651	0.703	1.128	1.134	−0.047
74	1cpn	Glucan-4-glucanohydrolase	CA	23345	1.80	42.69	0.490	0.496	0.605	0.572	0.625	1.347	1.435	−0.524
75	1csh	Citrate synthase	AMX OAA	48124	1.60	49.67	0.620	0.450	0.461	0.654	0.709	0.727	1.121	−2.834
76	1ddt	Diphtheria toxin	APU	58343	2.00	55.09	0.468	0.000	0.055	0.277	0.602	1.062	1.437	−3.362
77	1ede	Haloalkane dehalogenase		35145	1.90	39.31	0.709	0.626	0.671	0.671	0.674	0.664	0.969	−2.065
78	1frd	Heterocyst ferredoxin	Fe2S2	10818	1.70	41.45	0.501	0.831	0.904	0.869	0.880	1.039	1.274	−1.492
79	1gia	Gi alpha 1	GSP MG	40216	2.00	50.47	0.668	0.662	0.669	0.656	0.612	0.831	0.907	−0.573
80	1gky	Guanylate kinase	5GP SO4	20507	2.00	49.65	0.355	0.549	0.528	0.594	0.620	0.691	0.688	0.028
81	1gof	Galactose oxidase	ACY CU(II)	68523	1.70	49.83	0.662	0.761	0.798	0.597	0.540	1.438	1.501	−0.351
82	1gpr	Glucose permease		17381	1.90	37.36	0.576	0.599	0.749	0.668	0.796	0.828	1.410	−3.382
83	1iab	Astactin	Co(II)	22603	1.79	49.11	0.367	0.388	0.275	0.380	0.369	0.461	0.621	−1.196
84	1iag	Adamalysin II	CA SO4 Zn	23182	2.00	62.17	0.370	0.518	0.590	0.564	0.662	0.833	1.173	−2.421

(*continued*)

TABLE 3 (*continued*)

	Symbol	Protein	Cofactor	Molecular weight	R(Å)	% Sol	Libration	GNM	GNM omit	GNM contact	GNM neighbor	$\frac{k_BT}{\gamma}$	$\frac{k_BT}{\gamma'}$ Scale	$\frac{k_BT}{\gamma'}$ Offset
85	1lct	Lactoferrin	CO3 Fe(III)	37029	2.00	55.64	0.350	0.521	0.589	0.511	0.537	1.156	0.731	3.456
86	1lis	Lysin		16268	1.90	59.07	0.483	0.442	0.748	0.562	0.771	0.831	0.919	−0.674
87	1lst	Lao binding protein		26154	1.80	50.74	0.666	0.759	0.744	0.769	0.715	1.000	1.032	−0.232
88	1mjc	Major cold shock protein		7272	2.00	38.48	0.659	0.662	0.799	0.695	0.697	1.863	2.699	−4.613
89	1nar	Norbonin		33100	1.80	47.57	0.674	0.757	0.777	0.768	0.774	0.686	0.783	−0.701
90	1npk	Nucleoside diph. kinase		16664	1.80	50.94	0.646	0.556	0.783	0.761	0.770	1.061	0.910	1.130
91	1omp	D-maltodextrin b.p.		40709	1.80	45.42	0.532	0.653	0.673	0.664	0.708	0.989	1.182	−1.442
92	1onc	p-30 protein	PCA SO4	11717	1.70	37.19	0.327	0.659	0.714	0.677	0.649	0.615	0.615	−0.004
93	1oyc	Old yellow enzyme	FMN	45017	2.00	49.55	0.720	0.718	0.723	0.732	0.712	1.068	1.180	−0.764
94	1pbe	p-hydroxybenzoate hydrase	FAD PHB	44324	1.90	52.56	0.493	0.609	0.645	0.630	0.679	1.015	0.862	1.096
95	1pda	Porphobilinogen deaminase	ACY DPM	33853	1.76	87.65	0.397	0.742	0.755	0.774	0.804	1.059	1.254	−1.434
96	1pii	*N*- . . . anthranilate isomerase	PO4	49363	2.00	67.42	0.288	0.453	0.493	0.503	0.550	0.762	1.586	−6.525
97	1poc	Phospholipase A2	Ca GEL	15250	2.00	71.72	0.416	0.603	0.659	0.618	0.684	0.901	1.126	−2.105
98	1ppn	Papain cys-25	MOH	23429	1.60	44.47	0.623	0.654	0.713	0.678	0.695	0.613	0.832	−1.437
99	1rec	Recoverin	Ca	23203	1.90	47.45	0.355	0.523	0.599	0.587	0.653	1.375	1.687	−2.205
100	1ris	Ribosomal protein S6		11973	2.00	49.12	0.420	0.266	0.688	0.410	0.724	0.963	0.903	0.423
101	1sbp	Sulfate binding protein	SO4	34486	1.70	40.23	0.577	0.762	0.790	0.783	0.789	0.688	0.839	−1.020
102	1thg	Lipase triacylglycerol hydrase	NAG PCA	59566	1.80	46.75	0.411	0.510	0.539	0.531	0.527	0.867	0.864	0.021
103	1tml	Endo-1,4-B-D glucanase	SO4	30412	1.80	36.72	0.515	0.666	0.614	0.665	0.789	0.608	0.161	2.890
104	1ubi	Ubiquitin		8565	1.80	33.05	0.726	0.676	0.516	0.694	0.396	0.623	0.965	−2.138
105	2cmd	Malate dehydrogenase	CIT	32423	1.87	50.46	0.475	0.600	0.659	0.643	0.721	0.731	1.001	−1.862
106	2cy3	Cytochrome C3 (+heme)		12622	1.70	56.34	0.703	0.759	0.755	0.760	0.733	0.524	0.825	−2.246
107	2mnr	Mandelate racemase	Mn SO4	38348	1.90	54.52	0.408	0.493	0.569	0.544	0.666	0.689	1.230	−3.357
108	2ran	Annexin V	Ca SO4	35385	1.90	82.70	−0.084	0.431	0.516	0.510	0.766	1.196	1.560	−3.106
109	2sil	Sialidase		41944	1.60	42.70	0.567	0.585	0.525	0.588	0.552	0.724	0.677	0.274
110	2tgi	Transforming g.f.-btwo		12720	1.80	60.80	0.753	0.645	0.697	0.692	0.716	1.348	2.851	−9.990
111	3cox	Cholesterol oxide	FAD	54839	1.80	46.55	0.603	0.699	0.710	0.707	0.718	0.526	0.705	−1.170
112	4gcr	Gamma-B crystallin		20966	1.47	36.36	0.250	0.797	0.822	0.804	0.801	0.007	0.011	−0.030
113	4mt2	Metallothionein isoformII	Cd Na Zn	6145	2.00	47.40	0.349	0.319	0.778	0.587	0.623	1.199	0.833	2.666

the virtual springs connecting C_α atoms and sets the overall scale factor. We determined optimal values for k_BT/γ and also define k_BT/γ' as the scaling factor including a constant additive offset for each PDB entry (Table 3). The constant was added because of previous evidence that both static lattice and dynamic sources of displacements exist in crystals (Kuriyan et al., 1986).

The mean and standard deviation of k_BT/γ, 0.87 ± 0.46 Å^2, suggest that there is some relative variability in the basic spring constant of the proteins. The temperature dependence in the theory suggests that those crystal structures determined at lower temperatures might have smaller k_BT/γ values, although it is well known that crystal structures solved from rapidly quenched samples retain much of their dynamic disorder as static disorder. In fact, the mean k_BT/γ for the five crystal structures determined at ~100–150 K is 0.62, which is smaller than the overall average.

The mean k_BT/γ' and offset are 0.96 ± 0.50 Å^2 and −0.71 ± 2.39 Å^2, respectively. The offset, which can absorb several types of crystallographic artifacts such as lattice disorder and other $\sin(\theta/\lambda)$-dependent data processing errors, has a mean value very close to zero, implying that there is no large systematic contribution of lattice disorder to crystallographic B-factor. The standard deviation of 2 Å^2 suggests, however, that each crystal structure may have circumstances that lead to the need for such an offset. It is also likely that the simplifying nature of the model itself introduces some errors that are accommodated by this variable.

CONCLUSION

It appears that the GNM model is better suited for estimating protein motions than the libration model, especially for highly irregular or nonspherical structures. Furthermore, it is able to compute cross-correlations between different atoms. These cross-correlations are the determinants of directed motions. Having a theoretical method to test these cross-correlations against experimental data is extremely valuable. With further research to determine a good method for including cofactors in the protein structures, the method could be even more useful.

Biological implication

The function of a protein depends on both its structure and dynamics. Crystallographic analysis routinely provides estimates of the amplitudes of motions of atoms, but the effect of the surrounding lattice on the motions is always an uncertainty. Here we show that a simplified molecular mechanics model can effectively describe protein motions, including the effects of crystal contacts. Because the added neighbors improve our results, our confidence in the method is increased. The results further suggest that GNM calculations on a single protein molecule may give a different and more accurate picture than crystallographic temperature factors give on the dynamics of an isolated protein molecule because the constraining effects of the lattice on the crystallographic result can be factored out.

Funding for this work was provided by the Arnold and Mabel Beckman Foundation (J.M.), The Robert A. Welch Foundation (C-1142), the Wisconsin Alumni Research Foundation, the National Science Foundation (ACI-0082645), and National Institutes of Health grants AR40252 and GM64598.

REFERENCES

Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.

Bahar, I., A. R. Atilgan, M. C. Demirel, and B. Erman. 1998. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* 80:2733–2736.

Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Design.* 2:173–181.

ben-Avraham, D., and M. M. Tirion. 1998. Normal modes analyses of macromolecules. *Physica A.* 249:415–423.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Myer, Jr., M. D. Brice, J. R. Rodgers, Jr., O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structure. *J. Mol. Biol.* 112:535–542.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. Delano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta. Crystallogr.* D54:905–921.

Doruker, P., A. R. Atilgan, and I. Bahar. 2000. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to α-amylase inhibitor. *Proteins.* 40:512–524.

Eichinger, B. E. 1972. Elasticity theory. I. Distribution functions for perfect phantom networks. *Macromolecules.* 5:496–505.

Haliloglu, T., and I. Bahar. 1999. Structure-based analysis of protein dynamics: comparison of theoretical results for hen lysozyme with x-ray diffraction and NMR Relaxation data. *Proteins Struct. Funct. Genet.* 37:654–667.

Haliloglu, T., I. Bahar, and B. Erman. 1997. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090–3093.

Harata, K., Y. Abe, and M. Muraki. 1999. Crystallographic evaluation of internal motion of human α-lactalbumin refined by full-matrix least-squares method. *J. Mol. Biol.* 287:347–358.

Higo, J., and H. Umeyama. 1997. Protein dynamics determined by backbone conformation and atom packing. *Protein Eng.* 10:373–380.

Hinsen, K., and G. R. Kneller. 1999. A simplified force field for describing vibrational protein dynamics over the whole frequency range. *J. Chem. Phys.* 111:10766–10769.

Kloczkowski, A., and J. E. Mark. 1989. Chain dimensions and fluctuations in random elastomeric networks. I. Phantom Gaussian networks in the undeformed state. *Macromolecules.* 22:1423–1432.

Kuriyan, J., G. A. Petsko, R. M. Levy, and M. Karplus. 1986. Effect of anisotropy and anharmonicity on protein crystallographic refinement: an evolution by molecular dynamics. *J. Mol. Biol.* 190:227–254.

Kuriyan, J., and W. I. Weis. 1991. Rigid protein motion as a model for crystallographic temperature factors. *Proc. Nat. Acad. Sci. U.S.A.* 88: 2273–2277.

Levitt, M., C. Sander, and P. S. Stern. 1985. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* 181:423–447.

MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem.* 102:3586–3616.

Schomaker, V., and K. N. Trueblood. 1968. On the rigid-body motions of molecules in crystals. *Acta Crystallogr.* B24:63–76.

Singh, A. P., and D. L. Brutlag. 1997. Hierarchical Protein structure alignment using both secondary structure and atomic representations. *ISMB Proc.* 4:284–293.

Sternberg, M. J. E., D. E. P. Grace, and D. C. Phillips. 1979. Dynamic information from protein crystallography: an analysis of temperature factors from refinement of the hen egg white lysozyme structure. *J. Mol. Biol.* 130:231–253.

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

Word, J. M., S. C. Lovell, T. H. LaBean, H. C. Taylor, M. E. Zalis, B. K. Presley, and J. S. Richardson. 1999. Visualizing and quantifying molecular goodness-of-fit: small-probe contacts dots with explicit hydrogen atoms. *J. Mol. Biol.* 285:1711–1733.