Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 23 (2004) 167–174
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Comparative estimation of vibrational entropy changes in proteins through normal modes analysis

Benjamin J. Carrington [a], Ricardo L. Mancera [b,*]

[a] *Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1PD, UK*
[b] *De Novo Pharmaceuticals, Compass House, Vision Park, Chivers Way, Histon, Cambridge CB4 9ZR, UK*

Received 27 November 2003; received in revised form 8 April 2004; accepted 7 May 2004

Available online 15 June 2004

Abstract

We compare the vibrational entropy changes of proteins calculated using a full and a number of approximate normal modes analysis methods. The vibrational entropy differences for three conformational changes and three protein binding interactions were computed. In general, the approximate methods yield good estimates of the vibrational entropy change in a fraction of the time required by full normal modes analysis. The absolute entropies are either overestimated or greatly underestimated, but the difference is sufficiently accurate for some methods. This indicates that some of the approximate methods can give reasonable estimates of the associated vibrational entropy changes, making them suitable for inclusion in free energy calculations.

Keywords: Normal modes analysis; Vibrational entropy; Energy minimisation; Protein conformational changes

1. Introduction

The internal motion of a protein can be seen as the superposition of all its vibrations. At a given temperature, the mean square amplitude of motion in a given frequency is inversely proportional to the effective associated force constant. As a consequence, the motion tends to be dominated by low frequency vibrations. In the case of proteins, the lowest frequency vibrations occur at frequencies below 200 cm^{-1}. These low frequency vibrations have been shown to represent collective motions with the largest amplitude, implicated in the rearrangement of protein domains and large-scale conformational transitions in allosteric proteins [1–10]. In fact, low frequency vibrations dominate the overall motion of a protein's backbone, giving rise to characteristic correlated segmental rigid-body motions of several residues that can be described in terms of bending, twisting and swinging motions [7].

Normal modes analysis is a multi-dimensional treatment of coupled harmonic oscillators used to evaluate vibrational motion. It is usually applied to molecules based on a number of assumptions: atoms can be treated as point masses, the molecular conformation is at a local energy minimum, the potential energy is a quadratic function of the internal coordinates, and interactions with the solvent can be neglected [1]. In the case of proteins and macromolecules in general, only the first assumption can be taken as reasonable. The conformational potential energy surface of a protein has a very complex shape with multiple minima, protein dynamics is highly anharmonic, and low frequency collective motions are often highly overdamped and diffusive. Despite all the approximations involved, normal modes analysis is widely used because exact analytic solutions can be derived and semi-quantitative estimates can be obtained for many structural, dynamic and thermodynamic properties, such as the magnitude of atomic fluctuations [7], the displacement covariance matrix [7] and vibrational entropies [11,12].

The frequencies of the normal modes may be calculated from a molecular mechanics forcefield using the Hessian matrix of second energy derivatives in terms of the force constants for each type of interaction [13]. In proteins, there are many non-bond interactions that are non-quadratic functions of interatomic distances, which in turn are complicated non-quadratic functions of bond lengths, bond angles and torsion angles. Under such circumstances, the potential energy function will only approximate a quadratic function when the first derivatives are zero after an energy minimisation. In normal modes analysis this is achieved by expanding the potential energy surface as a Taylor series and

* Corresponding author. Fax: +44 1223 238088.
E-mail address: ricardo.mancera@denovopharma.com (R.L. Mancera).

1093-3263/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2004.05.003

truncating after the second term. The resulting matrix is then mass-weighted and diagonalised. The eigenvalues provide the vibrational frequencies and the eigenvectors give the details of the corresponding motions [1–10]. It is not difficult to realise that the normal modes analysis of large molecules such as proteins becomes a computationally demanding problem because of the large matrices involved. This can be tackled by reducing the number of degrees of freedom with models in which groups of atoms are joined into virtual extended atoms and simplified approximate potentials are used to describe the interactions between these pseudoatoms [14–24]. Special techniques have also been used to study selected normal modes rather than the full spectrum [8,25–28].

Tirion [14] introduced a single-parameter pairwise harmonic potential to describe all interatomic interactions in a protein within a specific cut-off, and where the reference configuration is the initial configuration. Consequently, no initial energy minimisation is required and only a universal force constant for all interacting atoms is needed. Normal modes analyses carried out with such an underlying simplified potential function were able to describe the low frequency motion of large globular proteins, including time scales and frequencies, as well as the displacements of atoms as predicted by the eigenvectors [14].

A Gaussian network model (GNM) using a harmonic potential consisting of only the Cα atoms has been used to predict the temperature factors obtained during crystallographic structure determination [15]. A very good agreement with experiment was obtained but the method required fitting one parameter for each specific protein. An extension, called the anisotropic network model (ANM), allowed for the calculation of the directions of collective motions [16] and was used with a more coarse-grained approach where only a fraction of the residues along a protein's backbone were retained [17]. Hinsen introduced a similar simplified pairwise harmonic potential where only one point mass per residue is used and is located at the Cα position, while the pair force constant is an exponentially decaying function of the square of the interatomic distance [18]. This approximation was used successfully to reproduce the low frequency modes of large proteins and to identify and describe domain motions [18,19]. The success of these methods suggests that a good description of the mass distribution of the molecular structure is all that is needed to model the collective motions of a protein.

Sanejouand and coworkers [20–22] introduced the rotations-translations of blocks (RTB) method, where a protein is divided into rigid-body blocks, each block being made of one or more consecutive residues. The lowest frequency normal modes of a protein are then obtained as a linear combination of the rotations and translations of these blocks. A linear relationship was observed between the "exact" and the approximate frequencies obtained for blocks of up to six residues [21], An approximate pairwise harmonic potential considering only the Cα atoms was also introduced and the resulting elastic network model (ENM) gave good agreement with the standard RTB approach and predicted accurately the low frequency normal modes, whose direction compared well with the conformational change of a number of proteins [22]. This method was also used to explain the open/closed transition of proteins by mapping a few of the lowest frequency normal modes determined directly from the structure of the open form [23]. A vector quantisation method has been recently introduced to produce a discrete reduced description of a continuous shape/mass distribution within the ENM framework, allowing the global distortions and motions of large biomolecular assemblies to be modelled [24].

The above-mentioned normal modes analysis methods have been designed with the purpose of making tractable the study of concerted motions in large biomolecules. However, no attempt has been made to elucidate their accuracy in the calculation of vibrational entropies and, in particular, vibrational entropy changes due to biomolecular association or conformational changes. The calculation of vibrational entropy changes upon binding or complexation is very important for the elucidation of the relative importance of translational and rotational entropy loss and vibrational entropy gain in such molecular events [12]. Moreover, approximate methods for the calculation of free energies of biomolecular association such as MM-PBSA [29] contain a contribution from the vibrational entropy changes involved. Although a more accurate evaluation of the vibrational entropy could be carried out by averaging over a representative ensemble of energy minima (obtained, for example, by performing a molecular dynamics simulation), most approximate free energy methods (including MM-PBSA) rely on an estimate of the vibrational entropy after a single energy minimisation (usually of the initial crystal or NMR structure).

We have investigated whether the approximations made in the normal modes analysis of vibrational motions produce also reasonable estimates of the relative vibrational entropies of large molecules such as proteins. This would make the use of such approximate methods a valid approach for the calculation of vibrational entropy changes in protein–protein interactions and large protein conformational changes. This paper reports our comparative findings for six model protein systems for an exact and several approximate normal modes analysis methods.

2. Materials and methods

Six different protein systems were considered for our comparative estimation of vibrational entropies. These consisted of three conformational changes and three protein–protein interactions. The structures involved are shown in Table 1.

The first of the conformational changes presented here is the trp repressor/aporepressor structural transition caused by the binding or loss of the L-tryptophan ligand. This structural

Table 1
Structures used in the vibrational entropy calculations

Protein system	Resolution (Å)	PDB entry	Reference
Trp repressor (trigonal form)	2.20	1WRP	[30]
Tip aporepressor	1.80	3WRP	[30]
CheY + Mg^{2+}	1.76	1CHN	[31]
CheY − Mg^{2+}	1.66	3CHY	[32]
Calmodulin + Ca^{2+}	1.70	1CLL	[33]
Calmodulin − Ca^{2+}	NMR	1CFD	[34]
D1.3–HEL	1.80	1VFB	[35]
D1.3	1.80	1VFA	[35]
HEL	1.65	1DPX	[36]
Trypsin–BPTI	1.90	2PTC	[37]
Trypsin	1.55	2PTN	[38]
BPTI	1.10	1BPI	[39]
xMdm2–p53	2.30	1YCQ	[40]
xMdm2	2.30	1YCQ	[40]
p53	2.30	1YCQ	[40]

shift occurs mainly in the DNA-binding domain. This system was small enough for a full normal modes calculation and contained only the L-tryptophan ligand in one case, which was removed together with any water molecules prior to the calculation.

The second conformational change considered is in the bacterial chemotaxis protein CheY on binding of a magnesium ion in the active site. Again, the lack of a large ligand made this system suitable for a full normal modes calculation on the protein only.

Thirdly, the two forms of calmodulin were considered when estimating the vibrational entropy contribution to the conformational change upon calcium binding. The calcium-free form was taken from a minimised average NMR structure and the calcium-bound form was taken from an X-ray crystal structure. The calcium ions were not included in the calculations described below, for consistency with a previously reported calculation of this system [41]. However, only residues Thr-5 to Ala-147 in the NMR structure were considered, as all other residues had not been fully resolved in the corresponding crystal structure and were thus removed from the NMR structure. This ensured that both structures had the same number of atoms and, consequently, the same number of normal modes.

The first two protein–protein complex examples were selected for comparison with a previously published vibrational analysis [42]. The Fv fragment of mouse monoclonal antibody D1.3 complex with its antigen hen egg lysozyme (HEL),and the β-trypsin complex with pancreatic trypsin inhibitor are two of the few examples where high-resolution structures of both the free and bound proteins exist. In this study, we have estimated the vibrational entropy change for the complex formed from the separately solved structures and for the protein–protein complex itself. In other words, one calculation computes the vibrational entropy contribution from each of the separately solved structures, and from the other calculation computes the contribution from each subunit as extracted from the structure of the complex. This was done in order to test whether there is any difference between using the structure of the complex only and taking the separately solved structures. It should be noted here that we were unable to carry out a full normal modes calculation for these two systems (see below). Therefore, our vibrational entropy estimates are compared solely with a previously reported full normal modes calculation [42] (see Section 3).

The third system considered the complex between the *X. laevis* oncoprotein Mdm2 (xMdm2) and a 15–29 residue stretch (the transactivation domain) of the tumor suppressor p53. This system allowed for the vibrational entropy contribution to this protein–peptide association to be estimated. For consistency with a previously reported calculation of this system [43], only the 12-residue stretch of p53 from Gln-16 to Pro-27 was considered (Ac-QETFSDLWKLLP-NH_2). Gln 16 was modelled in the complex using the Biopolymer module in Insight 2000 (Accelrys), since its structure was not resolved in the crystal structure due to the absence of electron density.

For the full normal modes analysis, all the protein structures were energy-minimised using the Discover3 module in Insight 2000 (Accelrys). In the case of the protein complexes, energy minimisations were also carried out for the separate subunits from both the individually solved structures and those extracted from the structure of the complex. Prior to all energy minimisations, all water molecules were removed and hydrogens were added to the proteins assuming a pH of 7.0 and standard aminoacid pK_a using the Biopolymer module in Insight 2000. The AMBER forcefield [44] was used for these energy minimisations. No cut-offs were used to treat non-bonded interactions, while electrostatic interactions were computed using a distant-dependent dielectric constant of 4.0 to mimic solvent screening.

The energy minimisations were carried out in a staged manner by initially reducing the van der Waals interactions to minimise the effect of strong repulsive steric clashes, and then gradually scaling them up. In our experience this procedure ensures that the lowest energy conformations are obtained while preserving the original structure as much as possible. At the beginning of the energy minimisations van der Waals interactions were thus scaled down to 10% of their full value. An initial minimisation of 100 steps using steepest descents [45] was performed. Van der Waals interactions were scaled up to 30% of their full value and a second minimisation of 100 steps using steepest descents was carried out. Van der Waals interactions were then scaled up to 50% of their full value and a third minimisation of 100 steps using steepest descents was then carried out. At this point, van der Waals interactions were again reduced to 10% of their full value and a fourth minimisation of 100 steps using the Fletcher–Reeves method for conjugate gradients [45] was performed. This was followed by a gradual increase in the van der Waals interactions to 30, 50, 75 and 100% of their full value in blocks of 100 steps of conjugate gradients,

as before. A sixth and final minimisation using conjugate gradients as before was then performed until convergence was achieved when the energy gradient was smaller than 0.00001 kcal/mol/Å.

A full normal modes analysis was then performed by diagonalising the full mass-weighted Hessian matrix obtained from the above minimisation. The vibrational entropies (TS_{vib}) at 298 K were then calculated using Eq. (1) [46–48]:

$$-TS_{vib} = \sum \left\{ \frac{kT \ln(1 - \exp[-h\upsilon_i/kT]) - h\upsilon_i}{\exp[h\upsilon_i/kT] - 1} \right\} \quad (1)$$

where υ_i are the vibrational frequencies, h the Planck's constant, T the absolute temperature and k Boltzmann's constant.

Vibrational entropy differences ($T\Delta S_{vib}$) in the conformational change cases were computed by subtracting the vibrational entropy of the ligand-free form from that of the ligand-bound form. Vibrational entropy changes upon protein–protein complexation were computed by subtracting the vibrational entropies of each subunit from that of the complex.

A number of approximate normal modes analysis methods were also used. For consistency the energy minimised crystal structures (or the NMR average structure for calcium-free calmodulin) of the above-mentioned proteins were taken as input for these methods, even though an energy minimisation is not strictly required [14,18,22]. The following approximate normal modes analysis methods were used:

(a) M1: Tirion's all-atom harmonic potential method with equal masses for all atoms [14]. For these calculations, the programmes PDBMAT and DIAGSTD [22,23] were used.
(b) M2: A variation of Sanejouand's RTB method [21], where all atoms are considered but a harmonic potential is used to represent all pairwise atomic interactions. For these calculations, the programmes PDBMAT and DIAGRTB [22,23] were used. Block sizes of one, two and three residues were considered.
(b) M3: Hinsen's harmonic potential forcefield for residues represented by point masses centred at the Cα atoms [18]. For these calculations, the Molecular Modelling Toolkit (MMTK) [49] was used.

In methods M1 and M2 above, default values for the potential cut-off (8 Å) and the force constant (10.0 kcal Å^2/mol) were used. It is important to mention that in all of the above calculations with approximate methods, all vibrations were computed and not only the low frequency ones. Checks were also carried out to ensure that in all cases the calculated entropies and entropy differences had indeed converged by inspecting plots of these entropies for the frequency ranges computed [12]. Fig. 1 shows the plots for the entropy difference using various methods in the case of calmodulin.

3. Results and discussion

Table 2 contains the results of all our calculations for each protein system considered. The first set of results (second column) refers to the full normal modes analysis calculations using the AMBER forcefield and a complete diagonalisation of the mass-weighted Hessian after a full energy minimisation. These results constitute our benchmark calculations against which the other approximate normal modes analysis methods (M1–M3) were compared to. In two of the protein–protein complex cases we were unable to carry out a fall normal modes calculation due to the large size of the system. The benchmark vibrational entropies are taken from the literature [42] where an iterative diagonalisation scheme within CHARMM [50] was used.

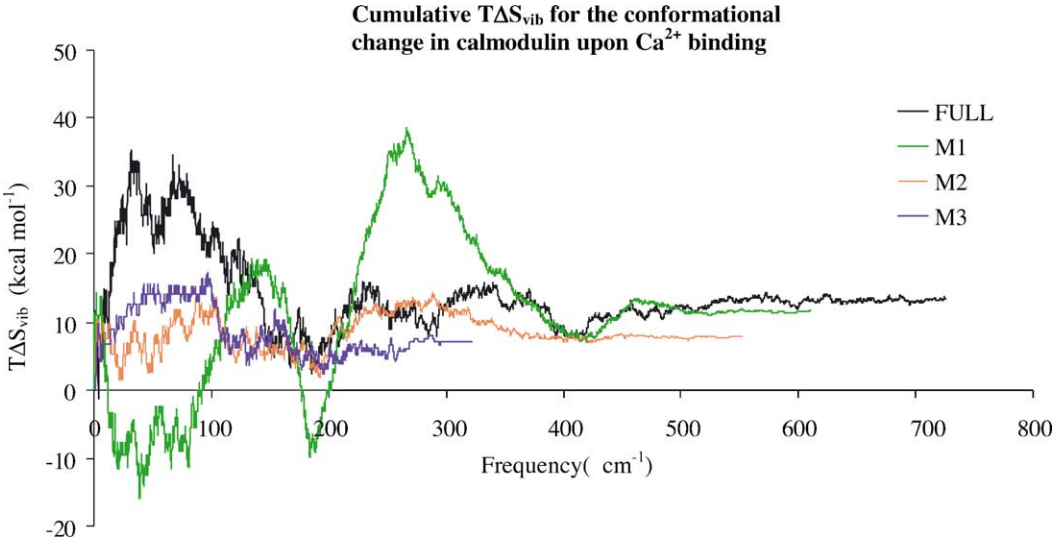

Fig. 1. Cumulative vibrational entropy difference for the conformational change in calmodulin upon Ca^{2+} binding. This is an example of how the entropy difference converges for both the exact and approximate normal modes methods considered. M2 is shown for bs = 3.

Table 2
Vibrational entropy estimates using different methods

Protein(s)	Full	M1	M2 (bs* = l)	M2 (bs = 2)	M2 (bs = 3)	M3
Trp repressor 5 (trigonal form)	1189.6	1475.3	604.6	361.0	270.0	247.9
Trp aporepressor	1177.7	1398.7	578.3	344.1	256.2	237.3
$T\Delta S_{vib}$	11.9	76.6	26.3	16.9	13.8	10.6
CheY + Mg^{2+}	1423.1	1597.5	689.3	407.9	301.3	286.4
CheY − Mg^{2+}	1426.2	1604.0	690.4	408.6	301.7	285.0
$T\Delta S_{vib}$	−3.1	−6.5	−1.1	−0.7	−0.4	1.4
Calmodulin + Ca^{2+}	1690.1	2075.5	861.8	520.9	389.2	360.3
Calmodulin − Ca^{2+}	1677.1	2063.8	853.6	513.4	381.2	353.2
$T\Delta S_{vib}$	13.0	11.7	8.2	7.5	8.0	7.1
D1.3–HEL	(2281.1)	3835.1	1818.6	1095.2	821.2	785.9
D1.3	(1569.2)	2651.3	1171.9	705.6	526.1	502.7
HEL	(718.9)	1560.6	684.3	409.3	298.9	288.7
$T\Delta S_{vib}$	(−7.0)	−376.8	−37.6	−19.7	−3.8	−5.5
D1.3–HEL	–	3835.1	1818.6	1095.2	821.2	785.9
D1.3 (from complex)	–	2614.9	1159.0	701.0	523.8	497.0
HEL (from complex)	-	1600.7	695.0	415.7	303.6	290.6
$T\Delta S_{vib}$	–	−380.5	−35.4	−21.5	−6.2	−1.7
Trypsin–BPTI	(1882.1)	3145.1	1424.5	865.2	643.1	600.4
Trypsin	(1476.0)	2539.3	1138.7	683.3	509.9	472.2
BPTI	(401.2)	797.2	328.0	192.1	144.9	132.8
$T\Delta S_{vib}$	(4.9)	−191.4	−42.2	−10.2	−11.7	−4.6
Trypsin–BPTI	–	3145.1	1424.5	865.2	643.1	600.4
Trypsin (from complex)	–	2505.0	1137.2	689.3	514.7	473.2
BPTI (from complex)	–	804.9	328.9	192.4	144.5	130.8
$T\Delta S_{vib}$	–	−164.8	−41.6	−16.5	−16.1	−3.6
XMdm2–p53	1201.5	1451.4	599.3	339.1	255.8	240.1
XMdm2	1047.7	1291.4	514.6	304.5	228.6	212.7
p53	158.2	283.5	88.0	44.9	28.8	31.2
$T\Delta S_{vib}$	−4.4	−123.5	−3.3	−10.3	−1.6	−3.8

All entropies are reported in kcal/mol at 298 K. Values in brackets taken from Noskov and Lim [42].
* bs: block size in number of residues.

We would like to point out here that the trypsin–BPTI complexation was estimated by all of the approximate methods as having a negative entropy change. This seemingly erroneous result was the only case where all such methods failed to predict the same sign as given by a fall normal modes analysis. However, it has been suggested that trypsin–BPTI is a difficult case and varying the conditions of the full normal modes calculation for this system, such as using an increased non-bond cut-off or a distant dependent dielectric, can yield a change of sign for the entropy change (S.Y. Noskov, personal communication). This leads us to assume that the true value of the entropy change for this system is indeed negative, consistent with all our estimates using the approximate methods.

The first approximate normal modes method, M1, overestimates the individual entropies of all the proteins analysed. However, it provides better estimates of the entropy differences, in particular for the conformational changes, where it predicts a reasonable entropy difference on two out of three systems (overestimating the tip repressor entropy change). In the protein–protein complexes this method provides the right sign for the entropy change but clearly with the wrong magnitude. This method is closest in spirit to a fall normal modes calculation and is almost as computationally demanding. In comparison to the other approximate methods, it seems that the extra cost of the calculation does not yield any better results. We were unable to improve all these results after experimenting with the values of the force constant and/or potential cut-off.

The second approximate method, M2, has an overall better performance than M1. The best results with this method are obtained using three residues per block: predictions of the values for the entropy change have the right sign and a reasonable magnitude in all cases considered. Such a simplified treatment is accompanied by a considerable speed-up in the calculation timings (in the order of minutes). We also investigated the use of five residues per block but found that all entropy change estimates were wrong (results not shown), suggesting that a farther increase in block size leads to the computation of wrong vibrational entropies with this method.

The last approximate method, M3, shows a similar overall performance to M2 with bs = 3. Nonetheless, the time scale for an M3 calculation shows yet again a considerable

speed-up, as the calculations are now in the order of seconds. It should be noted, however, that in the CheY calculations M3 fails to predict even the correct sign of the entropy change. The values of the other approximate methods (including M1) for this system are all very close to zero and the full normal modes calculation itself gives the lowest entropy change of the test set. This could be taken as an indication that M3 could fail in those cases where the entropy change converges to a value close to zero. For all other cases, M3 could be the method of choice given its speed and accuracy.

We would like to comment further on one case each of the protein conformational and binding test sets. We have chosen calmodulin and the xMdm2–p53 complex because they are two well studied systems.

Let us first consider the vibrational entropies associated with the conformational changes of calmodulin upon calcium binding. Our full normal modes calculations show that the calcium-bound form of calmodulin has a significantly larger vibrational entropy than the calcium-free form. The vibrational entropy change ($T\Delta S_{vib}$) is 13.0 kcal/mol. This is likely to be related to the observed formation of a long and flexible central α-helix connecting the two globular domains of calmodulin, which in turn might allow the protein to adopt a large number of conformations [41]. Such a connection between the structural flexibility of a protein and their functional conformational transitions has also been observed for a number of other proteins [22,51–53]. Our calculation is in excellent agreement with previously reported full and a block normal modes (BNM, a modification of the original RTB method [21]) calculations, which reported values of 13.2 kcal/mol for the full calculation and 6.6 kcal/mol for the BNM method [41], except for the fact that these values were reported with the wrong sign.

The various approximate normal modes analysis methods that we have used predict a positive vibrational entropy contribution to the conformational change in calmodulin upon calcium binding. Importantly, all of the estimates of the vibrational entropy change are in reasonably good agreement with the above result from the full normal modes analysis calculation. It is necessary to point out that the absolute vibrational entropies of each of the conformations of calmodulin are significantly underestimated by all of the approximate normal modes analysis methods. In the case of M2, increasing the block size reduces the absolute vibrational entropies, while the vibrational entropy change oscillates around a nearly correct value.

Let us now consider the vibrational entropies associated with the association of xMdm2 and p53. Our full normal modes calculations show that the protein–peptide interaction of xMdm2 and the Gln-16 to Pro-27 stretch of p53 is not favoured vibrationally due to a negative contribution to the entropy. The vibrational entropy change ($T\Delta S_{vib}$) upon complexation is −4.4 kcal/mol. This is likely to be the result of a rigidification of the complex with respect to the separate subunits, particularly the peptide. Our calculation provides an estimate of the vibrational entropy change that is somewhat higher than a previously reported full normal modes calculation, which reported a value of −28.4 kcal/mol [43].

All the approximate normal modes analysis methods that we have used predict a negative vibrational entropy contribution to the interaction of xMdm2 with p53. Once again, the estimates of the vibrational entropy change are all in reasonably good agreement with the above result from the full normal modes calculation. As before, the absolute vibrational entropies of the complex, the protein and the peptide are all clearly underestimated by all the approximate methods. With respect to M2, increasing the block size again reduces the absolute vibrational entropies while the vibrational entropy change oscillates around a nearly correct value.

It is important to point out that a number of these approximate normal modes analysis methods have been developed and optimised to calculate only the low frequency normal modes, as these are responsible for the collective motions seen in proteins [8,14,18,22,25–28]. At the same time, the vibrational entropy of proteins converges relatively quickly due to the fact that its largest contribution comes from the low frequency modes [12]. This is to be expected from the exponential behaviour of Eq. (1) since small variations of the frequency value in the low frequency range have a larger effect on the estimated entropy. Interestingly, vibrational entropies of molecular association have also been shown to converge at frequencies of between 600 and 1000 cm^{-1} [12]. We can see in Fig. 1 that, in the case of calmodulin, entropy differences converge in a frequency range between 300 and 700 cm^{-1}, depending on the method used. It can also be confirmed that, for all methods, changes in the frequency values in the low frequency range give rise to larger fluctuations in the entropy difference and that it is necessary to include higher frequency vibrations to obtain convergence in the estimated entropy difference. Our results seem to indicate that the approximations that the methods considered here lead to some important vibrations (which contribute significantly to the absolute vibrational entropy) being left out; however, this clearly does not seem to affect the estimation of vibrational entropy changes and the identification of the main normal modes responsible for the collective motions of proteins.

4. Conclusions

We have used a number of approximate normal modes analysis methods and compared them where possible with a full normal modes analysis calculation for the estimation of vibrational entropies in a number of systems where protein conformational changes and protein binding are observed.

Both in the case of conformational changes and protein binding, all approximate normal modes analysis methods fail to reproduce accurately the magnitude of absolute vibrational entropies. This is due to the fact that these methods and their underlying protein models were not formulated to compute frequencies in the whole range required for entropy calculations. However, in the calculation of the vibrational

entropy differences associated with the above processes, all these approximate methods give good estimates that are in reasonably good agreement with a full normal modes analysis calculation. These approximate methods provide a considerable speed-up without seriously compromising the accuracy.

These results indicate that approximate normal modes analysis methods, which are particularly necessary for treating large biomolecular systems, are useful not only for determining the existing collective motions but also for obtaining relatively good estimates of the vibrational entropy differences associated with molecular events such as protein–protein interactions and large protein conformational changes, where full normal modes calculations are not always tractable. Furthermore, these approximate methods provide vibrational entropy differences that could easily be incorporated into free energy calculations of large biomolecular systems.

Acknowledgements

BJC gratefully acknowledges De Novo Pharmaceuticals for the award of a studentship and for computational resources. RLM is also a Research Fellow of Hughes Hall, Cambridge, UK. Useful discussions with K. Hinsen and S.Y. Noskov are gratefully acknowledged. We would like to thank Prof. Y.-H. Sanejouand for making the programmes PDBMAT, DIAGSTD and DIAGRTB available to us.

References

[1] J.A. McCammon, S.C. Harvey, Dynamics of Proteins and Nucleic Acids, Cambridge University Press, Cambridge, 1987.
[2] M. Tasumi, H. Takeuchi, S. Ataka, A.M. Dwivedi, S. Krimm, Normal vibrations of proteins: glucagon, Biopolymers 21 (1982) 711–714.
[3] T. Noguti, N. Go, Collective variable description of small-amplitude conformational fluctuations in a globular protein, Nature 296 (1982) 776–778.
[4] N. Go, T. Noguti, T. Nishikawa, Dynamics of a small globular protein in terms of low-frequency vibrational modes, Proc. Natl. Acad. Sci. U.S.A. 80 (1983) 3696–3700.
[5] M. Levitt, C. Sander, P.S. Stern, Normal-mode dynamics of a protein: bovine pancreatic trypsin inhibitor, Int. J. Quant. Chem.: Quant. Biol. Symp. 10 (1983) 181–199.
[6] B. Brooks, M. Karplus, Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor, Proc. Natl. Acad. Sci. U.S.A. 80 (1983) 6571–6575.
[7] M. Levitt, C. Sander, P.S. Stern, Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme, J. Mol. Biol. 181 (1985) 423–447.
[8] O. Marques, Y.-H. Sanejouand, Hinge-bending motion in citrate synthase arising from normal mode calculations, Proteins Struct. Funct. Genet. 23 (1995) 557–560.
[9] C. Guilbert, F. Pecorari, D. Perahia, L. Mouawad, Low frequency motions in phosphoglycerate kinase. A normal mode analysis, Chem. Phys. 204 (1995) 327–336.
[10] D.A. Case, Normal mode analysis of protein dynamics, Curr. Opin. Struct. Biol. 4 (1994) 285–290.
[11] K.K. Irikura, B. Tidor, B.R. Brooks, M. Karplus, Transition from B to Z DNA: contribution of internal fluctuations to the configurational entropy difference, Science 229 (1985) 571–572.
[12] B. Tidor, M. Karplus, The contribution of vibrational entropy to molecular association. The dimerization of insulin, J. Mol. Biol. 238 (1994) 405–414.
[13] B.R. Brooks, D. Janezic, M.K. Karplus, Harmonic analysis of large systems I. Methodology, J. Comput. Chem. 16 (1995) 1522–1542.
[14] M.M. Tirion, Large amplitude elastic motions in proteins from a single-parameter, atomic analysis, Phys. Rev. Lett. 77 (1996) 1905–1908.
[15] I. Bahar, A.R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, Fold Des. 2 (1997) 173–181.
[16] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, Anisotropy of fluctuation dynamics of proteins with an elastic network model, Biophys. J. 80 (2001) 505–515.
[17] P. Doruker, R.L. Jernigan, I. Bahar, Dynamics of large proteins through hierarchical levels of coarse-grained structures, J. Comput. Chem. 23 (2002) 119–127.
[18] K. Hinsen, Analysis of domain motions by approximate normal mode calculations, Proteins Struct. Funct. Genet. 33 (1998) 417–429.
[19] K. Hinsen, A. Thomas, M.J. Field, Analysis of domain motions in large proteins, Proteins Struct. Funct. Genet. 34 (1999) 369–382.
[20] P. Durand, G. Trinquier, Y.-H. Sanejouand, New approach for determining low-frequency normal-modes in macro molecules, Biopolymers 34 (1994) 759–771.
[21] F. Tama, F.X. Gadea, O. Marques, Y.-H. Sanejouand, A building-block approach for determining low frequency normal modes of macromolecules, Proteins Struct. Funct. Genet. 41 (2000) 1–7.
[22] F. Tama, Y.-H. Sanejouand, Conformational change of proteins arising from normal mode calculations, Protein Eng. 14 (2001) 1–6.
[23] M. Delarae, Y.-H. Sanejouand, Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model, J. Mol. Biol. 320 (2002) 1011–1024.
[24] F. Tama, W. Wriggers, C.L. Brooks III, Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory, J. Mol. Biol. 321 (2002) 297–305.
[25] M.H. Hao, H.A. Scheraga, Analyzing the normal mode dynamics of macromolecules by the component synthesis method: residue clustering and multiple-component approach, Biopolymers 34 (1994) 321–335.
[26] L. Mouawad, D. Perahia, Diagonalization in a mixed basis: a method to compute low-frequency normal modes for large macromolecules, Biopolymers 33 (1993) 599–611.
[27] D. Perahia, L. Mouawad, Computation of low-frequency normal modes in macromolecules: improvements to the method of diagonalization in a mixed basis and application to hemoglobin, Comput. Chem. 19 (1995) 241–246.
[28] A. Thomas, J. Field, L. Mouawad, D. Perahia, Analysis of the low frequency normal modes of the T-state of aspartate transcarbamylase, J. Mol. Biol. 257 (1996) 1070–1087.
[29] J. Srinivasan, T.E. Cheatham, P. Cieplak, P.A. Kollman, D.A. Case, Continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices, J. Am. Chem. Soc. 120 (1998) 9401–9409.
[30] C.L. Lawson, R.G. Zhang, R.W. Schevitz, Z. Otwinowski, A. Joachimiak, P.B. Sigler, Flexibility of the DNA-binding domains of trp repressor, Proteins 3 (1988) 18–31.
[31] L. Bellsolell, J. Prieto, L. Serrano, M. Coll, Magnesium binding to the bacterial chemotaxis protein CheY results in large conformational changes involving its functional surface, J. Mol. Biol. 238 (1994) 489–495.

[32] K. Volz, P. Matsumura, Crystal structure of *Escherichia coli* CheY refined at 1.7 Å resolution, J. Biol. Chem. 266 (1991) 15511–15519.
[33] R. Chattopadhyaya, W.E. Meador, A.R. Means, F.A. Quiocho, Calmodulin structure refined at 1.7 Å resolution, J. Mol. Biol. 228 (1992) 1177–1192.
[34] H. Kuboniwa, S. Tjandra, S. Grzesiek, H. Ren, C.B. Klee, A. Bax, Solution structure of calcium-free calmodulin, Nat. Struct. Biol. 2 (1995) 768–776.
[35] T.N. Bhat, G.A. Bentley, G. Boulot, M.I. Greene, D. Tello, W. Dall'acqua, H. Souchon, F.P. Schwarz, R.A. Mariuzza, R.J. Poljak, Bound water molecules and conformational stabilization help mediate an antigen–antibody association, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 1089–1093.
[36] M.S. Weiss, G.J. Palm, R. Hilgenfeld, Crystallization, structure solution and refinement of hen egg-white lysozyme at pH 8.0 in the presence of MPD, Acta Crystallogr. Sect. D: Biol. Crystallogr. 56 (2000) 952–958.
[37] M. Marquart, J. Walter, J. Deisenhofer, W. Bode, R. Huber, The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors, Acta Crystallogr. Sect. B 39 (1983) 480–484.
[38] J. Walter, W. Steigemann, T.P. Singh, H. Bartunik, W. Bode, R. Huber, On the disordered activation domain in trypsinogen: chemical labelling and low-temperature crystallography, Acta Crystallogr. Sect. B 38 (1982) 1462–1472.
[39] S. Parkin, B. Rupp, H. Hope, The Structure of bovine pancreatic trypsin inhibitor (BPTI) at 125 K: definition of carboxyl-residues glycine-57 and alanine-58, Acta Crystallogr. Sect. D: Biol. Crystallogr. 52 (1996) 18–29.
[40] P.H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A.J. Levine, N.P. Pavletich, Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain, Science 274 (1996) 948–953.
[41] G. Li, Q. Cui, A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca(2+)-ATPase, Biophys. J. 83 (2002) 2457–2474.
[42] S.Y. Noskov, C. Lim, Free energy decomposition of protein–protein interactions, Biophys. J. 81 (2001) 737–750.
[43] I. Massova, P.A. Kollman, Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies, J. Am. Chem. Soc. 121 (1999) 8133–8143.
[44] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins and nucleic acids, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[45] A.R. Leach, Molecular Modelling: Principles and Applications, Longman, Harlow, 1996.
[46] T.L. Hill, An Introduction to Statistical Thermodynamics, Dover, New York, 1986.
[47] A.T. Hagler, P.S. Stern, R. Sharon, J.M. Becker, F. Naider, Computer simulation of the conformational properties of oligopeptides. Comparison of theoretical methods and analysis of experimental results, J. Am. Chem. Soc. 101 (1979) 6842–6852.
[48] D.A. McQuarrie, Statistical Mechanics, Harper & Row, New York, 1976.
[49] K. Hinsen, The molecular modeling toolkit: a new approach to molecular simulations, J. Comput. Chem. 21 (2000) 79–85.
[50] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.
[51] A. Thomas, M.J. Field, D. Perahia, Analysis of the low-frequency normal modes of the R state of aspartate transcarbamylase and a comparison with the T state modes, J. Mol. Biol. 261 (1996) 490–506.
[52] A. Thomas, K. Hinsen, M.J. Field, D. Perahia, Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study, Proteins Struct. Funct. Genet. 34 (1999) 96–112.
[53] J. Ma, M. Karplus, The allosteric mechanism of the chaperonin GroEL: a dynamic analysis, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 8502–8507.