Amino Acids (2012) 42:1773–1781
DOI 10.1007/s00726-011-0895-1

ORIGINAL ARTICLE

Wavelet transform analysis of NMR structure ensembles to reveal internal fluctuations of enzymes

Mei Hu · Yizhou Li · Gang Yang · Gongbing Li · Menglong Li · Zhining Wen

Received: 27 September 2010 / Accepted: 21 March 2011 / Published online: 9 April 2011

Abstract Internal motions and flexibility are essential for biological functions in proteins. To assess the internal fluctuations and conformational flexibility of proteins, reliable computational methods are needed. In this study, wavelet transformation was used to filter out the noise and facilitate investigating the internal positional fluctuations of enzymes within nuclear magnetic resonance (NMR) structure ensembles. Moreover, potential active sites were identified by combining with positional fluctuation score, sequence conservation, and solvent accessible surface area. Among the total 107 catalytic residues in 44 examined enzymes, 69 residues were identified correctly. Our results suggest that wavelet transform analysis of structure ensemble is applicable to extract essential fluctuation information of proteins; furthermore, analysis of positional fluctuations is helpful for the identification of catalytic residues.

Keywords Structures alignment · Mean square fluctuation · Potential active sites

Source code is available from http://cic.scu.edu.cn/bioinformatics/wt_nmrst.zip.

Electronic supplementary material The online version of this article (doi:10.1007/s00726-011-0895-1) contains supplementary material, which is available to authorized users.

M. Hu · Y. Li · G. Yang · G. Li · M. Li (✉) · Z. Wen
Key Laboratory of Green Chemistry and Technology, Ministry of Education, College of Chemistry, Sichuan University, Chengdu 610064, People's Republic of China
e-mail: liml@scu.edu.cn

Introduction

Internal motions and flexibility of proteins play a critical role in keeping the right conformation to carry out biological functions (Berendsen and Hayward 2000; Chou 1988; Erkip and Erman 2004; Hub and de Groot 2009; Yang and Bahar 2005). The dynamic information can be obtained experimentally in terms of temperature factors (B factors) measured by X-ray crystallography and the NMR-derived order parameters extracted from relaxation data (Karplus and McCammon 1981; Yang and Kay 1996). Computational methods are also adopted to characterize the internal large-scale motions in proteins, including molecular dynamics (MD) simulation (Karplus and McCammon 2002), normal mode analysis (NMA) (Bahar and Rader 2005) and elastic network model (ENM) (Yang and Bahar 2005). Besides, analysis of the mean square variations in the coordinates of amino acids about their mean (native) positions provides a measure of equilibrium dynamics of proteins (Berendsen and Hayward 2000; Yang et al. 2007). To assess the conformational flexibility and fluctuations, principal component analysis (PCA) is commonly used to extract the collective motions with the largest contribution to the variance of the atomic fluctuations (Bakan and Bahar 2009; Balsera et al. 1996; Berendsen and Hayward 2000; Howe 2001; Hub and de Groot 2009; Lange et al. 2008; Maisuradze et al. 2009; Yang et al. 2009).

Nuclear magnetic resonance (NMR) spectroscopy provides atomic-resolution snapshots of proteins (Henzler-Wildman and Kern 2007). An NMR-determined structure ensemble usually contains multiple models that solved in NMR determination protocols, and are agree with the experimental constraints (Best et al. 2006). Information about the molecular flexibility can be obtained by

systematic comparison of all models. Analysis carried out by Ramanathan et al. showed that the conformation flexibility associated with structural deviations in NMR ensembles of ubiquitin characterized using quasi-harmonic analysis (QHA) is in agreement with the large-scale motions identified by computational methods (Ramanathan and Agarwal 2009). Bakan et al. also confirmed the structural variability observed in NMR ensembles exhibits remarkable correlations with top-ranking ANM modes (Bakan and Bahar 2009). Yang et al. compared the variations in atomic coordinates among the NMR models, B factors, and the fluctuation dynamics predicted by the Gaussian Network Model (GNM). They noted that NMR data appears to provide a better measure of equilibrium dynamics compared with X-ray crystallographic B factors, and the root mean square deviation (RMSD) of NMR ensemble might contain physically meaningful contributions of equilibrium fluctuations (Yang et al. 2007).

In this work, we investigated the internal fluctuations of enzymes by analyzing the variations in atomic coordinates from the well-defined average structure within NMR ensembles. Variations from real differences between the structures might be contaminated with noise. To reveal the meaningful information of internal positional fluctuations, noise should be removed. Wavelet transformation (WT) has been proved to be a powerful tool for discarding the unimportant portion of the signal (Fadili and Boubchir 2005; Leung et al. 1998). Consequently, WT was used to effectively filter out the noise from a set of atomic coordinates (after a best fit to discard overall translation and rotation). Then, the internal positional fluctuation scores were calculated. Combining with information of sequence and structure, the positions of catalytic residues in the fluctuation profiles were investigated to identify the potential active sites (PASs). Among the total 107 catalytic residues in the 44 examined enzymes, 69 were found. Our results suggest that WT analysis of NMR structure ensembles could provide useful insights into internal positional fluctuations, more importantly, identifying catalytic residues.

Materials and methods

Dataset

Our dataset contains NMR structure ensembles of 44 enzymes. Ten of them were derived from previously study of Yang (Yang et al. 2009), while the others were selected from Protein Data Bank (PDB). The corresponding catalytic sites were obtained from Catalytic Site Atlas (Porter et al. 2004).

WT analysis of NMR ensemble

Atoms, both in backbone and side chain, participate in internal fluctuations. The alignment considering backbone atoms alone might give different fluctuations of proteins from the one taking side chain atoms for the further consideration (Wu 2006). Here, we focused on the coarse-grained level (Bahar and Rader 2005) and measured the backbone atoms as representative markers to reduce the computational task. For the enzymes with multiple chains, one chain was measured.

Optimal alignment of structures in NMR ensemble

To detect internal positional fluctuations of a molecular, rigid body translation and molecular rotation have to be removed by optimal alignment of the NMR structures (Okan et al. 2009). To minimize the RMSD between structures, we employed Kabsch algorithm to superimpose structures by finding the optimal rotation and translation (Kabsch 1978; Kabsch and Sander 1983).

For a protein which contains N residues and M models, each model k is represented by a $3N$-dimensional vector.

$$\{p^k\} = \begin{bmatrix} p_1^k & p_2^k & p_3^k & \cdots & p_N^k \end{bmatrix}^T \quad (1)$$

where $p_i^k = \begin{bmatrix} x_i^k & y_i^k & z_i^k \end{bmatrix}$ is the position vector of the residue i in model k. Here, the coordination of C_α atom is used for residue position. $\overline{p_i}$ is defined as the position vector of the residue i (represented by C_α atom) averaged over the ensemble of M models. The mean structure, an average over the coordinates matrix of the models, is treated as reference structure to avoid different orientations caused by different reference structures (Hess 2000). Each model from the NMR ensemble is superimposed on the reference structure. This procedure is iterated until the averaged RMSD between successive average structures is converged (Yang et al. 2009). The mean structure of final best fitted structures is considered as the equilibrium positions which residues fluctuate around. More details about RMSD and optimal alignment produce were described in Yang's work (Yang et al. 2009).

Wavelet transformation

WT is an effective signal-processing method and extensively used in bioinformatics and chemometrics for de-noising, variable reduction, and signal compression (Ehrentreich 2002). It has been used to calculate vibration frequency (Rahaman and Wheeler 2005), detect the molecular conformational change in MD simulation and refine the numerous data of dynamics (Askar et al. 1996; Otsuka and Nakai 2007). Wavelets are capable of

de-noising a signal without appreciable degradation of the original signal (Moore 2002). Wavelet analysis is especially appropriate for signals that are intermittent, non-stationary, or noisy.

WT is defined as the projection of a function or a signal $f(t)$ onto the wavelet function.

$$W_f(a,b) = \langle f(t), \psi_{a,b}(t) \rangle = \int f(t)\psi_{a,b}(t)\mathrm{d}t \tag{2}$$

$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right) \quad (a,b \in R; a \neq 0) \tag{3}$$

where a is the scale (dilation) and b is the position (translation) parameter, $\psi(t)$ is the basis function and $\psi_{a,b}(t)$ is the basis wavelet function at a particular scale a and a translation b.

$$\psi_{m,n}(t) = a_0^{-m/2}\psi(a_0^{-m_t} - nb_0) \ (m,n \in Z; a_0 \neq 0) \tag{4}$$

In this work, discrete wavelet transform (DWT) was used and thus Eq. (2) can be written as Eq. (3), where $a_0^m = a$ and $nb_0a_0^m = b$. DWT uses $a_0 = 2$ and $b_0 = 1$, and therefore

$$\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n) \tag{5}$$

For more details of the fundamentals of Wavelet analysis, the reader is referred to articles of Ehrentreich et al. (Ehrentreich 2002).

WT projects a signal onto a wavelet basis and decomposes it into scales of coefficients. Approximation and detail coefficients represent the smooth and detail of signal respectively (Li et al. 2008). At the deeper scale of decomposition, less information is contained in the coefficients. To analyze the internal fluctuations within an NMR structure ensemble, three steps were adopted: (1) Calculate the coordinate deviations of residues into numeral matrix; (2) Decompose the signal by wavelet; (3) Reconstruct the de-noised signal.

For an ensemble containing M models ($1 \leq k \leq M$) and N residues ($1 \leq i \leq N$) per model, we built a matrix of M rows each consisting of $3N$-dimensional vectors for N residues.

$$C = \begin{pmatrix} \Delta p_1^1 & \Delta p_2^1 & \Delta p_3^1 & \ldots & \Delta p_N^1 \\ \Delta p_1^2 & \Delta p_2^2 & \Delta p_3^2 & \ldots & \Delta p_N^2 \\ \Delta p_1^3 & \Delta p_2^3 & \Delta p_3^3 & \ldots & \Delta p_N^3 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \Delta p_1^M & \Delta p_2^M & \Delta p_3^M & \ldots & \Delta p_N^M \end{pmatrix} \tag{6}$$

The corresponding Δp_i^k describes the deviations in the x-, y- and z-coordinates of the residue i from its mean position $\overline{p_i}$.

$$\Delta p_i^k = p_i^k - \overline{p_i} = \left(\Delta x_i^k, \Delta y_i^k, \Delta z_i^k\right) \tag{7}$$

Each row C_k is considered to be the original signal. DWT based on Daubechies wavelet of order 8 (db8) was used to decompose C_k. To separate the noise from the information of internal fluctuations, we calculated approximation coefficients and detail coefficients at different decomposition scales. The latter were discarded while the former were kept intact to reconstruct the fluctuation profile (Fadili and Boubchir 2005). R_k is the reconstructed signal.

$$R_k = \begin{bmatrix} \Delta \mathrm{Pr}_1^k & \Delta \mathrm{Pr}_2^k & \Delta \mathrm{Pr}_3^k & \ldots & \Delta \mathrm{Pr}_N^k \end{bmatrix} \tag{8}$$

where $\Delta \mathrm{Pr}_i^k = \left[\Delta Xr_i^k, \Delta Yr_i^k, \Delta Zr_i^k\right]$ is the refined deviation of residue i in model k from its mean position.

Analysis of potential active sites (PASs)

Enzyme catalysis is an intrinsically dynamic process (Boehr et al. 2006; Eisenmesser et al. 2002; Tousignant and Pelletier 2004). Previous studies suggested that catalytic residues are highly constrained to exhibit low flexibility and prone to occupy the key positions corresponding to the global hinge regions (Yang and Bahar 2005; Yang et al. 2009). In this part, PASs of enzymes were determined by combining with positional fluctuation, sequence conservation and solvent accessibility.

Positional fluctuation scores

The mean square fluctuation (MSF) in the coordinate of residue about its mean position provides a way to detect protein internal fluctuations (Yang et al. 2007). In this work, the maximal mean square fluctuation (mMSF) of the particular residue over the whole ensemble was used to character the internal fluctuation amplitude of the residue.

Original fluctuation score was deduced from Δp_i^k while the fluctuation score refined by WT was deduced from $\Delta \mathrm{Pr}_i^k$.

$$f_i^k = \left\| \Delta \overset{k}{\underset{i}{\mathrm{Pr}}} \right\|^2 = (\Delta Xr_i^k)^2 + (\Delta Yr_i^k)^2 + (\Delta Zr_i^k)^2 \tag{9}$$

$$F = \begin{pmatrix} f_1^1 & f_2^1 & f_3^1 & \ldots & f_N^1 \\ f_1^2 & f_2^2 & f_3^2 & \ldots & f_N^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ f_1^M & f_2^M & f_3^M & \ldots & f_N^M \end{pmatrix} \tag{10}$$

$$F_k = \begin{pmatrix} f_1^k & f_2^k & f_3^k & \ldots & f_N^k \end{pmatrix} \tag{11}$$

$$Fm = [fm_1, fm_2, fm_3, \ldots fm_N] \tag{12}$$

The mMSF was standardized over the whole enzyme. The most mobile residue was assigned 1 and the least mobile one was 0. The normalized mMSF was represented by Fm and the element fm_i was the normalized mMSF of the residue i.

Conservation score

Highly sequenced conservation is one of the most significant features of catalytic residues (Tang et al. 2008). Conversation scores ranging from 1 (most variable) to 9 (most conserved) were assigned using the Consurf Server (Landau et al. 2005).

Solvent accessible surface area

Solvent accessible surface areas of residues in each model were calculated by DSSP (Kabsch and Sander 1983). The values of solvent accessible surface area of a particular residue differ from one model to another. We took the mean solvent accessible surface area through the NMR ensemble for representation.

Detection of the potential active sites

According to the study of Yang and Bahar, the tendency to locate in the key mechanical sites of the molecule and the low fluctuation scores of catalytic residues provide us a pathway to identify PASs (Yang and Bahar 2005). Based on the low fluctuations of catalytic residues, we detected PASs by incorporating the information of sequence conservation and solvent accessibility.

First, in accordance with the previous statistical result (Yang and Bahar 2005), we selected the sites whose mMSF was lower than 0.1 (Yang and Bahar 2005; Yang et al. 2009). Then we picked the most conserved ones from the first selection due to the highly conservation of catalytic residues (Bartlett et al. 2002; Tang et al. 2008). Finally, we ranked the selected sites by their mean solvent accessibility surface areas. The top-ranking ten (or less) PASs for the examined enzymes are listed in supplementary material.

Results and discussion

The positional fluctuation profiles of the 44 examined enzymes were deduced from WT. The PASs were selected based on the positional fluctuation score. 69 of the total 107 catalytic residues in these 44 examined enzymes were identified. Partial detected PASs are listed in Table 1. More results are obtainable in supplemental material.

Wavelet and decomposition scales

There are large number of known wavelet families which can be selected to decompose the signals, such as Biorthogonal, Coiflet, Harr, Symmlet and Daubechies wavelets (Mahmoodabadi et al. 2005). Different wavelet families are applicable to different signals, and different decomposition scales have different results in analyzing signals (Wen et al. 2005). In this work, Daubechies wavelets were used due to their simplicity and ability to yield better approximations to smooth signal (Lina and Mayrand 1995; Mahmoodabadi et al. 2005). For the sake of choosing the appropriate decomposition scale and Daubechies' basis type, the mean correlation coefficients between original fluctuation scores and those refined by WT based on

Table 1 Identification of PASs based on WT of NMR structure ensembles

NMR	Catalytic residues[a]	Top-ranking PASs[b]	α^c
1AH2	32, 62(1), 153, 215(4)	**62**, 38, 94, **215**, 123, 213, 216, 195, 150, 201	0.83
1AO8	19, 26(4)	57, 30, 54, **26**, 21, 6, 99, 14, 116, 42	0.83
1APS	23(2), 41(3)	17, **23**, **41**, 69, 19, 20, 15, 47, 37, 75	0.88
1AX3	66(7), 68, 83, 85	162, 116, 89, 135, 106, 91, **66**, 31, 79, 81	0.97
1AYK	119, 136	75, 128, 140, 160, 122, 9, 118, 76, 84, 83	0.97
1BC4	10, 35(1), 103(4), 104	**35**, 50, 105, **103**, 42, 102, 39, 83, 34, 81	0.85
1BM6	202, 219	223, 146, 170, 211, 151, 184, 205, 158, 166, 247	0.96
1BVE	25(7), 26(9)	29, 28, 87, 49, 88, 23, **25**, 84, **26**, 9	0.84
1BVM	30(4), 48, 99	43, 49, 111, **30**, 39, 105, 5, 4, 52, 84	0.87
1C54	54, 65(1), 85(3)	**65**, 84, **85**, 51, 22, 69, 59, 12, 11, 39	0.57
1EQ0	82, 92(1)	**92**, 121, 1, 10, 123, 95, 97, 124, 114, 37	0.61

[a] Catalytic residues derived from Catalytic Site Atlas, the order numbers of the catalytic residues in the Top-ranking PASs are written in the brackets

[b] Top-ranking ten (or less) minima predicted by our method, residues in boldface are catalytic residues

[c] Correlation coefficients between pairs of original fluctuation scores and that deduced from by WT

different Daubechies' basis type at scale 1–3 were calculated respectively (Table 2). For the 44 examined enzymes, they are 0.86, 0.83 and 0.78 at scale 1, 2 and 3 based on db8, respectively. These strong correlations clarify that db8 is applicable to extract the essential internal positional fluctuations of proteins.

Figure 1 shows the reconstructed fluctuation profiles of phospholipase A2 (PDB ID: 1BVM) (Yuan et al. 1999), with approximation coefficients only, at scales 1–3. The *y*-axis indicates the mean square fluctuation scores and the *x*-axis indicates the residue positions along the sequence. The reconstructed fluctuation profile at scale 1 has the strongest correlation with the original fluctuation profile. However, the reconstructed fluctuation profiles at scale 2 and scale 3 are over smoothed. As a result, the fluctuation magnitudes of some fragments are weakened. For instance, the peaks between residues 30 and 40 at scale 2 and scale 3 are more flattened than that at scale 1. It is clearly that the crudest approximate coefficients contain most important information of internal fluctuations. Therefore we chose the reconstructed fluctuation profile at scale 1 based on db8 as the representation of internal fluctuations.

Table 2 Correlation coefficients of fluctuation scores deduced by WT

Wavelet	C1	C2	C3
db2	0.56	0.16	−0.01
db4	0.67	0.31	0.06
db6	0.81	0.59	0.35
db8	0.86	0.83	0.78
db10	0.80	0.59	0.34

C1, C2, C3 are mean correlation coefficients between original fluctuation scores and that deduced from WT based on Daubechies family at scale 1, scale 2 and scale 3, respectively

Comparison of WT and PCA

PCA is also considered to be a useful tool for finding large-scale motions in proteins. The first few principal components (PCs) describe the collective motions of the system (Boehr et al. 2006; Eisenmesser et al. 2002; Hess 2000). In order to compare the performance of WT and PCA, positional fluctuation profiles of the 44 examined enzymes were also deduced from PCA.

The mean correlation between the fluctuations deduced from our methods and that from the first 3 PCs, 5 PCs and 10 PCs are 0.73, 0.84 and 0.85, respectively. The fluctuation scores deduced from our methods are in good agreement with those predicted by PCA except a few of enzymes. For example, the fluctuation scores of angiogenin (PDBID:1GIO) (Lequin et al. 1996) deduced from the two methods are highly correlated (Table 3), however, there is unlikeness in the fluctuation profiles. We examined the fluctuation dynamics predicted by Gaussian Network Model (GNM), which was proposed by Bahar et al. to explore the intrinsic dynamics of proteins (Bahar et al. 1997; Yang et al. 2005). The slowest GNM mode usually describes the largest amplitude fluctuations of the structure. Figure 2 shows the positional fluctuation profiles of angiogenin calculated by GNM for the first NMR model, WT and PCA. The fluctuation profile deduced from WT is in accordance with the slowest GNM mode. Even though the first 10 PCs roughly explained 100% of the total

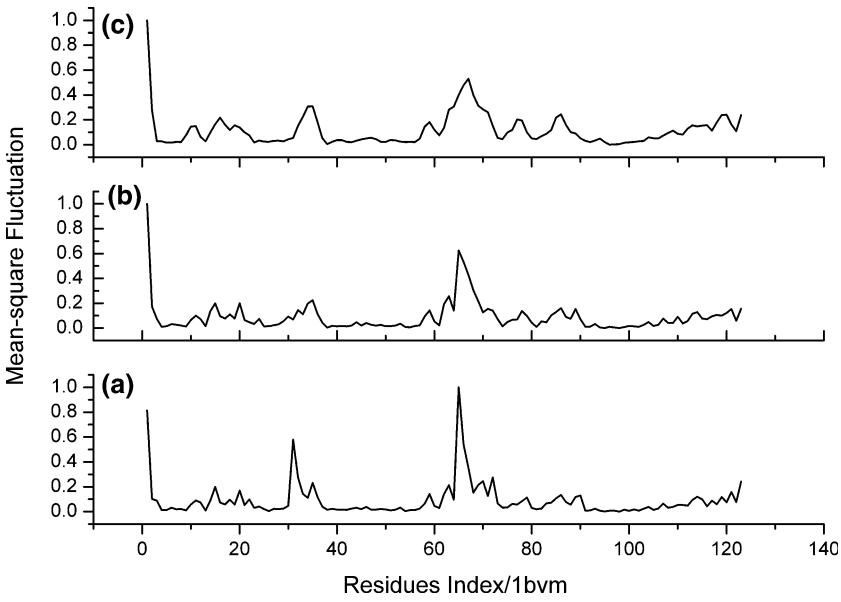

Fig. 1 Fluctuation profile of phospholipase A2 deduced by WT based on db8 at different decomposition scale:
a reconstructed fluctuation profile at scale 1;
b reconstructed fluctuation profile at scale 2;
c reconstructed fluctuation profile at scale 3

Table 3 Comparison the Fluctuation Scores of Angiogenin (PDB: 1GIO) obtained from WT and PCA

PCs	α (%)	δ_{WT}	δ_{GNM}
PC_3	78.34	0.69	0.04
PC_5	88.25	0.72	0.08
PC_{10}	100	0.73	0.10

PC_3, PC_5 and PC_{10} are the first 3, 5 and 10 principal components of PCA

"α" is the percent variability explained by the first n principal components (PCs)

δ_{WT} is the correlation between fluctuation scores deduced from WT at scale 1 based on db8 and that obtained from the first n PCs

δ_{GNM} is the correlation between fluctuation scores deduced from the slowest model of GNM and that obtained from the first n PCs

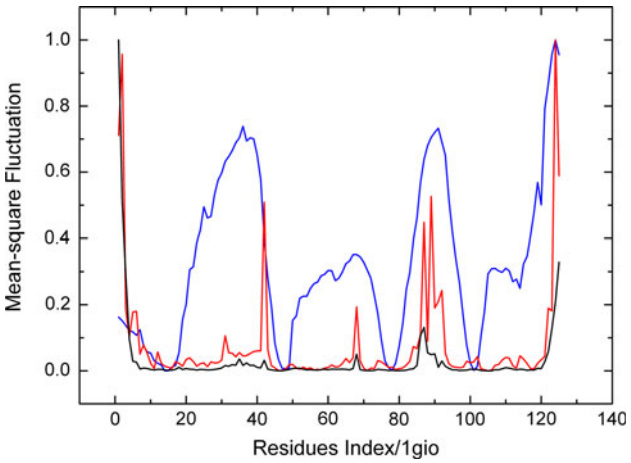

Fig. 2 Fluctuation profiles of angiogenin (PDBID: 1GIO) deduced from WT, PCA and GNM. The *blue curve* corresponds to the slowest GNM mode for the first NMR model. The *red curve* corresponds to the fluctuations deduced from WT based on db8 at scale 1 and the *black curve* corresponds to the fluctuations predicted by the first 10 PCs

variability (Table 3), the curve of fluctuations predicted by the first 10 PCs can hardly reflect the fluctuation amplitude of the residues. We also examined the fluctuation profiles predicted by the slowest GNM modes for different NMR models. It could yield the same results. For angiogenin, WT is more competent in analyzing the deviations to reveal internal positional fluctuations.

Determination of potential active sites

We investigated the positions of catalytic residues in the fluctuation profiles, and confirmed that catalytic residues exhibit low fluctuations and preferentially occupy the key regions of highly constrained (Fig. 3). This result was also established in other applications (Chen and Bahar 2004; Yang and Bahar 2005; Yang et al. 2009). Tang et al. reported that closeness centrality plays an important role in predicting catalytic residues, and active sites usually have more interactions with other residues (Tang et al. 2008). The interactions between catalytic residues and the other residues restrain their fluctuations. Catalytic residues are apt to be located in the minima regions of the fluctuation profiles. However, it is a challenge to identify catalytic residues from the minima regions (Fig. 3). For determining PASs, sequence conservation and solvent accessible surface area were used. Several residues adjoining catalytic residues were also top-ranked for their similarity with catalytic residues.

We further calculated the secondary structure and surface binding pocket using GOR (Garnier et al. 1996) and EPOS_BP (http://gepard.bioinformatik.uni-saarland.de/software/epos-bp), respectively. Figure 4 shows the secondary structure and binding pocket distribution of the PASs determined by our methods. Similar frequencies are observed for these two types of residues in β-sheets and coil regions. Majority of the detected PASs are either at the end of β-sheets or in the loops. Most PASs are located in the largest three pockets. Non-catalytic residues are spatial neighbors of catalytic residues, though they are not close to each other in sequence (Fig. 5). It is reasonable to presume that they probably collaborate with active residues for the biological functions.

Conclusion

In this paper, we employed wavelet transformation to effectively filter out the noise and investigated the important internal positional fluctuations of enzymes within NMR structure ensembles. The deviations from mean positions of residues reconstructed by crudest approximate coefficients are shown to contain the most important information of internal fluctuations. Our results are in good agreement with those predicted by PCA except a few of enzymes (e.g. angiogenin). The fluctuations deduced from WT appear to provide a better measure of internal fluctuations. This may be due to the ability of WT to reveal characteristics of the signals at multi-resolution. We concentrated on fluctuation amplitudes of residues, so it should be noted that this method is unable to indicate the fluctuation directions of residues. However, our results suggest that as a widely used signal-processing method, WT could be employed to extract essential positional fluctuations within proteins as well.

To identify the catalytic residues, we investigated the fluctuation score, conservation and solvent accessibility.

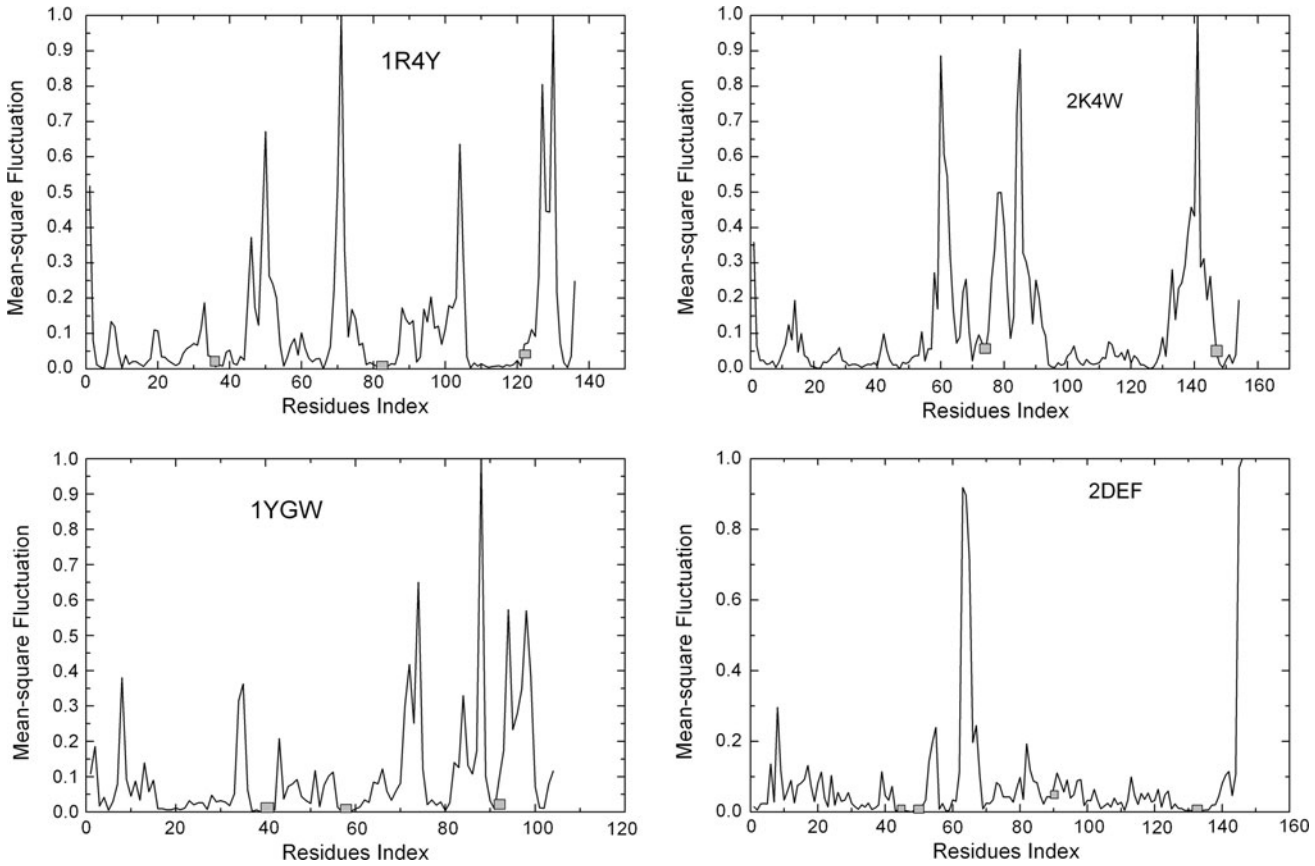

Fig. 3 Fluctuation profiles deduced from WT. The catalytic residues are marked with a *closed square*

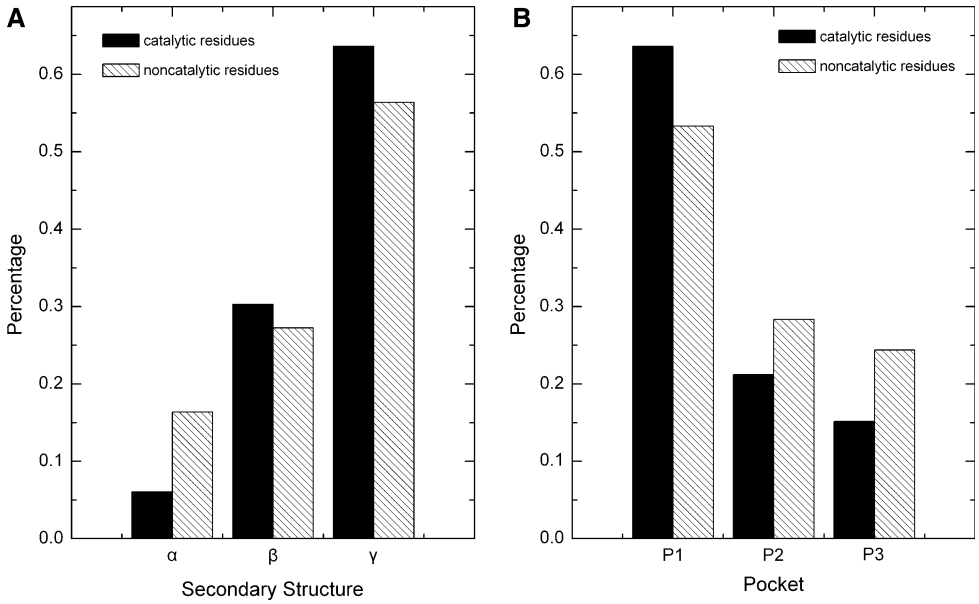

Fig. 4 Secondary structure and binding pocket distribution of detected PASs. **a** Distribution of secondary structure, α,β,γ correspond to α-helices, β-sheets and coil (loop) regions, respectively. **b** Distribution of binding pocket, p1, p2, p3 correspond to the largest three binding pockets, the fourth to ninth largest binding pockets and none pocket, respectively

We confirmed that catalytic residues tend to occupy key positions of highly constrained and solvent accessible. The analysis of enzyme internal motions gives an insight into catalytic mechanism. However, greater progress is still needed to unveil the relationship between protein dynamics and their biological functions.

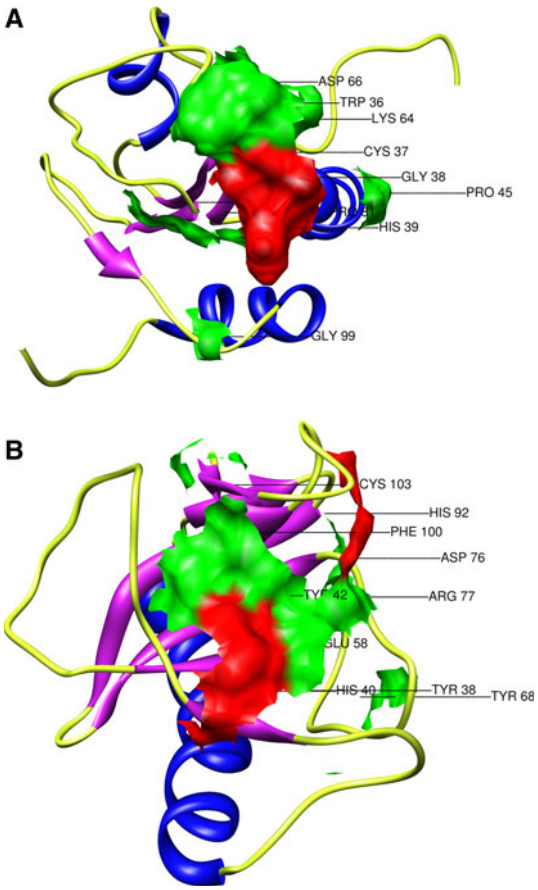

Fig. 5 Secondary structure and surfaces of PASs detected for **a** Protein disulfide-isomerase A6 (PDBID: 1X5C) and **b** Ribonuclease T1 (PDBID: 1YGW). The *blue*, *pink* and *yellow* regions correspond to α-helices, β-sheets and coils (loops), respectively. The *red* regions are surfaces of the detected catalytic residues, and the green regions are surfaces of non-catalytic residues comprised in the PASs detected by our methods

Acknowledgments This work was funded by the National Natural Science Foundation of China (No.20972103).

References

Askar A, Cetin AE, Rabitz H (1996) Wavelet transform for analysis of molecular dynamics. J Phys Chem 100:19165–19173

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

Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 2:173–181

Bakan A, Bahar I (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proc Natl Acad Sci USA 106: 14349–14354

Balsera MA, Wriggers W, Oono Y, Schulten K (1996) Principal component analysis and long time protein dynamics. J Phys Chem 100:2567–2572

Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. J Mol Biol 324:105–121

Berendsen HJC, Hayward S (2000) Collective protein dynamics in relation to function. Curr Opin Struct Biol 10:165–169

Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M (2006) Relation between native ensembles and experimental structures of proteins. Proc Natl Acad Sci USA 103:10901–10906

Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. Chem Rev 106:3055–3079

Chen SC, Bahar I (2004) Mining frequent patterns in protein structures: a study of protease families. Bioinformatics 20:i77–i85

Chou KC (1988) Low-frequency collective motion in biomacromolecules and its biological functions. Biophys Chem 30:3–48

Ehrentreich F (2002) Wavelet transform applications in analytical chemistry. Anal Bioanal Chem 372:115–121

Eisenmesser EZ, Bosco DA, Akke M, Kern D (2002) Enzyme dynamics during catalysis. Science 295:1520–1523

Erkip A, Erman B (2004) Dynamics of large-scale fluctuations in native proteins. Analysis based on harmonic inter-residue potentials and random external noise. Polymer 45:641–648

Fadili JM, Boubchir L (2005) Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities. IEEE T IMAGE PROCESS 14:231–240

Garnier J, Gibrat JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. Comp Methods Macromol Seq Anal 266:540–553

Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. Nature 450:964–972

Hess B (2000) Similarities between principal components of protein dynamics and random diffusion. Phys Rev E 62:8438–8448

Howe PWA (2001) Principal components analysis of protein structure ensembles calculated using NMR data. J Biomol NMR 20:61–70

Hub JS, de Groot BL (2009) Detection of functional modes in protein dynamics. Plos Comput Biol 5:e1000480

Kabsch W (1978) A discussion of the solution for the best rotation to relate two sets of vectors. Acta Cryst 34:827–828

Kabsch W, Sander C (1983) Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Karplus M, McCammon JA (1981) The internal dynamics of globular-proteins. CRC Crit Rev Biochem 9:293–349

Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9:646–652

Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E et al (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 33:W299–W302

Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KFA et al (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320:1471–1475

Lequin O, Albaret C, Bontems F, Spik G, Lallemand JY (1996) Solution structure of bovine angiogenin by H-1 nuclear magnetic resonance spectroscopy. Biochemistry 35:8870–8880

Leung AKM, Chau FT, Gao JB (1998) A review on applications of wavelet transform techniques in chemical analysis: 1989–1997. Chemom Intell Lab Syst 43:165–184

Li YZ, Wen ZN, Zhou CS, Tan FY, Li ML (2008) Effects of neighboring sequence environment in predicting cleavage sites of signal peptides. Peptides 29:1498–1504

Lina JM, Mayrand M (1995) Complex Daubechies wavelets. Appl Comput Harmon A 2:219–229

Mahmoodabadi SZ, Ahmadian A, Abolhasani MD (2005) ECG feature extraction using Daubechies wavelets. Proceedings of the

fifth IASTED international conference on visualization, imaging, and image processing: 343–348
Maisuradze GG, Liwo A, Scheraga HA (2009) Principal component analysis for protein folding dynamics. J Mol Biol 385:312–329
Moore M (2002) The use of Wavelets for Determining Wing Flexure in Airborne GPS Multi-Antenna Attitude Determination Systems. Proceedings of ION GPS-2002: 1022–1029
Okan OB, Atilgan AR, Atilgan C (2009) Nanosecond motions in proteins impose bounds on the timescale distributions of local dynamics. Biophys J 97:2080–2088
Otsuka T, Nakai H (2007) Wavelet transform analysis of ab initio molecular dynamics simulation: application to core-excitation dynamics of BF3. J Comput Chem 28:1137–1144
Porter CT, Bartlett GJ, Thornton JM (2004) The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucleic Acids Res 32:D129–D133
Rahaman A, Wheeler RA (2005) Wavelet transforms for determining time-dependent vibrational frequencies. J Chem Theory Comput 1:769–771
Ramanathan A, Agarwal PK (2009) Computational identification of slow conformational fluctuations in proteins. J Phys Chem B 113:16669–16680
Tang YR, Sheng ZY, Chen YZ, Zhang ZD (2008) An improved prediction of catalytic residues in enzyme structures. Protein Eng Des Sel 21:295–302
Tousignant A, Pelletier JN (2004) Protein motions promote catalysis. Chem Biol 11:1037–1042
Wen ZN, Wang KL, Li ML, Nie FS, Yang Y (2005) Analyzing functional similarity of protein sequences with discrete wavelet transform. Comput Biol Chem 29:220–228
Wu D (2006) Distance-based protain structure modeling. Ph.D. thesis, Program on Bioinformatics and Computational Biology and Department of Mathematics, Iowa State University
Yang LW, Bahar I (2005) Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure 13:893–904
Yang DW, Kay LE (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. J Mol Biol 263:369–382
Yang LW, Liu X, Jursa CJ, Holliman M, Rader A et al (2005) iGNM: a database of protein functional motions based on Gaussian Network Model. Bioinformatics 21:2978–2987
Yang LW, Eyal E, Chennubhotla C, Jee J, Gronenborn AM et al (2007) Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. Structure 15:741–749
Yang LW, Eyal E, Bahar I, Kitao A (2009) Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. Bioinformatics 25:606–614
Yuan CH, Byeon IJL, Li YS, Tsai MD (1999) Structural analysis of phospholipase A(2) from functional perspective. 1. Functionally relevant solution structure and roles of the hydrogen-bonding network. Biochemistry 38:2909–2918