ALADYN: a web server for aligning proteins by matching their large-scale motion

R. Potestio[1], T. Aleksiev[1], F. Pontiggia[2], S. Cozzini[3,1] and C. Micheletti[1,3,*]

[1]Scuola Internazionale Superiore di Studi Avanzati and eLab, via Bonomea 265, 34136 Trieste, Italy, [2]Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA and [3]Democritos CNR-IOM, via Bonomea 265, 34136 Trieste, Italy

ABSTRACT

The ALADYN web server aligns pairs of protein structures by comparing their internal dynamics and detecting regions that sustain similar large-scale movements. The latter often accompany functional conformational changes in proteins and enzymes. The ALADYN dynamics-based alignment can therefore highlight functionally-oriented correspondences that could be more elusive to sequence- or structure-based comparisons. The ALADYN server takes the structure files of the two proteins as input. The optimal relative positioning of the molecules is found by maximizing the similarity of the pattern of structural fluctuations which are calculated via an elastic network model. The resulting alignment is presented via an interactive graphical Java applet and is accompanied by a number of quantitative indicators and downloadable data files. The ALADYN web server is freely accessible at the http://aladyn.escience-lab.org address.

INTRODUCTION

The characterization of proteins and enzymes is usually articulated along the logical cascade *sequence → structure → function*. The current understanding of the connection between the various terms of this tripartite ladder has been much shaped by the availability of quantitative comparative schemes. Indeed, sequence and structure comparative (alignment) methods have been used to clarify the extent to which similarities at the level of primary sequence reverberate at the level of native conformation (1–5). The same methods, in addition, lend naturally to be used to classify proteins and to detect evolutionary relationships among them (6–11).

In recent years, computational schemes ranging from atomistic simulations to coarse-grained models (12–17) have aptly complemented single molecule experiments by showing that for several enzymes the native structure gives a specific imprinting to the molecule's internal dynamics. The latter, in turn, can directly impact on the functionality of many, though not all, enzymes by favoring the interconversion between biologically relevant conformers, such as the rest and catalytically potent forms.

Based on this perspective, valuable insight into the structure–function relationship was provided by investigations where the large-scale internal dynamics was compared for proteins with a substantial degree of structural similarity (12,18). The good spatial superposability of the proteins of interest was essential to identify their structurally equivalent amino acids, whose large-scale functional motion could be finally compared (12,19,20).

While this dynamics-oriented comparative scheme is valuable, the necessity to identify structurally equivalent pairs of amino acids prior to measuring their dynamical consistency rules out, *a priori*, the possibility to detect similar large-scale movements in proteins lacking an overall fold similarity. This limitation can be overcome by alignment strategies that are tolerant from the structural point of view and that directly promote the identification of common internal-dynamics patterns in two proteins. The so-called 'dynamics-based alignment' that was recently introduced by some of us (21) is a general quantitative method to perform such comparisons and was used to highlight pervasive funtional-oriented relationships between proteins that differ, according to the CATH classification (10), at the level of topology, and even class or architecture (21,22).

The dynamics-based alignment is now offered, after a major algorithmic redesign, in the form of a web server named ALADYN. With respect to the method originally formulated in (21), the alignment search implemented in ALADYN is more general and efficient. In particular, the constraint that segments of aligned amino acids in the two proteins had to have the same sequence order and directionality has been removed.

---

*To whom correspondence should be addressed. Tel: +39 040 3787 300; Fax: +39 040 3787 528; Email: michelet@sissa.it

A more computationally effective scoring function and stochastic optimization of the alignment have also been adopted. The resulting method is therefore much faster than its original formulation: an alignment of two proteins of 250 amino acids is typically returned in ~1 min on modern workstations.

The online server allows users to submit freely (without registration) jobs that require up to 20 min of CPU time. The allowed CPU limit is sufficient to align two proteins of about 1000 amino acids. Stand-alone LINUX and MAC-OSX versions of the ALADYN executable are made freely available upon request.

MATERIALS AND METHODS

Background

The large-scale structural rearrangements that accompany or assist the biological function of several proteins and enzymes are known to occur along generalized directions corresponding to the lowest energy modes of the system. Due to the collective character of these modes, which entail the concerted displacement of several amino acids, it can be expected *a priori* that their salient features can be adequately captured using simplified, coarse-grained protein models, such as elastic networks (23–25).

In this study, we adopt the β-Gaussian elastic network model (25) where amino acids are described with a two-centroid representation: one centroid for the mainchain, corresponding to the CA atom and a second one for the sidechains, whose location is controlled by the degrees of freedom of the mainchain, CA, centroids. In the general spirit of elastic network models, structural fluctuations around the native conformation are penalized by means of an effective quadratic potential energy,

$$H = \frac{1}{2} \sum_{ij} \delta \vec{x}_i \, M_{ij} \, \delta \vec{x}_j \tag{1}$$

where $\delta \vec{x}_i$ denotes the displacement of the *i*-th CA centroid from the native position. The lowest-energy modes correspond to the eigenvectors of M having the smallest, non-zero, eigenvalues. These modes dominate the equilibrium fluctuation dynamics of the system (26). In fact, indicating with $\vec{v}^l$ the *l*-th eigenvector of M and with λ_l the associated eigenvalue one has that the mean square fluctuation (or mobility) of the *i*-th CA centroid is, $\langle |\vec{x}_i|^2 \rangle = \sum_l' \frac{1}{\lambda_l} |\vec{v}_i^l|^2$, where the prime indicates that the sum is restricted to non-zero eigenvalues and $\vec{v}_i^l$ indicates the displacement of the *i*-th centroid entailed by the *l*-th mode.

Optimal dynamics-based alignment

The search for significant correspondences of the fluctuation dynamics in two proteins is performed by minimizing over relative translations and rotations of the two proteins, a scoring function that rewards the spatial superposition of protein regions that have similar motion. Specifically, the minimized function is

given by:

$$s = -\frac{N_1 N_2}{n} \sum_{l,m=1}^{n} \left[\sideset{}{'}\sum_{\substack{i=1...N_1 \\ j=1...N_2}} \vec{v}_i^l \cdot \vec{w}_j^m f(\Delta_{ij}) \right]^2 \tag{2}$$

where $f(d) = \frac{1}{2}[1.0 - \tanh((d - R_0)/1\,\text{Å})]$ and the indices i and j run respectively over the N_1 amino acids of the first protein and the N_2 of the second one. The primed sum denotes that score contributions are further restricted to pairs where the square mobility of amino acids i and j does not exceed by a factor of 4 the average one per amino acid. This restriction is introduced to avoid artifacts resulting from the presence of highly mobile loops or termini. The *l*-th lowest energy eigenmodes of the first and second protein is indicated as $\vec{v}^l$ and $\vec{w}^l$, respectively. As customary, only the $n = 10$ lowest energy modes are considered. The sigmoidal function f is used to restrict the score contribution to amino acid pairs i and j at a distance smaller than the interaction range $R_0 = 7\,\text{Å}$: specifically, the effective amino acid distance, Δ_{ij}, used in the sigmoidal function, measures the spatial separation of the fragments $[i-1,i,i+1]$ and $[j-1,j,j+1]$. *A priori* the latter could be matched with either the same or opposite sequence orientation. For the two cases, the segments distance is defined respectively as: $d^+ = \max\{d_{i-1,j-1}, d_{i,j}, d_{i+1,j+1}\}$ and $d^- = \max\{d_{i-1,j+1}, d_{i,j}, d_{i+1,j-1}\}$, with $d_{i,j}$ being the Euclidean distance of amino acids i and j. The most appropriate sequence orientation is chosen *a posteriori* by setting $\Delta_{ij} = \min(d^+, d^-)$.

The function in Equation 2 rewards those superpositions of the proteins having high scalar product among the fluctuation modes of every amino acid pair within the cutoff distance. This quantifies the consensus of the fluctuation of regions in spatial proximity.

The minimization of the score, s, over the relative rotations and translations of the two molecules of interest is carried similarly to the MISTRAL structural alignment method (27). The two proteins are first superposed by optimally aligning segments of up to 50 amino acids. This initial superposition is next optimized by minimizing s over the possible relative orientations of the molecules. The list of equivalent amino acids is finally computed using a 'seed and grow' search for matching segments (seed threshold equal to 4.5 Å and tolerance equal to 5 Å) (27,28).

Finally, the statistical significance of the returned alignment is computed by comparing its score, s, against a reference probability distribution of alignment scores of unrelated protein pairs. This reference distribution was obtained starting from the representative protein data set of Sierk and Pearson (29). From this set, we randomly picked 10^5 pairs of non-homologous and structurally dissimilar proteins (differing at the level of CATH topology) and computed the distribution of their alignment scores in dependence of the length of the longest protein of each pair. As customary (29), based on the high level of dissimilarity of these pairs, it is expected *a priori* that only a negligible fraction of the random

alignments will correspond to true positive correspondences; and the distributions are used as 'gold standard' for other queries.

The tails of the length-regularized score distributions were found to be well-described by the extremal Gumbel statistics which was accordingly used to compute the statistical significance of a specific alignment. The latter is quantified by means of a *P*-value or, equivalently, through a *z*-score. The former is the probability that an alignment of unrelated proteins returns a score as high as the observed one, while the second measures by how many standard deviations the observed score exceeds the one expected for random alignments. Statistically significant matches are therefore associated to small *P*-values and large *z*-scores.

Integration of non-aligned degrees of freedom and RMSIP

To quantify the accord of the internal dynamics of the matching regions, it is necessary to compute the effective (non-zero) lowest-energy modes of the amino acids marked for alignment after the thermodynamic integration of the degrees of freedom of the non-marked amino acids [see e.g. supplementary material of (21)]. The consistency of the effective modes of the aligned protein is aptly measured through the root mean square inner product (RMSIP). Indicating with $\vec{u}^l$ and $\vec{t}^l$, the *l*-th effective modes of the proteins, one defines

$$\text{RMSIP} = \sqrt{\frac{1}{10}\sum_{l,m=1}^{10}\left|\sum_{i=1...q}\vec{u}_i^l \cdot \vec{t}_i^m\right|^2} \tag{3}$$

where the index *i* labels the *q* corresponding amino acids. The RMSIP takes on the value 1 in case of perfect correspondence of the spaces, and 0 in case of their complete orthogonality. When the number of compared amino acids, *q*, is of the order of 100 amino acids, RMSIP values equal to 0.7 or higher are typically deemed as statistically significant (30).

Finally, in addition to the RMSIP value, the root mean square distance, RMSD, of the matching amino acid pairs is also used to convey the quality of the alignment.

Web server: input and output

In the input form, users are asked to provide the two proteins to align: this can be done either by uploading PDB coordinate files or by entering their PDBids (and, optionally, the chain identifier). The algorithm's running time scales approximately proportionally to the product of the lengths of the input proteins. In fact, the time required for the alignment of two proteins of about 250 amino acids is typically <1 min on the modern multicore server that hosts ALADYN, while two proteins of about 500 amino acids are completed in ~4 min. Clearly, the run time can vary depending on the number of jobs submitted at the same time.

Upon successful completion, users are finally directed to an interactive graphical representation of the superposed proteins, based on the Jmol (31) applet, which is complemented by a summary of the salient properties of the alignment, number of aligned amino acids, RMSIP, RMSD and the statistical significance conveyed by the *z*-score and *P*-value.

The applet controls can be used to visualize the matching regions and/or the matching modes ranked for decreasing mutual similarity. This ranking, which entails a redefinition of the basis of the low energy modes (and, as such, does not affect the alignment score nor the RMSIP), is carried out with the linear optimization procedure introduced in ref. (16).

The links provided at the bottom of the results page allow users to download data files containing all details of the alignment output, namely: the coordinate files of the optimally aligned structures, the list of corresponding amino acids, the corresponding ten lowest energy modes of the aligned amino acids and a VMD (32) state file for a convenient off-line visualization of the results.

RESULTS AND DISCUSSION

We discuss here two test cases in order to illustrate the performance of the ALADYN alignment tool, namely human β-secretase (BACE) versus HIV-1 protease and exonuclease III versus human adenovirus proteinase.

HIV-1 PR and β-secretase. The additional insight offered by the dynamics-based alignment with respect to 'static' alignment approaches is aptly illustrated by the comparison of HIV-1 PR (PDBid: 1aid) and human β-secretase (PDBid: 3hvgA). The two enzymes, which are both aspartic proteases, present major structural differences. In fact, HIV-1 PR is a 198-amino acid long homo dimer, and is almost entirely composed of β sheets. On the contrary, β-secretase is a monomeric enzyme consisting of 379 amino acids and rich in α helices. Despite the differences in symmetry, oligomeric state, length and secondary structure content, the two enzymes share several segments of the primary sequence and are hence believed to be evolutionarily related (33). In fact, they admit a partial, but significant, structural superposition: their DALIlite alignment (34) returns 94 corresponding residues with an associated RMSD of 3.4 Å, while the MISTRAL alignment returns 128 equivalent amino acids at 2.4 Å RMSD. In addition to the partial structural correspondence previous studies, based on atomistic molecular dynamics (MD) simulations, had highlighted the similarity of the low-energy modes of the two molecules (35,36).

The dynamics-based alignment returned by ALADYN is statistically significant, as the associated *P*-value is appreciably smaller than the conventional threshold of 0.05. and is fully consistent with the above-mentioned findings. The alignment consists of more than 140 amino acid pairs at an RMSD <4 Å. The good correspondence of the modes is highlighted by the large RMSIP value of the matching modes, which is ~0.8.

The functional relevance of the alignment returned by ALADYN is underscored by the following facts. First, the returned alignment superposes the catalytic dyads of the
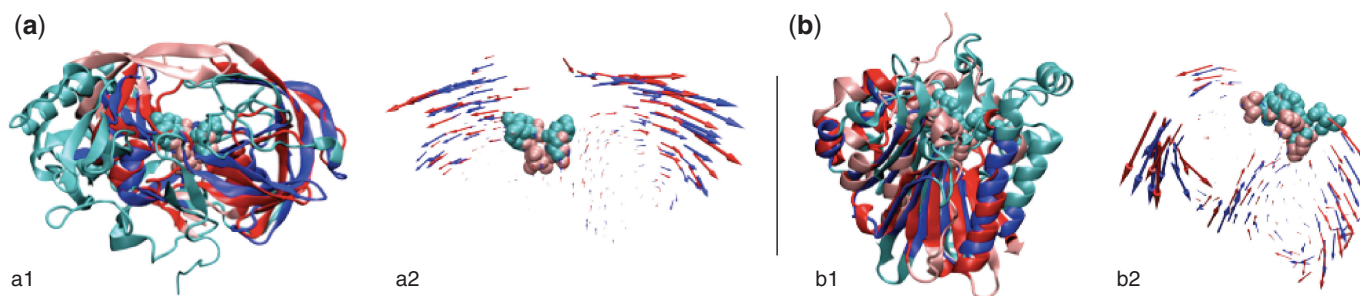
Figure 1. Examples of alignments returned by the ALADYN web server. The structural correspondences and the consistency of the fluctuation dynamics of the aligned regions are shown side-by-side for each of the test cases discussed in the Results section. (**a**) The alignment of HIV-1 protease (pink/red) and β-secretase (cyan/blue) are shown in subpanels a1 and a2. (**b**) The alignment of human adenovirus proteinase (pink/red) and exonuclease III (cyan/blue) are shown in subpanels b1 and b2. Aligned regions are shown with saturated colors (i.e. red and blue), while the active sites are highlighted using a Van der Waals representation.

two enzymes. This is a non-trivial aspect in consideration that no information about the chemical composition (such as the primary sequence) was used. The second observation regards the consensus movements in the two proteins, which entail the modulation of the region accommodating the peptide chain to be cleaved. It is known that in order for the proteolytic reaction to occur, both BACE and HIV-1 PR must 'stretch' the substrate in a β-extended conformation (35,36), and the consensus motion captured by ALADYN (see Figure 1) is consistent with the required deformation (37).

The dynamics-based alignment therefore vividly illustrates the existence of a fundamental similarity underlying the internal dynamics of these enzymes, which is instrumental to produce analogous, functionally oriented deformation patterns in spite of the overall structural differences.

Exonuclease III and human adenovirus proteinase. Exonuclease III (PDB: 1ako) and the human adenovirus proteinase (PDB: 1avp) are not evolutionarily related and are structurally dissimilar at the CATH architecture level. Their structural alignment has a *P*-value >0.1 according to MISTRAL and, similarly, it is ruled out as 'not significant' by DALIlite.

Despite these differences, the enzymes process chemically-similar substrates. In fact, both exonuclease III and human adenovirus proteinase bind DNA (in double- and single-stranded forms, respectively). In the study of Zen *et al.* (21) the dynamics-based alignment of the enzymes was found to have a good statistical significance. As for the case of BACE and HIV-1 PR, the functional relevance of the dynamical correspondence was underscored by the fact that the known active sites of the proteins (38) were spatially superposed by the alignment and by the fact that the consensus motion was compatible with the expected functionally oriented structural changes (39,40).

All the above established results are reproduced by the ALADYN alignment that employs a more general search scheme than the method of Zen *et al.* (21) (on which is conceptually-based). As visible in Figure 1, the two proteins align over more than 90 amino acids, at an RMSD <4 Å. The consistency of the dynamics of the aligned regions is high (RMSIP value ~0.7). It is readily noticed that the alignment yields a good spatial overlap of the active sites of the two enzymes. In accordance with the previous findings (21), the latter are located in a region at the interface between two oppositely moving 'domains'. As suggested for other enzymes (41), this characteristic ought to preserve the catalytic geometry at the active site, while facilitating the accommodation/processing of the substrate.

CONCLUSION

We have presented the ALADYN server that can be used to establish significant pairwise correspondences in proteins based on similarities of their large-scale internal dynamics, which expectedly assists or accompanies their biological functionality. The server is conceptually related to the dynamics-based alignment first introduced in (21). With respect to the original method, the ALADYN alignment scheme is both more general (being non-sequential) and faster and hence lends naturally to be interactively used through a web server.

The input required from the user is kept at a minimum and merely consists of the PDBid's of the proteins to be compared or, alternatively, of their structural coordinates (in PDB format). Alignments of proteins of up to 250 amino acids are typically completed in <1 min. The results are returned through a graphical interface based on the Jmol applet which allows users to interactively visualize the aligned regions and the associated large-scale motion (computed via an elastic network model). The graphical summary is accompanied by quantitative details about the quality and significance of the alignment. Further quantitative data, such as the list of corresponding amino acids and the deformations entailed by the low-energy modes of the matching regions, are provided as downloadable data files.

FUNDING

REFERENCES

1. Lesk,A.M. and Chothia,C. (1980) How different amino acid sequences determine similar protein structures. *J. Mol. Biol.*, **136**, 225–270.
2. Holm,L. and Sander,C. (1996) Mapping the protein universe. *Science*, **273**, 595–603.
3. Levitt,M. and Gerstein,M. (1998) A unified statistical framework for sequence comparison and structure comparison. *Proc. Natl Acad. Sci. USA*, **95**, 5913–5920.
4. Koehl,P. (2001) Protein structure similarities. *Curr Opin Struct Biol*, **11**, 348–353.
5. Gan,H.H., Perlow,R.A., Roy,S., Ko,J., Wu,M., Huang,J., Yan,S., Nicoletta,A., Vafai,J., Sun,D. *et al.* (2002) Analysis of protein sequence/structure similarity relationships. *Biophys J*, **83**, 2781–2791.
6. Chothia,C. (1992) One thousand families for the molecular biologist. *Nature*, **357**, 543–544.
7. Tang,J., James,M.N., Hsu,I.N., Jenkins,J.A. and Blundell,T.L. (1978) Structural evidence for gene duplication in the evolution of the acid proteases. *Nature*, **271**, 618–621.
8. Scheeff,E.D. and Bourne,P.E. (2005) Structural evolution of the protein kinase-like superfamily. *PLoS Comput. Biol.*, **1**, e49.
9. Murzin,A., Brenner,S., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
10. Orengo,C.A., Michie,A.D., Jones,S., Jones,D.T., Swindells,M.B. and Thornton,J.T. (1997) CATH – a hierarchical classification of protein domain structures. *Structure*, **5**, 1093–1109.
11. Xie,L. and Bourne,P.E. (2008) Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc. Natl Acad. Sci. USA*, **105**, 5441–5446.
12. Pang,A., Arinaminpathy,Y., Sansom,M.S.P. and Biggin,P.C. (2005) Comparative molecular dynamics–similar folds and similar motions? *Proteins*, **61**, 809–822.
13. Maguid,S., Fernández-Alberti,S., Parisi,G. and Echave,J. (2006) Evolutionary conservation of protein backbone flexibility. *J. Mol. Evol.*, **63**, 448–457.
14. Henzler-Wildman,K.A., Thai,V., Lei,M., Ott,M., Wolf-Watz,M., Fenn,T., Pozharski,E., Wilson,M.A., Petsko,G.A., Karplus,M. *et al.* (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature*, **450**, 838–844.
15. Lange,O.F., Lakomek,N.A., Farès,C., Schröder,G.F., Walter,K.F., Becker,S., Meiler,J., Grubmüller,H., Griesinger,C. and deGroot,B.L. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, **320**, 1471–1475.
16. Pontiggia,F., Zen,A. and Micheletti,C. (2008) Small- and large-scale conformational changes of adenylate kinase: a molecular dynamics study of the subdomain motion and mechanics. *Biophys. J.*, **95**, 5901–5912.
17. Velázquez-Muriel,J.A., Rueda,M., Cuesta,I., Pascual-Montano,A., Orozco,M. and Carazo,J.M. (2009) Comparison of molecular dynamics and superfamily spaces of protein domain deformation. *BMC Struct. Biol.*, **9**, 6.
18. Keskin,O., Jernigan,R.L. and Bahar,I. (2000) Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.*, **78**, 2093–2106.
19. Carnevale,V., Raugei,S., Micheletti,C. and Carloni,P. (2006) Convergent dynamics in the protease enzymatic superfamily. *J. Am. Chem. Soc.*, **128**, 9766–9772.
20. Capozzi,F., Luchinat,C., Micheletti,C. and Pontiggia,F. (2007) Essential dynamics of helices provide a functional classification of EF-hand proteins. *J. Proteome. Res.*, **6**, 4245–4255.
21. Zen,A., Carnevale,V., Lesk,A.M. and Micheletti,C. (2008) Correspondences between low-energy modes in enzymes: dynamics-based alignment of enzymatic functional families. *Protein Sci.*, **17**, 918–929.
22. Zen,A., deChiara,C., Pastore,A. and Micheletti,C. (2009) Using dynamics-based comparisons to predict nucleic acid binding sites in proteins: an application to OB-fold domains. *Bioinformatics*, **25**, 1876–1883.
23. Hinsen,K., Petrescu,A.J., Dellerue,S., Bellisent-Funel,M.C. and Kneller,G. (2000) Harmonicity in slow protein dynamics. *Chem. Phys.*, **261**, 25–37.
24. Atilgan,A.R., Durell,S.R., Jernigan,R.L., Demirel,M.C., Keskin,O. and Bahar,I. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, **80**, 505–515.
25. Micheletti,C., Carloni,P. and Maritan,A. (2004) Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. *Proteins*, **55**, 635–645.
26. Potestio,R., Caccioli,F. and Vivo,P. (2009) Random matrix approach to collective behavior and bulk universality in protein dynamics. *Phys. Rev. Lett.*, **103**, 268101.
27. Micheletti,C. and Orland,H. (2009) MISTRAL: a tool for energy-based multiple structural alignment of proteins. *Bioinformatics*, **25**, 2663–2669.
28. Shatsky,M., Nussinov,R. and Wolfson,H.J. (2004) A method for simultaneous alignment of multiple protein structures. *Proteins*, **56**, 143–156.
29. Sierk,M.L. and Pearson,W.R. (2004) Sensitivity and selectivity in protein structure comparison. *Protein Sci.*, **13**, 773–785.
30. Amadei,A., Ceruso,M.A. and Di Nola,A. (1999) On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins*, **36**, 419–424.
31. Jmol: an open-source java viewer for chemical structures in 3D. http://www.jmol.org/.
32. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
33. Blundell,T. and Srinivasan,N. (1996) Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. *Proc. Natl Acad. Sci. USA*, **93**, 14243–14248.
34. Holm,L., Kääriäinen,S., Rosenström,P. and Schenkel,A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 2780–2781.
35. Cascella,M., Micheletti,C., Rothlisberger,U. and Carloni,P. (2005) Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *J. Am. Chem. Soc.*, **127**, 3734–3742.
36. Carnevale,V., Raugei,S., Micheletti,C. and Carloni,P. (2007) Large-scale motions and electrostatic properties of furin and HIV-1 protease. *J. Phys. Chem. A*, **111**, 12327–12332.
37. Tyndall,J.D., Nall,T. and Fairlie,D.P. (2005) Proteases universally recognize beta strands in their active sites. *Chem. Rev.*, **105**, 973–999.
38. Porter,C.T., Bartlett,G.J. and Thornton,J.M. (2004) The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic. Acids Res.*, **32**, D129–D133.
39. Mol,C.D., Kuo,C.F., Thayer,M.M., Cunningham,R.P. and Tainer,J.A. (1995) Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature*, **374**, 381–386.
40. Gupta,S., Mangel,W.F., McGrath,W.J., Perek,J.L., Lee,D.W., Takamoto,K. and Chance,M.R. (2004) DNA binding provides a molecular strap activating the adenovirus proteinase. *Mol. Cell Proteomics*, **3**, 950–959.
41. Potestio,R., Pontiggia,F. and Micheletti,C. (2009) Coarse-grained description of protein internal dynamics: an optimal strategy for decomposing proteins in rigid subunits. *Biophys. J.*, **96**, 4993–5002.