ELSEVIER

Biophysical Chemistry 73 (1998) 1–5

Biophysical Chemistry

Letter

Internal motions of native lysozyme are more organized than those of mutants: a principal component analysis of molecular dynamics data

Reino Laatikainen*, Janne Saarela, Kari Tuppurainen, Tommi Hassinen

University of Kuopio, Department of Chemistry, P.O.B. 1627, 70211, Kuopio, Finland

Received 10 February 1998; accepted 2 March 1998

Abstract

Principal component analysis (PCA) of molecular dynamics simulations of hen egg white lysozyme and its mutants indicate that even small changes in the amino acid sequence alter considerably the internal molecular motions and that the internal motions are more organized in the native enzyme than in the mutants.

Keywords: Lysozyme; Principal component analysis; Internal motions; Molecular dynamics

1. Introduction

Proteins are often thought of as mechanical devices [1,2]. This means that in addition to exact key-to-lock action, the motional freedoms are also important. These atomic fluctuations may be used by proteins to maintain the catalysis. This property has been called the time dimension of protein structure [3]. A deep understanding of the role of internal motions would be essential to understand the mode of action of the biomolecules and would improve the success of the modification of the structures. It would also have an impact on the design of fully synthetic active enzyme-like molecules.

If the internal motions have a role in the action of a biomolecule, it should be reflected in the evolution of the molecular structures [4]. This leads to a hypothesis that the molecular motions of the native structure should be more organized than those of its mutants. The hypothesis can be tested by using the principal component analysis (PCA) of molecular dynamics data [5,6]. The technique, often called the essential dynamics method

* Corresponding author. Tel.: +358 17 163248; fax: +358 17 163259; e-mail: reino.laatikainen@uku.fi

0301-4622/98/$19.00
PII S0301-4622(98)00141-0

[6] is well-documented and has been applied to a number of systems [6–9], including lysozyme [6].

In this work we report PCA for molecular dynamics data of HEWL (Hen Egg White Lysozyme), which is a well-known example of proteins for the action of which the molecular motions are important [10,11], especially, its low frequency motions. In order to avoid problems arising from large mutation induced structural changes, we studied some randomly chosen Ser → Ala and Ala → Ser mutants; the other mutants were those for which the activity information was available. All the calculations were done in vacuo; this point is to be discussed below.

2. Methods

The modelling was carried out using the Quanta versions 3.3 and 4.1 (Molecular Simulations Inc.) and 4.1/CHARMm versions 22 and 23rl [21] on IBM RISC/6000 work stations. The crystallographical coordinates (entry 1lzt, triclinic form [22] of native HEWL (E.C.3.2.1.17) were retrieved from the Brookhaven Protein Data Bank. The crystallographical water molecules were removed, hydrogens added and incorrect bonds removed.

The structure was minimized by the steepest descent and adopted-basis Newton Raphson method using 4000 steps. The non-bonded cutoff value was 15 Å, the non-bonded energies and forces were smoothly truncated using a van der Waals switching function and an electrostatic switching function. A time step value of 1 fs was used in all the MD runs, which were done with SHAKE [23] on using defaults and updating the non-bonded energies at every 20 steps. The MD runs for producing the data consisted of three phases: (1) a 10-ps heating (from 0 to 310 K); (2) a 20-ps equilibration; and (3) the simulation (2–100 ps at 310 K). The output frequency was modified to produce 1000 or 2000 data points for the analysis. The drift values were an order of 10^{-5}. The mutations were done by Protein Design procedure of CHARMm. The mutants were minimized 1000 steps by the adopted-basis Newton Raphson method — the simulations were done as above. The Cartesian coordinates were obtained by CHARMm from the dynamic trajectories. The PCA was performed using the SPSS program package (version 6.01). The eigenvalues and vectors of the 387×387 matrices were computed with an NIPALS algorithm [24].

3. Results

HEWL is divided into two domains by a deep cleft [12,13] which has been identified as the active site. The cleft consists of amino acids participating in three different types of interactions: hydrogen bonding (Asp-101, Trp-62, Trp-63), van der Waals interactions (Trp-57) and actual catalytic active sites (Asp-52, Glu-35). In addition, Phe-3, Gly-4, Gly-16, Gly-22, Ser-24, Cys-30, Ser-36 and Gly-117 clustered around the active site amino acids [13]. It seems reasonable to suggest that this box provides optimum conditions for transfer of mechanistic energy and that the motions of the corresponding amino acids are probably the most important. Accordingly, PCA was carried out for the Cartesian coordinates of the C_α carbons of the whole enzyme (3×129 variables) and the above box (3×14 variables).

If P is the Cartesian coordinate matrix (P_{ij} is the deviation of the jth coordinate from its average value at time i) obtained by MD calculation, the diagonalization of the correlation matrix C ($C_{ij} = D_{ij}/(D_{ii}\ D_{jj})^{1/2}$ where D_{ij} is an element of matrix $P^{\mathrm{T}}\mathrm{P}$ and P^{T} stands for the transpose of the P matrix [14]) yields principal components (PCs) which define orthogonal sets of dynamic states, so that any MD data point can be expressed as their linear combination. The eigenvalue of a 'state' gives the amplitude of the total variance explained by the state. The PC can be also understood as a motion: a large value of the loading means a large contribution of the coordinate to the motion represented by the PC. If only a few PCs are needed to explain, say, 90% of the total variance, the motions corresponding to the variations of the coordinates are highly correlated. Instead, if many PCs are needed, the variations of the coordinates do not correlate and the motions are more random. If the motions are random, every PC equals to $100/N\%$, where N is the number of variables.

To study the convergence speed of the eigenvalue profile we performed the MD simulation of

Table 1
The distribution of the five largest eigenvalues (EV1–EV5) and the corresponding cumulative percentages (i.e. the variances explained by the first five principal components) for native lysozyme and its mutants. Numbers in parentheses give the corresponding numbers of the active site and its neighborhood

	EV1	EV2	EV3	EV4	EV5	%
Lysozyme (100 ps)	79.6 (11.2)	26.9 (3.76)	14.5 (2.39)	13.2 (1.96)	9.2 (1.74)	35.9 (49.3)
Lysozyme (10 ps)	77.5 (11.4)	26.3 (3.73)	20.4 (2.85)	13.2 (1.96)	11.7 (1.90)	38.5 (52.0)
Lysozyme (2 ps)	57.5 (7.39)	48.0 (5.79)	36.5 (4.14)	33.5 (3.29)	26.4 (2.99)	52.1 (56.2)
Ser-24[a] → Ala	27.8 (4.55)	20.1 (3.41)	16.3 (2.86)	14.8 (2.47)	13.7 (2.25)	24.0 (37.0)
Glu-35 → Ala[b]	62.0 (5.88)	22.9 (3.93)	17.2 (2.90)	15.6 (2.30)	12.4 (2.07)	33.6 (40.7)
Asn-37 → Glu	23.8 (3.88)	19.1 (3.47)	16.8 (2.70)	14.4 (2.44)	12.4 (2.21)	22.4 (35.0)
Trp-108 → Tyr[c]	50.8 (5.61)	25.3 (3.95)	16.8 (2.80)	12.9 (2.39)	11.6 (2.04)	30.3 (40.0)
Asn-37 → Gly[d]	55.1 (8.16)	23.5 (3.36)	14.5 (2.68)	12.4 (2.51)	11.5 (2.10)	30.2 (44.8)
Asp-52 → Asn[e]	33.5 (4.09)	21.9 (3.46)	17.7 (3.24)	14.1 (2.50)	12.4 (2.34)	25.7 (37.2)
Ser-36 → Ala	26.1 (4.05)	22.2 (3.23)	14.9 (2.64)	13.5 (2.38)	13.0 (2.13)	23.2 (34.4)
Ser-50 → Ala	40.1 (5.58)	18.9 (3.52)	17.0 (2.66)	14.1 (2.38)	12.9 (2.21)	26.6 (38.9)
Ser-60 → Ala	23.4 (3.44)	16.0 (2.85)	15.3 (2.63)	12.3 (2.57)	12.3 (2.38)	20.5 (33.0)
Ser-72 → Ala	29.7 (3.18)	24.8 (3.01)	14.7 (2.72)	13.7 (2.58)	11.9 (2.23)	24.5 (32.7)
Ser-81 → Ala	30.7 (4.33)	21.4 (3.20)	15.5 (2.93)	14.8 (2.62)	13.8 (2.28)	24.8 (36.6)
Ser-85 → Ala	23.4 (3.89)	19.1 (3.34)	16.3 (2.83)	13.6 (2.36)	11.0 (2.20)	21.5 (34.8)
Ser-86 → Ala	29.5 (5.16)	21.2 (2.78)	16.8 (2.51)	13.6 (2.30)	11.9 (1.96)	24.0 (35.0)
Ala-82 → Ser	35.1 (4.45)	23.2 (3.65)	16.2 (2.71)	13.1 (2.38)	12.7 (2.20)	26.0 (36.7)
Ala-42 → Ser	38.9 (5.35)	20.3 (3.18)	19.0 (2.90)	15.9 (2.63)	12.3 (2.54)	27.5 (39.5)
Ala-90 → Ser	56.0 (7.09)	26.9 (3.84)	17.3 (2.77)	13.3 (2.43)	12.6 (2.22)	32.6 (43.7)
Ala-31 → Ser	91.5 (9.39)	28.2 (3.63)	18.0 (2.18)	12.7 (2.17)	12.4 (2.05)	42.1 (46.3)
37,62,101[f]	82.3 (9.66)	29.9 (3.24)	17.9 (2.99)	15.3 (2.34)	13.0 (2.05)	40.9 (48.3)

[a] The residues underlined belong to the active region.
[b] Not active [25].
[c] Activity 17% of that of the native [26].
[d] Activity the same as that of the native [16].
[e] Activity 5% of that of the native form [25].
[f] Asn-37–Trp-62–Asn-101 → Gly–Tyr–Gly; activity 300% of that of native [16].

native lysozyme for 2, 10 and 100 ps. The results are shown in Table 1: the values of the first few eigenvalues converge properly in 10 ps, a full convergence of all the eigenvalues would demand a much longer time [6,15]. The eigenvalue profiles for the native lysozyme and a few mutants are shown in Fig. 1.

The eigenvalues (Table 1), especially for the active box, show that the motions of the native lysozyme are more organized than most of the mutants. For the whole enzyme, only two mutants (Ala-31 → Ser and Asn-37–Trp-62–Asp-101 → Gly–Tyr–Gly) with a higher level of organization were found; the activity of the former is not known, the activity of the latter is three times higher than that of the native [16]. Also the eigenvalues of the active Asn-37 → Gly mutant [16] are clearly higher than the average values. However, the results indicate that even small changes in the amino acid sequence may alter significantly the eigenvalue profile of the MD-data and accordingly, if the internal motions are important for the activity of the molecule, it can be changed considerably. In general, it seems that the native structure has the highest degree of organization.

4. Discussion

Although there is rather little systematic data available about the effects of the mutations that are out of the vicinity of the active site, it is known that mutations around His-15 of human lysozyme, far away from the active site stimulate activity; the effect accounted for some sort of fluctuations [11]. It has also been proposed that

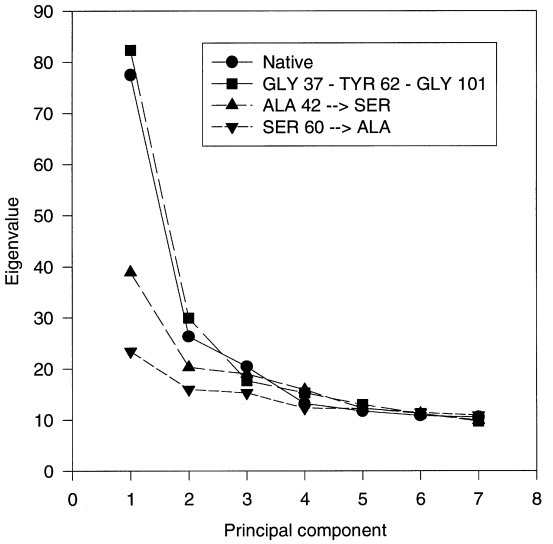

Fig. 1. The eigenvalue profiles of native and mutant lysozymes for a simulation time of 10 ps. ● = Native, ■ = the active mutant Asn-37–Trp-62–Asp-101 → Gly–Tyr–Gly, ▲ = Ala-42 → Ser representing an average non-active or unknown activity mutant, ▼ = Ser-60 → Ala the mutant with the lowest eigenvalue profile.

mutations in certain domain interfaces decrease the activity and that the exposed loops are tolerant to them [17]. However, from the experimental data it is difficult to decide whether the changes in the activity arise from changes in the internal motions or from changes in the structure or in the steric and electrostatic interactions.

It should be emphasized that all the above calculations were performed in vacuo. Although the solvent certainly plays an important role in the motions of the molecule [18], it is obvious that the most significant factors are the steric strain of the molecular backbone, the van der Waals interactions in the tertiary structure and the hydrogen bonds. We also found it too time consuming to perform sufficiently long MD simulations in solution for all the cases and, moreover, in the cases tested, the PC-statistics did not converge properly. Previous calculations show that the atomic displacement profiles of lysozyme in solution and in vacuo are topologically rather similar [18]. Furthermore, the solvent damps the Coulombic intermolecular interactions and, thus, the in vacuo calculations in fact may enhance the interactions and their effects to the internal motions. Therefore, for the above reasons, the results should be considered with due caution.

As to the mechanism of how the mutations affect the internal motions, most of the present mutations affect mostly Coulombic interactions: serine and alanine have very similar steric requirements and, thus, the mechanism of the effects of the above mutations is not based so much on steric interactions as the change of electrostatic interactions.

The loadings of the coordinates represent the importance of the coordinates to the motion related to a certain PC and, thus, to the function of the motion. The low amplitude opening motion of lysozyme is assumed to be connected to its catalytic activity [10,11]. However, when the loading profiles of the first PCAs of the mutants and the native molecule are compared, very little similarity is seen in the profiles. This means that the opening motion of the cleft may take place in many possible ways or that the loading profiles converge very slowly. The latter is supported by a study where it was found that the B-factors computed by MD dynamics converge slowly [15]. It has also been found that the loading profiles of four lysozyme molecules in the same unit cell do not converge well even during a nanosecond MD run [19].

5. Conclusion

In conclusion, the above analysis leads to a hypothesis analogous to the ‘thermodynamic hypothesis’ [20]: the motions of the native lysozyme are more organized than those of mutants. The generalization of the result to other enzymes, proteins and biomolecules would be of great interest and, depending on the generality of the rule, have considerable consequences on the structural chemistry of biomolecules. Even though the rule may not be general and the activity and the motions not generally related, the above results show that the internal motions of protein can be sensitive to minor changes in the structure and that the changes can be rather easily explored by the essential motion analysis.

Acknowledgements

This work was supported by the Academy of Finland.

References

[1] R.J.P. Williams, TIBS 18 (1993) 115.
[2] M. Kurzynski, Biophys. Chem. 65 (1997) 1.
[3] S.W. Englander, N.R. Kallenbach, Q. Rev. Biophys. 16 (1984) 521.
[4] M. Karplus, G.A. Petsko, Nature 347 (1990) 631.
[5] A.E. Garcia, Phys. Rev. Lett. 68 (1992) 2696.
[6] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Proteins Struct. Funct. Genet. 17 (1993) 412.
[7] D.M.F. van Aalten, A. Amadei, A.B.M. Linssen, V.G.H. Eijsink, G. Vriend, H.J.C. Berendsen, Proteins Struct. Funct. Genet. 22 (1993) 45.
[8] D.M.F. van Aalten, J.B.C. Findlay, A. Amadei, H.J.C. Berendsen, Protein Eng. 8 (1995) 1129.
[9] D.M.F. van Aalten, A. Amadei, J.B.C. Findlay, H.J.C. Berendsen, C. Sander, P.G.W. Stouten, Biophys. J. 70 (1996) 684.
[10] J. Emsley, New Scientist 16 (1987) 35.
[11] T. Imoto, T. Ueda, T. Tamura, Y. Isakari, Y. Abe, M. Inoue, T. Miki, K. Kawano, M. Yamada, Protein Eng. 7 (1994) 743.
[12] L.J. Smith, M.J. Sutcliffe, C. Redfield, C.M. Dobson, J. Mol. Biol. 229 (1993) 930.
[13] I. Cosic, The resonant recognition model of macromolecular bioactivity, theory and applications. In: T. Meir, H.P. Saluz (Eds.), BioMethods, Vol. 8, Birkhäuser Verlag, Basel, 1997, p. 54.
[14] E.R. Malinowski, Factor Analysis in Chemistry, 2nd Ed., John Wiley and Sons, Inc. 1991, p. 40.
[15] P.H. Hunenberger, A.E. Mark, W.F. van Gunsteren, J. Mol. Biol. 252 (1995) 492.
[16] I. Kumagai, F. Sunada, S. Takeda, K. Miura, J. Biol. Chem. 267 (1992) 4608.
[17] L. Holm, A.K. Koivula, P.M. Lehtovaara, A. Hemminki, J.K.C. Knowles, Protein Eng. 3 (1990) 181.
[18] C.L. Brooks, M. Karplus, J. Mol. Biol. 208 (1989) 159.
[19] S. Hery, D. Genest, J.C. Smith, J. Chem. Inf. Comput. Sci. 37 (1997) 1011.
[20] C.B. Anfinsen, Science 181 (1973) 223.
[21] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, F.J. States, S. Saminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187.
[22] J.M. Hodsdon, G.M. Brown, L.C. Sieker, Acta Crystallogr. B46 (1990) 54.
[23] J.P. Ryckaert, G. Ciccotti, H.J. Berendsen, J. Comput. Phys. 107 (1977) 327.
[24] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 1.
[25] B.A. Malcom, S. Rosenberg, M.J. Corey, J.S. Allen, A. de Baetselier, J.F. Kirsch, Proc. Natl. Acad. Sci. USA 86 (1989) 133.
[26] M. Inoue, H. Yamada, Y. Yasukochi, R. Kuroki, T. Miki, T. Horiuchi, T. Imoto, Biochemistry 31 (1992) 5545.