Journal of Computer-Aided Molecular Design, **16:** 755–765, 2002.
KLUWER/ESCOM

Analysis methods for identifying coordinated movements during ligand unbinding

P.-L. Chau[1,*] & P.W.A. Howe[2]
[1]*Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom;* [2]*Syngenta, Jealott's Hill, Bracknell RG42 6EY, United Kingdom*

Received 1 July 2002; accepted in final form 1 November 2002

Key words: essential dynamics, ED, projection to latent structures, PLS, molecular dynamics simulations, ligand-receptor interactions, molecular recognition

Summary

Molecular dynamics simulations have been applied to unbind biological ligands from their receptors [1–5]. Conformation changes are observed in the biomolecules during unbinding, but there exists no systematic method to detect these conformation changes. In this work, we have used 'essential dynamics' (ED) [6–7] and projection to latent structures (PLS) [8] to investigate the conformation changes of the bovine serum retinol-binding protein when retinol unbinds from its receptor site. The results of these analyses characterise a large proportion of the movements that occur during unbinding. We find that the loop regions of retinol-binding protein exhibit the largest movements during unbinding. The sudden changes in unbinding speed during the unbinding process appear not to be caused by sudden changes in protein structure.

Introduction

Ligand-receptor interaction is the first step to many basic processes of life, e.g., enzyme catalysis, neurotransmitter and hormone action and antibody-antigen recognition. In order to elucidate the detailed mechanisms of ligand-receptor interactions, experiments such as crystallography, nuclear magnetic resonance and electron microscopy are invaluable. Nevertheless, they only provide limited information about the dynamics of the ligand-receptor complex and the solvent molecules, which are crucial to understanding the binding process. Furthermore, it is difficult to isolate the factors that contribute to binding, and quantitate their relative importances. Simulations complement theory and experiments by making it possible to study each factor in depth and to obtain structural and dynamical details simultaneously at picosecond resolutions.

*To whom correspondence should be addressed: Bioinformatique Structurale, Institut Pasteur, 75724 Paris, France. E-mail: pc104@pasteur.fr

One of the early simulation studies of the mechanism of ligand-receptor interaction was by Kern *et al.*, who performed a molecular dynamics (MD) simulation of 300 ps on adenylate kinase complexed to its transition-state inhibitor in water [9]. They identified secondary structure transitions and domain closures, and thus the induced fit movement of the enzyme. They also concluded that reliable results were achievable only if water was explicitly included in the simulation.

Rognan *et al.* performed MD simulations on a different ligand-receptor complex system, the Class I MHC HLA-B 2075 protein bound to three different artificial peptides and two natural peptides in explicit water [10]. The trajectory lasted 150 ps, and they found a higher stability of the MHC-ligand complexes with the artificial peptides than with the natural peptides. These results agreed with the experimental results of a semi-quantitative assay, which showed that the artificial ligands bind with an affinity constant of 0.5 μM, while the natural homologues bind with an affinity constant of 40 μM. This study demonstrates

that MD simulations can be used to study ligand-receptor interactions on a semi-quantitative basis.

Subsequently, steered molecular dynamics was developed to simulate ligand unbinding from the receptor [11], and this method has been applied to various ligand-receptor systems successfully [1–4]. Steered molecular dynamics requires the pre-determination of the unbinding trajectory. Previous work used molecular complexes where the unbinding trajectory is determined from experiments, or where various putative unbinding trajectories are used for the simulation. To overcome this problem, one of us developed the mutual repulsion method [5] to allow the ligand to explore its own unbinding trajectory.

In all these simulations, conformation changes are observed in the biomolecules during unbinding, but these exists no systematic method to detect these conformation changes. One has to know in advance where the major changes will be, or one has to go through a lot of data to locate the major changes. In this work, we have used 'essential dynamics' (ED) [6–7] and projection to latent structures (PLS; also known as partial least squares) [8] to investigate the conformation changes of the bovine plasma retinol-binding protein when retinol unbinds from its receptor site. This is the first study of the conformational changes involved *during* the process of ligand unbinding. We have compared the results given by ED and PLS, and we conclude that PLS can be used as a general method for identifying the major conformational changes involved in ligand unbinding.

Methods

Molecular dynamics simulation

Atomic coordinates of the holo bovine serum retinol-binding protein were obtained from the Brookhaven Protein Databank (PDB code: 1HBP) [12]. The model used for the protein and retinol is the GROMOS87 potential [13]. With conservation of crystallographically determined water molecules, the protein was solvated in truncated octahedral boxes filled with 5537 SPC/E [14] water molecules. Four sodium ions were added to ensure total electronic neutrality.

All minimisations and simulations were performed using version 2 of DL_POLY [15]. The solvated structure was minimised, and simulations were started with an NVT run at 30 K, taking initial velocities from a Maxwellian distribution. The temperature was gradually increased to 310 K over a period of 100 ps, using a Nosé-Hoover thermostat [16–17] with a time constant of 0.1 ps. SHAKE was used to constrain bond lengths [18]. The time step was 1 fs. A cutoff of 10 Å was used with simple truncation. The simulations were then continued for 400 ps in an NPT ensemble using the Nosé-Hoover thermostat and the Melchionna barostat [19]. The last 200 ps of this NPT simulation was used to determine the optimal box size for a pressure of 101325 Pa (1 atm). This box size was then used for a subsequent 100-ps NVT simulation to equilibrate the system, followed by a mutual repulsion [5] run of 850 ps. Configurational data were output every 20 fs during these 850 ps, but in the essential dynamics analysis, only one configuration per 10 ps was used.

Essential dynamics

Essential dynamics (ED) is the term given by Amadei *et al.* [7] to the method they developed for identifying large-scale movements in structures produced by molecular dynamics simulations. The same method was independently suggested by García [6].

ED is the application of Principal Components Analysis (PCA) or Single Value Decomposition to the Cα co-ordinates of a set of structures. The result of these calculations are eigenvectors, which represent movements of Cα co-ordinates that are correlated with the movements of other Cα co-ordinates. All eigenvectors are orthogonal with one another, so that the movements they represent are independent of (i.e., uncorrelated with) one another.

If we assume that $\mathbf{X}$ is the matrix of observations, $\mathbf{T}$ is the matrix of projections onto the eigenvectors (or scores), $\mathbf{P}$ is the matrix of eigenvectors (or principal components) and $\mathbf{E}$ is the error matrix, then PCA assumes a data model where [20]:

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{1}$$

where the symbol $'$ denotes the transpose. In expanded notation, for a protein with $m/3$ residues, and where n configurations are under study, the equation is:

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} (p_1\ p_2\ \cdots\ p_m) + \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ e_{21} & e_{22} & \cdots & e_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nm} \end{pmatrix}$$

In PCA, the values of **P** are those that are orthogonal to one another (uncorrelated) and that minimise the error terms. The goodness of fit to a PCA model can be assessed from the error matrix **E**. If the eigenvectors being considered explain the variation in the population, then the elements of **E** will be near zero. As other sources of variation become more significant, the elements of **E** will increase.

ED has been applied to molecular dynamics simulations of a range of proteins including *ras* p21 [21] and bovine serum retinol-binding protein itself [22]. It has also been used to compare the results of NMR structure determination with simulations [23] and to identify correlated differences between individual structures within NMR structure ensembles [24]. One of us has recently suggested using an extension of essential dynamics for use in selecting converged structures when determining structures by NMR [25].

Projection to latent structures (PLS)

The model underlying PCA assumes a population where there is correlated variation among variables that all follow the same distribution (usually assumed to be the normal distribution). This makes it ideal for checking that a population is homogeneous or for identifying correlated variations. Structures produced by molecular dynamics usually conform to both of the assumptions of PCA. The data presented here, however, would not necessarily be expected to contain variables with the same distribution; they are a series of structures over an unbinding trajectory, so co-ordinates taking part in key conformational changes are likely to have different distributions from others. This difference would not be important if the deviation from normality were small, but there is an additional problem in using PCA with the structures presented here. This problem is that the changes of interest are those that are related to the movement of the ligand. With PCA, there is the risk that changes that are correlated with ligand unbinding will be much smaller or less highly correlated than other changes in the structure. If this is the case, then such changes will not be highlighted by PCA.

Identifying movements that are related to ligand unbinding requires an analysis method which can relate changes in a multivariate data set (the co-ordinates of the protein structure) to an external variable (the ligand unbinding). One method that is suitable for such an analysis is PLS, which was developed almost 30 years ago [8] to deal with similar situations that commonly arise in chemometrics.

PLS is closely related to PCA, but considers the presence of one or more external variables, usually referred to as the response matrix (**Y**). Values of **P** are calculated to be orthogonal and to minimise the error terms for both the **X** and **Y** matrices:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \tag{2}$$

$$\mathbf{Y} = \mathbf{TC}' + \mathbf{F} \tag{3}$$

where **F** is the error matrix for **Y** and **C** are the weightings for the **Y** variables. The PLS factors (**P**) are analogous to the eigenvectors in PCA, and the PLS scores (**T**) are analogous to the projections. In the remainder of this paper, the components of **P** will be referred to as loadings, complete rows of **P** will be referred to as factors and the components of **T** will be referred to as scores.

For a protein with $m/3$ residues, where only one factor is being considered, and where n configurations are under study, the equations are equivalent to:

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} (p_1\ p_2\ \cdots\ p_m) + \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ e_{21} & e_{22} & \cdots & e_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nm} \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} . \ c_1 + \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}$$

These can be extended to more than one factors by expanding **T**, **C** and **P**.

In this case, PLS was applied using the Cα co-ordinates as the **X** matrix and the centre-of-mass separation between retinol and the protein as a single **Y** variable. This identifies correlated movements of the protein which are related to movement of the ligand.

Assessing PLS and PCA models of protein structures

It is very important to assess the differences between the observed data and the models produced by PLS and PCA. This is because PLS and PCA both assume correlated variation across a whole population. Movements which only occur in a few structures, or which

are uncorrelated, will not be highlighted by PLS or PCA.

The difference between the movements which occur, and those which are captured by the model, can be assessed by analysing the error matrix **E**. If the model were perfect and the only movements occuring were the factors, then there would be no difference between the observed co-ordinates and those produced by multiplying the factors by the scores. All the elements of **E** would be zero. In reality, however, some correction to the model will be needed to produce the original data, so the elements of **E** are non-zero.

E can assessed in two ways. Ths first is to sum the squares of each element in a row. This produces one value for each structure in the population, referred to as the Sample Residual. If a structure exhibits movements which are not captured by the factors, then there will be a difference between its co-ordinates and those predicted from the factors, and it will have a high Sample Residual.

The second way of assessing **E** is to sum the squares of each element in a column. This produces a Variable Residual for each co-ordinate. If the changes in a co-ordinate are not captured by the factors being considered, then it will have a high Variable Residual.

Essential dynamics and PLS analysis

86 structures (one every 10 ps) were analysed by essential dynamics. The structures were superimposed onto the starting structure as described above to ensure that all structures were in the same coordinate frame. The Cα co-ordinates were extracted from the co-ordinate files and were then referenced to the starting structure by subtracting the co-ordinates of the starting structure from all the remaining ones.

PCA and PLS were performed using the program Pirouette 2.0 (Infometrix Inc., Seattle, Washington, USA). Ten factors were calculated leaving out 10% of the data for cross-validation. The significance of PCA and PLS factors was assessed using a hybrid of the IND function and an F-test of the remaining factors [26] as implemented within Pirouette.

Visualisation of eigenvectors

Displaying the loadings (**T**) calculated by PLS and PCA can be difficult because of the large number of variables involved. In the case of essential dynamics, the loadings are usually visualised by projecting them back onto the protein structure. This reduces the number of variables that have to be visualised three-fold because each Cα atom then represents three variables. Once projected onto the protein structure, each factor represents a straight-line vector in 3D space for each Cα atom. This is because all the loadings within a factor are multiplied by the same score for each structure when PCA or PLS reconstruction is calculated.

This paper presents PLS and PCA loadings by projection of the maximum and minimum of each factor onto the protein structure, i.e., for the i-th factor, the minimum and maximum shown on the graph are:

$$x(i)_{min} = x(t = 0) + t(min)p_i$$

$$x(i)_{max} = x(t = 0) + t(max)p_i$$

where $x(t = 0)$ is the starting co-ordinate of one atom, $p(i)$ is the loading of that co-ordinate, and where t_{max} and t_{min} are, respectively, the maximum and minimum scores for the factor under consideration across the whole population.

Results

Details of the simulation has been previously reported [5]; here we briefly describe the results and concentrate on the analysis of the configurations generated during the unbinding process using ED.

Unbinding process

During the molecular dynamics simulation, mutual repulsion was applied to induce the ligand to unbind from the receptor protein from 50 ps. Figure 1 shows the evolution of the distance between the centres of mass of the ligand and the protein (top panel), the radius of gyration of the protein (middle panel), and the root-mean-square deviation (RMSD) of the protein to the starting protein structure (lower panel) over the whole 850-ps unbinding simulation. Overall, the distance between the protein and ligand increases gradually during the separation, showing that the ligand disengages from the protein and moves away relatively smoothly. However, at 370 ps, 440 ps and 760 ps there are three distinct steps where the distance increases more rapidly while the unbinding force decreases. These two characteristics suggest that at these times, changes occur to reduce the favourable interactions that bind the ligand to the protein and so allow the protein and the ligand to separate more rapidly.

The plot of root-mean-square deviation (RMSD) over the simulation (Figure 1, lower panel) shows that

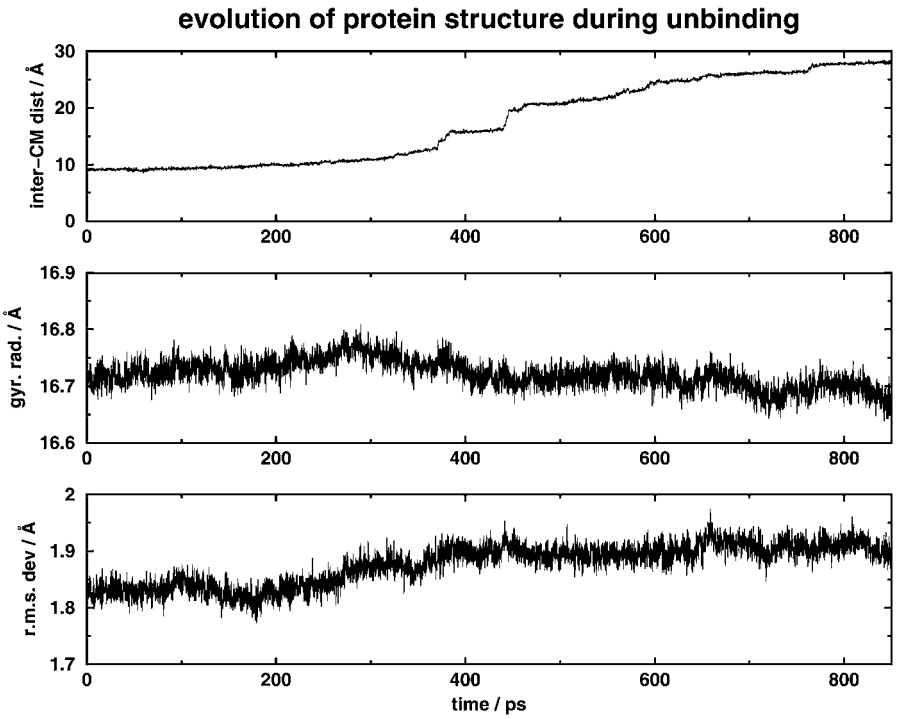

Figure 1. The evolution of the distance between the centres of mass of retinol and of the retinol-binding protein (top panel), the radius of gyration (middle panel), and the root-mean-square deviation of the protein to the starting protein structure (lower panel) over the whole 850-ps unbinding simulation.

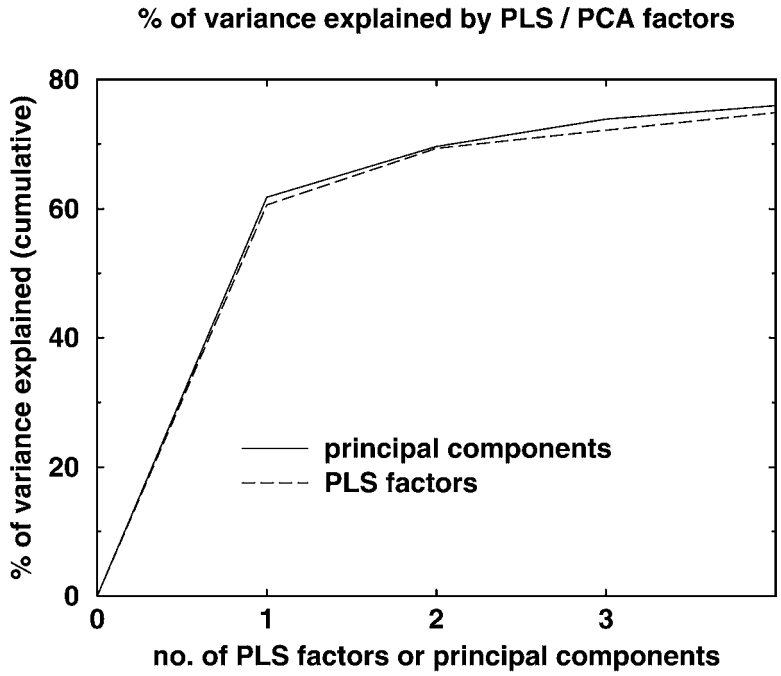

Figure 2. The proportion of variance (sum-of-squares) explained by the PLS factors and Principal Components.

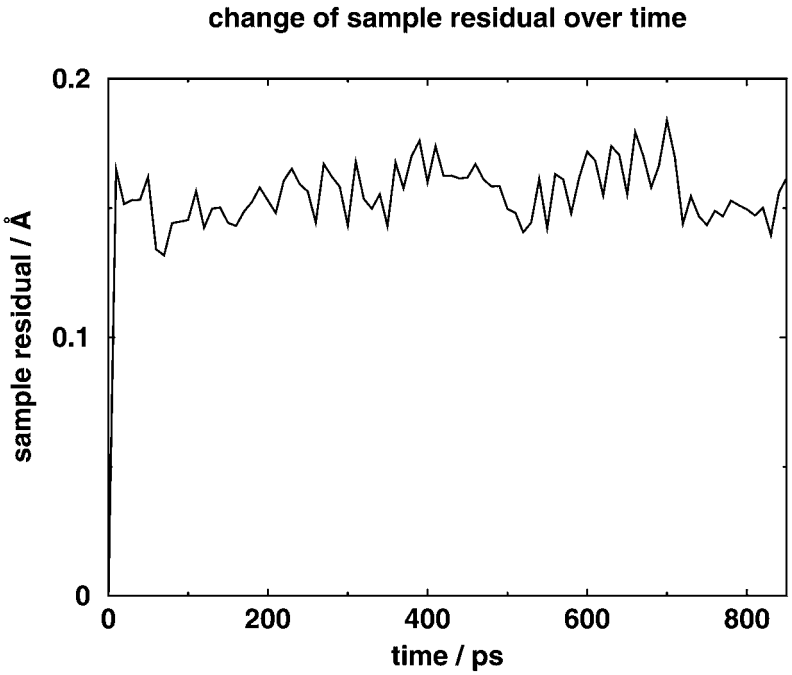

Figure 3. The distribution of sample residuals for the PCA model during the unbinding simulation.

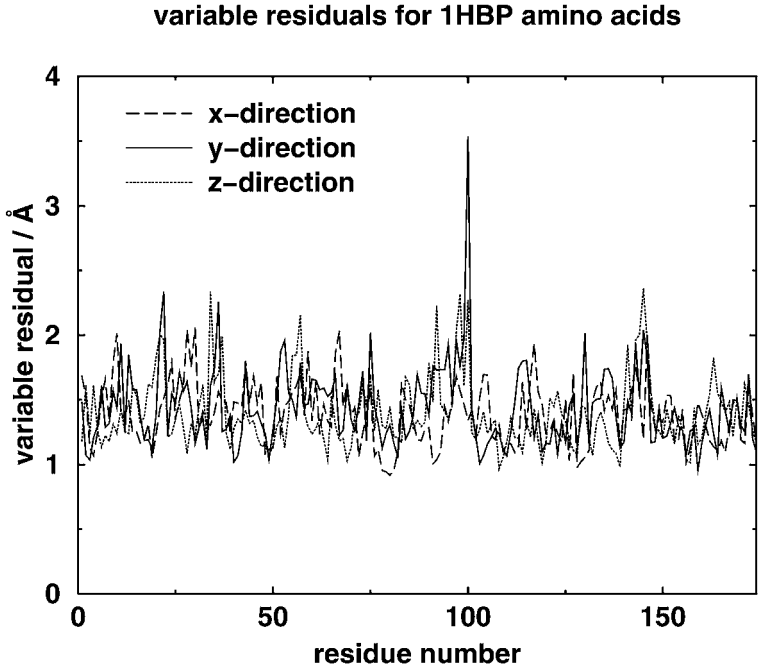

Figure 4. Variable residual of the amino acids, in all three directions, averaged over the whole simulation.

there are no large-scale changes in protein structure during the simulation. Although this might perhaps be expected from the crystal structures of apo- and holo-sRBP, which are very similar (0.31 Å Cα RMSD) [12], it is still surprising that the ligand is able to exit from a binding cleft deep within the protein without disrupting the fold. It should also be noted that no steps in RMSD occur at the three distinct unbinding steps mentioned above, indicating there are no large-scale changes of protein structure at these points.

Validation of model

Since it is an average over the whole structure, RMSD is a good indication of overall fold, but there is the risk that it may conceal interesting changes within particular regions of the protein. For this reason, analysis of the RMSD was supplemented by Essential Dynamics analysis using PCA and PLS. Figure 2 shows the proportion of variance (sum-of-squares) explained by the components of these models. These make it clear that many of the movements in the protein backbone are correlated. The first factor (eigenvector) accounts for approximately 60% of the structural variation in the protein over the course of the simulation. The remaining factors each account for less than 10% of the variation, and cross-validation suggests that only 4 PCA factors and 3 PLS factors are significant.

When interpreting the results of PLS and PCA, it is important to consider the structural variation that the methods do not model. This is because PLS and PCA identify variation in a population as a whole, so may ignore movements which only occur in a small number of structures within the population. The extent of this problem can be evaluated from the Sample Residual, as described in Methods. The Sample Residual for the structures is plotted in Figure 3. This shows that the amount of movement not explained by the model is about the same for all the structures. Therefore, no structure exhibits movements which are much more poorly modelled than any of the others.

A supplementary check is to examine the Variable Residual, which is the change in each co-ordinate that is not explained by the model summed across the population of structures. Figure 4 shows that one co-ordinate (from residue Gly 100) does have a considerably larger residual than others. This co-ordinate is the one which changes most during the simulation so the PLS model still explains about 32% of its total variance. However, the large residual movement of this residue should be investigated further. The distri-

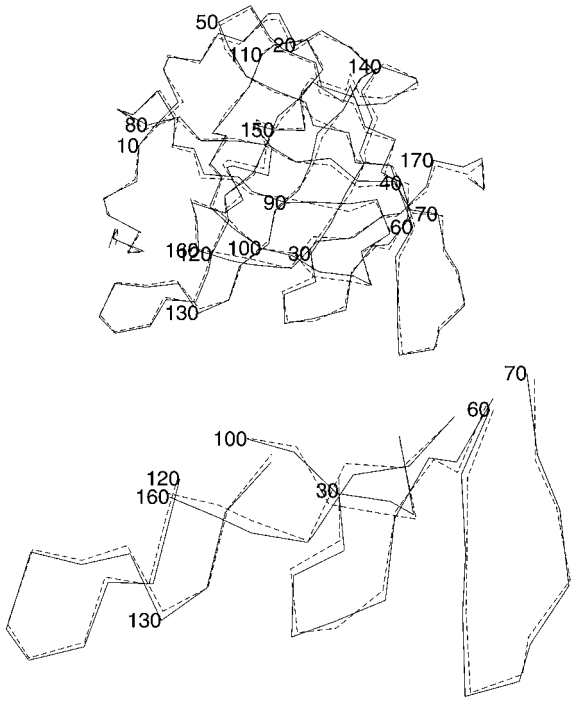

Figure 5. Correlated movement represented by PLS factor 1. Lower diagram shows the structure of the mouth of the binding site at extremes of the correlated movement.

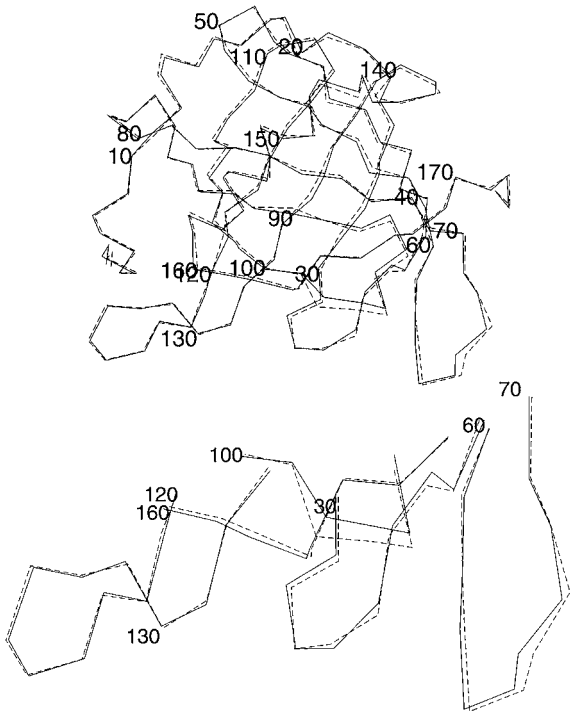

Figure 6. Correlated movement represented by PLS factor 2. Lower diagram shows the structure of the mouth of the binding site at extremes of the correlated movement.

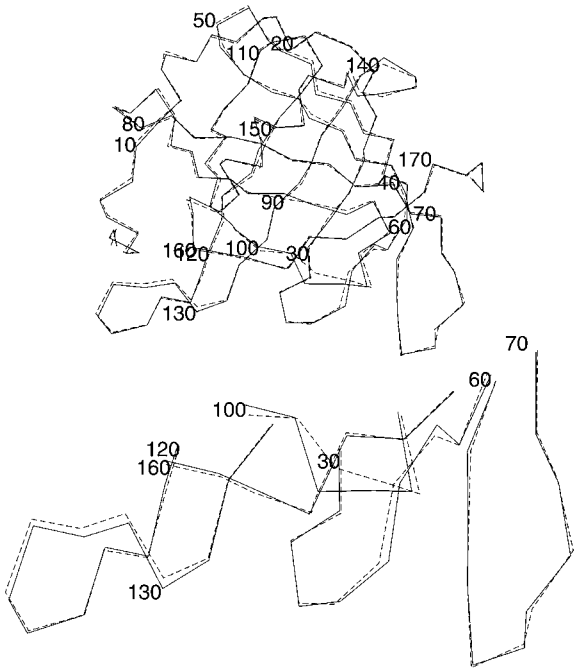

Figure 7. Correlated movement represented by PLS factor 3. Lower diagram shows the structure of the mouth of the binding site at extremes of the correlated movement.

bution of residuals for all other co-ordinates appear reasonably even, suggesting that the assumptions of PLS and PCA are valid for the data being analysed here.

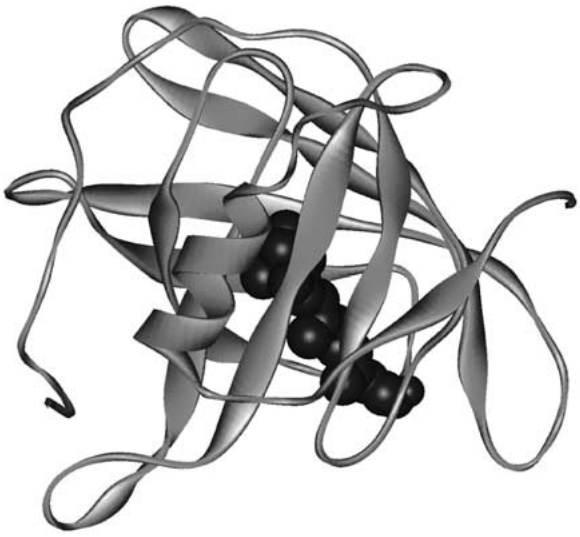

Figure 8. Structure of the mouth of the binding site of bovine serum retinol-binding protein.

Table 1. Correlation of the PCA factors and PLS factors.

	PCA1	PCA2	PCA3	PCA4
PLS 1	−0.99	−0.14	0.03	−0.02
PLS 2	0.15	−0.96	0.18	−0.13
PLS 3	0.00	0.23	0.57	−0.60

To turn to the results of the two different calculation methods, comparison shows that the factors they produce are closely related to one another (Table 1). PLS factor 1 is nearly identical to the reflection of PCA factor 1 ($r = -0.99$), and PLS factor 2 is closely related to the reflection of PCA factor 2 ($r = -0.96$).

However, the remaining factors are much more weakly correlated. This paper will only consider the structural implications of the PLS factors because, as explained above, PLS explicitly identifies correlated motions that are related to ligand movement. PCA will identifies correlated movement which need not be related to unbinding.

Correlating PLS factors with structural changes

Figure 5 shows the correlated movement captured by PLS factor 1, where the structures shown in solid and dotted lines are the extremes of the correlated movement. To aid orientation, Figure 8 is a schematic view of the protein from the same angle as Figure 5. This diagram shows clearly that, even though PLS factor 1 represents approximately 60% of the structural variation during the simulation, the movements are relatively small. The main regions which move with PLS factor 1 are one end of the β-sheet between residues 20 and 30 and one of the surface loops, between residues 60 and 75. Most of the remaining movement are of individual Cα atoms (e.g., residue 88) rather than regions.

The remaining two factors represent a much smaller proportion of the structural variation than PLS factor 1; PLS factor 2 only represents 8% and PLS factor 3 only represents 3%. Figure 6 shows that movement represented by PLS factor 2 is more evenly distributed throughout the structure of the protein but that particularly large changes are seen in the loop regions between residues 60 and 70, 90 and 100, and in the region between 140 and 160. PLS factor 3 is much more localised than factor 2. The largest movements involve residues in the loop regions between residues

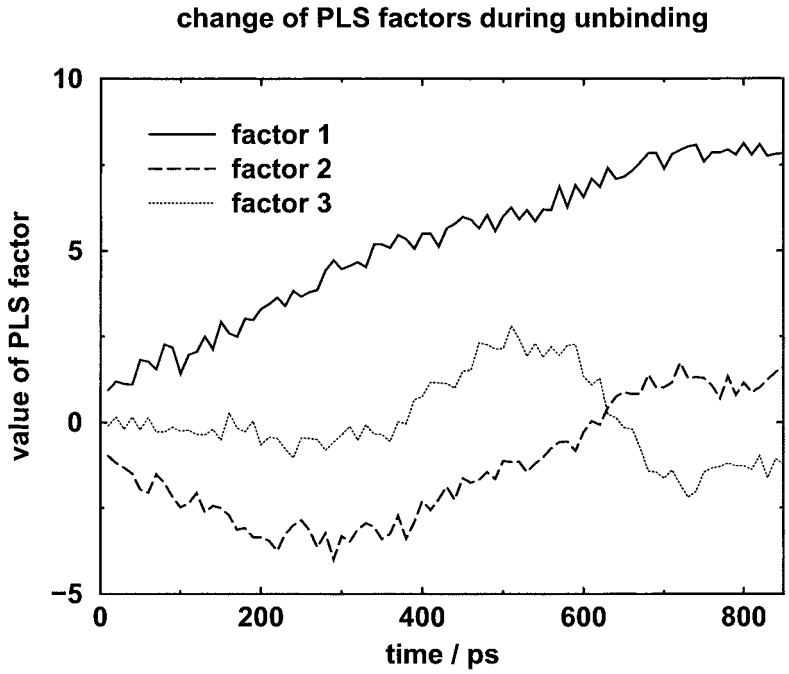

Figure 9. Change of PLS factors in the course of the unbinding simulation.

30 and 40, 60 and 70, and 90 and 100. Smaller movements are observed in the region between residues 140 and 150.

Having identified these correlated movements, their changes during the time course can be considered (Figure 9). PLS factor 1 (the dominant factor) changes almost linearly over the course of the simulation, from one value at the start to another at the end. This should represent the conformational change from holo- to apo-sRBP. PLS factor 2 initially decreases, but begins to increase at about 300 ps and returns to higher than its initial value. Therefore, it captures a conformational change that moves one way, then back the other until it is beyond its inital position. The third component changes little in the first 400 ps, and then sharply increases to a maximum at 500 ps before decreasing to near to its original value by 700 ps. This suggests it represents a conformational change which is important later in the unbinding process.

The mouth of the retinol-binding site is formed by three loops, one consisting of amino acids 30–40, another one of residues 60–70 and the last one of amino acids 90–100 (see Figures 5 and 8). Each of these loop regions are highlighted by the PLS factors. Previous work [27] has shown that in the first half of the unbinding simulation, Asn 65 exhibits large changes in its dihedral ϕ- and ψ-angles, while in the latter half of the dissociation process, large ϕ- and ψ-angle changes are observed in Phe 96 and Leu 97 (see Figure 10). These observations are consistent with the loops changing conformation to allow retinol to unbind. However, none of these dihedral angle changes correspond in time to the sudden increases in the distance between the centres of mass of the ligand and of the receptor (shown in the top panel of Figure 10 for comparison).

Perhaps the most surprising feature of the PLS factors is that none of them change especially quickly at any of the 3 steps in unbinding that were mentioned above. This suggests that the steps do not correspond to structural changes within the protein, which is in agreement with the backbone RMSD plot (Figure 1).

To our knowledge, this is the first published use of PLS to analyse protein structures produced by molecular dynamics. In this case PCA and PLS give identical results for the two most significant movements (the higher factors). However, the third factor calculated by PLS is different from that calculated by PCA. This suggests that PLS is identifying a correlated protein movement that is related to unbinding, while PCA is identifying other movements which are significant over the population but are not related to unbinding. If this is the case, then PLS is preferable to PCA because it is identifying only correlated move-

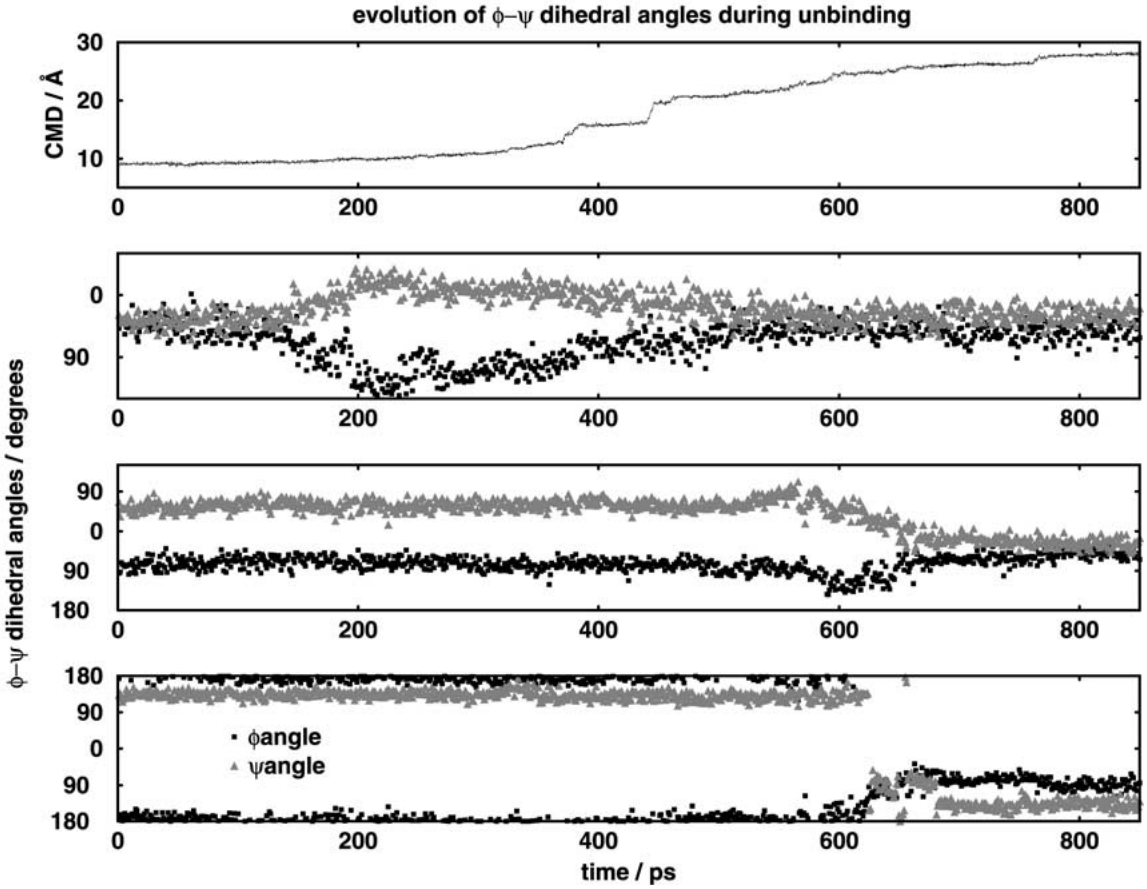

Figure 10. Distance between the centres of mass of the ligand and of the protein, and ϕ- and ψ-dihedral angles of Asn 65, Phe 96 and Leu 97, over the course of the unbinding simulation.

ments of interest and so reduces the number of factors that need to be assessed.

Although this is an advantage, the use of the centre-of-mass separation in the PLS approach does have disadvantages. Once the protein and ligand have unbound, the centre-of-mass separation will increase independently of any structural change in the protein, so changes unrelated to binding might begin to contribute to the model. This is unlikely to be a problem in the simulation here, where the ligand cannot move far from the protein, but it may be possible to find a better variable related to binding that does not have this disadvantage, such as the sum of van der Waals contacts between protein and ligand.

Conclusion

The oldest model for ligand-receptor interactions is the lock-and-key model of Fischer [28], where the ligand is regarded as rigid and fits a rigid receptor. This model was not successful in explaining part of the experimental data concerning substrate binding to enzymes, so it was updated by the induced-fit theory, namely that proteins adapt their shape to fit the ligand upon binding [29]. This theory has been quite useful in the study of ligand-receptor interactions [30], and its predictions have been borne out by experiments involving the binding of substrates to enzymes [31], or of neurotransmitters to their receptors [32].

Simulation studies allow us to examine protein motions in detail. Recently, using a novel simulation method, one of us has succeeded in unbinding retinol from bovine serum retinol-binding protein [5]. From various structural analyses, we know that there are three loops at the mouth of the binding site of the bovine serum retinol-binding protein, and that there three loops are important in the binding of serum retinol-binding protein to retinol, and also to

transthyretin [12, 33]. However, there has been no systematic study of the motion involved in retinol unbinding from retinol-binding protein.

This work represents a new application of two established methods, Principal Components Analysis (PCA) and Projection to Latent Structures (PLS), to examine the movement of protein atoms as a function of retinol unbinding. Analysing the agreement between the data and the PLS model shows that PLS is able to characterise much of the movement occurring during unbinding. However, the movement of one residue (Glycine 100) is not fully explained.

We have discovered that the loop regions in the mouth of the binding site exhibit the largest movement, consistent with results from dihedral analysis [27]. The results lead us to suggest that the sudden changes in unbinding speed during the unbinding process are probably a result of changes in retinol-protein interaction rather than sudden changes in protein structure. Further work is being performed to elucidate the detailed changes in the retinol-protein interaction during the unbinding process.

Acknowledgements

The authors thank Stephen Cornell for useful comments. PLC thanks the UK Biotechnology and Biological Sciences Research Council for the award of a David Phillips Research Fellowship, and New Hall, Cambridge, for a Senior Research Fellowship. Computational work carried out at the University of Cambridge High Performance Computing Facility is gratefully acknowledged.

References

1. Grubmüller, H., Heymann, B. and Tavan, P., Science, 271 (1996) 997.
2. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. and Schulten, K., Biophys. J., 72 (1997) 1568.
3. Marrink, S.-J., Berger, O., Tieleman, P. and Jähnig, F., Biophys. J., 74 (1998) 931.
4. Kosztin, D., Izrailev, S. and Schulten, K., Biophys. J., 76 (1999) 188.
5. Chau, P.-L., Chem. Phys. Lett., 334 (2001) 343.
6. García, A.E., Phys. Rev. Lett., 68 (1992) 26967.
7. Amadei, A., Linssen, A.B.M. and Berendsen, H.J.C., (1993) Proteins: struct., funct. genet., 17 (1993) 412.
8. Wold, S., Chem. Scripta, 5 (1974) 97.
9. Kern, P., Brunne, R.M. and Folkers, G., J. Comput.-Aid. Mol. Des., 8 (1994) 367.
10. Rognan, D., Scapozza, L., Folkers, G. and Daser, A., Biochemistry, 33 (1994) 11476.
11. Leech, J., Prins, J. and Hermans, J., IEEE Comput. Sci. Eng., 3 (1996) 38.
12. Zanotti, G., Berni, R. and Monaco, H.L., J. Biol. Chem., 268 (1993) 10728.
13. van Gunsteren, W.F. and Berendsen, H.J.C., *Groningen Molecular Simulation Library Manual*, Biomos B.V., Groningen, Netherlands, 1987.
14. Berendsen, H.J.C., Grigera, J.R. and Straatsma, T.P., J. Phys. Chem., 91 (1987) 6269.
15. Forester, T.R. and Smith, W., J. Mol. Graph., 14 (1996) 136.
16. Nosé, S., J. Chem. Phys., 81 (1984) 511.
17. Hoover, W.G., Phys. Rev. A, 31 (1985) 1695.
18. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C., J. Comput. Phys., 23 (1977) 327.
19. Melchionna, S., Ciccotti, G. and Holian, B.L., Mol. Phys., 78 (1993) 533.
20. Malinowski, E.R., *Factor Analysis in Chemistry. 2nd edition* John Wiley & Sons, New York, 1991.
21. Mello, L.V., van Aalten, D.M.F. and Findlay, J.B.C., Biochemistry, 37 (1998) 3137.
22. Chau, P.-L., van Aalten, D.M.F., Bywater, R.P. and Findlay, J.B.C., J. Comput.-Aid. Mol. Des., 13 (1999) 11.
23. Abseher, R., Horstink, L., Hilbers, C.W. and Nilges, M., Proteins: struct., funct. gen., 31 (1998) 370.
24. O'Donoghue, S.I., Chang, X., Abseher, R., Nilges, M. and Led, J.J., J. Biomol. NMR, 16 (2000) 93.
25. Howe, P.W.A., J. Biomol. NMR, 20 (2001) 61.
26. Eriksson, L., Johansson, E., Kettaneh-Wold, N. and Wold, S., *Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA & PLS)* UMETRICS AB, Umeå, Sweden, 1999.
27. Chau, P.-L., J. Biol. Phys. (2002) (in press).
28. Fischer, E., Berichte Deutsch. Chem. Gesell., 27 (1894) 2985.
29. Koshland, D.E. Proc. Nat. Acad. Sci. U.S.A., 44 (1958) 98.
30. Miller, D.W. and Dill, K.A., Prot. Sci., 6 (1997) 2166.
31. Ford, L.O., Johnson, L.N., Mackin, P.A., Phillips, D.C., Tijan, R., J. Mol. Biol., 88 (1974) 349.
32. Miyazawa, A., Fujiyoshi, Y., Stowell, M. and Unwin, N., J. Mol. Biol., 288 (1999) 765.
33. Taylor, H.M. and Newcomer, M.E., Biochemistry, 38 (1999) 2647.