Molecular Dynamics Simulations of Protein-Tyrosine Phosphatase 1B. I. Ligand-Induced Changes in the Protein Motions

Günther H. Peters,* Thomas M. Frimurer,# Jannik N. Andersen,§ and Ole H. Olsen*
#MedChem Research IV, Novo Nordisk A/S, DK-2760 Måløv; *Department of Chemistry, Technical University of Denmark, DK-2800 Lyngby; and §Target Cell Biology, Novo Nordisk A/S, DK-2800 Bagsvård, Denmark

ABSTRACT Activity of enzymes, such as protein tyrosine phosphatases (PTPs), is often associated with structural changes in the enzyme, resulting in selective and stereospecific reactions with the substrate. To investigate the effect of a substrate on the motions occurring in PTPs, we have performed molecular dynamics simulations of PTP1B and PTP1B complexed with a high-affinity peptide DADEpYL, where pY stands for phosphorylated tyrosine. The peptide sequence is derived from the epidermal growth factor receptor ($EGFR_{988-993}$). Simulations were performed in water for 1 ns, and the concerted motions in the protein were analyzed using the essential dynamics technique. Our results indicate that the predominately internal motions in PTP1B occur in a subspace of only a few degrees of freedom. Upon substrate binding, the flexibility of the protein is reduced by ~10%. The largest effect is found in the protein region, where the N-terminal of the substrate is located, and in the loop region Val^{198}-Gly^{209}. Displacements in the latter loop are associated with the motions in the WPD loop, which contains a catalytically important aspartic acid. Estimation of the pK_a of the active-site cysteine along the trajectory indicates that structural inhomogeneity causes the pK_a to vary by approximately ± 1 pK_a unit. In agreement with experimental observations, the active-site cysteine is negatively charged at physiological pH.

INTRODUCTION

Phosphorylation and dephosphorylation are recognized to be important regulatory events, particularly in those pathways that control cell growth and differentiation. Proper cellular functions are regulated by a balanced action of protein phosphatases and kinases (Fisher et al., 1991; Stone and Dixon, 1994; Walton and Dixon, 1993). Protein phosphatases (PPs) belong to a diverse group of proteins that is classified according to their substrate specificities for phosphorylated serine, threonine, or tyrosine residues (Krueger et al., 1990; Barford, 1995; Fauman and Saper, 1996). Understanding of the physiological functions of phosphatases has made much progress as a result of the determination of the three-dimensional crystal structures, including the open (substrate not bound, hereafter referred to as "uncomplexed" structure) and closed (substrate bound; hereafter called "complexed" structure) conformers. These three-dimensional structures of PPs revealed that although very diverse in size, sequence, and structural organization, the catalytic PPs domain is highly conserved and usually contained within a ~250-residue segment (Mourey and Dixon, 1994). This domain is the only structural element that has amino acid sequence identity among all PTPs from bacteria to mammals (Zhang et al., 1994a,b) and contains the essential residues of the substrate binding site (Stuckey et al., 1994; Barford et al., 1994). This binding pocket is primarily composed of the phosphate binding loop (P-loop) with the common active site motif $(H/V)CX_5R(S/T)$ (X denotes any amino acid residue). Biochemical investigations using chemical modifications and site-directed mutagenesis have established that the cysteinyl residue within the consensus motif is essential in the catalytic reaction and carries out the nucleophilic attack on the phosphate moiety. Active-site thiolates are also found in other hydrolases. For instance, in papain (Lewis et al., 1981) and dienelactone hydrolase (Pathak and Ollis, 1990), the active Cys is stabilized by ion pairing to nearby histidines, a structural arrangement that is not observed in tyrosine and serine/threonine phosphatases. The only invariant and positively charged residue in the vicinity of the active cysteine (Cys^{215} in PTP1B), which has an important function in stabilizing the binding of the phosphate moiety, is the consensus sequence arginine (Arg^{221} in PTP1B). Generally, this residue is immediately followed by a Ser or Thr residue, which facilitates the breakdown of the phosphotyrosine-enzyme transition state (Zhang et al., 1995; Zhao and Zhang, 1996; Pannifer et al., 1998). Adjacent to the consensus sequence is a flexible binding loop, which upon binding of the substrate moves several Ångstroms toward the active site to trap the oxyanion and to bring an aspartic acid in position (closed conformation). The amino acid sequence of this loop is quite diverse, except for the WPD sequence that includes the catalytic aspartic acid residue and a tryptophan residue near the hinge region of the loop.

The hydrolysis of phosphotyrosine involves several steps, including binding of the substrate, activation of the protein tyrosine phosphatase (PTP) (i.e., closing of the binding loop), formation of a covalent phosphoenzyme intermediate, and release of the product (Zhang, 1990; Guan and Dixon, 1991; Wo et al., 1992; Cho et al., 1992). The present

Received for publication 9 September 1998 and in final form 14 April 1999.
Address reprint requests to Dr. Ole H. Olsen, MedChem Research IV, Novo Nordisk A/S, Novo Nordisk Park, DK 2760 Malov, Denmark. Tel.: 45-4443-4511; Fax: 45-4443-4547; E-mail: oho@novo.dk.

0006-3495/99/07/505/11 $2.00

study focuses on the first two steps, and the objective of this study is to examine the dynamics and structural changes in the protein tyrosine phosphatase PTP1B upon binding of a peptide-based substrate, DADEpYL. The DADEpYL sequence is derived from the epidermal growth factor receptor ($EGFR_{988-993}$), which is dephosphorylated by PTP1B in cells. The analysis does not consider the transition state, where the active Cys is covalently bound to the phosphate and the oxygens of the phosphate group form a planar arrangement. To determine the protein dynamics in the free and bound states, we have applied the essential dynamics analysis technique (Amadei et al., 1993), which allows us to analyze molecular dynamics trajectories and to identify collective motions in proteins. The essential dynamics analysis technique has previously been used to correlate essential motions in lysozyme (Amadei et al., 1993), in thermolysin (van Aalten et al., 1995), in SH3 binding domain (van Aalten et al., 1996), in cellular retinol binding protein (van Aalten et al., 1997), and in lipases (Peters et al., 1996, 1997) with their biological function. Several common features emerged from these studies. Generally, there are only a few large concerted motions in proteins originating from the flexibility of loops and/or secondary structural elements such as hinge-bend motions that facilitate binding and release of the substrate.

Although much information about the location and the nature of the catalytic center as well as the substrate binding pocket and the structural basis for activation has been obtained from crystallograhic studies and kinetic measurements (Barford, 1995; Barford et al., 1994; Zhang et al., 1994b), no attempt has been made to describe the large collective motions in PTP1B. To investigate the importance of protein flexibility for substrate binding, we have performed molecular dynamics simulations of PTP1B and PTP1B complexed with a hexapeptide. The results presented here focus on the collective motions occurring in the protein, using the C_α coordinates (Amadei et al., 1993) of PTP1B and the PTP1B-complex. Thus substrate atoms are not included in the essential dynamics analysis. This analysis allows us to identify regions in the protein, which are affected by substrate binding and are of interest for substrate/inhibitor design. Our results indicate that there are pronounced differences in the dynamics of the binding pocket and in a region opposite the binding cleft. Some regions in the binding pocket show changes in flexibility when the dynamics of the uncomplexed enzyme is compared to that of the complexed enzyme. These regions have been further analyzed on an atomic level, and the dynamic and energetic properties of the binding pocket and the collective motions between residues in the substrate and the protein are presented in part II (Peters et al., unpublished observations). As mentioned above, the nucleophilicity of the active-site cysteine is important for the enzymatic activity; therefore we have estimated the pK_a for this side chain along the molecular dynamics trajectories.

METHODS

High-resolution crystal structures of the PTP1B structure complexed with tungstate at 2.8 Å resolution and the PTP1B/DADEpYL-NH_2 complex (active cysteine mutated to serine) at 2.6-Å resolution (Jia et al., 1995; Barford et al., 1994) were used as models for the open ("uncomplexed," no substrate bound) and closed ("complexed," substrate bound) structures, respectively. Both structures were obtained from the Protein Data Bank at Brookhaven (Bernstein et al., 1977). The entry codes are 2hnq and 1ptu for the uncomplexed and complexed PTPs, respectively. Before the simulations were performed, the serine (215) was replaced with an ionized cysteine, using the WHAT IF modeling program (Vriend, 1990). The reason for modeling the cysteine side chain as a thiolate is based on the experimentally observed low pK_a value (Peters et al., unpublished observations). The N-terminal of the 2hnq structure was extended using the 1ptu structure as a template. We have modified the peptide in the crystal structure from DADEpYL-NH_2 to DADEpYL by adding a negatively charged C-terminus (COO^-).

Molecular dynamics simulations were performed for PTP1B (2hnq) and for the PTP1B-peptide complex (1ptu) in periodic boundary conditions, using explicit SPC water, which was taken from a liquid equilibrium configuration (Berendsen et al., 1981). Simulations were run for 1 ns, by applying the charged version of the GROMOS force field (Van Gunsteren and Berendsen, 1987). To neutralize the systems, water molecules at the lowest electrostatic potential were replaced by sodium ions. The systems were composed of 8079 (2hnq) and 7222 (1ptu) atoms. Simulations were started from the crystallographic PTP1B structures and subjected to a steepest descents energy minimization until no significant energy change could be detected (<0.01 kJ/mol). This was followed by a 5-ps startup run, taking the initial velocities from a Maxwellian distribution and using a temperature coupling constant of 0.01 ps and a pressure coupling constant of 0.05 ps (Van Gunsteren and Berendsen, 1987). The nonbonded pair list was updated every 10 fs, and the nonbonded interactions were truncated at 8 Å and 10 Å, respectively. The SHAKE algorithm (Ryckaert et al., 1977) was applied to constrain the bond lengths to their equilibrium positions, and the equations of motion were solved using the Verlet algorithm (Allen and Tildesley, 1989). Simulations were performed with a time step of 2 fs, and coordinates were saved every 0.05 ps. Examinations of the molecular structures and analyses of the trajectories were carried out using the WHAT IF modeling program (Vriend, 1990). Evaluation of several geometric properties (radius of gyration, number of hydrogen bonds, root mean square displacement, accessibility, number of residues in random coil conformation, number of strained dihedrals) indicated that 300 ps (uncomplexed PTP1B) and 150 ps (PTP1B-peptide complex) were required for equilibration. The remaining 700 ps of the trajectories (i.e., from 300 to 1000 ps) were used for the subsequent essential dynamics analysis. The essential dynamics method has been described

several times in the literature, and the reader is referred to Amadei et al. (1993) and Balsera et al. (1996). Briefly, the method is based on the diagonalization of the covariance matrix (Ichiye and Karplus, 1991), which yields a set of eigenvectors and corresponding eigenvalues. The eigenvectors represent a direction in a high-dimensional space, describing concerted displacements of atoms. The eigenvalues represent the mean square fluctuation of the total displacement along these eigenvectors. Motion within the subspace can be studied by projecting the trajectory onto the individual eigenvectors. This dot product provides information about the time dependence of the conformational changes.

pK_as were computed using the so-called single-site titration model modified by Antosiewicz et al. (1994). In this macroscopic approach, ionization is represented as the addition of ± 1 proton charge to a single atom in each titratable group. The electrostatic field at the titrating site is calculated by numerically solving the finite-difference linearized Poisson-Boltzmann equation, using an incomplete Cholesky preconditioned conjugate gradient method as implemented in the UHBD program (Madura et al., 1995; Davis et al., 1991). The apparent pK_as and protonation state as a function of pH were computed by employing the "hybrid procedure" developed by Gilson (1993), which is based on separation of ionizable groups into clusters. The interaction between ionizing charges within a cluster is treated exactly, whereas the intercluster interactions are approximated. This method yields both accurate energies and fractional charges (Gilson, 1993). A detailed description of the parameters used in the calculations has appeared elsewhere (Peters et al., unpublished observations). Recently, an improvement in the accuracy of pK_a predictions was proposed (Demchuk and Wade, 1996). This technique operates with two classes of ionizable sites, which can be distinguished using a criterion based on desolvation energy. A Z_{rf} value given by

$$Z_{rf} = \text{sign}(\Delta pK_{a,Coulomb}(\epsilon_{interior}))\Delta pK_a(\epsilon_{interior}) \qquad (1)$$

is computed for each ionizable site of interest from four pK_a calculations (using the single-site titration method described above) at two extreme protein dielectric constants, $\epsilon_{interior}$ (78.5 and 3; Demchuk and Wade, 1996). The magnitude of Z_{rf} is derived from $\Delta pK_a(\epsilon_{interior})$, the difference between the computed pK_as at a high and a low dielectric constant. $\Delta pK_{a,Coulomb}(\epsilon_{interior})$ is the corresponding pK_a difference when the charges are considered to be immersed in a homogeneous dielectric of zero ionic strength. Thus one class of mostly solvent exposed residues was identified ($Z_{rf} > -1.6$) for which the best pK_as are calculated if the protein dielectric constant is close to that of the aqueous solvent. For another set of mostly buried residues it was shown ($Z_{rf} < -1.6$) that a relatively low dielectric constant (in the range 10–20) gave the best results. According to the criterion defined above and applied to the x-ray structures of PTP1B, the active-site cysteine belongs to the first class of residues, i.e., the group for which the best pK_as are calculated with a high dielectric constant. The resulting pK_a is 1.6 pK_a units above the experimentally determined value. For comparison we have calculated the pK_a for this residue, using a low dielectric constant. The resulting pK_a is closer to the experimentally determined pK_a, but this methodology underestimates the experimental pK_a. Clearly, both approaches result in a pK_a that is over- or underestimated, yielding pK_a's with similar accuracy. We therefore have applied both methodologies to calculate the pK_a along the trajectories and to estimate the influence of structural inhomogeneity on the pK_a of the active Cys.

RESULTS AND DISCUSSION

Molecular dynamics simulations of the uncomplexed PTP1B structure and the PTP1B-DADEpYL complex (Fig. 1) were performed using explicit SPC water and applying periodic boundary conditions. Simulations were carried out for 1 ns, where 300 ps (uncomplexed PTP1B) and 150 ps (PTP1B-peptide complex) were used for equilibration. Root mean square displacement (rmsd) with respect to the starting structure, accessible surface area (ASA), and radius of gyration are displayed in Fig. 2. As shown in Fig. 2, the rmsd for the uncomplexed PTP1B is larger than for the complexed structure. A slight increase is observed, but the other geometrical properties are stable and fluctuating around a constant value. Relatively large fluctuations in the ASA data are observed in both structures, indicating that the overall flexibility of side chains at the protein surface is similar.

The resulting trajectories were used in the essential dynamics analysis (Amadei et al., 1993; van Aalten et al., 1995; Peters et al., 1996). Note that no substrate atoms were included in the analysis because we only focus on the changes in enzyme dynamics resulting from the presence of the substrate. Motions along eigenvectors corresponding to large eigenvalues are mainly large-amplitude, anharmonic fluctuations. The harmonicity of the motion along an eigenvector can be estimated by comparing the probability distribution for the displacements along these eigenvectors with an ideal Gaussian distribution derived from the eigenvalue of the corresponding eigenvector (Amadei et al., 1993). The correlation coefficients resulting from this comparison and the eigenvalues as a function of eigenvector index are displayed in Figs. 3 and 4, respectively. As shown in Fig. 3, ~85% of the total positional fluctuations are described by the first 50 eigenvectors. This number of eigenvectors is an order of magnitude larger than observed for other proteins; e.g., for thermolysin it has been shown that taking the first 10 eigenvectors represents ~95% of the total motion of the protein (Amadei et al., 1993; van Aalten et al., 1995). This indicates that the fluctuations in PTP1B are more complex, so that more eigenvectors are needed to describe the motions. However, as observed for several other proteins (Amadei et al., 1993; van Aalten et al., 1995; Peters et al., 1996), the actual number of eigenvectors is much smaller than the original $3N$-dimensional configura-

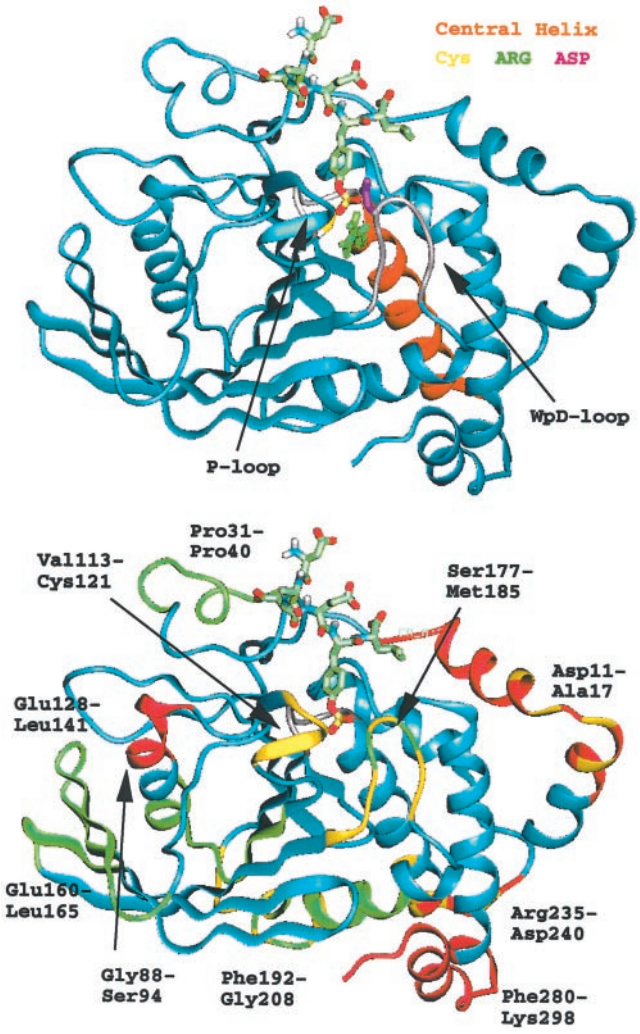
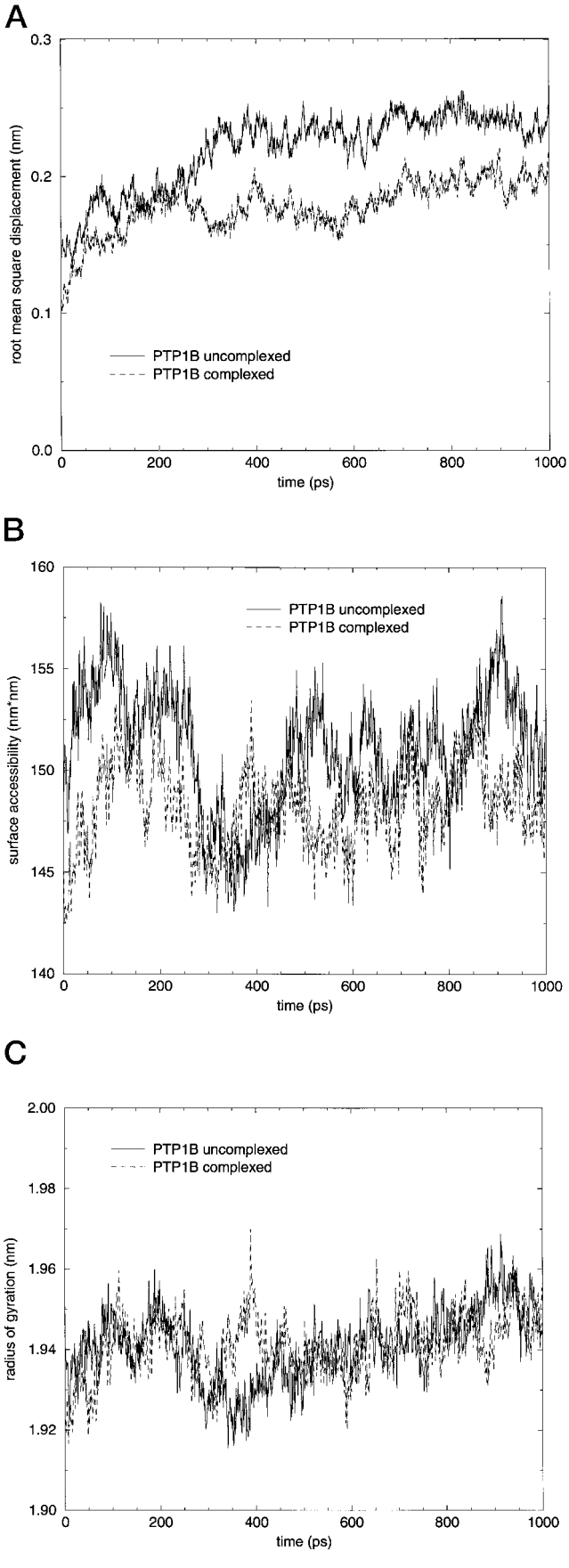


FIGURE 1 Ribbon structure of PTP1B complexed with the peptide DADEpYL. The color coding corresponds to the important components of the catalytic site. The active site motif, the HCX_5R loop, and the WPD loop are colored gray. The central helix is colored brown. The active Cys, the conserved Arg, and the essential acidic Asp are shown in stick modus and are colored yellow, green, and red, respectively. Atoms of the substrate are displayed in stick model and colored as follows: carbons, green; nitrogens, blue; oxygens, red; hydrogens, white. (*Top*) The active Cys, the conserved Arg, and the essential acidic Asp are shown in stick models and colored yellow, green, and red, respectively. (*Bottom*) Regions (not including the WPD loop) that show relatively large displacements (*yellow*), increase (*red*) and decrease (*green*) fluctuations upon substrate binding. See text for more details.

tional space formed by the Cα coordinates ($3N$ Cα; N = 297). This shows that most of the internal motions of these enzymes are confined within a subspace of very small dimensions. This is also reflected in the correlation coefficients shown in Fig. 4. The coefficient reveals that most of the motions can be described by Gaussian distributions (i.e., fluctuations are harmonic). Nonharmonic (non-Gaussian) fluctuations are only observed at small eigenvector indices, and the coefficient approaches ~0.95 after 10 eigenvectors.

To investigate the effect of a bound peptide substrate on the dynamics of the PTP1B, we have performed a combined analysis of the trajectories (van Aalten et al., 1995) of the uncomplexed PTP1B and the PTP1B-substrate complex. The covariance matrix was constructed from the C_α coordinates from the two trajectories and subsequently diagonalized. The analysis yields a set of eigenvalues and corresponding eigenvectors, in which each pair describes independent motions in the subspace. The fluctuations within these subspaces can be studied by projecting the trajectory onto the individual eigenvectors. This dot product provides information about the time dependence of the conformational changes. The average and mean square fluctuations of these projections as a function of eigenvector indices are shown in Fig. 5, and the corresponding absolute values of the components of the eigenvectors are displayed in Fig. 6. The average value shown in Fig. 5 *A* provides information about the structural changes (i.e., static shifts) upon binding of the substrate. The two curves are symmetrical around zero, because the analysis was performed on the combined trajectories of the two simulations. Structural differences are computed with respect to an average structure calculated from the combined trajectories. Significant differences are only observed in the subspace spanned by the first eigenvector. Visualization of the motions along eigenvector 1 revealed that the structural changes originated from the displacement of the WPD loop, which moves toward the active Cys as the substrate binds. As indicated by the absolute value of components of the first eigenvector (Fig. 6 *A*), this movement results in several structural changes in PTP1B. The most significant shifts (absolute values > 0.09 nm) are observed in the loop regions Asp^{11}-Ala^{17}, Val^{113}-Ser^{118}, Thr^{177}-Pro^{185}, Val^{198}-Gly^{209}, and Asp^{236}-Pro^{241}. As the WPD-loop (Thr^{177}-Pro^{185}) moves toward the substrate, the loop Thr^{263}-Asp^{265} located opposite the WPD loop moves backward causing the displacement of the segment Asp^{11}-Ala^{17}. The displacement observed in the flap Val^{113}-Ser^{118} widens the binding groove and provides more space for the substrate to the binding pocket. The most pronounced effect is seen in the region Val^{198}-Gly^{209}, which is located between the α-helix following the WPD loop and the β-strand, where the active cysteine is located. Both the α-helix and the β-strand show only small fluctuations, indicating that the displacement of the WPD loop is mediated by the structural changes in the loop region Val^{198}-Gly^{209}. This segment, which is well exposed to the solvent, contains several polar residues and two glycines contributing to the flexibility of the segment.

For eigenvector indices greater than 1, significant differences are found in mean square fluctuations (Fig. 5 *B*), which reflects the differences in the protein dynamics observed in the uncomplexed PTP1B and in the PTP1B-substrate complex. Motions in the subspaces spanned by eigenvectors 3 and 4 are predominantly found in the uncomplexed PTP1B, whereas motions along eigenvectors 2 and 5 are mainly observed in PTP1B complexed with the peptide. Visualization of the motions along several eigenvectors indicates that as observed for other proteins, the catalytic residue Cys^{215} (Fig. 1) is relatively rigid (Fig. 6 *B*).



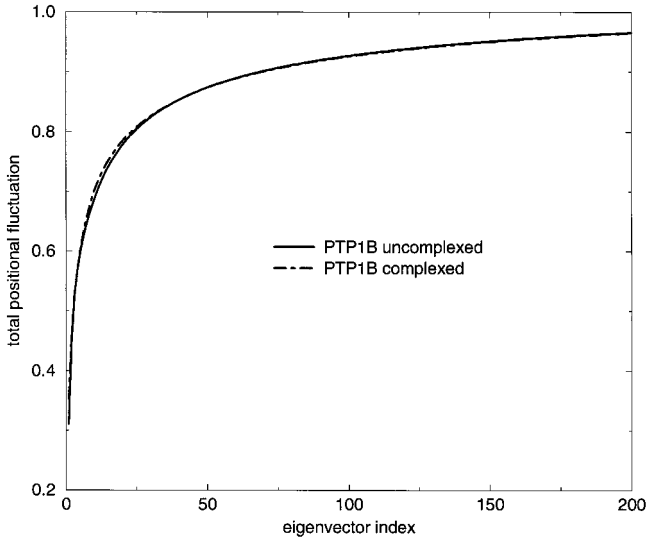


FIGURE 3 Cumulative normalized eigenvectors determined from the C_α coordinate covariance matrix of the trajectory obtain from the simulation of the uncomplexed PTP1B and PTP1B-peptide complex.

Fluctuations of the WPD loop are seen along several eigenvectors (Fig. 6 *B*), but these are relatively small in comparison to the other fluctuations. In the uncomplexed PTP1B, fluctuations in the WPD loop are oscillatory in nature, and no full opening and closing of the loop were observed, as suggested by recent experiments (Lohse et al., 1997). However, this is not surprising, because typical motions of peptide loops occur on a much longer time scale (10^{-9} to 10^{-1} s) (Wade et al., 1993, 1994; Kempner, 1993; Philippopoulos et al., 1995; Falzone et al., 1994; Williams and McDermott, 1995). Fig. 6 *B* shows the absolute values of the components of the eigenvectors 2–6. Both the uncomplexed PTP1B and PTP1B complexed with the substrate show relatively large fluctuations, and the differences in the protein dynamics are localized in certain regions of the enzyme. This is surprising because larger effects of inhibitors bound to proteins have been found, for instance, for lipases (Peters et al., 1997). As indicated by the magnitude of the mean square fluctuations shown in Fig. 5 *B*, the most pronounced effects are seen in the subspaces spanned by eigenvectors 2 (complexed PTP1B) and 3 (uncomplexed PTP1B). In Fig. 6 *C*, the absolute values of the components of these two eigenvectors are compared. Substrate binding reduces the motions in the Pro^{31-40}, Glu^{97}-Val^{108}, Lys^{128}-Lys^{141}, Leu^{160}-Thr^{165}, Pro^{179}-Gly^{182}, and Leu^{192}-Ser^{204}, whereas an increase in fluctuations is observed in the segments Leu^{88}-His^{94} and Asp^{235}-Asp^{240}, as well as in the N-terminal and C-terminal regions. The region Leu^{88}-His^{94}

FIGURE 2 Selected geometrical properties. (*A*) Root mean square displacement, (*B*) radius of gyration, and (*C*) accessible surface area as a function of simulation time for the uncomplexed PTP1B and the PTP1B-peptide complex. Surface accessibilities were calculated with DSSP software (Kabsch and Sander, 1983).

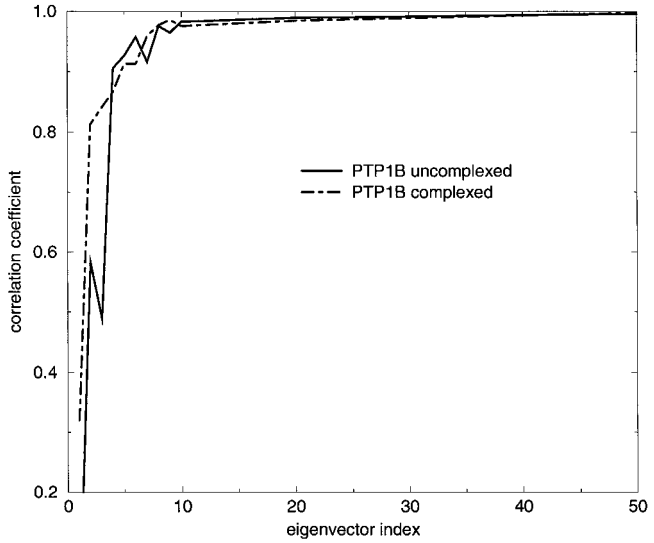


FIGURE 4 Correlation coefficient as a function of eigenvector index determined from the simulation of the uncomplexed PTP1B compared to an ideal Gaussian distribution derived from the eigenvalue.

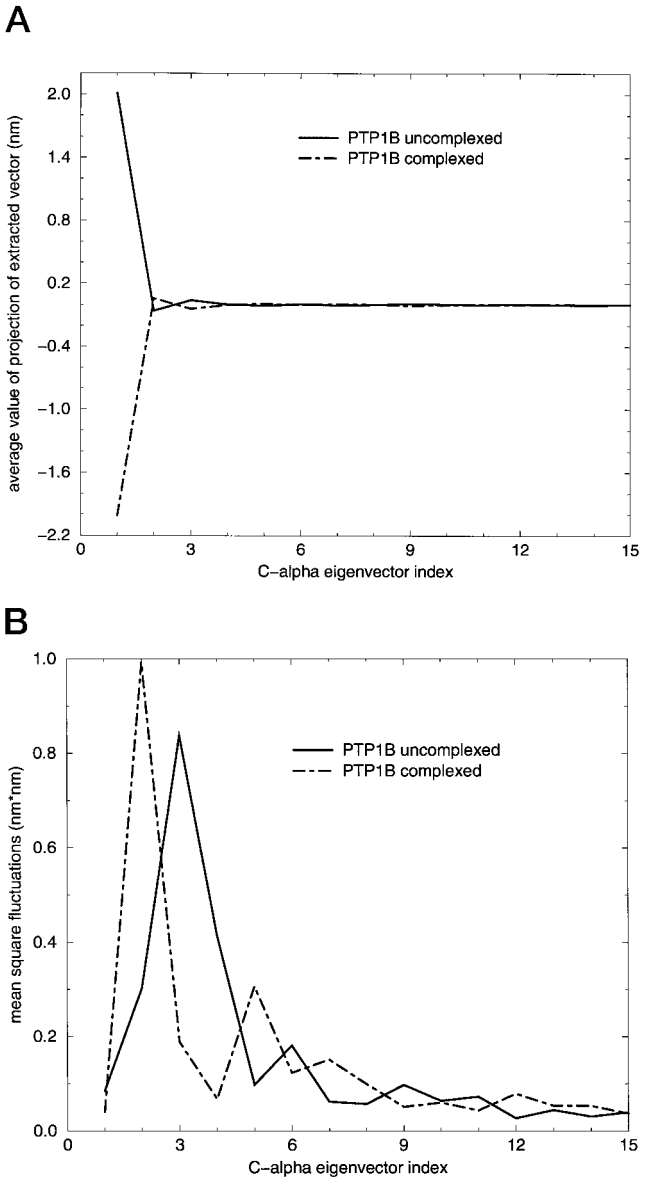


FIGURE 5 (*A*) Average values and (*B*) mean square fluctuations of the projections of the separate trajectories onto the eigenvectors extracted from the C_α coordinate covariance matrix of the concatenated trajectories as a function of the eigenvector index.

is located close to the N-terminal of the peptide-based substrate, and motions are increased because of binding of the substrate. The WPD loop and the the loop Val^{198}-Gly^{209} show larger flexibility in the uncomplexed PTP1B. Interestingly, large conformational changes have already been seen along eigenvector 1, which describes the static shift between the uncomplexed PTP1B and PTP1B complexed with the peptide. To elucidate the effect of substrate binding on the flexibility of the protein, we have calculated the total fluctuation in the essential space defined by the first 20 eigenvectors. The results reveal that the fluctuations decrease by ~10%; i.e., that (as expected) binding of the substrate causes an induced fit (Koshland, 1958).

Protein flexibility may be an important factor in substrate recognition and binding. Internal protein motions are generally governed by flexible loops or links and are correlated with the positions of glycines in the protein structure. To locate hinge bending regions in the PTP1B, we have applied a moving window superposition method (van Aalten et al., 1996). Rmsds calculated from the minimum and maximum structures of the motions occurring in different subspaces as a function of residue number are shown in Fig. 7. The positions of glycines and prolines in PTP1B are indicated by the filled or open circles. There is no apparent correlation between the flexibility of the protein and the position of the prolines. However, noticeable rmsd values are observed close to glycines. In particular, the glycines 106, 183, 218, 220, 223, and 259 are closely located to hinge bending regions. Structural sequence alignment of PTP1B, *Yersinia* PTP, VHR, and low-molecular-weight PTP (Yuvaniyama et al., 1996) reveals that most of the glycines are not conserved in these structures, suggesting that glycines may or may not have an important role in the function of these proteins. Important functional roles of glycines have been observed earlier, in which single-site mutations of glycines or of nonglycine residues at other sites to glycines can have remarkable effects on the enzyme action and can be lethal. Mutations involving glycines have been linked to protein stability (Matthews, 1987; Hecht et al., 1986; Berndt et al., 1993), reduced catalytic activity (Jancso and Szent-Györgyi, 1994; Wilkinson et al., 1983), changes in enzyme specificity (Vermersch et al., 1990), or diseases such as lipoprotein lipase deficiency (Henderson et al., 1992; Busca et al., 1996), hemolytic anemia (Vulliamy et al., 1988), Alzheimer's disease (Mann et al., 1992), or insulin resistance (Almind et al., 1996). This suggests that the absence of conserved glycines in the different structures causes different protein flexibility and hence different modes of substrate recognition. The only conserved glycines are

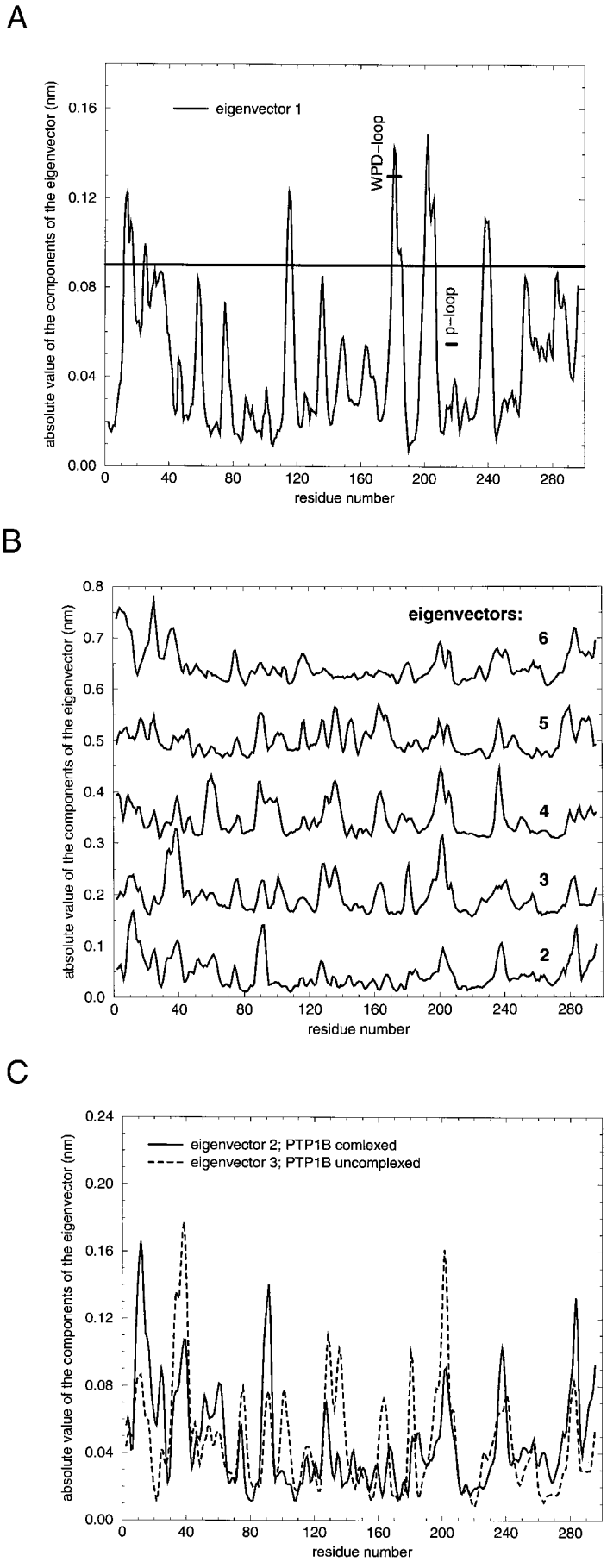


FIGURE 6 Absolute value of the components of eigenvectors (*A*) 1, (*B*) 2–6, and (*C*) 2 and 3. Eigenvectors were computed from the C_α coordinate covariance matrix of the concatenated trajectories as a function of coordinate number. (*A*) Structural changes between the uncomplexed and complexed structures. (*B*) Motions observed in the subspaces spanned by eigenvectors 2–6. (*C*) Motions predominantly found in the complexed (eigenvector 2) and uncomplexed (eigenvector 3) structures. The location of the WPD- and p-loops are shown in *A*. The horizontal solid line in *A* corresponds to an absolute value of 0.07. See text for more details.

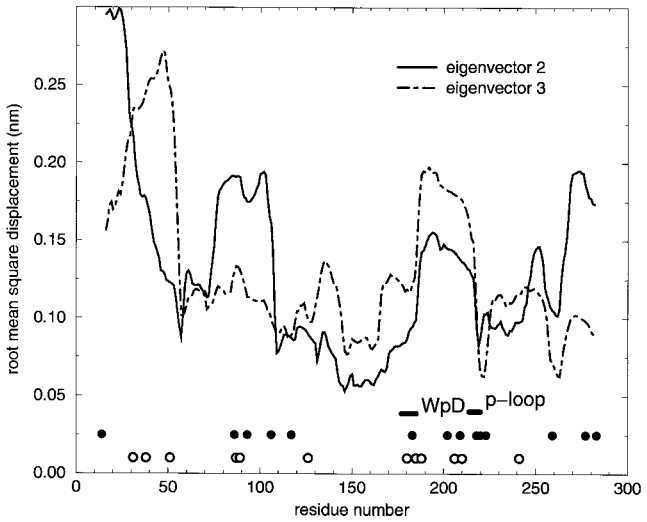


FIGURE 7 Output from the moving window superposition method, using the minimum and maximum structures observed in the subspaces spanned by eigenvectors 2 and 3. Positions of glycines and prolines are indicated by the filled and open circles, respectively.

found in the vicinity of the active Cys, which probably provide a certain flexibility of the active Cys to shift it to the correct positions relative to the phosphate moiety of the substrate, so that the catalyzed reaction may take place.

The structural inhomogeneity may not only influence substrate binding, but may also change the electrostatic environment of an ionizable side chain, causing the protonation state (pK_a) of that site to vary over a wide range. This variation could be important because pK_a shifts have been linked to enzymatic catalysis, substrate binding, and protein-protein association (Warshel and Russel, 1984; Warshel and Åqvist, 1991; Antosiewicz et al., 1994; Honig and Nicholls, 1995). We have applied the single-site titration method (Antosiewicz et al., 1994) and calculated the pK_as along the two trajectories. Furthermore, we have applied the Z_{rf} criterion (Eq. 1) to the x-ray structures to determine which value of the dielectric constant is most appropriate. For the x-ray structures in the present investigation, the Z_{rf} values are close to 20, which indicates that a high dielectric constant should be applied. However, the resulting pK_a is overestimated when compared to the experimentally determined value. Therefore, we have performed two calculations of the pK_a's, using a low as well as a high dielectric constant. The ionization state of titratable groups may change upon substrate binding or be sensitive to structural inhomogeniety. To estimate the effects of both factors, we have calculated the pK_a values along the two molecular dynamics trajectories. The pK_as were estimated by calculating the pK_a every 20 ps of the 600-ps trajectory, using the single-site titration model (Antosiewicz et al. 1994). Calculations were performed with and without substrate present to estimate 1) the effect of the substrate on the pK_a of the active cysteine and 2) the influence of flexibility on the pK_a of the active cysteine. The time variations of the pK_a data for 1ptu with and without substrate present are shown in Fig. 8. The average pK_a's for the active-site Cys are 5.5 ± 0.8 ($\epsilon = 20$, with substrate), 2.4 ± 1.1 ($\epsilon = 20$, without substrate), 8.0 ± 0.2 ($\epsilon = 78.5$, with substrate), and 6.4 ± 0.2 ($\epsilon = 78.5$, without substrate). The pK_a computed in the presence of the substrate is in good agreement with the experimentally determined pK_a of 5.57, which is based on the k_{cat}/K_m versus pH rate profile (Lohse et al., 1997). The observed change in pK_a due to substrate binding stems from the electrostatic contribution of the substrate. In our calculations the methodology based on the Z_{rf} criterion overestimates the pK_a values. We would expect that the pK_a of the active Cys during the formation of the transition state is further reduced, because in the structure used in the simulations, the oxygens in the phosphate moiety are in a tetrahedral conformation, whereas upon the formation of the transition state, the oxygens take a planar conformation and form hydrogen bonds with the backbone hydrogen atoms of the p-loop (Jia et al., 1995). This effectively stabilizes the thiolate, further causing a lowering of the pK_a. Excluding the substrate from the calculations and varying the dielectric constant (Demchuk and Wade, 1996) lead to an over- or underestimation of the experimentally determined pK_a. Similar effects were found for the open structure. The calculations yield average pK_a's of 1.7 ± 1.3 ($\epsilon = 20$) and 6.5 ± 0.2 ($\epsilon = 78.5$). As observed for different proteins (Bashford et al., 1993; Warshel and Åqvist, 1991), protein dynamics can have a significant influence on the charge state of a amino acid. Structural variation close to the active Cys has been observed in *Yersinia* PTP (Juszczak et al., 1997), where the WPD loop alternates between the open and

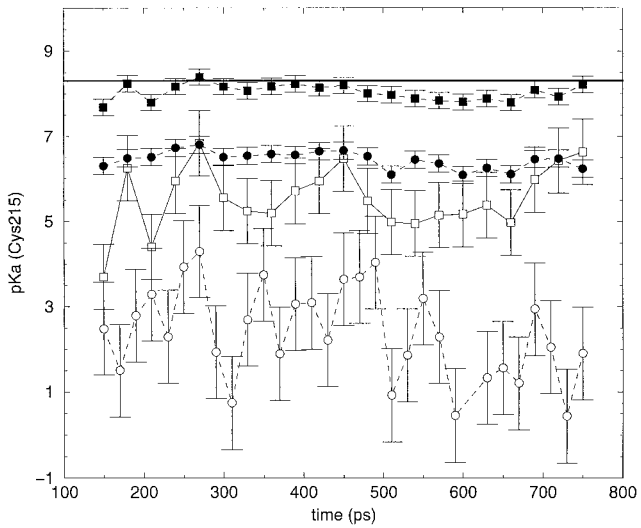


FIGURE 8 pK_a data of the active Cys in the 1ptu structure as a function of the simulation time. Data are shown for (□) substrate present, interior $\epsilon = 20$; (○) substrate excluded, interior $\epsilon = 20$; (●) substrate excluded, interior $\epsilon = 78.5$. The solvent dielectric constant was 80 in all calculations. The dielectric constant for the protein was chosen according to the single-site titration method (Antosiewicz et al., 1994) or the methodology proposed by Demchuk and Wade (1996). The solid line indicates the pK_a of a free Cys residue.

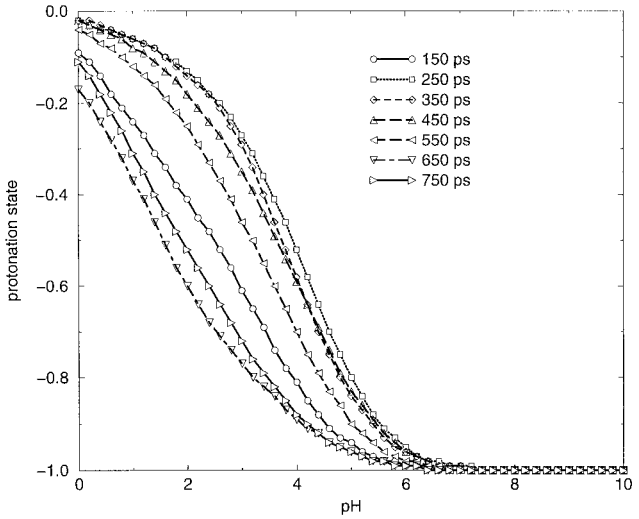


FIGURE 9 Protonation state of the active Cys as a function of pH computed at selected simulation time, as indicated in the inset. See Fig. 8 for more details.

closed forms with a rate constant of $\sim 2.6 \times 10^{-8}$ s^{-1} corresponding to a low energy barrier between the two loop conformations (Juszczak et al., 1997). As shown in Fig. 8, the pK_a fluctuates substantially at a relatively low dielectric constant, and the estimated standard deviation is ~ 1 pK_a unit. Similar pK_a shifts have also been found in a recent study of bacteriorhodopsin (Sandberg and Edholm, 1997), where it was demonstrated that the shifts were caused by the flexibility of the protein. Variations are also observed in the titrating behavior of the active cysteine. The protonation state as a function of pH is shown in Fig. 9 for selected structures extracted from the trajectory. The titration curves follow the usual S-shaped characteristics, indicating that there is no strong coupling (i.e., interactions) between the active cysteine and other titratable sites. Interactions larger than ~ 3 kcal/mol between titratable residues lead to complex titrating behavior, indicated by deviations from the typical S-shape curves (Sandberg and Edholm, 1997; Yang et al., 1993). As illustrated in Fig. 9, the fluctuations in the titration curves are significant. Changes in the protonation state of the active cysteine are mainly found at relatively low pH, and at a physiological pH the cysteine is negatively charged, as found experimentally (Lohse et al., 1997) and theoretically (Peters et al., unpublished observations; Hansson et al., 1997).

CONCLUSION

In the present study, we have investigated the concerted motions in the uncomplexed PTP1B and studied the effect of substrate binding (high-affinity peptide DADEpYL) on these motions by using molecular dynamics simulations. Simulations were carried out for 1 ns, and the resulting trajectories were used to analyze the change in concerted motions upon binding of the substrate as well as the electrostatic properties of the enzyme. The overall fluctuations as deduced from the geometrical properties are similar in PTP1B and the PTP1B-substrate complex. The essential dynamics analysis identified several regions in the protein that are affected by substrate binding. Conformational (static) changes are observed in several regions in the protein and are associated with the displacement of the WPD loop, which on activation moves toward the active Cys^{215}. Relatively large static shifts are seen in the surroundings of the WPD loop, causing pronounced structural changes in the loop region Val^{198}-Gly^{209}, located between the α-helix following the WPD loop and the β-strand at whose end the Cys^{215} is located. It appears that the movement of the WPD loop is facilitated by the structural rearrangements in the segment Val^{198}-Gly^{209}. Furthermore, analysis of the concerted motions in the subspaces spanned by the different eigenvectors revealed that the fluctuations in the distinct subspaces are similar and differences are localized in certain regions of the protein. The most pronounced effects are found along eigenvectors 2 and 3. Upon binding of the substrate, fluctuations in several regions are reduced, which can be regarded as an induced fit mechanism (Koshland, 1958). Considering the motions defined by the first 20 eigenvectors, the flexibility of the enzyme is reduced by $\sim 10\%$.

Calculation of the pK_a of the titratable amino acids along the trajectory shows that protein fluctuations ("structural inhomogeneity") cause variation of the pK_as on the order of 1 unit. The pK_a calculated in the presence of the substrate is in good agreement with the experimental determined pK_a, while excluding the substrate from the calculation results (depending on the dielectric constant) in an under- or overestimation of the active Cys pK_a. This could be due to the approximations used in the macroscopic approach, which neglects explicit water and counterions. The effect of these contributions is dependent on the enzyme (Bashford et al., 1993; Warshel and Åqvist, 1991), and in a recent study, it has been shown that in the pK_a calculation of bacteriorhodopsin counterions, membrane and the choice of partial charges did not influence the results significantly (Sandberg and Edholm, 1997).

Helpful discussions with Drs. J. Antosiewicz and N. P. H. Møller are acknowledged.

GHP was supported under grant 97 100 05 from the Danish Cancer Society.

REFERENCES

Allen, M. P., and D. J. Tildesley. 1989. Computer Simulation of Liquids. Clarendon, Oxford.

Almind, K., G. Inoue, O. Pedersen, and C. R. Kahn. 1996. A common amino acid polymorphism in insulin receptor substrate-1 causes impaired insulin signaling. *J. Clin. Invest.* 97:2569–2575.

Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins Struct. Funct. Genet.* 17:412–425.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238:415–436.

Balsera, M. A., W. Wriggers, Y. Oono, and K. Schulten. 1996. *J. Phys. Chem.* 100:2567–2572.

Barford, D. 1995. Protein phosphatases. *Curr. Opin. Struct. Biol.* 5:728–734.

Barford, D., A. J. Flint, and N. K. Tonks. 1994. Crystal structure of human protein tyrosine phosphatase 1B. *Science.* 263:1397–1404.

Bashford, D., D. A. Case, C. Dalvit, L. Tennant, and P. E. Wright. 1993. Electrostatic calculations of the side-chain pK_a values in myoglobin and comparison with NMR data for histidines. *Biochemistry.* 32:8045–8056.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Membrane Proteins: Structures, Interactions and Models. A. Pullman et al., editors. Kluwer Academic Publishers, Leiden University, Leiden, the Netherlands. 457–470.

Berndt, K. D., J. Beunink, W. Schröder, and K. Wüthrich. 1993. Designed replacement of an internal hydration water molecule in BPTI: structural and functional implications of a glycine-to-serine mutation. *Biochemistry.* 32:4564–4570.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rogers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer based archival file for macromolecular structure. *J. Mol. Biol.* 112:535–542.

Busca, R., M. Martinez, E. Vilella, P. Pognonec, S. Deeb, J. Auwerx, M. Reina, and S. Vilaro. 1996. The mutation $Gly^{142} \rightarrow Glu$ in human lipoprotein lipase produces a missorted protein that is diverted to lysosomes. *J. Biol. Chem.* 271:2139–2146.

Cho, H., R. Krishnaraj, E. Kitas, W. Bannwarth, C. T. Walsh, and K. S. Anderson. 1992. Isolation and structural elucidation of a novel phosphocysteine intermediate in the LAR protein tyrosine phosphatase enzymatic pathway. *J. Am. Chem. Soc.* 114:7296–7298.

Davis, M. E., J. D. Madura, B. A. Luty, and J. A. McCammon. 1991. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comp. Phys. Comm.* 62:187–197.

Demchuk, E., and R. C. Wade. 1996. Improving the continuum dielectric approach to calculating pKas of ionizable groups in proteins. *J. Phys. Chem.* 100:17373–17387.

Falzone, C. J., P. E. Wright, and S. J. Benkovic. 1994. Dynamics of a flexible loop in dihydrofolate reductase from *Escherichia coli* and its implication for catalysis. *Biochemistry.* 33:439–442.

Fauman, E. B., and M. A. Saper. 1996. Structure an function of the protein tyrosine phosphatases. *Trends Biochem. Sci.* 21:413–417.

Fisher, E. H., H. Charbonneau, and N. K. Tonks. 1991. Protein tyrosine phosphatases. *Science.* 253:401–406.

Gilson, M. K. 1993. Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins Struct. Funct. Genet.* 15:266–282.

Guan, K. L., and J. E. Dixon. 1991. Evidence for protein-tyrosine-phosphatase catalysis proceeding via a cysteine-phosphate intermediate. *J. Biol. Chem.* 266:17026–17030.

Hansson, T., P. Nordlund, and J. Åqvist. 1997. Energetics of nucleophile activation in a protein tyrosine phosphatase. *J. Mol. Biol.* 265:118–127.

Hecht, M. H., J. M. Sturtevant, and R. T. Sauer. 1986. Stabilization of λ repressor against thermal denaturation by site-directed $Gly \rightarrow Ala$ changes in α-helix 3. *Proteins Struct. Funct. Genet.* 1:43–46.

Henderson, H. E., F. Hassan, G. M. B. Berger, and M. R. Hayden. 1992. The lipoprotein lipase $Gly^{188} \rightarrow Glu$ mutation in South Africans of Indian descent: evidence suggesting common origins and an increased frequency. *J. Med. Gen.* 29:119–122.

Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.

Ichiye, T., and M. Karplus. 1991. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins Struct. Funct. Genet.* 11:205–217.

Jancso, A., and A. G. Szent-Györgyi. 1994. Regulation of scallop myosin by the regulatory light chain depends on a single glycine residue. *Proc. Natl. Acad. Sci. USA.* 91:8762–8766.

Jia, Z., D. Barford, A. J. Flint, and N. K. Tonks. 1995. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science.* 268:1754–1758.

Juszczak, L. J., Z.-Y. Zhang, L. Wu, D. S. Gottfried, and D. D. Eads. 1997. Rapid loop dynamics of *Yersinia* protein tyrosine phosphatases. *Biochemistry.* 36:2227–2236.

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

Kempner, E. S. 1993. Movable lobes and flexible loops in proteins. *FEBS Lett.* 326:4–10.

Koshland, D. E., Jr. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA.* 44:98–106.

Krueger, N. X., M. Streuli, and H. Saito. 1990. Structural diversity and evolution of human receptor-like protein tyrosine phosphatases. *EMBO J.* 9:3241–3252.

Lewis, S. D., F. A. Johnson, and J. A. Shafer. 1981. Effect of cysteine-25 on the ionization of histidine-159 in papain as Determined by proton nuclear magnetic resonance spectroscopy. Evidence for a His-Cys-25 ion pair and its possible role in catalysis. *Biochemistry.* 20:48–51.

Lohse, D. L., J. M. Denu, M. Santoro, and J. E. Dixon. 1997. Roles of aspartic acid-181 and serine-222 in intermediate formation and hydrolysis of the mammalian protein-tyrosine-phosphatase PTP1. *Biochemistry.* 36:4568–4575.

Madura, J. D., J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comp. Phys. Comm.* 91:57–67.

Mann, D. M. A., D. Jones, J. S. Snowden, D. Neary, and J. Hardy. 1992. Pathological changes in the brain of a patient with familial Alzheimer's disease having a missense mutation at codon 717 in the amyloid precursor protein gene. *Neurosci. Lett.* 137:225–228.

Matthews, B. W. 1987. Genetic and structural analysis of the protein stability problem. *Biochemistry.* 26:6885–6888.

Mourey, R. J., and J. E. Dixon. 1994. Protein tyrosine phosphatases of extracellular and intracellular domains. *Curr. Opin. Genet. Dev.* 4:31–39.

Pannifer, A. D. B., A. J. Flint, N. K. Tonks, and D. Barford. 1998. Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by x-ray crystallography. *J. Biol. Chem.* 273: 10454–10462.

Pathak, D., and D. J. Ollis. 1990. Refined structure of dienelactone hydrolase at 1.8 Å. *J. Mol. Biol.* 214:497–525.

Peters, G. H., D. M. F. van Aalten, O. Edholm, S. Toxvaerd, and R. Bywater. 1996. Dynamics of proteins in different solvent systems: analysis of essential motion in lipases. *Biophys. J.* 71:2245–2255.

Peters, G. H., D. M. F. van Aalten, A. Svendsen, and R. Bywater. 1997. Essential motions in enzyme binding sites and bound inhibitors: the effect of inhibitors of different chain length. *Protein Eng.* 10:148–156.

Philippopoulos, M., Y. Xiang, and C. Lim. 1995. Identifying the mechanism of protein loop closure: a molecular dynamics simulation of the *Bacillus stearothermophilus* LDH loop in solution. *Protein Eng.* 8:565–573.

Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comp. Physiol.* 23: 327–341.

Sandberg, L., and O. Edholm. 1997. pK_a calculations along a bacteriorhodopsin molecular dynamics trajectory. *Biophys. Chem.* 65:189–204.

Stone, R. L., and J. E. Dixon. 1994. Protein-tyrosine phosphatases. *J. Biol. Chem.* 269:31323–31326.

Stuckey, J. A., H. L. Schubert, E. B. Fauman, Z.-Y. Zhang, J. E. Dixon, and M. A. Saper. 1994. Crystal structure of *Yersinia* protein tyrosine phosphatase at 2.5 Å and the complex with tungstate. *Nature.* 370:571–574.

van Aalten, D. M. F., A. Amadei, R. Bywater, J. B. C. Findlay, H. J. C. Berendsen, C. Sander, and P. F. W. Stouten. 1996. A comparison of structural and dynamic properties of different simulation methods applied to SH3. *Biophys. J.* 70:684–692.

van Aalten, D. M. F., A. Amadei, A. B. M. Linssen, V. G. H. Eijsink, and G. Vriend. 1995. The essential dynamics of thermolysin—conformation of hinge-bending motion and comparison of simulations in vacuum and water. *Proteins Struct. Funct. Genet.* 22:45–54.

van Aalten, D. M. F., P. C. Jones, M. De Sousa, and J. B. C. Findlay. 1997. Engineering protein mechanics: inhibition of concerted motions of the cellular retinol binding protein by site-directed mutagenesis. *Protein Eng.* 10:31–37.

Van Gunsteren, W. F., and H. J. C. Berendsen. 1987. GROMOS: Groningen Molecular Simulation Computer Program Package. University of Groningen, Groningen, the Netherlands.

Vermersch, P. S., J. J. G. Tesmer, D. D. Lemon, and F. A. Quiocho. 1990. A Pro to Gly mutation in the hinge of the arabinose-binding protein enhances binding and alters specificity. *J. Biol. Chem.* 265: 16592–16603.

Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8:52–56.

Vulliamy, T. J., M. D'Urso, G. Battistuzzi, M. Estrada, N. S. Foulkes, G. Martini, V. Calabro, V. Poggi, R. Giordana, M. Town, L. Luzzatto, and M. G. Persico. 1988. Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. *Proc. Natl. Acad. Sci. USA.* 85:5171–5175.

Wade, R. C., M. E. Davis, B. A. Luty, J. D. Madura, and J. A. McCammon. 1993. Gating of the active site of triose phosphate isomerase: Brownian dynamics simulations of flexible peptide loops in the enzyme. *Biophys. J.* 64:9–15.

Wade, R. C., B. A. Luty, E. Demchuk, J. D. Madura, M. E. Davies, J. M. Briggs, and J. A. McCammon. 1994. Simulation of enzyme-substrate encounter with gated active sites. *Nature Struct. Biol.* 1:63–67.

Walton, K. M., and J. E. Dixon. 1993. Protein tyrosine phosphatases. *Annu. Rev. Biochem.* 62:101–120.

Warshel, A., and J. Åqvist. 1991. Electrostatic energy and macromolecular function. *Annu. Rev. Biophys. Biophys. Chem.* 20:267–298.

Warshel, A., and S. T. Russel. 1984. Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.* 17: 283–422.

Wilkinson, A. J., A. R. Fersht, D. M. Blow, and G. Winter. 1983. Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation. *Biochemistry.* 22:3581–3586.

Williams, J. C., and A. E. McDermott. 1995. Dynamics of the flexible loop of triosephosphate isomerase: the loop motion is not ligand gated. *Biochemistry.* 34:8309–8319.

Wo, Y.-Y., M.-M. Zhou, P. Stevis, J. P. Davis, Z.-Y. Zhang, and R. L. Van Etten. 1992. Cloning, expression, and catalytic mechanism of the low molecular weight phosphotyrosyl protein phosphatase from bovine heart. *Biochemistry.* 31:1712–1721.

Yang, A. S., M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig. 1993. On the calculation of pK_as in proteins. *Proteins Struct. Funct. Genet.* 15:252–265.

Yuvaniyama, J., J. M. Denu, J. E. Dixon, and M. A. Saper. 1996. Crystal-structure of the dual-specificity protein phosphatase vhr. *Science.* 272: 1328–1333.

Zhang, Z.-Y. 1990. Mechanistic and kinetic studies of the bovine heart low molecular weight phosphotyrosyl protein phosphatase. Ph.D. thesis. Purdue University, West Lafayette, IN.

Zhang, Z.-Y., D. Maclean, D. J. McNamara, T. K. Sawyer, and J. E. Dixon. 1994b. Protein tyrosine phosphatase substrate specificity: size and phosphotyrosine positioning requirements in peptide substrates. *Biochemistry.* 33:2285–2290.

Zhang, Z.-Y., B. A. Palfey, L. Wu, and Y. Zhao. 1995. Catalytic function of the conserved hydroxyl group in the protein tyrosine phosphatase signature motif. *Biochemistry.* 34:16389–16396.

Zhang, Z.-Y., Y. Wang, and J. E. Dixon. 1994a. Dissecting the catalytic mechanism of protein-tyrosine phosphatases. *Proc. Natl. Acad. Sci. USA.* 91:1624–1627.

Zhao, Y., and Z.-Y. Zhang. 1996. Reactivity of alcohols toward the phosphoenzyme intermediate in the protein-tyrosine phosphatase-catalyzed reaction: probing the transition state of the dephosphorylation step. *Biochemistry.* 35:11797–11804.