THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

The Influence of Ser-154, Cys-113, and the Phosphorylated Threonine Residue on the Catalytic Reaction Mechanism of Pin1

Esteban Vöhringer-Martinez,*[,†] Toon Verstraelen,[‡] and Paul W. Ayers[§]

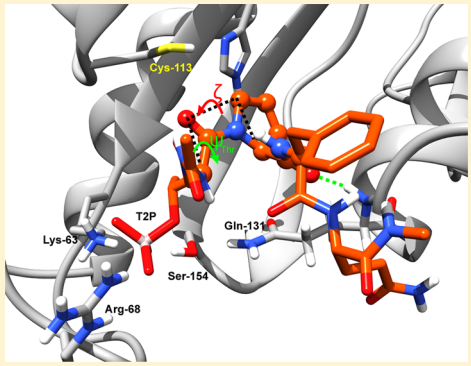
[†]Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, 4030000 Concepción, Chile

[‡]Center for Molecular Modeling (CMM), Ghent University, 9000 Ghent, Belgium (Member of the QCMM Ghent−Brussels Alliance)

[§]Department of Chemistry, McMaster University, 1280 Main Street West, Hamilton, Ontario L8 S4L8, Canada

Supporting Information

ABSTRACT: Pin1 is an enzyme that specifically catalyzes the *cis*−*trans* isomerization of proline amide bonds in peptides that contain a phosphorylated threonine or serine residue in the position preceding proline. In the cell, the isomerization reaction is associated with cellular signaling and has been related to diseases such as Alzheimer and cancer. The catalytic mechanism by which Pin1 accelerates the isomerization reaction, however, is still unknown. In this study, we use molecular dynamics simulation in combination with the QM/MM methodology to disclose the influence of the residues Ser-154 and Cys-113 in the enzyme and the phosphorylated threonine residue in the peptide on the reaction mechanism. To account for the correct electrostatic interaction between the three residues and the reactive center, we derive atomic charges that account for the varying electrostatic field in the catalytic cavity. Different methods based on reproducing the molecular electrostatic potential or an atoms in molecules approach were investigated. Finally, the reaction mechanism is analyzed with the mean reaction force and the influence of the three residues is disclosed. Our results show that Pin1 specifically catalyzes the isomerization of the *trans* conformer in a jump-rope type of motion, as suggested by us and confirmed experimentally by others. This is accomplished by anchoring the threonine phosphate residue on one end of the peptide through electrostatic interactions with the basic triad of the enzyme and at the other end through specific enzyme−peptide hydrogen bonds. Cys-113 reduces the structural contribution to the activation free energy through the stabilization of the *cis* conformer, and Ser-154 in combination with Gln-131 assist in the isomerization reaction of the *trans* isomer.



INTRODUCTION

Amide bonds in peptides that involve the nitrogen and carbon−carbonyl atom are produced by the ribosomes almost exclusively in the *trans* configuration in the cell. Crystal structures of proteins in combination with other experimental techniques, however, have provided conclusive evidence that some of these bonds—especially the ones in proline residues—present a *cis* conformation. Therefore, the newly formed proline peptide bonds have to isomerize from the *trans* to the *cis* conformation before the protein can adopt its biological function in the cell.

From experimental studies in bulk water, it is known that at room temperature this process is very slow owing to the large free energy barrier that separates the conformations.[1] The free energy barrier is even larger when a phosphorylated serine or threonine residue (pSer, pThr) is found in the position preceding the proline.[2]

Therefore, one might expect the cell to contain prolyl-isomerase enzymes that accelerate the interconversion. Pin1 has been identified to catalyze the isomerization reaction, being highly selective for peptides with the aforementioned pThr/pSer-Pro motif.[3,4] The ability of Pin1 to switch between the two conformations has been associated with its function as a molecular timer for cellular signaling[5] and also with various diseases such as Alzheimer[6,7] and cancer.[8] Recently, the Pin1 catalyzed isomerization reaction of the C-terminal end of RNA polymerase II, the molecule for eukaryotic transcription and most significant substrate of Pin1 in humans, has been found to be coupled with the phosphorylation step, thereby contributing to the mechanism by which cells regulate gene expression.[9]

Despite the importance of Pin1 in the regulatory mechanism of the cell, the exact catalytic mechanism of the isomerization reaction has only been studied experimentally and theoretically recently. After the pioneering work of Prof. Fischer on the characterization of Pin1 and its role,[2−4,10,11] the group of Prof. Etzkorn has unveiled important interactions between reactant, transition state, and product peptide analogues and Pin1 by X-ray crystallography and NMR spectroscopy.[9,12−15] Recently, new insights into the mechanism were provided by kinetic isotope studies.[16] Earlier mutational studies addressed the effect

Received: June 6, 2014
Revised: July 24, 2014
Published: July 24, 2014

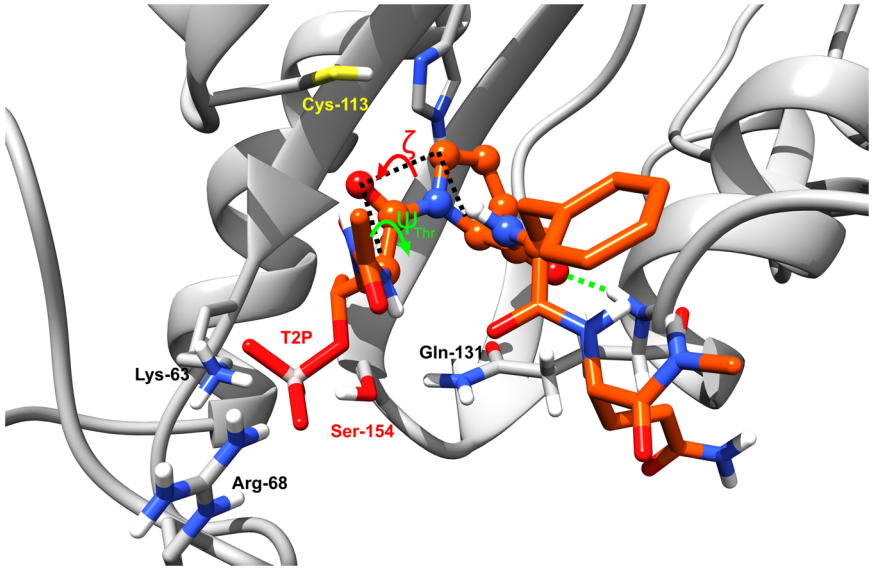


Figure 1. Snapshot of the catalytic active site of human Pin1 with the Ace-Gly-TPO-Pro-Phe-Gln-Nme peptide (orange) and the dihedral angle ζ representing the reaction coordinate together with the ψ_{TPO} bond which has to be rotated for the activation of the *trans* form. Atoms shown as spheres represent the QM region in the QM/MM calculations.

of a special residue (Cys-113) that was proposed in the first X-ray structure.[17,18] From the theory side, accelerated molecular dynamics simulations have also contributed to the qualitative understanding of the reaction mechanism in Pin1 and in solution, yielding a guiding map on how the isomerization may occur.[19,20] Recently, we refined this map by extensive QM/MM MD simulations in combination with the mean reaction force providing quantitative free energy barriers for the reaction in solution and in the enzyme and also insights on their origins.[21]

The selectivity of Pin1 and important aspects of its catalytic reaction mechanism are reflected in its catalytic cavity located in the peptidyl-prolyl isomerase (PPIase) domain, as shown in Figure 1 (a schematic poseview representation of the relevant enzyme−peptide interactions is shown in Figure 1 in the Supporting Information). The basic triad formed by residues Lys-63, Arg-68, and Arg-69 (not shown because of its flexible character) stabilizes the negative charged phosphate group of the pThr-Pro motif (TPO). Hydrogen bonds between the proline carbonyl oxygen atom and the backbone amide group of Gln-131 in the enzyme (green dotted line) and the amino backbone group of the glutamine residue of the peptide and the carbonyl oxygen atom of the Gln-129 residue keep this part of the peptide essentially fixed, allowing for the rotation only of the C-terminal part of the peptide (left side of the peptide in Figure 1). The isomerizing peptide bond forms an intramolecular hydrogen bond between its nitrogen atom and the amide hydrogen atom of the next but one residue (phenylalanine). This interaction has already been proposed as one factor contributing to catalysis due to the stabilization of the sp^3 character of the nitrogen atom at the transition state.[22,23] In our previous study, we were able to show that this interaction is also present in bulk water, excluding it as the main origin of the enzyme catalysis.[21]

On the basis of the first reported crystal structure,[24] Cys-113 was proposed to catalyze the reaction in the enzyme by a nucleophilic attack on the carbonyl carbon atom. However, mutational studies and the recent kinetic isotope study have ruled out this mechanism.[16−18] In our recent contribution, we have identified Cys-113 to play a role in the first part of the reaction stabilizing the carbonyl group in the initial rotation. This interaction was also confirmed experimentally by the crystal structures of *cis* isosteric compounds and the loss of flexibility of Cys-113 upon binding.[9,12] In our previous work, we were able to identify one additional residue, Ser-154, that interacts with the carbonyl oxygen atom via a hydrogen bond near the transition state.[21] This interaction was also reported in a previous crystal structure of a reduced amide inhibitor that resembled the twisted amide transition state.[15] More recently, Velazquez et al. found by accelerated MD simulations that the interaction of Ser-154 in the transition state and in the *cis* form is predominantly with the phosphate group of the threonine residue but in the *trans* form with the carbonyl oxygen atom.[19]

In this study, we focus on the specific role of Cys-113, Ser-154, and the phosphorylated (TPO) residue in the catalytic mechanism. We use QM/MM MD simulations to explicitly account for the electron reorganization during the isomerization reaction and follow the reaction coordinate assuring proper equilibration of the system from the *cis* reactant through the transition state to the *trans* product.

The influence of the residues on the reaction mechanism depends critically on how they interact with the rotating bond. The largest contribution originates in the QM/MM simulations from the electrostatic interaction represented by their atomic charges taken from the employed force field. Therefore, in this contribution, we propose a new method where the atomic charges of the residues are first calculated from their electron density in the protein environment to reproduce the molecular electrastic potential or in an atoms in molecules (AIM) approach. Additionally, to account for the varying electrostatic field from the dynamics of the neighboring solvent and protein, the charges are updated "on the fly" in molecular dynamics simulation until their convergence.

Finally, the influence of the residues on the reaction mechanism is addressed by the mean reaction force which, as we have shown previously, is able to provide structural and electronic contributions to activation free energies in enzyme catalysis and other molecular environments.[21,25]

dx.doi.org/10.1021/jp505638w | J. Phys. Chem. B 2014, 118, 9871−9880

METHODS

1. Methods for the Calculation of Atomic Charges. The performance of different procedures to derive atomic charges for force fields was analyzed distinguishing two kinds of widely used methods. In the first one, the atomic charges are calculated to reproduce the molecular electrostatic potential (MEP) of small model systems in vacuum. The MEP is obtained from electronic structure calculations employing different molecular surfaces (Chelpg surface and the surface by Conolly, ESP-C1, ESP-A1 in the Supporting Information). In the past, however, it has been shown that these methods produce unphysical high charges on "buried atoms"—atoms that are distant from the surface—since the electrostatic potential on the molecular surface used for their derivation becomes independent of their values. To avoid this problem, it has been proposed to apply an additional hyperbolic restraint on the charges of the heavy atoms in a second step of the charge derivation procedure. These methods have been used and are recommended for new parametrizations of the well-known Amber and Gaff force fields.[26] The atomic charges for these methods were calculated with the Red-III.5 software package[27] (for further details, see the Supporting Information).

In the second category, the atomic charges are computed with different atoms in molecules (AIM) approaches. Hirshfeld suggested to decompose the molecular electron density, $\rho(\mathbf{r})$, into atomic contributions by means of given pro-atomic densities, ρ_A^0, using the Stockholder formula:[28]

$$\rho_A(\mathbf{r}) = \rho(\mathbf{r})\frac{\rho_A^0(\mathbf{r})}{\sum_{B=1}^{N}\rho_B^0(\mathbf{r})} \tag{1}$$

The AIM charge is then obtained as usual, by adding the charge of the nucleus and the electrons assigned to the atom:

$$q_A = Z_A - \int \rho_A(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{2}$$

One can show that this definition corresponds to the minimum of the Kullback−Leibler divergence between the AIM density and the pro-atom;[29,30] i.e., the AIM densities are maximally similar to the pro-atom densities. Also, other AIM methods using geometric considerations (e.g., DMA[31]) or the topology of the electron density (e.g., QTAIM[32]) have been proposed.

Hirshfeld suggested to use spherically averaged densities of isolated atoms as pro-atoms. This has been shown to yield charges that are relatively small.[33] Furthermore, one may as well use other pro-atoms, e.g., derived from isolated cations or anions, leading to vastly different numbers. These weaknesses were addressed by Bultinck et al.,[34] who proposed an iterative approach, Hirshfeld-I, where the charge of each pro-atom is updated until it becomes consistent with the AIM charge. Pro-atoms with fractional charges are obtained from a linear interpolation between the densities of two isolated atoms with an integer charge.[35,36] The Hirshfeld-I method has been tested over the last years and has presented some remarkable characteristics: the charges are only weakly dependent on the basis set,[37] and they reproduce the electrostatic properties of the molecule.[38] Additionally, the charges are almost independent of the conformation which is a major advantage over some other methods, such as unrestrained ESP-fitting and Mulliken population analysis.[39] All atomic charges of the Hirshfeld derived methods were calculated with the HORTON 1.2.0 software package.[40]

The atomic charges of the model systems, ethanol (Ser-148), ethanthiol (Cys-113), and 2-propanylphosphate (TPO), were calculated with these methods at the HF/6-31G(d) level. Bigger basis sets and other methods including electron correlation were not considered to maintain the compatibility with charges of the Amber99sb force field which were derived at the same level of theory. Initial structures were taken from the side chains of the residues in the crystal structure (PBD-ID: 2Q5A), saturated with hydrogen atoms and optimized prior to the derivation of their charges.

2. Atomic Charges in the Protein Environment. For the calculation of the atomic charges in the protein environment for each residue, an independent calculation was performed in which the residue up to the α-carbon was described quantum mechanically at the HF/6-31G(d) level and the rest of the protein with the Amber99sb force field. For the covalent bond connecting the QM α-carbon and the MM atoms, the link atom method was used. The methods for the derivation of the atomic charges applied were the ESP method as implemented in the ORCA software package with the COSMO radii, a grid spacing of 0.3 Å, and a cutoff of 2.8 Å and the Hirshfeld-I[34] and Hirshfeld-E methods[41] as implemented in the HORTON 1.2.0 software package.[40] ESP charges to represent the interactions between the QM region and the MM atoms were already proposed in previous studies to speed up QM/MM simulations.[42,43]

Initially, the minimized, solvated protein X-ray structure was used without the inclusion of the force field charges (mechanical embedding). In a second step, the surroundings were considered including the atomic charges of the protein and the solvent in a radius of 1.5 nm (electrostatic embedding). Finally, to account for the solvent and protein dynamics which produce a varying electrostatic field around each residue, the charges were obtained as an average from molecular dynamics simulations. This average value was obtained by first equilibrating the protein and the solvent employing the standard Amber99sb force field and the Sticht[44] parameters for the threonine phosphate. Then, new charges were calculated for each method "on the fly" in a QM/MM calculation and the old charges of the respective QM atoms were replaced prior to the continuation of the simulations. This procedure was repeated after a variable number of integration steps in molecular dynamics simulation. This routine was implemented in a modified version of the GROMACS package that performs the QM/MM calculation and replaces the charges automatically in a MD simulation. The convergence of the charges was investigated by varying the total simulation time from 500 ps to 3 ns (see Tables 11−13 in the Supporting Information). Also, the number of integration steps between new charge calculations was varied but no dependence for time steps longer than 1 ps was observed. The obtained charges were named **D-ESP**, **D-HI**, and **D-HE** to account for the dynamics used for their derivation.

3. Molecular Dynamics Simulations. All MD simulations were performed with the Gromacs 4.5.3 software package.[45] For a detailed overview of the employed parameters, the reader is referred to our previous study.[21]

The QM/MM molecular dynamics simulations were performed in combination with the ORCA package[46] in an electrostatic embedding, as incorporated in the Gromacs 4.5.3 package. In these simulations, the electrostatic interactions in the MM region were treated with the reaction field method (cutoff radius = 1.0 nm, switching radius = 0.9 nm, ε_r = 78) and

dx.doi.org/10.1021/jp505638w | J. Phys. Chem. B 2014, 118, 9871−9880

the time step was reduced to 1 fs. The neighbor list was updated at every step.

4. System Setup. A detailed description of the system setup can be found in ref 21. To avoid any bias from the absence of the missing residues 39−50 in the crystal structure, here they were modeled with the xleap program of the antechamber module of AMBER12. The threonine phosphate residue was described by Craft et al.[47] parameters as in our previous study but also compared to parameters provided by Sticht[44] that only introduce a change in the OS-P bond in the bonded part of the force field. The atomic charges derived by Sticht et al., however, follow the general procedure recommended for the parametrization of new molecules for the Amber99 force field (RESP HF/6-31G*) and therefore should be more compatible with the rest of the peptide and the enzyme.

For each charge set (Craft, Sticht and D-HI/E), the whole system consisting of the Pin1 enzyme, the peptide, and the surrounding water and ions was equilibrated for 100−150 ns starting from the minimized crystal structure until a constant backbone RMSD was reached.[21]

5. Reaction Mechanism. The reaction mechanism was studied starting from the respective equilibrated structure in the *cis* form and following the reaction coordinate represented by the angle ζ (see Figure 1), as described in our previous study.[21] It is important to note that the rotation of the amide bond was first followed at the force-field level equilibrating the system always 5 ns after 10 degrees of rotation. This should result in a proper relaxation of the perpendicular degrees of freedom and proper sampling.

As in our previous study and in accordance with the results of Velazquez et al.,[19,20] a second reaction coordinate, described by the ψ_{Thr} angle of the threonine phosphate residue (see Figure 1), was introduced once 140° was reached. The second reaction coordinate was needed to avoid unfavorable interactions with the enzyme. The simulation procedure was the same and the chosen ζ and ψ_{Thr} values were 140/90, 150/70, 160/50, 160/30, 160/10, 160/−10, 160/−30, 170/−30, 180/−30, and 180/−50.

From the MM simulations, three structures were taken from the last 4 ns along the reaction coordinate and served as input for the QM/MM simulations. The QM region was the same as in our previous study involving all atoms in the amino acid sequence starting at the C_α atom of threonine phosphate until the N−H group of phenylalanine (in total 20 atoms represented as balls in Figure 1). Each window was simulated 15 ps saving a frame every 0.05 ps and the last 10 ps were used to obtain the mean reaction force (MRF). The QM calculations were performed with the HF/3-21G and the link atom method in an electrostatic embedding. To test the accuracy of the QM method and the possible impact of electron correlation, the MRF of one of the three starting structures was also calculated with the BLYP/SVP method.

To elucidate the mechanistic features and the impact of the electrostatics on the reaction mechanism, the mean reaction force (MRF) that has proven to unveil valuable mechanistic insights in catalysis[25] was used. It is defined as the negative derivative of the free energy along the reaction path (at constant pressure and temperature):

$$\langle F(\xi)\rangle = -\frac{dG}{d\xi} \tag{3}$$

As we have shown previously, the mean reaction force represents a sensitive tool to study the influence of the environment on reaction mechanisms.[21,25] It provides a direct access to the characterization of different processes taking place along the path. The minimum and maximum of the mean reaction force divide the reaction path in three regions: the reactant, transition state, and product region. The reactant and product region involve mostly structural rearrangements of the atoms to reach the transition state from either the reactant or the product, whereas the transition state region involves mainly changes in the electron configuration required to make or break chemical bonds.

These regions allow one to divide the activation energy $\Delta G^{\ddagger}$ into two contributions, $\Delta G^{\ddagger} = [G(\xi_{TS}) - G(\xi_R)] = W_1 + W_2$. W_1 is mostly structural and W_2 encompasses mostly electronic contributions.

The MRF was obtained along the reaction coordinate with the umbrella integration method of Kästner et al.,[48,49] as reported previously.[21]

RESULTS AND DISCUSSION

To study the influence of Cys-113, Ser-154, and the threonine phosphate (TPO) residue on the reaction mechanism, we first have to guarantee that their interaction with the reactive center (amide bond) is described correctly. The largest contribution stems from the electrostatic interaction, and this is described by the atomic charges that are included in the QM/MM calculations. Because the atomic charge is not a proper quantum-mechanical observable, different methods are used for their derivation.[50] In this first part, we compared a selection of methods on small model systems of the residues in the gas phase and in the protein environment.

Atomic Charges of Cys-113, Ser-154, and the Threonine Phosphate (TPO) Residue. For the derivation of the atomic charges, two kinds of methods were studied. In the first one, the atomic charges are calculated to reproduce the molecular electrostatic potential (MEP), and in the second one, the charges are obtained with different atoms in molecules (AIM) approaches (for details, see the Methods section). Specifically, in the first category, different molecular surfaces were considered and additional hyperbolic restraints on the heavy atoms were added in the charge evaluation procedure to reduce the unphysical high charges on "buried atoms", whereas for the second kind the original Hirshfeld[28] and the Hirshfeld-I[37] were taken into account.

Recently, a variant of the original Hirshfeld method was proposed by Truhlar as CM5.[51] This method corrects the original Hirshfeld scheme to reproduce molecular dipoles of organic molecules in the gas phase. Albeit the good performance in the reproduction of the dipole moment, it also displayed conformational independence and a small basis set dependence in agreement with the Hirshfeld-I method. Additionally, in a recent study, small organic molecules described by scaled CM5 charges were able to produce solvation free energies with errors of approximately 1 kcal/mol with respect to the experimental values.[52]

Therefore, we expect Hirshfeld derived charges to be adequate for the derivation of atomic charges of classical force fields and also for their application in QM/MM simulations.[53]

First, we tested the performance of the mentioned methods to compute charges on small model systems in the gas phase: ethanol (Ser-109), ethanthiol (Cys-113), and 2-propylphos-

dx.doi.org/10.1021/jp505638w | J. Phys. Chem. B 2014, 118, 9871−9880

phate (TPO) (for details, see the Methods section and the Supporting Information). For the methods employing the molecular electrostatic potential (MEP), the main conclusions are that the atomic charges varied only marginally with the molecular surfaces employed for their calculation, showing the largest variations on carbon atoms. This may be related to the "buried atom" problem that stems from ill-conditioned derivation of charges distant from the MEP. When a restraint on the carbon atoms is applied (RESP-A1, RESP-C1), large changes in the atomic charges of polar groups as the hydroxyl, sulfuryl, or the phosphate group are observed. As a consequence, the resulting dipole moments (ethanol and ethanthiol) deviate considerably from the one obtained from the electronic structure calculation, raising doubts about the ability of this method to reproduce higher order moments of intermolecular interactions. For the second category, where the atomic charges are derived directly from the molecular electron density, we could confirm the too small values for the Hirshfeld atomic charges. The Hirshfeld-I charges, however, reproduced very nicely the trend observed in the charges from the electrostatic potential and also the dipole moments. In the case of 2-propylphosphate, the Hirshfeld-I method predicted large negative charges on the oxygen atoms and large positive ones on the phosphorus atom. Verstraelen et al.[41] had already shown in their work on silica clusters that Hirshfeld-I results in large negative charges on the oxygen atoms, because the atomic density of the O^{2-} anion is used as a pro-atom. This dianion, however, is unstable, and the atomic density is only bounded by the basis set employed in the calculation. Therefore, for this model system, we also tested the Hirshfeld-E method proposed by Verstraelen et al.[41] As expected, the Hirshfeld-E method reduced the magnitude of the atomic charges in the phosphate group.

From the results of the model systems, we decided to focus only on the three methods: ESP where the charges are derived directly from the electrostatic potential without applying restraints on the charges of heavy atoms and the Hirshfeld-I (HI) and Hirshfeld-E (HE) methods.

To evaluate the charges for Cys-113, Ser-154, and TPO inside the protein, we only took the atoms in the amino acid side chain into account starting at the α-carbon and employing the link atom technique to saturate its valence. The atomic charges of the side chain atoms of each residue were derived from HF/6-31G(d) electronic structure calculations using the methods described above in three different ways: (i) employing the position of the atoms as found in the solvated, minimized X-ray structure of the protein neglecting all interactions with the surroundings; (ii) for the same structure but including the point charges of the surrounding atoms (QM/MM, electrostatic embedding); (iii) with a new implemented method where the charges are calculated as an average from molecular dynamics (MD) simulations. In these MD simulations after a variable number of steps, a QM/MM calculation is invoked to derive new atomic charges that replace the old ones prior to the continuation of the dynamics (for details, see the Methods section and the Supporting Information). All charges obtained with this procedure will be labeled with a **D-** standing for dynamics, e.g., D-ESP, D-HI, or D-HE. We are aware that updating the charges in a MD simulation with a classical force field does not produce a correct ensemble, since the Hamiltonian is changed constantly. However, assuming that the dynamics of the system is not coupled to the variation of the charges and that the procedure serves only for a refinement of the charges rather than producing a correct ensemble from which to obtain physical observables, we do think that the average represents the average charge of the atom in the specific environment.

The results for each residue and method are summarized in the Supporting Information. Comparison with the atomic charges of the model systems shows that the charges derived neglecting interactions to the surrounding atoms (i) match the previous results. This confirms that the introduced link atom and the structure in the protein correspond to the minimum structure. However, when the surrounding atoms are added as point charges in the QM/MM methodology, an increase in the absolute value is observed due to a polarization of the electron density. This polarization is largest at the link atoms (D1 and D2) and the polar groups as the hydroxyl, sulfuryl, and phosphate group. For the hydrogen link atoms, the origin of the polarization is the neighboring amine or carboxyl group along the backbone. The polarization of the polar groups, however, is mostly ascribed to the surrounding water molecules and protein atoms.

It is also interesting to see how the charges obtained as an average from the MD simulation (case iii) compare to the charges from only one minimized structure (case ii) (see the Supporting Information). The results in all cases clearly show that one structure is not enough to capture the average electrostatics of the protein and the solvent. The charges of only one structure even do not lie in the range of the standard deviation of the average charges. This confirms the fact that different structures have to be taken into account to describe the correct average electric field to which the electron density of the atoms is exposed.

To discuss the method dependence of the obtained averaged charges over the MD simulation, we show in Table 1 the atomic charges obtained for the threonine phosphate residue. From the table, one can conclude that D-ESP charges depend on the conformation which is reflected in the higher standard deviation but also in unphysical values for the structures of the minimized solvated X-ray structure (see the Supporting Information). The D-ESP charges also suffer from the "buried"

Table 1. Atomic Charges (±Standard Deviation) in Elementary Units of the Side Chain Atoms of Threonine Phosphate Calculated with Different Methods from a 500 ps MD Simulation in the Protein Environment (D1 and D2 Are the Link Atoms in the QM/MM Calculation)

	D-ESP	D-Hirshfeld-I	D-Hirshfeld-E
C_α	−0.48 ± 0.08	−0.43 ± 0.03	−0.17 ± 0.02
H(D1)	0.35 ± 0.04	0.38 ± 0.03	0.32 ± 0.02
H(D2)	−0.17 ± 0.03	−0.21 ± 0.02	−0.27 ± 0.02
H	0.11 ± 0.03	0.11 ± 0.01	0.03 ± 0.02
C_β	0.89 ± 0.08	0.60 ± 0.03	0.48 ± 0.02
H	−0.08 ± 0.03	−0.05 ± 0.02	−0.06 ± 0.02
C	−0.64 ± 0.07	−0.47 ± 0.02	−0.25 ± 0.02
H	0.13 ± 0.03	0.10 ± 0.01	0.04 ± 0.01
H	0.15 ± 0.03	0.10 ± 0.01	0.05 ± 0.02
H	0.11 ± 0.02	0.10 ± 0.02	0.04 ± 0.01
O	−0.70 ± 0.05	−0.90 ± 0.03	−0.73 ± 0.02
P	1.75 ± 0.08	3.14 ± 0.12	2.46 ± 0.08
O	−1.16 ± 0.03	−1.51 ± 0.04	−1.31 ± 0.03
O	−1.14 ± 0.03	−1.47 ± 0.05	−1.34 ± 0.03
O	−1.12 ± 0.03	−1.48 ± 0.04	−1.28 ± 0.03

dx.doi.org/10.1021/jp505638w | J. Phys. Chem. B 2014, 118, 9871−9880

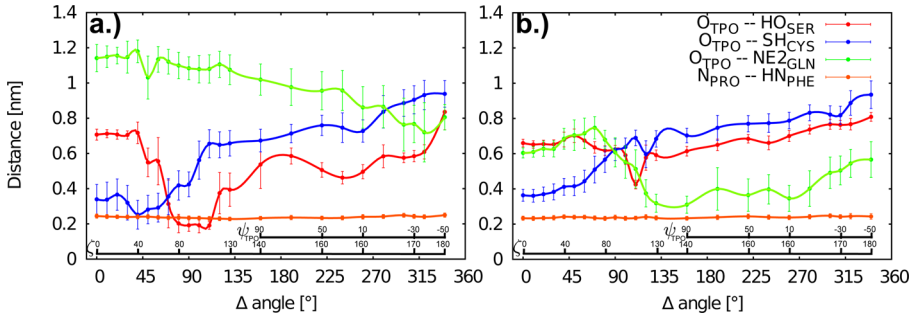


Figure 2. Relevant enzyme−substrate interactions (Ser-154, Cys-113, and Gln-131 in the enzyme and the carbonyl oxygen atom of the rotating amide bond in the peptide, see Figure 1) and the intramolecular hydrogen bond in the peptide for the simulation setup (a) Craft and (b) with the optimized charges with the D-HI/E method. The interactions are shown as a function of the reaction coordinate for the isomerization reaction that includes the rotation of the ζ and ψ angles shown in Figure 1.

atom problem for the carbon and the phosphorus atom. The charges of these atoms that are distant from the molecular surface where the electrostatic potential is evaluated are nearly arbitrary, and very sensitive to unimportant details in the fitting procedure.

The problem of the Hirshfeld-I charges with respect to the phosphate group (already described in the previous section) persists and is even enhanced due to the polarization. Despite the phosphate group, however, one can observe by analysis of the last two columns that both methods, D-HI and D-HE, present very small standard deviations on all atoms. Additionally, hydrogen atoms as, e.g., in the CH_3 group have the same charge in accordance with chemical intuition. The small difference in the oxygen atoms of the phosphate group is due to specific interactions in the catalytic cavity that make the three atoms inequivalent with respect to their surroundings.

For the phosphorylated threonine residue, one can conclude that the D-HE method presents the best method, since it corrects the failures of D-HI and it does not suffer from the buried atom problem due to its derivation from the local electron density. Additionally, it also yields equal charges for equivalent atoms, as can be noted in the CH_3 group.

To test if the charges are converged with respect to the length of the simulation, we extended the simulation time to 3 ns but only a small deviation of the charges could be observed that lay in the standard deviation (see Tables 11−13 in the Supporting Information).

According to the results described above, we decided to take the D-HI charges for Cys-113 and Ser-154 and the D-HE ones for the threonine phosphate residue from a 3 ns trajectory for the study of the influence of the residues on the reaction mechanism. However, since the link atoms suffer from an unphysical polarization due to the presence of the covalent bonded MM atoms included as point charges, we decided to only replace the charges of the polar groups (−SH or −OH) in the case of Cys-113 and Ser-154. The difference between the sum of these charges and the charges provided by the Amber99sb force field was distributed equally between the oxygen/sulfur and hydrogen atom to retain the neutrality of the system. Only changing the charges on the polar group has an additional benefit, since the internal dihedral rotation of the residues is not altered due to the presence of the new charges. Threonine phosphate, however, is not part of the standard Amber force field. Therefore, we decided to replace all charges from the β-carbon onward and the difference in the sum of the old and new charges was distributed equally between the β carbon and its hydrogen atom. The complete set of charges is provided in the Supporting Information.

Influence of Cys-113, Ser-154, and Threonine Phosphate on the Catalytic Reaction Mechanism of Pin1. The influence of the three residues on the catalytic reaction mechanism of Pin1 was studied employing three different setups for the molecular dynamics simulations. The first one uses the parameters in our previous study which were given by Craft et al. for the threonine phosphate residue.[21] The second one uses the parameters of Sticht,[44] who followed the standard parametrization procedure employing the RESP method, as suggested for the Amber force field. Finally, the last setup replaces the RESP charges of Sticht et al. using the values reported in the previous section and also changes the atomic charges of the sulfuryl and hydroxyl group for Cys-113 and Ser-154. Accordingly, in the following, the three setups will be called *Craft*, *Sticht*, and *D-HI/E*.

For each setup, the entire solvated protein with the peptide was first equilibrated in the *cis* form for 100−150 ns until a constant backbone RMSD was obtained. Then, applying a harmonic restraint on the ζ angle (see Figure 1), the peptide bond was rotated from the *cis* conformation until 140° where the ψ angle of the threonine phosphate was also restrained as described in the Methods section and in accordance with the minimum free energy path reported by Velazquez et al.[19,20] The whole rotation until the *trans* form was first performed at the force-field level with 5 ns equilibration after a 10° rotation. This ensures a proper equilibration of the system during the rotation.

All relevant interactions between the peptide and the enzyme were analyzed along the reaction path for each setup. Figure 2 shows the average distances between the oxygen and sulfur atoms of Cys-113 and Ser-154 and the oxygen atom of the rotating carbonyl group in the peptide for the setups *Craft* (a) and *D-HI/E* (b). Additionally, Figure 2 also displays the intramolecular distance between the nitrogen atom of the rotating bond and the amino backbone group of the phenylalanine residue in the peptide, which remains constant during the isomerization process in both systems. This intramolecular distance was already reported in our previous study to be important for the reaction mechanism but also present in solution. Therefore, it may not be the origin of the enzyme catalysis.

From the comparison of the distances involving Cys-113 for the two setups, one can conclude that in both cases this residue interacts with the oxygen atom of the rotating carbonyl bond in the beginning of the reaction, while the interaction is stronger

dx.doi.org/10.1021/jp505638w | J. Phys. Chem. B 2014, 118, 9871−9880

in the *Craft* system. For both systems, the interaction vanishes around 50°. The distance between the hydroxyl oxygen atom of Ser-154 and the rotating carbonyl group, however, differs in the two setups. For the *Craft* setup, there is a strong interaction after 50° which persists after the transition state, whereas this interaction is absent in the *D-HI/E* system. In this simulation setup, Ser-154 interacts rather with the phosphate group. Its position is taken over by Gln-131 but only after the transition state at 90°, as is also shown in Figure 2. The distances for the setup *Sticht* are in general similar to the ones in the *D-HI/E* setup but present some variations in the interaction of Ser-154 after 130° (see the Supporting Information). Therefore, one can conclude that the parameters from Sticht[44] describe the enzyme–substrate interactions better than the Craft parameters[47] but that the improved atomic charges of TPO, Ser-154, and Cys-113 are needed to describe them correctly.

One may ask what the effect of the weak interaction to Cys-113 and the absence of any interaction to Ser-154 in the *D-HI/E* system would be on the reaction mechanism. From our previous study and the mean reaction force analysis, one can anticipate that a weak interaction with Cys-113 would increase the structural contribution to the free energy barrier and, therefore, also the total free energy barrier. The influence of Ser-154 can be rationalized with the two resonance structures in Figure 3. The presence of a hydrogen bond to Ser-154 before

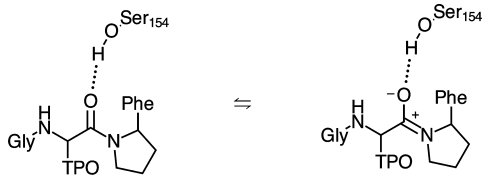


Figure 3. Scheme with resonance structures showing the influence of a hydrogen bond to Ser-154 on the peptide undergoing isomerization.

reaching the transition state would stabilize the double bond character of the rotating bond and increase the electronic contribution to the free energy barrier. Indeed, we found a larger electronic contribution to the total free energy barrier in our previous study with the *Craft* setup confirming our hypothesis. Therefore, the absence of this interaction in the *D-HI/E* system should decrease the electronic contribution to the total free energy barrier.

All these effects can now be analyzed in detail with the mean reaction force (MRF) that is shown for each setup in Figure 4. The increased structural contribution due to the weaker interaction to Cys-113 prior to the minimum is reflected by the larger MRF value of the *D-HI/E* setup in red with respect to the *Craft* setup in blue (for comparison with the MRF profile of the *Sticht* setup, see the Supporting Information). The minimum is slightly shifted to smaller values of the angle, resulting in an electronic contribution to the free energy barrier that equals the structural one for the *D-HI/E* system. This can be compared to the ratio of 61% to 39% electronic to structural contribution in the *Craft* setup. Therefore, the absent interaction with Ser-154 reduces the electronic contribution to the free energy barrier but at the expense of the structural contribution so that a larger total free energy barrier is obtained (see Figure 4, right side).

It is also clear from Figure 4 that for the reverse process the small barrier found for the *Craft* system almost disappears such that the *trans–cis* isomerization reaction would be a barrier-less process in the presence of the enzyme.

To exclude that the barrier and the energetics depend on the level of theory used for the QM calculation (HF/3-21G), we also calculated the mean reaction force starting from one starting structure employing the BLYP/SVP method that includes electron correlation and a bigger basis set (see the Supporting Information). However, the MRF profiles were the same despite a small variance that could be ascribed to slightly different dynamics and were also present in the other simulations.

Since we started from a crystal structure containing the *cis* form, we wanted to check if our *trans* product is correct. However, until now, no crystal structure of the *trans* form has been isolated. The absence of a crystal structure for the *trans* form may suggest that this form is only present in small population, which agrees with the reported lower free energy of the *cis* isomer and the smaller binding affinity of the *trans* form as noted below. Additionally, the almost barrier free isomerization process from the *trans* form would make the isolation of this bound isomer difficult to achieve. However, there is a crystal structure of a *trans* isosteric compound reported by Zhang et al.[9] (PDB-ID: 3tdb). To compare our simulation product, we took the last structure obtained from the isomerization reaction with the *Sticht* parameters and simulated it without any restraint for 100 ns ($\bar{\zeta}$ = 165°, $\bar{\psi}_{TPO}$ = −20°). In Figure 5, a superposition of one representative snapshot of the equilibrated *trans* product and the respective isosteric compound by Zhang et al. is shown. As can be seen, the match between the two is almost complete (backbone RMSD 0.9 Å) and provides us with the evidence that the obtained product is correct.

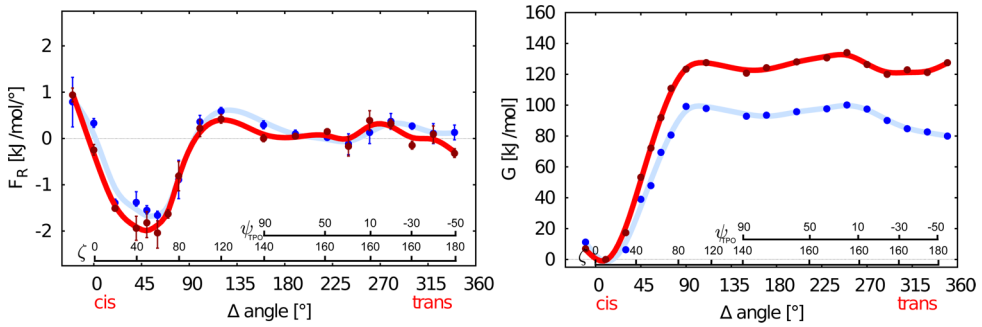


Figure 4. Mean reaction force (left) and free energy (right) for the setup Craft (blue) and D-HI/E (red) as a function of the reaction coordinate for the isomerization reaction that includes the rotation of the ζ and ψ angles shown in Figure 1.

dx.doi.org/10.1021/jp505638w | J. Phys. Chem. B 2014, 118, 9871–9880

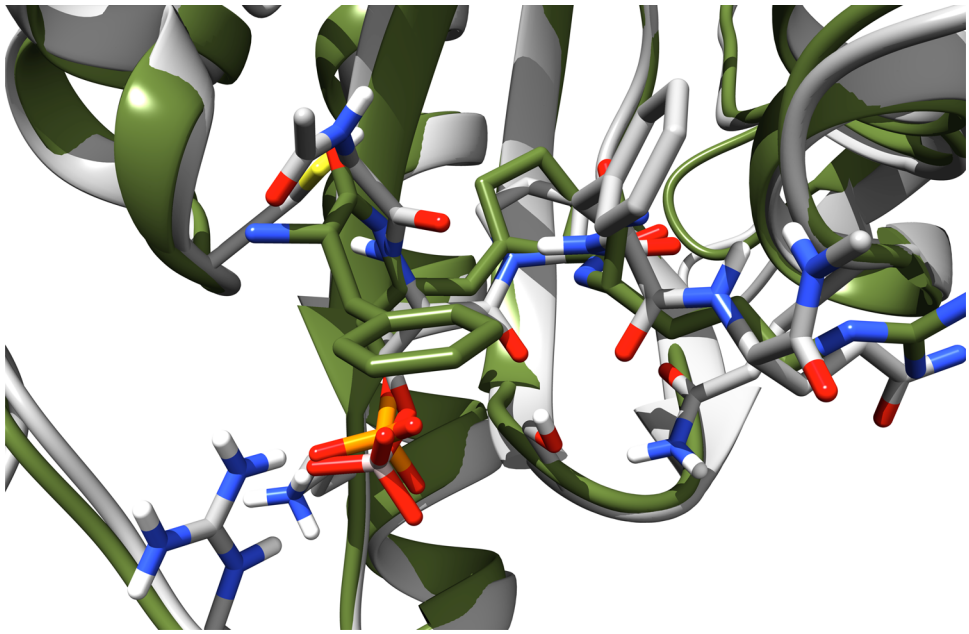


Figure 5. Equilibrated *trans* conformer (gray) using the *Sticht* setup and the crystal structure of the *trans* form analogue PDBID: 3tdb.

One might ask, however, what the *D-HI/E* setup yields as a product. The comparison in fact displays that this setup presented a product that was different in structure than the *trans* isoster (see the Supporting Information). However, the question is if the crystal structure is able to represent all binding modes in solution. The authors of the crystal structure already noted that the electron density of the *trans* isosteric compound at the alkene bond was lower than the one for the *cis* isosteric compound, whereas the electron density of the enzyme remains the same. This already indicates that the alkene bond was less ordered in the *trans* analogue than in the *cis* compound. The increased flexibility of the *trans* binding mode was further confirmed by NMR studies in solution, where larger side-chain flexibility in the interacting enzyme residues was observed for the *trans* analogue in comparison to the *cis* form.[12,13] The decrease in rigidity is also reflected in a binding constant which is 20−40 times larger for the *trans* isosteric compounds than for the *cis* one. The experimental evidence and our simulation result therefore suggest that the *trans* form indeed may present different binding modes which contribute to catalysis.

One question is yet to be answered: How does the enzyme catalyze the reaction from the *trans* to the *cis* form? Our results indicate that Gln-131 plays a double role: it fixes the N-terminal part of the peptide through hydrogen bonds of the backbone impeding the rotation (see the Introduction), and it also stabilizes the rotating carbonyl group through the amide group in its side chain (see Figure 2). Additionally, the peptide is anchored through interactions of the phosphate group with the basic triad (see Figure 1) such that the peptide has to rotate over the ζ angle and the ψ backbone angle of the phosphorylated treonine residue. Our results, therefore, confirm the experimental evidence of the recently reported crystal structures and the proposed mechanism of Mercedes-Chamacho et al.[16] However, with our results, we could refine their mechanism and assign their suggested "jump-rope" type of motion only to the rotation of the *trans* form. The enzyme lets the peptide rotate first only along the ψ bond which is almost barrier free. The beginning of this rotation is assisted by an intramolecular hydrogen bond between the carbonyl oxygen atom of the glycine residue and the amino group of the phenylalanine residue in the γ turn of the peptide. Once the ψ angle has adopted positive values, further rotation contributes to the rotation of the ζ bond with its partial double bond character.

Cys-113 plays only a minor role in the stabilization of the *cis* form, and Ser-154 will most likely interact with the phosphate group or may stabilize the transition state through specific hydrogen bonds.

If our mechanism is correct, the release of the *cis* and *trans* peptide from the enzyme has to be associated with different free energy changes, with the *cis* form presenting a higher binding affinity. This difference in free energy of binding has to be present to restore the almost equal free energies of the two isomeric forms in bulk water.[2] Indeed, the larger flexibility of the *trans* isomer and a smaller binding constant of the *cis* isosteric compounds support our results.[12,13]

CONCLUSIONS

In this study, we have shown that the combination of molecular dynamics simulations with atomic charges derived from the electron density and the QM/MM methodology are able to provide useful insights in the catalytic reaction mechanism of Pin1. Our results show that the influence of Cys-113 is restricted to the isomerization reaction from the *cis* form where it interacts with the substrate in the beginning of the rotation of the carbonyl bond. Ser-154 and Gln-131 stabilize the rotation from the *trans* form but are not involved in the transition state.

The main effect of Pin1 on the isomerization reaction is the reduction of the free energy barrier for the *trans* form to an extent where the reaction is almost barrier free. This is accomplished by a combined rotation of the ψ backbone angle of the threonine residue and the amide bond in a "jump-rope" type motion. This mechanism proposed by us in our previous study has recently been confirmed experimentally by Mercedes-Chamacho et al.[16] through kinetic isotope effects. The key for this mechanism is the electrostatic stabilization of the phosphate group of the threonine residue through the basic triad and the hydrogen bonds between the enzyme (Gln-131/

dx.doi.org/10.1021/jp505638w | J. Phys. Chem. B 2014, 118, 9871−9880

129) and the peptide which fix one part of the peptide in the catalytic site, hindering its rotation.

■ ASSOCIATED CONTENT

Supporting Information

Atomic charges of various methods in different environments, mean reaction force profile at the BLYP/SVP level, results for the *Sticht* setup, and comparison of the X-ray structure of a *trans*-peptide analogue with the product of the *D-HI/E* setup are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: evohringer@udec.cl.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

E.V.-M. is thankful for funding support provided by FONDECY 11121179 and Milenium Nucleus NC120082. T.V. acknowledges the Foundation of Scientific Research - Flanders (FWO), the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP/7/05 for their financial support. P.W.A. thanks NSERC for funding.

■ REFERENCES

(1) Dugave, C.; Demange, L. Cis-Trans Isomerization of Organic Molecules and Biomolecules: Implications and Applications. *Chem. Rev.* **2003**, *103*, 2475−2532.

(2) Schutkowski, M.; Bernhardt, A.; Zhou, X. Z.; Shen, M.; Reimer, U.; Rahfeld, J.-U.; Lu, K.-P.; Fischer, G. Role of Phosphorylation in Determining the Backbone Dynamics of the Serine/Threonine-Proline Motif and Pin1 Substrate Recognition. *Biochemistry* **1998**, *37*, 5566−5575.

(3) Rahfeld, J. U.; Schierhorn, A.; Mann, K.; Fischer, G. A Novel Peptidyl-Prolyl *cis/trans* Isomerase from *Escherichia Coli*. *FEBS Lett.* **1994**, *343*, 65−69.

(4) Yaffe, M. B. M.; Schutkowski, M. M.; Shen, M. M.; Zhou, X. Z. X.; Stukenberg, P. T. P.; Rahfeld, J. U. J.; Xu, J. J.; Kuang, J. J.; Kirschner, M. W. M.; Fischer, G. G.; Cantley, L. C. L.; Lu, K. P. K. Sequence-Specific and Phosphorylation-Dependent Proline Isomerization: a Potential Mitotic Regulatory Mechanism. *Science* **1997**, *278*, 1957−1960.

(5) Lu, K.; Hanes, S.; Hunter, T. A Human Peptidyl-Prolyl Isomerase Essential for Regulation of Mitosis. *Nature* **1996**, *380*, 544−547.

(6) Pastorino, L.; Sun, A.; Lu, P.-J.; Zhou, X. Z.; Balastik, M.; Finn, G.; Wulf, G.; Lim, J.; Li, S.-H.; Li, X.; Xia, W.; Nicholson, L. K.; Lu, K.-P. The Prolyl Isomerase Pin1 Regulates Amyloid Precursor Protein Processing And Amyloid-Beta Production. *Nature* **2006**, *440*, 528−534.

(7) Butterfield, D. A.; Abdul, H. M.; Opii, W.; Newman, S. F.; Joshi, G.; Ansari, M. A.; Sultana, R. Review: Pin1 in Alzheimer's disease. *J. Neurochem.* **2006**, *98*, 1697−1706.

(8) Yeh, E. S.; Means, A. R. PIN1, the Cell Cycle and Cancer. *Nat. Rev. Cancer* **2007**, *7*, 381−388.

(9) Zhang, M.; Wang, X. J.; Chen, X.; Bowman, M. E.; Luo, Y.; Noel, J. P.; Ellington, A. D.; Etzkorn, F. A.; Zhang, Y. Structural and Kinetic Analysis of Prolyl-Isomerization/phosphorylation Cross-Talk in the CTD Code. *ACS Chem. Biol.* **2012**, *7*, 1462−1470.

(10) Fischer, G. Chemical Aspects of Peptide Bond Isomerisation. *Chem. Soc. Rev.* **2000**, *29*, 119−127.

(11) Zhang, Y.; Daum, S.; Wildemann, D.; Zhou, X. Z.; Verdecia, M. A.; Bowman, M. E.; Luecke, C.; Hunter, T.; Lu, K.-P.; Fischer, G.; Noel, J. P. Structural basis for High-Affinity Peptide Inhibition of Human Pin1. *ACS Chem. Biol.* **2007**, *2*, 320−328.

(12) Namanja, A. T.; Wang, X. J.; Xu, B.; Mercedes-Camacho, A. Y.; Wilson, K. A.; Etzkorn, F. A.; Peng, J. W. Stereospecific Gating of Functional Motions in Pin1. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 12289−12294.

(13) Namanja, A. T.; Peng, T.; Zintsmaster, J. S.; Elson, A. C.; Shakour, M. G.; Peng, J. W. Substrate Recognition Reduces Side-Chain Flexibility for Conserved Hydrophobic Residues in Human Pin1. *Structure* **2007**, *15*, 313−327.

(14) Xu, G. G. G.; Slebodnick, C. C.; Etzkorn, F. A. F. Cyclohexyl Ketone Inhibitors of Pin1 Dock in a Trans-Diaxial Cyclohexane Conformation. *PLoS One* **2012**, *7*, e44226−e44226.

(15) Xu, G. G. G.; Zhang, Y. Y.; Mercedes-Camacho, A. Y. A.; Etzkorn, F. A. F. A Reduced-Amide Inhibitor of Pin1 Binds in a Conformation Resembling a Twisted-Amide Transition State. *Biochemistry* **2011**, *50*, 9545−9550.

(16) Mercedes-Camacho, A. Y.; Mullins, A. B.; Mason, M. D.; Xu, G. G.; Mahoney, B. J.; Wang, X.; Peng, J. W.; Etzkorn, F. A. Kinetic Isotope Effects Support the Twisted Amide Mechanism of Pin1 Peptidyl-Prolyl Isomerase. *Biochemistry* **2013**, *52*, 7707−7713.

(17) Behrsin, C. D. C.; Bailey, M. L. M.; Bateman, K. S. K.; Hamilton, K. S. K.; Wahl, L. M. L.; Brandl, C. J. C.; Shilton, B. H. B.; Litchfield, D. W. D. Functionally Important Residues in the Peptidyl-Prolyl Isomerase Pin1 Revealed by Unigenic Evolution. *J. Mol. Biol.* **2007**, *365*, 20.

(18) Bailey, M. L.; Shilton, B. H.; Brandl, C. J.; Litchfield, D. W. The Dual Histidine Motif in the Active Site of Pin1 has a Structural Rather than Catalytic Role. *Biochemistry* **2008**, *47*, 11481−11489.

(19) Velazquez, H. A.; Hamelberg, D. Conformation-Directed Catalysis and Coupled Enzyme-Substrate Dynamics in Pin1 Phosphorylation Dependent Cis-Trans Isomerase. *J. Phys. Chem. B* **2013**, *117*, 11509−11517.

(20) Velazquez, H. A.; Hamelberg, D. Conformational Selection in the Recognition of Phosphorylated Substrates by the Catalytic Domain of Human Pin1. *Biochemistry* **2011**, *50*, 9605−9615.

(21) Vöhringer-Martinez, E.; Duarte, F.; Toro-Labbé, A. How Does Pin1 Catalyze the Cis-Trans Prolyl Peptide Bond Isomerization? A QM/MM and Mean Reaction Force Study. *J. Phys. Chem. B* **2012**, *116*, 12972−12979.

(22) Cox, C.; Young, V.; Lectka, T. Intramolecular Catalysis of Amide Isomerization. *J. Am. Chem. Soc.* **1997**, *119*, 2307−2308.

(23) Yonezawa, Y.; Nakata, K.; Sakakura, K.; Takada, T.; Nakamura, H. Intra- and Intermolecular Interaction Inducing Pyramidalization on Both Sides of a Proline Dipeptide during Isomerization: An Ab Initio QM/MM Molecular Dynamics Simulation Study in Explicit Water. *J. Am. Chem. Soc.* **2009**, *131*, 4535−4540.

(24) Ranganathan, R.; Lu, K. P.; Hunter, T.; Noel, J. P. Structural and Functional Analysis of the Mitotic Rotamase Pin1 Suggests Substrate Recognition is Phosphorylation Dependent. *Cell* **1997**, *89*, 875−886.

(25) Vöhringer-Martinez, E.; Toro-Labbé, A. The Mean Reaction Force: a Method to Study the Influence of The Environment on Reaction Mechanisms. *J. Chem. Phys.* **2011**, *135*, 064505.

(26) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: the RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(27) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances In RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.

(28) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44*, 129−138.

(29) Nalewajski, R. F.; Parr, R. G. Information Theory, Atoms in Molecules, and Molecular Similarity. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8879−8882.

(30) Ayers, P. W. Information Theory, the Shape Function, and the Hirshfeld Atom. *Theor. Chim. Acta* **2006**, *115*, 370−378.

(31) Stone, A. J.; Alderton, M. Distributed Multipole Analysis. *Mol. Phys.* **1985**, *56*, 1047−1064.

(32) Bader, R.; Molecules, A. I. *A Quantum Theory*; Clarendon: Oxford, U.K., 1990.
(33) Davidson, E. R.; Chakravorty, S. A Test of the Hirshfeld Definition of Atomic Charges and Moments. *Theor. Chem. Acc.* **1992**, *83*, 319−330.
(34) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical Analysis and Extension of the Hirshfeld Atoms in Molecules. *J. Chem. Phys.* **2007**, *126*, 144111.
(35) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L., Jr. Density-Functional Theory for Fractional Particle Number:Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49*, 1691.
(36) Yang, W. T.; Zhang, Y. K.; Ayers, P. W. Degenerate Ground States and a Fractional Number of Electrons in Density and Reduced Density Matrix Functional Theory. *Phys. Rev. Lett.* **2000**, *84*, 5172−5175.
(37) Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and Basis Set Dependence of Iterative Hirshfeld Charges. *Chem. Phys. Lett.* **2007**, *444*, 205−208.
(38) Van Damme, S.; Bultinck, P.; Fias, S. Electrostatic Potentials from Self-Consistent Hirshfeld Atomic Charges. *J. Chem. Theory Comput.* **2009**, *5*, 334−340.
(39) Verstraelen, T.; Pauwels, E.; De Proft, F.; Van Speybroeck, V.; Geerlings, P.; Waroquier, M. Assessment of Atomic Charge Models for Gas-Phase Computations on Polypeptides. *J. Chem. Theory Comput.* **2012**, *8*, 661−676.
(40) HORTON 1.2.1, http://theochem.github.io/horton/index.html.
(41) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. Hirshfeld-E Partitioning: AIM Charges with an Improved Trade-Off between Robustness and Accurate Electrostatics. *J. Chem. Theory Comput.* **2013**, *9*, 2221−2225.
(42) Hu, H.; Lu, Z.; Parks, J. M.; Burger, S. K.; Yang, W. Quantum Mechanics/Molecular Mechanics Minimum Free-Energy Path for Accurate Reaction Energetics in Solution and Enzymes: Sequential Sampling and Optimization on the Potential of Mean Force Surface. *J. Chem. Phys.* **2008**, *128*, 034105−034118.
(43) Hu, H.; Yang, W. Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573−601.
(44) Homeyer, N.; Horn, A. H. C.; Lanig, H.; Sticht, H. AMBER Force-Field Parameters for Phosphorylated Amino Acids in Different Protonation States: Phosphoserine, Phosphothreonine, Phosphotyrosine, and Phosphohistidine. *J. Mol. Model.* **2005**, *12*, 281−289.
(45) Hess, B.; Kutzner, C.; Spoel, D. v. d. Gromacs 4: Algorithms for Highly Efficient, Load Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(46) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *2*, 73−78.
(47) Craft, J. W., Jr; Legge, G. B. An AMBER/DYANA/MOLMOL Phosphorylated Amino Acid Library Set and Incorporation into NMR Structure Calculations. *J. Biomol. NMR* **2005**, *33*, 15−24.
(48) Kästner, J.; Thiel, W. Bridging the Gap Between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: "Umbrella Integration". *J. Chem. Phys.* **2005**, *123*, 144104.
(49) Kästner, J.; Thiel, W. Analysis of the Statistical Error in Umbrella Sampling Simulations by Umbrella Integration. *J. Chem. Phys.* **2006**, *124*, 234106.
(50) Parr, R. G.; Ayers, P. W.; Nalewajski, R. F. What is an Atom in a Molecule? *J. Phys. Chem. A* **2005**, *109*, 3957−3959.
(51) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.
(52) Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation of CM5 Charges for Condensed-Phase Modeling. *J. Chem. Theory Comput.* **2014**, *10*, 2802−2812.
(53) Wang, B.; Truhlar, D. G. Tuned and Balanced Redistributed Charge Scheme for Combined Quantum Mechanical and Molecular Mechanical (QM/MM) Methods and Fragment Methods: Tuning Based on the CM5 Charge Model. *J. Chem. Theory Comput.* **2013**, *9*, 1036−1042.