A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems†

Chris Vanessa Sumowski and Christian Ochsenfeld*

Theoretische Chemie, Universität Tübingen, Auf der Morgenstelle 8, D-72076 Tübingen, Germany

Received: March 30, 2009; Revised Manuscript Received: June 10, 2009

A systematic study of the convergence of QM/MM results with respect to the chosen size of the QM region is presented for two examples of peptidic systems. For this purpose, we increased the QM region to up to 1637 atoms at the HF/SVP and 383 atoms at the SOS-AO-MP2/6-31G** level. While the convergence behavior is almost independent of the chosen method and basis set, the study clearly shows that for the considered proton-transfer energy the QM/MM treatment leads to a significantly faster convergence than the pure QM treatment. This behavior can be rationalized by the fair description of the surrounding of the active center using MM methods, even though the MM description of the active center is not adequate in our present case. At the same time, the observed convergence is quite insensitive to a variation of charge surroundings in the chosen model peptides. Although the QM/MM results do converge much quicker with the system size than the pure QM ones, the data show that even for the chosen simple model systems about 150−300 QM atoms are needed to achieve accuracies in the order of 10 kJ/mol and about 300−1000 atoms for an accuracy of 2 kJ/mol with respect to a convergence with the QM-region size.

1. Introduction

Over the last decades, quantum chemistry has evolved to become an important tool for gaining insights into molecular properties and processes. Although originally constrained to small molecules, the development of linear-scaling methodologies (for an overview see, e.g., refs 1−4 and references therein) coupled to ever faster computer technologies allows us to access, even on simple one-processor computers, nowadays molecular systems on the order of 1000 atoms with 10 000 basis functions. This is not only possible at the Hartree−Fock (HF) or density-functional theory (DFT) levels (see, e.g., refs 5−8) but also for wave function based electron-correlation approaches such as Møller−Plesset second-order perturbation theory (MP2).[9]

It is obvious that once a linear-scaling ab initio method is attained, any speed-up in computer technologies translates directly into treatable molecular size, which would have not been possible with traditional quantum-chemical approaches due to their steep polynomial increase of computational cost. Despite this success and the positive perspectives for the applicability of quantum-chemical methods (denoted as quantum-mechanical (QM) methods in the following), it is clear that a QM method cannot be as cheap as a simple molecular-mechanics (MM) approach. Of course such simple force-field schemes do not provide an ab initio description of the correct physics and parametrizations can be cumbersome (in particular if reliable reference data are hard to obtain), but pragmatically such approaches have reached widespread applicability due to their simplicity both in science and in industry. A central and difficult issue of MM schemes is of course the reliability and that no check is possible by systematically increasing the accuracy. This contrasts with the highly successful hierarchy of quantum-chemical methods, where, in principle, the pathway toward the exact solution of the Schrödinger equation is known.

Since many processes encountered in complex molecular systems are dominated by the mechanisms occurring in an active center, the QM/MM concept of combining the strengths of QM (for describing the active center) and MM approaches (for describing influences of the surrounding) has become a very popular choice. Although a physically correct coupling between the QM and the MM part is intrinsically difficult, if not impossible, many successful protocols for the coupling have been well established (see, e.g., ref 10−23 and references therein). While the capabilities are often impressive, also the limitations of QM/MM approximations have been widely discussed, and the four major error sources can be well separated: (1) the MM parametrization,[24−27] (2) the quality of the chosen QM approach[23,25,28−31] (although in principle systematic convergence to the exact result is possible), (3) the frontier description itself,[29,32−36] and (4) the question where to place the frontier and how large the QM region needs to be chosen.

In our present work we focus in particular on the fourth aspect listed above (while we also touch on the second and third aspects) and try to systematically investigate how large the QM region needs to be chosen for a reliable description of the isomerization energies in two example peptides. To the best of our knowledge such investigations of systematically enlarging the QM region and the influence on the energetics have so far been only made for small model systems.[32,34,35,37] For real case studies often only isolated QM cutouts have been selected and QM-region sizes only partially increased, so that no systematic convergence studies with the QM-region size have been performed (see, e.g., refs 29, 31, 38, and 39). An obvious reason for this lack is the very steep increase of the computational effort with the enlargement of the QM region, so that conventional ab initio methods quickly encounter the scaling wall and cannot be applied. In our study we employ linear-scaling methods both at HF[40−43] and atomic orbital-based MP2 (AO-MP2)[9,44,45] levels, so that calculations for systems with up to 1637 atoms at the HF/SVP and 383 atoms at the scaled opposite-spin AO-MP2 (SOS-AO-MP2) levels become feasible. As model systems for our present study, we focus on the proton transfer in

† Part of the "Walter Thiel Festschrift".

* To whom correspondence should be addressed. E-mail: christian.ochsenfeld@uni-tuebingen.de.

10.1021/jp902876n CCC: $40.75
Published on Web 07/08/2009

TABLE 1: Isomerization Energies in Ab Initio QM Calculations for Various Peptidic Systems and Different QM Methods and Basis Sets[a]

system	QM method	QM
peptide **mpC** (full system)	HF/SVP	365.0
TIM	HF/3-21G	−71.2
(residues K13, E97	HF/SVP	−54.7
and PGH250)	HF/aug-cc-pVDZ	−59.9
	MP2/aug-cc-pVDZ	−73.2
TIM ($S6_{TIM}$)	HF/3-21G	−76.0

[a] All energy differences are given in kJ/mol. Negative signs correspond to a more stable zwitterionic state.

two peptidic systems, since such processes are often of uttermost importance for enzymatic catalysis.

2. Methodological Aspects

2.1. Structural Model Systems. As model systems for our QM and QM/MM convergence studies, we focused on two systems:

• a model peptide using 32 residues with a total of 383 atoms and

• an enzyme with a total of 4195 atoms of which we considered at most 1787 atoms at the QM level.

Since proton transfer is of general importance in many molecular systems, we considered for both systems an artificial transfer model. Our focus is, however, not on the energetics of the artificial proton displacement itself but on the influences exerted by the surrounding of the active center and how well these influences are described by the QM/MM approach. Since the energetics themselves were not in the focus, we performed no or only partial structure optimizations limited to the proton positions, in order to minimize the number of parameters affecting the energetics for our systematic convergence study. In this way, the absolute energies of the isomerizations are not meaningful themselves.

For the MM description we chose standard force-field parameters as described in the corresponding sections.[46] As the isomers have different topologies and, accordingly, different energy expressions arise, force-field calculations are not suitable for determining the respective proton-transfer energies. This deficiency is partially illustrated by comparing pure QM calculations, summarized in Table 1, and pure MM data for selected model systems: For the model peptide the calculations using the CHARMM22 force field reveal an isomerization energy of 1 kJ/mol compared to 365 kJ/mol at the HF/SVP level. A similar behavior is observed for the TIM enzyme, where the proton-transfer energy calculated at the force-field level is −500 kJ/mol (residues K13, E97, and PGH250) and −473 kJ/mol ($S6_{TIM}$), while the ab initio results are in a range of −55 to −76 kJ/mol, depending on the chosen QM method and basis set. However, due to the reasons noted above, the major error in the MM description arises in the 'active center', which we will treat in the following by QM methods, so that the MM approach needs only to provide a fair description of the surrounding within the QM/MM scheme.

2.2. QM Methods and Basis Sets. For the description of the QM part, we employ both the HF[47−49] and SOS-MP2[50,51] methods, where the MP2 calculations serve as a reference, since they account for electron-correlation (also dispersion-type) effects. In order to be able to compute large molecules, we employ linear-scaling techniques throughout this work for forming the Fock matrix[6,40−43] and for calculating the MP2 energy by a linear-scaling AO-based formulation[9,44,45] as implemented in a development version of the Q-Chem program package.[52] We employ 3-21G,[53] 6-31G**,[54,55] 6-31++G**,[54−56] SVP,[57] and TZP[57] basis sets.

2.3. QM/MM Details. To check the influence of different QM/MM coupling schemes on the size convergence of isomerization energies, we have chosen three different QM/MM methods for comparison: First, we used a mechanical embedding scheme with a hydrogen link atom method for the QM/MM boundary treatment.[20,37] Here, the QM calculation is performed essentially in the gas phase and the QM/MM electrostatic interaction is introduced by the MM code using a classical point charge model for the QM charge distribution.

Furthermore, an electrostatic embedding scheme was chosen,[20,37] for which the static MM charges are included in the one-electron Hamiltonian of the QM part. Hence the QM/MM electrostatic interactions are calculated between the QM electrostatic potential and the MM partial charges. Also for the electrostatic embedding calculations, hydrogen link atoms were employed with either the *L2*[20,58] or the *charge shift* model[20,59] to adjust the electrostatics at the QM/MM boundary. In the *L2* scheme atoms of neutral charge groups next to the QM/MM boundary are deleted from the QM Hamiltonian. Hence, the overall charge of the MM part is conserved, but significant electrostatic interactions between QM and MM part may be removed. In addition, the dipole moment of the eliminated charge group is missing. The *charge shift* model overcomes the latter problem: Here, the charge of the MM atom at the boundary is shifted to its neighboring atoms, which preserves the overall charge of the MM fragment. At the same time, point charges of opposite signs are added to these neighboring atoms and, hence, also the dipole moment is conserved. For the nonbonding MM and QM/MM interactions no cutoffs were adapted. For treating the QM part in all QM and QM/MM calculations a development version of Q-Chem[52] was used. For the MM part the CHARMM22 force field[60] was employed both within the QM/MM approach and the classical structure optimizations (as described below). The CHARMM package[61,62] was run through the DL_POLY code[63] as integrated in the ChemShell package.[20]

As the respective QM regions of the peptides are generated by adding amino acids in a specified radius around the modification, the QM/MM boundaries go through peptide bonds. This homolytic separation of the peptide bonds is of course physically not correct, since such bonds are polar and have partial double-bond character. Nevertheless, we employ such homolytic separated peptide bonds in our present study since we do consider only relative energies where these effects cancel: Both isomers are treated in the same way, and the considered proton-transfer reaction does not involve the backbone. Hence, the perturbation in the electronic structure of the backbone does not influence the results significantly.

3. Model Peptide Sequence

3.1. Structural Details. As basic model peptide a poly-alanine sequence within an α-helical arrangement comprising 32 residues was employed

$$\underbrace{\underbrace{\underbrace{[AAAAA\,[EAAAA\,[KAAAA\,\underbrace{[EK]}_{S1_{mp}}\,AAAAE]\,AAAAK]\,AAAAA]}_{S2_{mp}}}_{S3_{mp}}}_{\mathrm{S4_{mp}}},$$

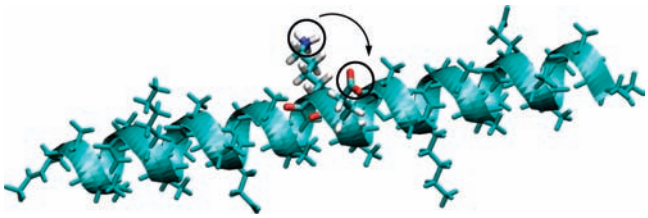

Figure 1. Illustration of the model peptide **mpC**.

in which some alanine (A) residues were replaced by charged glutamate (E) and lysine (K) residues (brackets indicate the chosen QM subregions for the QM/MM calculations). In this way, four subsystems denoted as $S1_{mp}$ to $S4_{mp}$ (for the total system) result. The glutamate and lysine residues in the helix allow to easily modify the charge situation, so that these influences can be studied. In addition, the chosen model describes a quasi one-dimensional extension, which has the advantage that by increasing the system size well separated units add on, which exhibit no direct interaction between each other, so that a particularly simple model-type system is obtained.

As model for an isomerization, we consider a proton transfer from the positively charged lysine group to the neighboring negatively charged glutamate, so that charge neutrality is preserved. For this isomerization three different structure models were generated by varying the protonation states of the surrounding glutamate and lysine groups: In model **mpA** all four residues carry no charge, in **mpB** the two glutamate residues are charged while the two lysine groups are not, and for **mpC** only charged species were present (for an illustration of model **mpC** see Figure 1).

An additional model was generated by starting from model **mpC**, in order to study the difference between bound and through-space interactions: removal of connecting units consisting each of three alanine residues results in a sequence with seven segments denoted as **mpC***: AA***EA***KA***EK***AE***AK***AA (here, each star stands for a removed alanine A; structural arrangements were otherwise unchanged). From this procedure an overall neutral charge situation results for models **mpA**, **mpC**, and **mpC***, while model **mpB** carries an overall charge of −2. All N-termini were acetylated and all C-termini saturated with *N*-methylamide, whereas for none of the model peptides a reoptimization was performed in order to minimize structural influences on the study described below (the orignal structure was generated using the Maestro 7.5 package[64] without any further optimization).

Employing these model systems for computing the isomerization energy, we study in the following the influences (1) of different coupling schemes between QM and MM regions, (2) of different QM approaches, and (3) of the charge and bonding situations upon increasing the QM size within the QM/MM calculation.

The observed large isomerization energies are caused by the artificial model structures used here. In these structures the distance between the ammonium group and the carboxylate is very large with approximately 11 Å. Hence the neutral situation is considerably more favored than the zwitterionic case as charge compensation is hindered by the large distance. These observations are in line with proton-isomerization energies observed for isolated species and other long-range transfers such as, e.g., discussed in refs 65 and 66.

3.2. Influence of Coupling Schemes between QM and MM Regions. The influence of different coupling models on the convergence of relative energies with increasing size of the QM region was first studied using structure model **mpC** (data listed in Table 2). The QM calculations were carried out at the HF/SVP level and deviations of the various QM/MM coupling models are discussed with respect to the QM calculation of the entire system comprising 383 atoms.

For *mechanical embedding* a deviation of 22 kJ/mol is found for the smallest QM region ($S1_{mp}$). Upon increasing the subsystem to $S2_{mp}$, the error changes signs and reduces to 13 kJ/mol. For fragment $S3_{mp}$ (with 273 atoms) the deviation from the full QM result is 2 kJ/mol. Clearly, the larger the QM region becomes, the less important for the isomerization energy is the missing polarization of the QM region within the mechanical QM/MM model, in which the QM calculation is performed in the gas phase and the coupling calculated by the MM part. Although mechanical embedding is certainly the simplest QM/MM model, its convergence is still significantly faster than the one of the pure QM calculation of subsystems $S1_{mp}/S2_{mp}/S3_{mp}$ with deviations of −28/18/9 kJ/mol.

In all *electrostatic embedding* QM/MM models, in contrast to the mechanical one, a recoupling between QM and MM parts is introduced, so that the QM part is also polarized by the MM charges and better results might be expected. For these electrostatic embedding models we employed two different boundary adjustment schemes *L2*[20,58] and *charge shift*.[20,59] Furthermore, we compare for both boundary adjustment schemes the QM calculation just in the field of point charges obtained from the MM part (denoted as QM_{pol} throughout this work) and the analogous calculation combined with the MM energies of the total system (denoted as QM/MM). Here, QM_{pol} includes the polarized QM energy as well as the energy of the nucleus-charge interaction. Since in our present example the structure is only locally changed in the $S1_{mp}$ part, the results for QM_{pol} and QM/MM are basically identical (see Table 2), apart from short-range interaction terms arising in the latter treatment between the QM and MM fragment. Once the chosen subsystem size is increased, the contribution of these short-range couplings rapidly decay, so that their influence onto the isomerization energy vanishes and QM/MM and QM_{pol} results become identical. Therefore we restrict ourselves in the following to the discussion of QM and QM/MM results.

While the deviations of the *L2* scheme to the full QM result start out to be even larger than the one found for mechanical embedding (29 vs 22 kJ/mol), the convergence of the results with fragment size seems faster, although for $S3_{mp}$ still a deviation of 5 kJ/mol remains. A possible explanation for the relatively poor performance for the $S1_{mp}$ subsystem might be that in the *L2* approach some important electrostatic interactions could have been neglected, since *L2* excludes all atoms of the "charge groups" next to the link atoms.

The *charge shift* method reduces for our smallest QM region ($S1_{mp}$) the deviation to 15 kJ/mol which can be compared to 28 kJ/mol for the pure QM calculation. Although for the $S2_{mp}$ fragment still 11 kJ/mol deviation is observed, the deviation reduces to 2 kJ/mol for $S3_{mp}$, which includes all atoms with at least 8 Å distance to the 'active site' (the corresponding QM deviation is still 9 kJ/mol).

These observations lead us in the following section to perform all QM/MM calculations with the *charge shift* method. Overall, the QM/MM values (besides *L2* for fragment $S1_{mp}$) lead in general to a significantly faster convergence toward the full QM result than the calculation of isolated QM fragments.

3.3. Influence of the QM Approach. Besides the QM/MM coupling schemes also the choice of the quantum-chemical method can influence the QM/MM results significantly. To ensure the reliability of the simple HF/SVP approach (used

TABLE 2: Comparison of QM/MM Coupling Schemes and the Dependence of the Isomerization Energy for a Modified Poly-Alanine Helix (Model mpC) upon the Chosen QM Region[a]

QM region	d_{new}[b] Å	# atoms (QM)	mech[c] QM/MM	*L2*[d] QM/MM	*L2*[d] QM_{pol}	shift[e] QM/MM	shift[e] QM_{pol}	QM
$S1_{mp}$		(39)	386.5	394.1	394.0	380.1	379.9	393.0
$S2_{mp}$	6.2	(156)	351.9	357.0	357.0	353.8	353.8	346.7
$S3_{mp}$	8.6	(273)	362.6	360.1	360.1	363.0	363.0	356.3
$S4_{mp}$	15.7	(383)						365.0
Δ_{S4-S1}[f]			−21.5	−29.1	−29.0	−15.1	−14.9	−28.0
Δ_{S4-S2}[f]			13.1	8.0	8.0	11.2	11.2	18.3
Δ_{S4-S3}[f]			2.4	4.9	4.9	2.0	2.0	8.7

[a] In addition, QM data are listed. The QM parts were treated at the HF/SVP level with the total system ($S4_{mp}$) comprising 383 atoms. All energies are given in kJ/mol. [b] d_{new} describes the minimal distance between the modification and the Cα atoms of newly comprised residues upon enlarging the QM region. [c] Mechanical embedding means no polarization of the QM by the MM part. Therefore no QM_{pol} values exist. [d] Electrostatic QM/MM embedding scheme using the *L2* boundary adjustment model. [e] Electrostatic QM/MM embedding scheme using the *charge shift* boundary adjustment model. [f] Δ_{S4-Sx} denotes the error for the isomerization energy calculated with QM region Sx compared to the full QM calculation of $S4_{mp}$.

TABLE 3: Influence of Electron-Correlation Effects for QM and QM/MM Calculations upon the Convergence with the QM-Region Size for the Isomerization Energy of a Modified Poly-Alanine Helix (Model mpC)[a]

QM region	QM/MM HF	QM/MM SOS-AO-MP2	QM_{pol} HF	QM_{pol} SOS-AO-MP2	QM HF	QM SOS-AO-MP2
$S1_{mp}$	382.2	383.9	382.0	383.7	395.7	396.2
$S2_{mp}$	356.6	355.8	356.6	355.8	349.5	348.6
$S3_{mp}$	365.6	363.8	365.6	363.8	358.9	357.2
$S4_{mp}$					367.5	370.0
Δ_{S4-S1}[b]	−14.7	−13.9	−14.5	−13.7	−28.2	−26.2
Δ_{S4-S2}[b]	10.9	14.2	10.9	14.2	18.0	21.4
Δ_{S4-S3}[b]	1.8	6.2	1.8	6.2	8.6	12.8

[a] The QM parts were treated at the HF and SOS-AO-MP2 levels using the basis set 6-31G**. The total system ($S4_{mp}$) comprises 383 atoms. All energy differences are given in kJ/mol. All SOS-AP-MP2 calculations were carried out using the frozen core approximation and a scaling factor of 1.3. [b] Δ_{S4-Sx} denotes the error for the isomerization energies calculated with QM region Sx compared to the full QM calculation of $S4_{mp}$.

throughout this work) for the present molecular systems, we performed SOS-AO-MP2 calculations for approximating electron-correlation effects (and also dispersion) for model system **mpC** with up to 383 atoms. This calculation became possible by using our recently developed linear-scaling AO-MP2 method, which allows to access molecules comprising 1000 atoms with 10 000 basis functions.[9] At the current development stage of our AO-MP2 method, we are constrained to basis sets of typically double-ζ polarization quality for such large molecules. Nevertheless, the SOS-AO-MP2/6-31G** method permits to obtain an estimate of the influences expected by electron-correlation effects.

Data listed in Table 3 indicate small electron-correlation influences of up to 3 kJ/mol on the total isomerization energy for all QM-region sizes. At the same time, the MP2 results converge slightly slower with increasing the QM-region size than the corresponding HF values. For subsystem $S3_{mp}$ a deviation of 6 kJ/mol remains to the SOS-AO-MP2 reference calculation of the full system, which is slightly larger than the 2 kJ/mol observed at the HF level. This has to be compared to the pure QM calculations of $S3_{mp}$ with deviations of 13 and 9 kJ/mol at the SOS-AO-MP2 and HF levels, respectively.

While electron correlation affects the isomerization energy of the present molecular system only relatively weakly, the choice of larger basis sets has considerable influence upon the absolute isomerization energies: in the order of 30 kJ/mol as estimated at the HF level (corresponding data using 6-31++G** and TZP bases are listed in Table 4). However, the larger bases

TABLE 4: Basis Set Influences upon the Computed Isomerization Energies for the Poly-Alanine Helix Model mpC[a]

	QM			
	$S1_{mp}$	$S2_{mp}$	$S3_{mp}$	$S4_{mp}$
HF/3-21G	387.1	341.0	350.8	359.6
HF/6-31G**	395.7	349.5	358.9	367.5
HF/6-31++G**	363.1	316.7	326.2	334.9
HF/SVP	393.0	346.7	356.3	365.0
HF/TZP	368.5			
	$\Delta_{S4(QM)-S1}$	$\Delta_{S4(QM)-S2}$	$\Delta_{S4(QM)-S3}$	
HF/3-21G	−27.5	18.7	8.8	
HF/6-31G**	−28.2	18.0	8.6	
HF/6-31++G**	−28.2	18.2	8.7	
HF/SVP	−28.0	18.3	8.7	
	QM/MM			
	$S1_{mp}$	$S2_{mp}$	$S3_{mp}$	
HF/3-21G	373.6	347.9	357.6	
HF/6-31G**	382.2	356.6	365.6	
HF/6-31++G**	349.7	323.9	332.9	
HF/SVP	380.1	353.8	363.0	
HF/TZP	355.1			
	$\Delta_{S4(QM)-S1}$	$\Delta_{S4(QM)-S2}$	$\Delta_{S4(QM)-S3}$	
HF/3-21G	−14.0	11.7	2.0	
HF/6-31G**	−14.7	10.9	1.8	
HF/6-31++G**	−14.8	11.0	2.0	
HF/SVP	−15.1	11.2	2.0	

[a] The complete system ($S4_{mp}$) comprises 383 atoms. All energies are given in kJ/mol. $\Delta_{S4(QM)-Sx}$ denotes the error for the calculation with QM region Sx compared to the full QM calculation of $S4_{mp}$ within the respective basis set.

have virtually no influence on the convergence of QM or QM/MM results upon increasing the QM region, which is the focus of our present work: deviations of the size convergence in using 3-21G, 6-31G**, SVP, and 6-31++G** basis sets are less than 1 kJ/mol.

Overall these results suggest that for our present molecular system, the convergence with the size of the QM region is mostly independent of both the choice of basis set and quantum-chemical approach, so that we select for the following study the HF/SVP approach for the peptidic systems.

3.4. Influence of the Surrounding Charge Situation. As many biological systems such as, e.g., peptides or RNA consist of charged parts, the charge dependence of the convergence behavior with QM fragment size within the QM/MM treatment for the calculation of relative energies is an important issue. Therefore, we compare in the following the isomerization energies for the three model peptides **mpA**, **mpB**, and **mpC**,

TABLE 5: Isomerization Energies for Poly-Alanine Helices Modified with Unequally Charged Amino Acids (Models mpA, mpB, mpC, and mpC*)[a]

	QM/MM				QM			
QM region	**mpA**	**mpB**	**mpC**	**mpC***	**mpA**	**mpB**	**mpC**	**mpC***
$S1_{mp}$	394.7	396.6	380.1	368.7	393.1	393.1	393.0	373.6
$S2_{mp}$	369.3	370.4	353.8	370.5	355.6	357.9	346.7	373.8
$S3_{mp}$	374.9	378.4	363.0	372.5	368.2	371.7	356.3	369.4
$S4_{mp}$					376.4	380.0	365.0	372.6
Δ_{S4-S1}[b]	−18.3	−16.6	−15.1	4.0	−16.7	−13.1	−28.0	−1.0
Δ_{S4-S2}[b]	7.1	9.6	11.2	2.1	20.9	22.1	18.3	−1.2
Δ_{S4-S3}[b]	1.5	1.6	2.0	0.1	8.2	8.4	8.7	3.2

[a] QM and QM/MM calculations with systematically enlarged QM regions were carried out treating the QM parts at the HF/SVP level. $S4_{mp}$ is the respective total system. Energy differences are given in kJ/mol. [b] Δ_{S4-Sx} denotes the error for the calculation with QM region Sx compared to the full QM calculation of $S4_{mp}$ within the respective peptide model.

which allow us to gain an impression of the influences caused by different charge situations in the QM part and the surrounding environment (see Table 5). The three model peptides differ in the protonation state of every fifth amino acid in the sequence, and thus, different charge situations arise for each structure type (see structural description above).

Pure QM calculations show a fairly slow convergence with the fragment size of models **mpA** to **mpC**, with errors on the order of 8 kJ/mol for the largest model fragment $S3_{mp}$ comprising 273 atoms of the total number of 383.

In contrast, the QM/MM calculations offer a more consistent convergence behavior. For all model structures the error decreases upon increasing the QM region. For the largest QM part $S3_{mp}$ the relative energies are converged within an error of about 2 kJ/mol for all three models (**mpA** to **mpC**), which indicates that clearly the QM part has to be chosen sufficiently large.

3.5. Influence of the Interconnection between Fragments. Typically the different fragments are interconnected by chemical bonds. In order to study this influence as compared to only through-space interactions, we compare model systems **mpC** and **mpC***: in the latter model connecting alanine residues have been removed (see structural description above).

The results listed in Table 5 indicate as expected a significantly faster convergence of the isomerization energy with the QM fragment size for structure **mpC***, since the coupling between the various units is not as pronounced anymore. It is worthwhile to note that the Coulomb-type energies (internuclear repulsion, nuclear-electron attraction, and the two-electron Coulomb part of HF) are not clearly better converged within the QM/MM as compared to the pure QM scheme (see Table S.1 in the Supporting Information). It is only the total isomerization energy which converges quicker in the decoupled system **mpC***.

Exploiting the information gained by the systematic study described above, we will employ in the following the HF/SVP approach in combination with the *charge shift* method as boundary adjustment for the electrostatic QM/MM scheme.

4. Proton Transfer in the TIM Enzyme

4.1. Structural Details. As a real case benchmark system, we used a monomeric subunit of the triosephosphate isomerase bound to phosphoglycolohydroxamate (PGH, enzyme complex denoted as TIM throughout this work; related work can be found in refs 11−13, 29, 32, and 67−71 and references therein). Here we focus on the proto-nation state by considering the proton-transfer energy between lysine (Lys13) and glutamate (Glu97), which is chosen as a model reaction and is of course not the function of the enzyme (for illustration of the structure and the isomerization process see Figure 2).

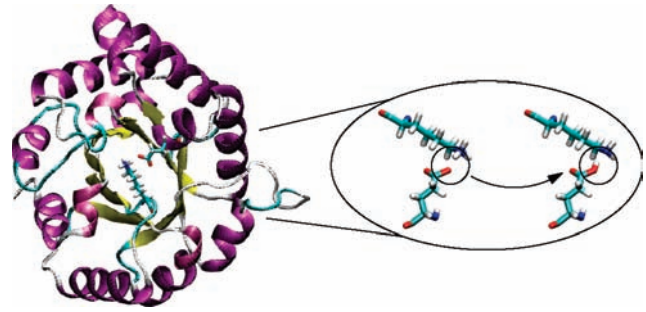

Figure 2. Illustration of the TIM complex and schematic picture of the isomerization process.

The initial coordinates for the TIM complex were taken from the work of Zhang et al.[69] (PDB code 1TPH with a resolution of 1.8 Å). Missing hydrogen atoms were added to the crystal structure and the protonation states of the titrable groups were always chosen as charged residues. A nonstandard assignment was selected only for residue Glu165 which was protonated as discussed in refs 67 and 69. While the crystal water present in the PDB entry was retained, no additional water molecules were added to the TIM complex. The N-terminus was saturated as ammonium group and the C-terminus as carboxylate. The proton positions for all residues were optimized by classical minimizations (for details see ref 72). The resulting structure was modified to build a glutamate-lysine isomer using the Maestro 7.5 program and only the hydrogen atoms at the modified positions (Glu97 HE2 and Lys13 HZ1/HZ2) were reoptimized.

For the TIM enzyme complex we chose six QM regions denoted as $S1_{tim}$ to $S6_{tim}$. The notation for the enclosed residues follows pdb-entry 1TPH.[69] $S1_{tim}$ comprises residues Glu97 and Lys13. To systematically enlarge the QM regions, spheres were placed around the center of the isomerization, and all residues with atoms in these spheres are included in the corresponding QM region. The radii of the spheres for $S2_{tim}$ to $S6_{tim}$ are 3, 5, 7, 9, and 11 Å. The largest QM region is denoted as $S6_{tim}$ comprising 1637 as compared to 4195 atoms in the total system. A complete list of the residues incorporated in the respective QM regions can be found in the Supporting Information.

4.2. QM-Region Size Convergence for the Total TIM Monomeric Unit. In Figure 3a the behavior of the isomerization energy upon increasing the size of the QM region (at the HF/SVP level) is studied using the six subsystems defined above, with the largest system comprising 1637 atoms (for detailed data see Supporting Information, Table S.2). Since for the full monomer unit with 4195 atoms the full QM calculation is not feasible with our present computer resources, we chose the

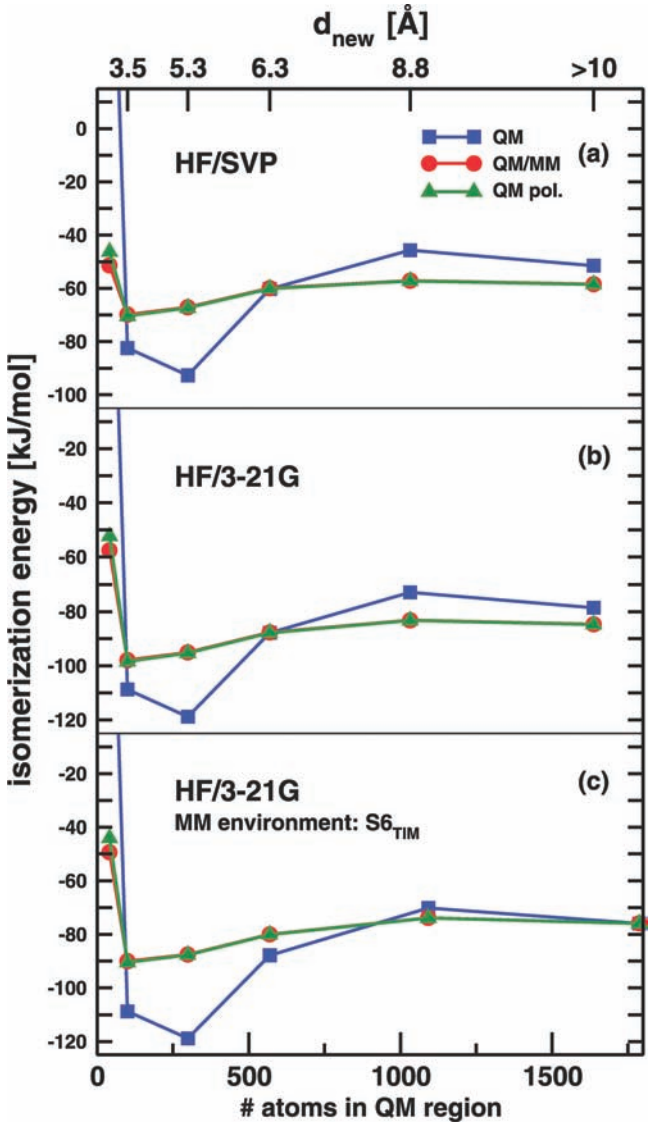

Figure 3. Convergence of isomerization energies for the TIM enzyme calculated with increasing QM-region size. In diagrams (a) and (b) the complete monomeric subunit was incorporated in the MM environment (4195 atoms). In diagram (c) a cutout of the monomeric subunit ($S6_{TIM}$, 1787 atoms) of TIM was used as MM environment. Also the full QM value for $S6_{TIM}$ is displayed. d_{new} describes the minimal distance between the modification and the Cα atoms of newly comprised residues upon enlarging the QM region.

isomerization energy obtained with $S6_{TIM}$ at the QM and QM/MM levels, respectively, as a reference for studying the convergence. Clearly the QM/MM isomerization energies converge much faster than the respective QM values: At the QM/MM level the relative energy was already converged within a range of 2 kJ/mol for QM region $S4_{TIM}$, that comprises 569 atoms and includes all residues within a distance of less than 6 Å around the modification. In constrast, the pure QM calculations reach convergence to slightly lower accuracies only if more than 1000 atoms are included, such as in $S5_{TIM}$. In passing we note that the QM/MM and QM_{pol} values are very similar for QM regions larger than $S1_{TIM}$, which is in accordance with the observations for the model peptide discussed above.

A strikingly large error was found for the smallest QM part $S1_{TIM}$ for the QM treatment. This can be understood as $S1_{TIM}$ contains only Lys13 and Glu97 while the electrostatic interaction between Lys13, Glu97, and the highly charged PGH is entirely neglected. Hence the error for the QM gas phase calculations of $S1_{TIM}$ was in the range of 200 kJ/mol.

At first sight, it may seem that the QM/MM value for $S1_{TIM}$ converges much faster with only 7 kJ/mol deviation. However, for $S2_{TIM}$ the value changes by 19 kJ/mol, which indicates that the agreement for the smaller fragment is fortuitous. Overall these kinds of errors for QM/MM calculations are in line with other studies for different systems,[29,32,34,37−39] in which, however, no extensive size convergence studies for large systems were perfomed.

Besides the HF/SVP results, Figure 3b also contains results obtained at the HF/3-21G level. Similar to the basis set influences observed for the model peptide above, the convergence of the isomerization energy is rather independent of the basis and only the absolute values change, which are, however, not in the focus of our present work. The only exception is found for the smallest and unrealistically small QM region $S1_{TIM}$ in which the deviation with respect to the corresponding result for $S6_{TIM}$ increases from 7 (HF/SVP) to 27 kJ/mol (HF/3-21G). This may be rationalized by the deficiency of the smaller basis set to describe polarization effects induced by the external point charges. In contrast, the pure QM convergence is virtually unaffected by the basis set change.

4.3. QM-Region Size Convergence for a Subsystem of the TIM Monomeric Unit. Since for the entire monomeric unit of TIM with 4195 atoms the full QM calculation is not possible due to limited computer resources, we had chosen above the subsystem $S6_{TIM}$ with 1637 atoms as a reference. However, this does not correspond to a true QM reference for the QM/MM system, since the remaining number of atoms up to the total of 4195 were always described at the MM level.

In order to obtain a true reference system, which can be treated entirely at the QM level, we chose in the following a subsystem of TIM with 1787 atoms obtained by saturating $S6_{TIM}$ which comprised originally 1637 atoms (where N-termini were saturated by acetylation and C-termini as carboxylates using Maestro 7.5 without further optimization). This 1787 atoms system (corresponding results in Figure 3c) is taken as the total system of which we again generated subsystems in the same way as described for TIM only modified by the saturations described above. In this way, we are able to perform the QM calculation for the full system. Since the convergence behavior observed using the 3-21G basis is expected to be very similar to the one for the SVP basis (see discussion above), we simplify our study for the present system to the 3-21G basis.

For the 1787 atoms system the behavior is very similar to the one observed for the total TIM system (compare Figure 3, panels c vs a and b). Only the QM/MM isomerization energies themselves are slightly shifted by roughly 5 kJ/mol, which is an indication for the relatively small influence caused by residues which are more than 10 Å apart from the isomerized amino acids.

Studying the error with respect to the QM value of the full 1787 atoms system shows that the QM/MM scheme leads already for $S4_{TIM}$ to a small deviation of 4 kJ/mol that decreases to 2 kJ/mol for the next larger QM region. Hence the inclusion of the residues in the range of 6−8 Å around the modification provides adequate QM/MM values for the isomerization. In contrast, for reliable QM values at least 8−10 Å of the surrounding need to be included in the gas phase calculation, while an error of 6 kJ/mol remains for $S5_{TIM}$, which comprises already more than 1000 atoms.

5. Conclusion

In our present work we have studied the convergence of QM/MM isomerization energies in dependence of the selected QM-region size for two example peptides with up to 1637 atoms at the HF/SVP and 383 atoms at the SOS-AO-MP2 level. The convergence toward the QM result for the full system is significantly improved by QM/MM as compared to pure QM calculations.

The isomerization energies of the present peptidic systems are only weakly influenced by electron-correlation effects, whereas basis-set effects upon the absolute isomerization energies are large. However, the convergence behavior is only very weakly influenced by them, which is in the focus of our present work. The small electron-correlation influence is expected, of course, to be different for many other biochemical systems in which electron correlation and in particular dispersion-type effects are of importance.

At the same time, the accelerated QM/MM convergence seems to be relatively insensitive to variations of the charge surrounding. Only if the chemical bonds responsible for the direct interconnection between the molecular fragments are removed, the convergence becomes much faster, since only through-space interactions occur.

Despite the faster convergence of QM/MM results, fairly large QM regions are required for good reliability even for the simple model peptides: systems of about 150−300 QM atoms are needed to achieve accuracies in the order of 10 kJ/mol and about 300−1000 atoms for an accuracy of 2 kJ/mol with respect to the converged full QM result for the total system. Such large QM fragments are nowadays accessible by modern linear-scaling methods both at the mean-field or correlated levels, so that the reliability of results obtained within the popular QM/MM approximation is expected to grow.

Acknowledgment. The authors thank Dr. Paul Sherwood (Daresbury Laboratory, U.K.) and Prof. Dr. Walter Thiel (MPI für Kohlenforschung, Germany) for the possibility to employ the ChemShell program suite as the basis for interfacing the MM approaches to the Q-Chem program package. Furthermore, we thank Bernd Doser (Universität Tübingen) for valuable support in performing the AO-MP2 calculations. C.O. acknowledges financial support by the VolkswagenStiftung within the funding initiative "New Conceptual Approaches to Modeling and Simulation of Complex Systems". C.V.S. thanks the FCI ("Fonds der Chemischen Industrie") for a graduate fellowship.

Supporting Information Available: The structures of the chosen model peptides will be publicly avaliable on the web under http://www.uni-tuebingen.de/qc/download. The Supporting Information contains a list of all residues incorporated in the respective QM regions of TIM and tables including details of isomerization energies for models **mpC**, **mpC***, and TIM. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. Linear-Scaling Methods in Quantum Chemistry. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Cundari, T. R., Eds.; VCH Publishers: New York, 2007; Vol. 23.

(2) Knowles, P.; Schütz, M.; Werner, H.-J. Ab Initio Methods for Electron Correlation in Molecules. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, 2000; Vol. 3.

(3) Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*, 4782.

(4) Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213.

(5) Ochsenfeld, C.; Kussmann, J.; Koziol, F. *Angew. Chem. Int. Ed.* **2004**, *43*, 4485.

(6) Lambrecht, D. S.; Ochsenfeld, C. *J. Chem. Phys.* **2005**, *123*, 184101.

(7) Sałek, P.; Høst, S.; Thøgersen, L.; Jørgensen, P.; Manninen, P.; Olsen, J.; Jansík, B.; Reine, S.; Pawłowski, F.; Tellgren, E.; Helgaker, T.; Coriani, S. *J. Chem. Phys.* **2007**, *126*, 114110.

(8) Rudberg, E.; Rubensson, E. H.; Sałek, P. *J. Chem. Phys.* **2008**, *128*, 184106.

(9) Doser, B.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2009**, *130*, 064107.

(10) Sherwood, P. Hybrid Quantum Mechanics/Molecular Mechanics Approaches. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, 2000; Vol. 3.

(11) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

(12) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173.

(13) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.

(14) Senn, H. M.; Thiel, W. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182.

(15) Mulholland, A. J. *Drug Discovery Today* **2005**, *10*, 1393.

(16) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.

(17) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425.

(18) Martí, S.; Roca, M.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Chem. Soc. Rev.* **2004**, *33*, 98.

(19) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.

(20) Sherwood, P.; et al. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1.

(21) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 15645.

(22) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem.* **1996**, *60*, 1189.

(23) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458.

(24) Sharma, R.; Thorley, M.; McNamara, J. P.; Watt, C. I. F.; Burton, N. A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2475.

(25) Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell, A. P. *J. Comput. Chem.* **2000**, *21*, 788.

(26) Ferré, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002**, *23*, 610.

(27) Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467.

(28) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angew. Chem.* **2006**, *118*, 7010.

(29) Lennartz, C.; Schäfer, A.; Terstegen, F.; Thiel, W. *J. Phys. Chem. B* **2002**, *106*, 1758.

(30) Mata, R. A.; Werner, H.-J.; Thiel, S.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 025104.

(31) Altun, A.; Thiel, W. *J. Phys. Chem. B* **2005**, *109*, 1268.

(32) König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.

(33) Hall, R. J.; Hindle, S. A.; Burton, N. A.; Hillier, I. H. *J. Comput. Chem.* **2000**, *21*, 1433.

(34) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.

(35) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.

(36) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.

(37) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.

(38) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejón, P.; Roitberg, A. E.; Estrin, D. A. *J. Phys. Chem. B* **2003**, *107*, 13728.

(39) Altun, A.; Shaik, S.; Thiel, W. *J. Comput. Chem.* **2006**, *27*, 1324.

(40) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663.

(41) Ochsenfeld, C. *Chem. Phys. Lett.* **2000**, *327*, 216.

(42) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8.

(43) White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1994**, *101*, 6593.

(44) Lambrecht, D. S.; Doser, B.; Ochsenfeld, C. *J. Chem. Phys.* **2005**, *123*, 184102.

(45) Doser, B.; Lambrecht, D. S.; Ochsenfeld, C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3335.

(46) The LSN and GLUP patches in the CHARMM22 force field were used as provided in the CHARMM package version c34b2.

(47) Hartree, D. R. *Proc. Cambridge Philos. Soc.* **1928**, *24*, 89.

(48) Fock, V. *Z. Phys.* **1930**, *61*, 126.

(49) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.

(50) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(51) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793.

(52) Development version of the program package Q-Chem. http://www.q-chem.com.

(53) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.

(54) Hehre, W.; Ditchfield, R.; Pople, J. *J. Chem. Phys.* **1972**, *56*, 2257.
(55) Hariharan, P. C.; Pople, J. A. *Theor. Chem. Acc.* **1973**, *28*, 213.
(56) Clark, T.; Chandrasekhar, J.; Spitznagel, G.; v, R. Schleyer, P. *J. Comput. Chem.* **1983**, *4*, 294.
(57) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.
(58) Antes, I.; Thiel, W. Hybrid Quantum Mechanical and Molecular Mechanical Methods. In *ACS Symposium Series*; Gao, J., Ed.; American Chemical Society: Washington, DC, 1998; Vol. 712.
(59) de Vries, A. H.; Sherwood, P.; Collins, S. J.; Rigby, A. M.; Rigutto, M.; Kramer, G. J. *J. Phys. Chem. B* **1999**, *103*, 6133.
(60) MacKerell Jr., A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
(61) Brooks, B. R.; Bruccoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(62) MacKerell Jr., A. D.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In *The Encyclopedia of Computational Chemistry*; v. R. Schleyer, P., Eds.; John Wiley & Sons: Chichester, 1998; Vol. 1.
(63) Smith, W.; Forester, T. *J. Mol. Graph.* **1996**, *14*, 136.
(64) Schrödinger Inc., Portland OR, Maestro 7.5, 2005.
(65) Asada, T.; Takahashi, T.; Koseki, S. *Theor. Chem. Acc.* **2008**, *120*, 263.
(66) Melo, A.; Ramos, M.; Floriano, W. B.; Gomes, J.; Leão, J.; Magalhães, A.; Maigret, B.; Nascimento, M. C.; Reuter, N. *J. Mol. Struct. (THEOCHEM)* **1999**, *463*, 81.
(67) Davenport, R. C.; Bash, P. A.; Seaton, B. A.; Karplus, M.; Petsko, G. A.; Ringe, D. *Biochemistry* **1991**, *30*, 5821.
(68) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826.
(69) Zhang, Z.; Sugio, S.; Komives, E. A.; Liu, K. D.; Knowles, J. R.; Petsko, G. A.; Ringe, D. *Biochemistry* **1994**, *33*, 2830.
(70) Neria, E.; Karplus, M. *Chem. Phys. Lett.* **1997**, *267*, 23.
(71) Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 1768.
(72) Classical minimizations were carried out using the CHARMM22 force field[60] for the model peptide and TIM. Additional parameters for the TIM inhibitor PGH were taken from the ChemShell tutorial section. The CHARMM package was run through the DL_POLY code[63] as integrated in ChemShell.[20]

JP902876N