Available online at www.sciencedirect.com
ScienceDirect
Polymer 48 (2007) 3726–3731
polymer
www.elsevier.com/locate/polymer

On the 3D structure and catalytic mechanism study of AmiF formamidase of *Helicobacter pylori*

Wei-Wei Han, Yi-Han Zhou, Quan Luo, Yuan Yao, Ze-Sheng Li*

Institute of Theoretical Chemistry, State Key Laboratory of Theoretical and Computational Chemistry, Jilin University, Changchun 130023, People's Republic of China

Received 8 March 2007; received in revised form 24 April 2007; accepted 24 April 2007
Available online 5 May 2007

Abstract

The 3D structure of the AmiF formamidase of *Helicobacter pylori* (denoted as AmiF(fhp)) is built by homology modeling. The docking studies show that AmiF(fhp) has restricted substrate specificity, as it only hydrolyzes formamide. In order to reveal the reaction mechanism and the catalytic role for Cys166, Lys133 and Asp168 in AmiF(fhp), three quantum mechanics' models are constructed based on the 3D structure of AmiF(fhp) and the reaction paths are obtained at B3LYP 6-31 + G* level. The calculated results show that (1) the reaction of Cys166–formamide anion in the enzyme active proceeds via a transition state without the intervention of tetrahedral intermediate; (2) the positive charge on Lys133 polarizes the formamide in the TS region to redistribute the electron and thence decreases the free energy barrier; (3) the active site residue of Asp168 increases the free energy barrier as the negative charge will affect the electron distribution.

Keywords: AmiF(fhp); Docking; Catalytic mechanism

1. Introduction

The human *Helicobacter pylori* is a microaerophilic, spiral-shaped, Gram-negative bacterium that colonizes the gastric mucosa [1,2]. It is a risk factor for the development of gastric cancer [1–3]. This pathogen produces several virulence factors. One of the major factors contributing to acid resistance of *H. pylori* is the production of ammonia by its urease enzyme, which is essential for gastric colonization in different animal models [2,4–7]. Ammonia is a key component of bacterial nitrogen metabolism, because it is the preferred source of nitrogen for the synthesis of amino acids, pyrimidines, and purines [8–10]. Urea is thought to be the main source of ammonia in the gastric environment, but *H. pylori* does have alternative pathways for the production of ammonia via amino acid catabolism and the activity of its two paralogous amidases, aliphatic amidase (AmiE, EC 3.5.1.49) and formamidase (AmiF(fhp) EC 3.5.1.49) [1,2,11,12]. Formamide, the smallest amide in existence, is the unique substrate of AmiF(fhp) [1]. Only a few bacteria have been described as being able to grow on formamide as a nitrogen source [1].

AmiF(fhp) is a thiol enzyme, converting formamide directly into the corresponding formyl acid plus ammonia through a transition state, without the intermediate formation of an amide. AmiF(fhp) belongs to the nitrilase superfamily, and Cys166 functions as the nucleophile in the catalytic mechanism [1,2].

To the best of our knowledge, the 3D structure of AmiF(fhp) is not known up to now. In the present investigation, the 3D model of AmiF(fhp) was built by a homology modeling procedure based on the crystal structure of *N*-carbamyl-D-amino acid amidohydrolase (PDB code 1ERZ) [13]. In order to describe the bond-breaking and bond-making processes in the enzyme, the reaction models were set up and the quantum mechanical (QM) calculation was carried out to obtain the equilibrium geometries and frequencies of stationary points (reactants, products, and transition states). Our

* Corresponding author. Tel.: +86 431 88498960; fax: +86 431 88998026.
E-mail address: zeshengli@mail.jlu.edu.cn (Z.-S. Li).

0032-3861/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2007.04.070

results may be helpful for understanding the mechanism of the formation of formamide.

2. Theory and methods

2.1. Target and template proteins

The amino acid sequence of the target protein, AmiF(fhp), was obtained from UniProtKB-Swiss-Prot (Accession No. O25836) with 334 residues involved [1,2]. The template protein was *N*-carbamyl-D-amino acid amidohydrolase, deposited in Protein Data Bank (PDB code 1ERZ) [13].

2.2. Flexible docking

Affinity [14] was used for docking. In the docking process, the potential function of the complex was assigned by using the CVFF force field and the non-bonding interaction was dealt with cell-multiple approach. To consider the solvent effect, the centered enzyme–ligand complexes were solvated in a sphere of TIP3P water molecules with radius 10 Å. All the default parameters in the affinity module were used. Ten conformations were presented from surface affinity (SA) docking and the generated conformations were clustered according to the RMS deviation (RMSD). The docked complexes were finally selected by the criterion of the total energy combined with the geometrical matching quality and favorable interaction energy.

2.3. Quantum mechanical calculation method

All of the QM calculations have been performed with Gaussian 03 program package. Density functional theory with the three-parameter hybrid exchange functional of Becke and Lee, Yang, and Parr correlation functional (B3LYP) [15,16] was employed. As it is known that the B3LYP functional is superior to the other current DFT functional [17]. And its accuracy is comparable to the more accurate ab initio methods [18,19]. In our study, the 6-31+G* basis set was employed for all geometry optimizations and intrinsic reaction coordinate (IRC) calculations. For the reactant, the transition state, and product, the single-point calculations with 6-311+G (2d, 2p) basis set were carried out on the 6-31+G* geometry. To consider the solvent effects of the enzyme environment on the energetic of reaction step, the polarizable-continuum model (CPCM) [17] was used to calculate the Gibbs free energy of salvation for each species using its gas-phase-optimized geometry. The first QM subsystem (denoted as model A) consists of the formamide and sidechain of Cys166, with a total of 11 atoms. The second subsystem (denoted as model B) includes model A, and sidechain of Lys133, leading to a total of 28 atoms. The third subsystem (denoted as model C) involves model B and sidechain of Asp168, with a total of 38 atoms. In our calculations, the heavy atoms for models B and C were fixed in their homology modeling presented positions and this may give rise to a few small negative eigenvalues for the optimized structures [17,19]. These are, however, very small, in the order from -10 to $-30\,cm^{-1}$, and do not affect the obtained results. In these models, a cavity around the system is surrounded by a polarizable dielectric continuum. The dielectric constant chosen is $\varepsilon = 4$, which is the standard value used to model the protein surrounding [17]. Furthermore, the thermal correction to Gibbs free energies was carried out at the temperature of 318 K, as it was reported that the optimum temperature for AmiF(fhp) is 318 K [1].

3. Results and discussion

3.1. Sequence alignment and homology models validation

BLAST [20] procedure was used for the on line search (http://pir.georgetown.edu/). The sequence identity between the AmiF(fhp) and the reference protein *N*-carbamyl-D-amino acid amidohydrolase (PDB code 1ERZ) [13] is 26% which allows for straightforward sequence alignment. It has been suggested that short-chain aliphatic amidases are structurally related to a family of carbon–nitrogen (C–N) hydrolases that includes nitrilases, cyanide hydratases and dihydratases and β-alanine synthases [1,21]. Since a number of structural motifs are conserved between AmiF(fhp) and *N*-carbamyl-D-amino acid amidohydrolase, and they all belong to carbon–nitrogen (C–N) hydrolase family. These proteins are susceptible to thiol reagents, and this led to their classification as sulphydryl enzymes [1,21,22]. Despite a relatively low sequence homology, one can confidently build a homology model of carbon–nitrogen (C–N) hydrolases due to their relatively model good structural homology [1,23]. So the chosen template is reasonable.

Following the alignment, the backbone coordinates of the residues were generated with the modeler [24] and the side-chain conformations were optimized by manually selecting lower-energy conformations from a rotamer library. Then the initial model was constructed and this model was further refined by using energy minimization and molecular dynamics simulation. The conformation at the 840 ps MD simulation is chosen as the final structure as presented in Fig. 1. When the structure is superimposed on 1ERZ, their root mean square deviation (RMSD) value is 0.94 Å. The compatibility scores are obtained by using Profile-3D [25] and the results are plotted in Fig. 2. Note that compatibility scores above zero correspond to an 'acceptable' side-chain environment. From Fig. 2, we know that 10 residues (Leu309–Lys318) with compatibility scores below zero are presented in 334 residues. Fortunately, these residues are far away from the active site of AmiF(fhp) in the 3D structure. Thus, they could not influence the catalytic function of the enzyme.

ProStat was also used to calculate the percentage of backbone $\phi-\varphi$ angles within the allowed Ramachandran region. The result shows that 75.9% of the $\phi-\varphi$ angles is in the Ramachandran plot. No significant difference between the calculated values and the known proteins is found by ProStat

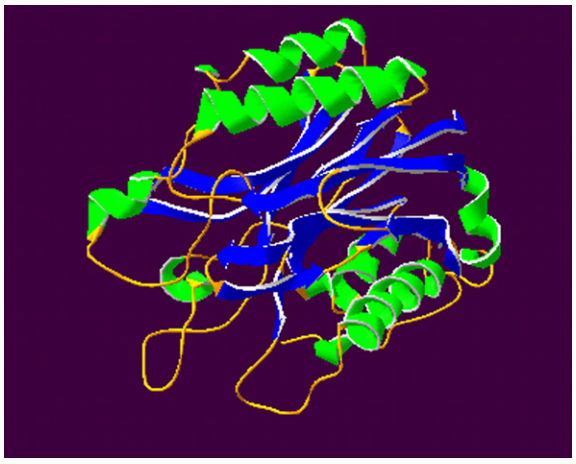

Fig. 1. The predicted structure of amidase based on homology-based modeling.

analysis. By checking with two different criteria mentioned above, we believe that the final model is reliable.

It should be pointed out that the crystallographic resolutions of AmiF(fhp) (PDB code 2DYV, 2DYU, 2E2K and 2E2L) were solved during our study. The crystallographic resolutions of this protein (PDB code 2DYV) can be employed to testify the rationality of our final model. It can be found that the root mean sequence deviation of the Cα atoms (Cα RMSD) between homology model and the crystallographic resolutions of this protein (PDB code 2DYV) is 1.29 Å for 334 residues. Furthermore, from the BLAST search, we knew that the domain from residue 14 to 287, which contains the catalytic triad (Glu60, Cys166, and Lys133), belonged to CN hydrolase. We also compared the homology model (from residue 14 to 287) with the crystallographic resolutions of AmiF(fhp), and found that the root mean sequence deviation of the Cα atoms (Cα RMSD) is 1.16 Å for 273 residues. These results further indicate that the homology model is reasonable.

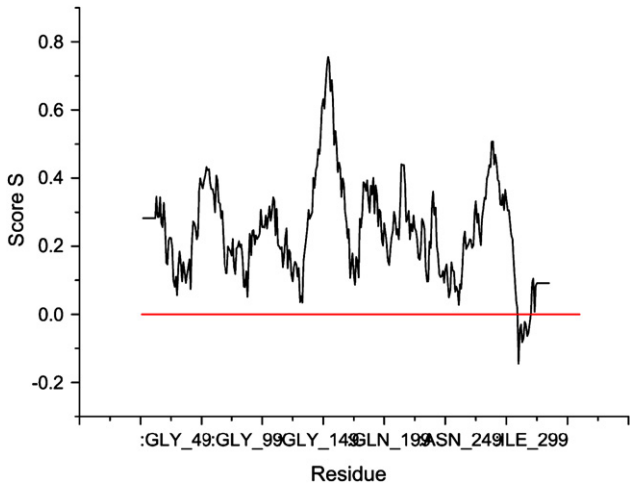

Fig. 2. 3D profiles verified results of AmiF model, residues with positive S are reasonably folded.

3.2. Identification of substrate-binding region

Among several binding sites given by Insight II/binding site module [26], the most similar to the binding site of *N*-carbamyl-D-amino acid amidohydrolase determined from the X-ray crystal structure was chosen. It was reported that AmiF(fhp) has catalytic triad (Cys166, Glu60 and Lys133) [1,2]. The cavity volume measured by binding site is dependent on the maximum distance between grid points exposed at the aperture to the cavity; the default value of 1.4 Å outlines a cavity of 288 $Å^3$ for AmiF(fhp). Whereas mutation of Asp168 in Ala of AmiF(fhp) (D168A) makes the default value of 1.4 Å outline a cavity of 212 $Å^3$. From these results, we can conjecture that mutation of Asp168 in Ala of AmiF(fhp) (D168A) makes the active site less spacious than that of AmiF(fhp), and it can lead to the enzyme inactive.

We also measure the cavity volume of the crystallographic resolutions of AmiF(fhp) (PDB code 2DYV), and find that the default value of 1.4 Å outlines a cavity of 280 $Å^3$ for the crystallographic resolutions of AmiF(fhp) (PDB code 2DYV). While the distance between Cα of Cys166 and Cα of Asp168 is 5.30 Å in the crystallographic resolutions of AmiF(fhp) (PDB code 2DYV) and distance between Cα of Cys166 and Cα of Lys133 is 9.57 Å. In the homology model, the distance between Cα of Cys166 and Cα of Asp168 is 5.46 Å, and the distance between Cα of Cys166 and Cα of Lys133 is 10.33 Å. From these results, we believe that the homology model is proximate to the crystallographic resolutions of AmiF(fhp). It is known that for the structural optimization in quantum chemical calculation, the sidechains of residues in selective amino acid in the active site are flexible and permit them adjustable to arrive at the optimal structure. Thus, the little difference between our homology model and the crystallographic resolutions of AmiF(fhp) (PDB code 2DYV) might not sufficiently influence the final calculated results to reveal the catalytic mechanism for this enzyme.

3.3. Docking study

It was reported that formamide is the unique and specific substrate of the AmiF(fhp) enzyme ($K_m = 32 \pm 8.7$ mM, $V_m = 1081 \pm 114$ μmol·NH_3 min^{-1} mg^{-1}), because no activity is detected with the other substrates, such as propinamide and acetamide [1,2]. This substrate specificity for AmiF(fhp) can be explained by the substrate affinity of AmiF(fhp) with formamide based on its 3D structure and docking experiment, and by comparative docking of propinamide and acetamide with AmiF(fhp). The results are shown in Table 1. From Table 1, we can see that the total interaction energy between AmiF(fhp) and formamide (−21.86 kcal mol^{-1}) is lower than those of AmiF(fhp)–acetamide (−8.33 kcal mol^{-1}) and AmiF(fhp)–propinamide (−8.80 kcal mol^{-1}). On one hand, compared with AmiF(fhp)–formamide, the AmiF(fhp)–acetamide and AmiF(fhp)–propinamide complexes are not thermodynamically stable. On the other hand, the volumes of acetamide and propinamide are larger than that of formamide's, and the crowding of the acetamide or propinamide

Table 1
The calculated energies (kcal mol^{-1}) of the ligand tested for AmiF binding

Ligand	E_{vdw} (kcal mol^{-1})	E_{ele} (kcal mol^{-1})	E_{total} (kcal mol^{-1})	K_m [a]
Formamide	−7.53	−14.33	−21.86	32 ± 8.7
Acetamide	−11.66	3.33	−8.33	–
Propinamide	−7.38	−1.42	−8.80	–
AmiF(D168A)–formamide	−4.37	−0.89	−5.26	–

[a] K_m proposed by Skouloubris et al.

at the active site may lead to the spatial obstacle and block the nucleophile attacking. Our docking results are in agreement with the kinetic experiment by Skouloubris et al. as they pointed out that formamide appears to be the unique and specific substrate of the AmiF(fhp) enzyme [1].

It was reported that comparison of amino acid sequences for the region surrounding the active site of the C–N hydrolases identified an aspartate residue that is conserved in all bacterial amidases and is replaced by glutamate residue in nitrilase, cyanide hydrolase and dihydratase enzymes [1]. We also dock formamide to AmiF(fhp) (D168A). Seen from Table 1, the total interaction energy of AmiF(fhp) (D168A)–formamide (−5.26 kcal mol^{-1}) is higher than that of AmiF(fhp)–formamide (−21.86 kcal mol^{-1}), and this means that this mutation makes the AmiF(fhp) thermodynamically less stable. From binding site search, we know that the active site of AmiF(fhp) (D168A) is less spacious than that of AmiF(fhp). As deduced from the active site model of the enzyme, the larger sulfur atom cannot fit into the less spacious pocket. So mutation of Asp168 to Ala can make the enzyme inactive. This result is in harmony with the mutation experiment by Skouloubris et al. as they pointed out that Asp residue is essential for the conformational stability of the AmiF and no formamidase is detected in the mutation protein [1,2].

3.4. Quantum mechanical calculation

Through the analysis of the final structure of AmiF(fhp) (EC 3.5.1.49), it is known that the active site of this enzyme is mainly composed of Cys166, Lys133, and Glu60 [1,2]. It might be expected that cysteine proteases have a similar mechanism to the serine proteases since the nature and orientation of the catalytic groups are similar. The proposed mechanism of *N*-carbamyl-D-amino acid amidohydrolase is outlined as follows: the activated nucleophile (Cys171) attacks the amide carbon of a substrate to form a tetrahedral intermediated, and

Table 2
calculated the free energy (kcal mol^{-1}) difference for the different models used

Model	Parts included	Reactant (in gas)	Transition state (in gas)	Product (in gas)
A	Sidechain of Cys166 + formamide	0	50.90	26.73
B	A + sidechain of Lys133	0	45.43	39.78
C	B + sidechain of Asp168	0	47.37	39.99
Model	Parts included	Reactant (in solution)	Transition state (in solution)	Product (in solution)
A	sidechain of Cys166 + formamide	0	47.10	30.17
B	A + sidechain of Lys133	0	43.94	28.40
C	B + sidechain of Asp168	0	44.79	34.68

then the intermediate collapses to form an acyl-enzyme complex and releases ammonia [13]. However, Cys166–formamide anion in the active site of AmiF(fhp) proceeds via a transition state without the intervention of tetrahedral intermediate (shown in Fig. 3) [21,27,28]. In the present study, three quantum chemical models with different sizes, ranging from 11 atoms up to 38 atoms, have been used to study the nucleophilic attack of SH^- on formamide step in AmiF(fhp).

In our calculation models, we first choose a small model consisting of formamide, and the sidechain of Cys166 (denoted as model A), and then add the other residues one at a time (model B and model C). The energetic results are summarized in Table 2, and the optimized transition state structures for model C are shown in Fig. 4.

Model A has a free energy barrier of 50.90 kcal mol^{-1} in gas and 47.10 kcal mol^{-1} in solution, and is exothermic by 30.17 kcal mol^{-1} in gas and 26.73 kcal mol^{-1} in solution. These results are in agreement with the kinetic experiment performed by Skouloubris et al. as they pointed out that Cys166 in AmiF(fhp) is essential for the catalytic activity in *H. pylori* [1].

Adding the sidechain of Lys133 (model B) to model A results in a decrease of the free energy barrier with 45.43 kcal mol^{-1} in gas and 43.94 kcal mol^{-1} in solution. This step is exothermic by 39.78 kcal mol^{-1}, and CPCM calculation shows that the protein environment will decrease the exothermicity to 28.40 kcal mol^{-1}. The function of Lys133 in the first step of acylation is to redistribute the electron of TS. By comparing the results of model A and model B, the free energy barrier of model B is lower than that of model

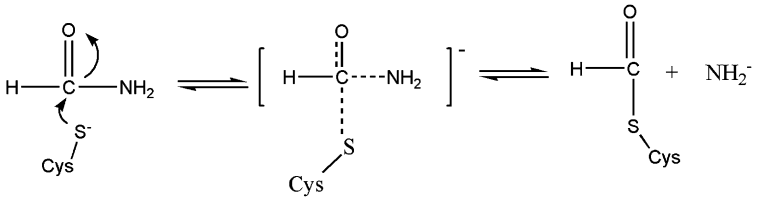

Fig. 3. Reaction catalyzed by AmiF.

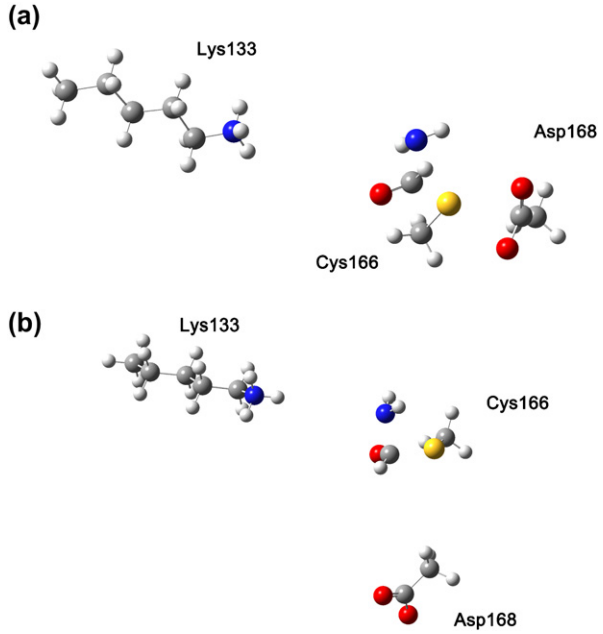

Fig. 4. The transition states located in active site: (a) model C in solution and (b) model C in gas.

A's (4.27 kcal mol^{-1} in gas and 3.16 kcal mol^{-1} in solution). This means that the ε-NH_2 group of Lys133 makes the electron transfer, polarizes the formamide in the TS region and thence decreases the free energy barrier either in gas or in solution. Mulliken population analysis confirms this: in model A at the transition state, oxygen of formamide bears a charge of −0.437 in gas, while in model B the charge is −0.433. In the solution, the difference is larger, from −0.587 to −0.523. Furthermore, Lys133 is in a good position to act as a proton shuttle which is very important in subsequent reaction steps when the leaving nitrogen is protonated by Glu60 [13,29]. The reduced activity of mutants Lys133 can lose this effect in acylation. This is in good agreement with relevant experimental findings that substitution of Lys by Asp is associated with a large loss of activity [29].

When both sidechains of Asp168 and Lys133 are added (model C), the free energy barrier is found to be higher (47.37 kcal mol^{-1} in gas, 44.79 kcal mol^{-1} in solution) compared with model B (45.43 kcal mol^{-1} in gas and 43.94 kcal mol^{-1} in solution). This step is exothermic by 39.99 kcal mol^{-1}, and CPCM calculation shows that the protein environment will decrease the exothermicity to 34.68 kcal mol^{-1}. These results are not inconstant with the suggestion that Asp168 is probably not involved in the catalytic mechanism, but is probably involved in maintenance of the structural integrity of the AmiF(fhp), as pointed out by Skouloubris et al. [1]. As also seen for model C discussed above (Fig. 4), the negative group of Asp168 is found to point towards the formamide, and may influence the electron transfer. Mulliken population analysis shows that in model C the Mulliken charges on oxygen of formamide are −0.421 in gas and −0.510 in solution at the transition state. As for model B at the transition state, the Mulliken charges on oxygen of formamide are −0.433 in gas and −0.523 in solution (Table 3). The difference of Mulliken charges between model B and model C indicate that the negative charge group of Asp168 can influence the electron transfer in the catalytic reaction.

Table 3
Mulliken atomic charges with oxygen of formamide with model A, model B and model C

Species	Transition state (in gas)	Transition state (in solution)
Model A	−0.437	−0.587
Model B	−0.433	−0.523
Model C	−0.421	−0.510

4. Conclusion

The 3D structure of AmiF(fhp) (EC 3.5.1.49) is obtained by homology modeling by taking the *N*-carbamyl-D-amino acid amidohydrolase as a template. The docking study AmiF(fhp) has restricted substrate specificity, as it only hydrolyzes formamide. Furthermore, the reaction models were proposed and the reaction mechanism was investigated by employing quantum mechanics' calculations in both gas phase and solution phase. The calculated results are in harmony with relevant experimental results. Although our reaction model is an approximation to reveal the reaction mechanism, it would be beneficial to investigate the enzymatic reactions.

Acknowledgment

This work was supported by the National Science Foundation of China (20333050, 20673044), Doctor Foundation by the Ministry of Education, Foundation for University Key Teacher by the Ministry of Education, Key subject of Science and Technology by the Ministry of Education of China, and Key subject of Science and Technology by Jilin Province.

References

[1] Skouloubris S, Labihne A, Reuse HD. Mol Microbiol 2001;40:596.
[2] Vliet AHV, Stoof J, Poppelaars SW, Bereswill S, Homuth G, Kist G, et al. J Biol Chem 2003;278:9052.
[3] Alm RA, Ling LS, Moir LDT, King BL, Brown ED, Doig PC. Nature 1999;397:176.
[4] Eaton KA, Krakowka S. Infect Immun 1994;62:3604.
[5] Tsuda M, Karita M, Morshed MG, Okita K, Nakazawa T. Infect Immun 1994;62:3586.
[6] Wirth HP, Beins MH, Yang M, Tham KT, Blaser MJ. Infect Immun 1998; 66:4856.
[7] Nolan KJ, McGee DJ, Mitchell HM, Kolesnikow T, Harro JM, ORourke J, et al. Infect Immun 2002;70:685.
[8] Williams CL, Preston TM, Hossack C, Slater C, McColl K. FEMS Immunol Med Microbiol 1997;13:87.
[9] Igarashi M, Kitada Y, Yoshiyama H, Takagi A, Koga Y. Infect Immun 2001;69:816.
[10] Nagahashi S, Suzuki H, Miyazawa M, Miura S, Ishii H. Free Radic Biol Med 2002;33:1073.
[11] Nakamura H, Yoshiyazawa M, Mizote T, Okita K, Nakazawa T. Infect Immun 1998;66:4832.

[12] McGee DJ, Radcliff FJ, Mendz GL, Ferrero RL, Mobley HL. J Bacteriol 1999;181:7314.
[13] Nakai T, Hasegawa T, Yamashita E, Yamamoto M, Kumasaka T, Ueki T, et al. Structure 2000;8:729.
[14] Affinity user guide. San Diego, USA: Accelrys Inc.; 1999.
[15] Wang JY, Dong H, Li SH, He HW. J Phys Chem B 2005;109:18644.
[16] Kravitz JY, Pecoraro V, Carlson HA. J Chem Theory Comput 2005; 1:1265.
[17] Himo F. Theor Chim Acta 2006;116:232.
[18] Guner V, Khuong KS, Leach AG, Lee PS, Bartberger MD, Houk KN. J Phys Chem A 2003;107:11445.
[19] Curtiss LA, Krishnan R, Redfern PC, Pople JA. J Chem Phys 1997;106: 1063.
[20] Altschul SF, Madden TL, Schfer AA, Zhang JZ, Miller DJ. Nucleic Acids Res 1997;25:3389.
[21] Bauschlicher Jr CW. Chem Phys Lett 1995;246:40.
[22] Bassan A, Zou WB, Reyes E, Himo F, Cordova A. Angew Chem Int Ed 2005;44:7028.
[23] Brenner C. Cur Opin Sci Biol 2002;12:775.
[24] Sali ATL, Blundell TL. J Mol Biol 1993;234:779.
[25] Profile-3D user guide. San Diego, USA: Accelrys Inc.; 1999.
[26] Binding site analysis user guide. San Diego, USA: Accelrys Inc.; 1999.
[27] Harrison MJ, Burton NA, Hillier IH. J Am Chem Soc 1997;119:12285.
[28] Byun K, Gao JL. J Mol Graph Model 2000;18:50.
[29] Novo CS, Farnaud S, Tata R, Clemente A, Brown PR. Biochem J 2002; 365:731.