Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemical Physics Letters 401 (2005) 24–29
CHEMICAL PHYSICS LETTERS
www.elsevier.com/locate/cplett

Homology models of main proteinase from coronavirus associated with SARS

Hsuan-Liang Liu [a,*], Jin-Chung Lin [a], Yih Ho [b], Chin-Wen Chen [a]

[a] *Department of Chemical Engineering, Graduate Institute of Biotechnology, National Taipei University of Technology, No. 1 Sec. 3, Chung-Hsiao E. Road, Taipei 10608, Taiwan*
[b] *School of Pharmacy, Taipei Medical University, No. 250 Wu-Hsing Street, Taipei 110, Taiwan*

Received 27 February 2004; in final form 3 November 2004
Available online 23 November 2004

Abstract

In this study, two homology models of the main proteinase (M^{pro}) from the novel coronavirus associated with severe acute respiratory syndrome (SARS-CoV) were constructed. These models reveal three distinct functional domains, in which an intervening loop connecting domains II and III as well as a catalytic cleft containing the substrate binding subsites S1 and S2 between domains I and II are observed. S2 exhibits structural variations more significantly than S1 during the 200 ps molecular dynamics simulations because it is located at the open mouth of the catalytic cleft and the amino acid residues lining up this subsite are least conserved. In addition, the higher structural variation of S2 makes it flexible enough to accommodate a bulky hydrophobic residue from the substrate.

1. Introduction

Coronaviruses belong to a diverse group of positive-stranded RNA viruses and share a similar genome organization and common transcriptional/translational processes as *Arteriviridae* [1,2]. The human coronavirus HcoV-229E replicase gene encodes two overlapping polyproteins [3], that mediate all the functions required for viral replication and transcription [4]. The functional polypeptides are released from the polyproteins by extensive proteolytic processing, which is primarily achieved by the 33.1-kDa main proteinase (M^{pro}) [5]. M^{pro} from HcoV-229E (M^{pro}H) has been biosynthesized in *Escherichia coli* and its enzyme properties have been well characterized [5,6].

Several studies have revealed significant differences in both the active sites and domain structures of M^{pro} from coronavirus and picornavirus [6–8]. Previous experimental data have shown that the differential cleavage kinetics of all coronaviruses is a conserved feature of M^{pro} [9]. Furthermore, the cleavage pattern appears to be conserved in M^{pro} from SARS-CoV (M^{pro}S) and from other coronaviruses [10], as deduced from the genome sequence [11,12]. The functional importance of M^{pro} in the viral life cycle has made it an attractive target for the development of drugs directed against SARS and other coronavirus infections. Thus, screening the known proteinase inhibitor libraries may be an appreciated shortcut to discover anti-SARS drugs [13]. Crystal structures of M^{pro}H [10] and M^{pro} from porcine coronavirus (transmissible gastroenteritis virus, TGEV) (M^{pro}T) complexed with its inhibitor [14] have been determined. Comparison of these structures reveals a remarkable degree of structural conservation.

Previously, several molecular dynamics (MD) simulations, homology modeling, and molecular docking

* Corresponding author. Fax: +886 2 2731711.
E-mail address: f10894@ntut.edu.tw (H.-L. Liu).

0009-2614/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2004.11.030

experiments have been conducted in our group [15–18]. In this Letter, two homology models of $M^{pro}S$ (denoted as $M^{pro}SH$ and $M^{pro}ST$) were constructed based on the crystal structures of $M^{pro}H$ [10] and $M^{pro}T$ [14], respectively. In addition, MD simulations were performed to investigate the dynamics behaviors of these structures.

2. Methods

2.1. Template proteins

The atomic coordinates of $M^{pro}T$ and $M^{pro}H$ were obtained from the protein data bank (PDB; 1lvo and 1p9u, respectively). Unfavorable non-physical contacts in these structures were eliminated using Biopolymer module of Insight II (Accelyrs, San Diego, CA, USA) with the CVFF forcefield [19] in the SGI $O2^+$ workstation with 64-bit MIPS RISC R12000 270 MHz CPU and PMC-Sierra RM7000A 350 MHz processor (Silicon Graphics, Inc., Mountain View, CA, USA), followed by 10000 energy minimization calculations using steepest descent method.

2.2. Structural homology

The procedures of amino acid sequence alignment and homology modeling were described previously [18]. The newly built homology models were substantially refined to avoid van der Waals radius overlapping, unfavorable atomic distances, and undesirable torsion angles using molecular mechanics and dynamics features in Discover module.

2.3. Molecular dynamics simulations

The present MD simulations were performed in the CVFF forcefield [19]. The crystal structures of $M^{pro}H$ and $M^{pro}T$ and the homology models of $M^{pro}SH$ and $M^{pro}ST$ were subjected to energy minimization calculations. Each energy-minimized structure was placed in the center of a lattice with the size of $50 \times 60 \times 85$ Å^3 full of 6222, 5866, 5836, and 5776 water molecules for the system of $M^{pro}H$, $M^{pro}T$, $M^{pro}SH$, and $M^{pro}ST$, respectively. In order to arrange the soaked water molecules randomly, water molecules alone were submitted to 10000 iterations by conjugate gradient minimization, keeping the protein atoms fixed. The system composed of the minimized structures of protein and water molecules was then used as the starting image. Finally, 200 ps MD simulation with 5 ps in equilibrium step was carried out for each system using the Discover module of Insight II. The explicit image periodic boundary condition (PBC) was used for solvent equilibrium. The temperature and pressure were maintained for each MD simulation at 300 K and one atmosphere, respectively, as described by Berendsen et al. [20]. Cut-off radius of 10 Å for the non-bonded interactions was applied. The time-step of the MD simulations was 1 fs. The trajectories and coordinates of these structures were recorded every 2 ps for further analysis.

3. Results and discussion

3.1. Amino acid sequence alignment

The results of amino acid sequence alignment of $M^{pro}S$ to $M^{pro}T$ and $M^{pro}H$ are given in Fig. 1. The residue corresponding to Ala46 in domain I of $M^{pro}S$ and those corresponding to Asp248, Ile249, and Gln273 in domain III of $M^{pro}S$ are missing in both $M^{pro}T$ and $M^{pro}H$. In addition, there are one and two extra residues at the C-terminus of $M^{pro}S$ comparing to $M^{pro}T$ and $M^{pro}H$, respectively. Domain III exhibits higher sequence variation among these three domains. Both

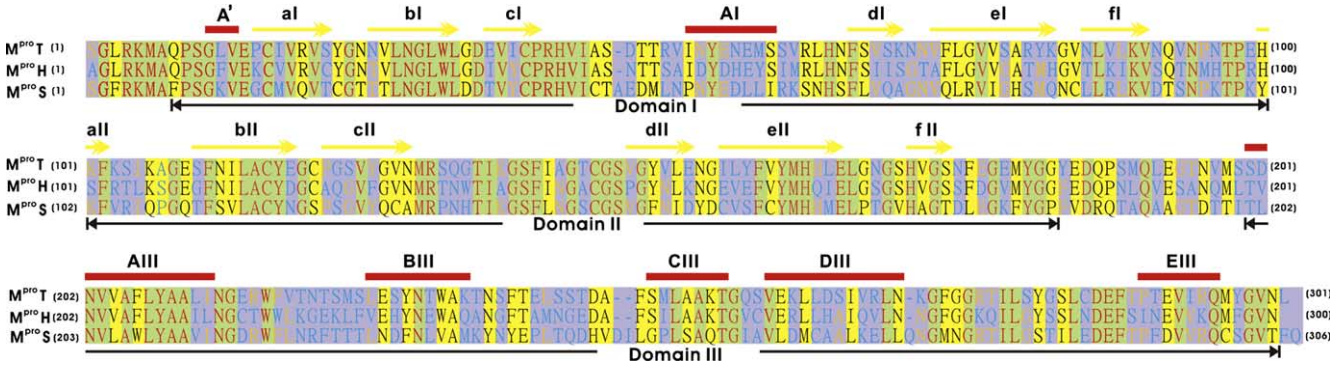

Fig. 1. Amino acid sequence alignment of $M^{pro}T$, $M^{pro}H$, and $M^{pro}S$. Secondary structures defined in the crystal structure of $M^{pro}T$ are shown on top.

the general acid–base catalyst (His residue in domain I) and the nucleophile (Cys residue in domain II) of these three proteins are totally conserved.

Table 1 lists the percentages of amino acid identity among these proteins. $M^{pro}T$ and $M^{pro}H$ show the highest total amino acid identity (60.80%), whereas $M^{pro}H$ and $M^{pro}S$ exhibit the lowest total amino acid identity (40.19%). In addition, domain II has the highest amino acid identity, whereas domain III shows the lowest amino acid identity among these three proteins. The low sequence identities between $M^{pro}S$ and $M^{pro}T$ and between $M^{pro}S$ and $M^{pro}H$ from the present study are in good agreement with the previous results [21], where SARS-CoV was classified as a new group of coronavirus based on the analysis of the deduced genome sequence.

Table 1
The amino acid sequence identities among $M^{pro}H$, $M^{pro}T$, and $M^{pro}S$

	Identity (%)			
	Total	Domain I	Domain II	Domain III
$M^{pro}H$ and $M^{pro}T$	60.80	63.44	65.06	55.45
$M^{pro}H$ and $M^{pro}S$	40.19	41.94	45.78	35.64
$M^{pro}T$ and $M^{pro}S$	43.85	44.09	49.40	39.22

3.2. The homology models of $M^{pro}ST$ and $M^{pro}SH$

The homology models of $M^{pro}ST$ and $M^{pro}SH$ are illustrated in Figs. 2a and b, respectively. Both $M^{pro}ST$ and $M^{pro}SH$ exhibit three distinct domains and adopt similar folds as $M^{pro}T$ and $M^{pro}H$, respectively. These models are in the similar order of magnitude comparing to the homology models constructed previously [10,13]. The quality of the geometry and of the stereochemistry of these homology models was further validated using Homology/ProStat/Struct_Check commend of Insight II. A total of 97% and 96% of the backbone dihedral angle (φ and ϕ) densities are located within the structurally favorable regions in Ramachandran plot for $M^{pro}ST$ and $M^{pro}SH$, respectively. The calculation of main chain torsion angles (χ_1 and χ_2) of these models showed no severe distorsion of the backbone geometry.

The putative substrate binding subsites S1 and S2 of $M^{pro}ST$ and $M^{pro}SH$ are located in a cleft between domains I and II, which are nearly identical to those of $M^{pro}T$ and $M^{pro}H$ (Fig. 2). It indicates that $M^{pro}S$ may follow the similar substrate binding mechanisms

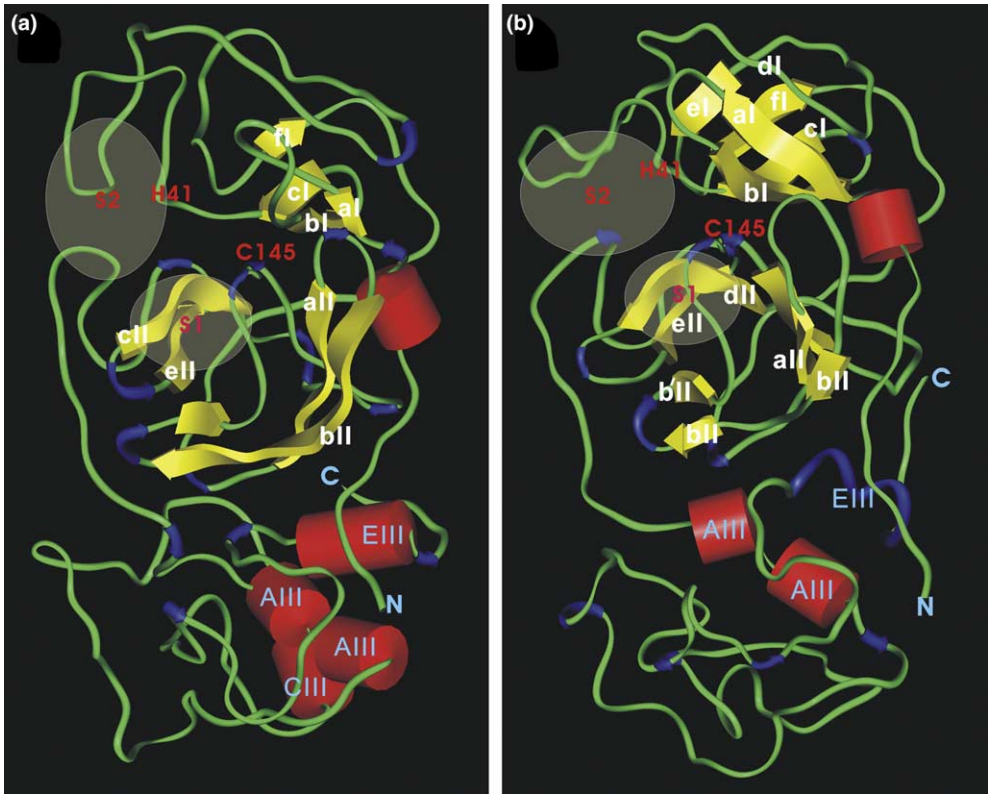

Fig. 2. The homology model of (a) $M^{pro}ST$ and (b) $M^{pro}SH$ visualized by Insight II. α-Helices and β-strands are shown in red cylinders and yellow arrows, respectively. The general acid–base catalyst His residue and the nucleophilic Cys residue are labeled. The locations of the putative substrate binding subsites S1 and S2 are indicated.

Table 2
The RMSDs between the template proteins, $M^{pro}H$ and $M^{pro}T$, and the homology models, $M^{pro}SH$ and $M^{pro}ST$

	RMSD (Å)		
	$M^{pro}H$	$M^{pro}T$	$M^{pro}SH$
$M^{pro}T$	2.01	–	–
$M^{pro}SH$	4.51	3.94	–
$M^{pro}ST$	4.84	4.37	5.78

of $M^{pro}T$ and $M^{pro}H$, allowing us to design anti-SARS drugs by simply screening the known proteinase inhibitors. The low sequence identity and secondary structure conservation in domain III among these proteins suggest that it may play a minor role in proteolytic activity. As shown in Table 2, the RMSDs of $M^{pro}SH$ and $M^{pro}ST$ are 4.84 and 3.94 Å, comparing to their corresponding templates, $M^{pro}H$ and $M^{pro}T$, respectively; while the RMSD between $M^{pro}SH$ and $M^{pro}ST$ is 5.78 Å. It indicates that the structure of $M^{pro}S$ is more similar to that of $M^{pro}T$.

3.3. Molecular dynamics simulations

As shown in Fig. 3, these structures remained considerably stable during the MD time course, with the root-mean-square deviations (RMSDs) remained within 3 Å. It is obvious that domain III exhibits higher structural variations than the other two domains in all cases. S1 was found to maintain its structural integrity, whereas S2 exhibits higher structural fluctuations during the entire MD simulations. It is attributed to that S2 is located on the open mouth of the catalytic cleft between domains I and II, whereas S1 is situated in the very bottom of this cleft and is well protected by the hydrophobic core. The higher structural variation of S2 makes it flexible enough to accommodate a bulky hydrophobic residue from the substrate.

In the crystal structures, the distance between the sulfur atom of Cys144 and the $N^{\varepsilon 2}$ of His41in $M^{pro}T$ is 4.05 Å [14], longer than the corresponding Cys–His distances in HAV $3C^{pro}$ (3.92 Å) [22], poliovirus (PV) $3C^{pro}$ (3.4 Å) [23], and papain (3.65 Å) [24]. From a dynamics point of view (Fig. 4), the Cys144–His41 distance of $M^{pro}H$ fluctuated more rapidly than that of $M^{pro}T$. In addition, the Cys145–His41 distances of $M^{pro}SH$ fluctuated more rapidly than that of $M^{pro}ST$ beyond 150 ps. These results indicate that both $M^{pro}T$ and $M^{pro}ST$ may exhibit more stable active site configurations than those of $M^{pro}S$ and $M^{pro}SH$. The large degree of fluctuation of these Cys–His distance may indicate that the structure of the catalytic site is not stable when it is

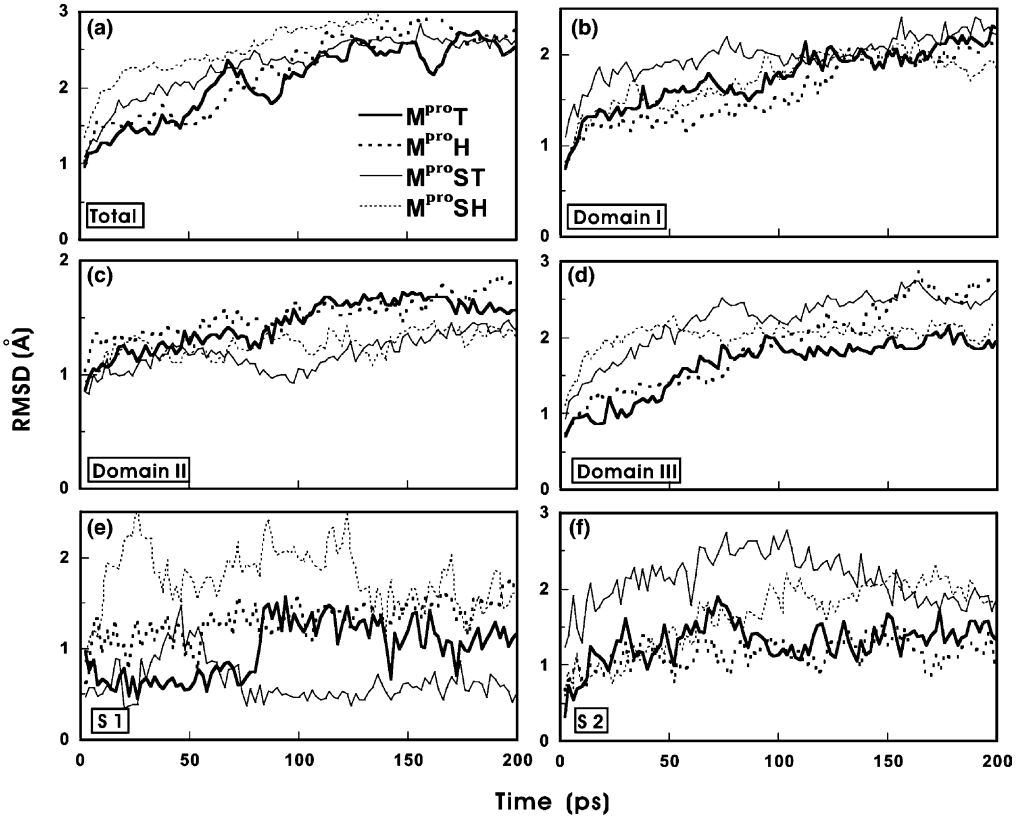

Fig. 3. The RMSDs of the backbone C_α for (a) the whole protein, (b) domain I, (c) domain II, (d) domain III, (e) S1, and (f) S2 of $M^{pro}T$, $M^{pro}H$, $M^{pro}ST$, and $M^{pro}SH$ during MD simulations.

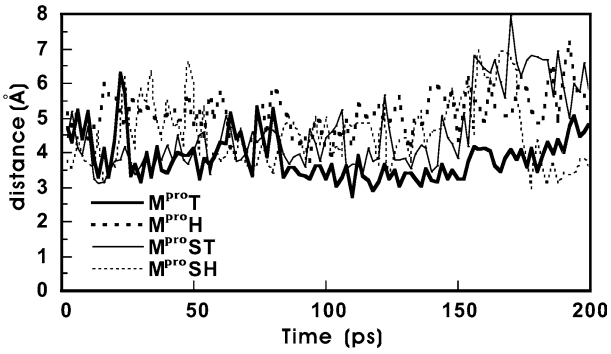

Fig. 4. The linear distance between the sulfur atom of the nucleophilic Cys residue and the N^2 of the general acid–base catalyst His residue as a function of MD simulation time.

not protected from substrate or ligand binding. This result is in very good agreement with the previous findings that there are significant differences in the flexibility in the active site of the SARS-CoV proteinase [25]. Furthermore, the high flexibility of the active site may allow these proteins to execute the catalytic process more efficiently.

It has been shown previously that, similarly to $3C^{pro}$ [23,24], specific substrate binding by M^{pro} is ensured by the well-defined S1 and S2 binding subsites [14]. In both $M^{pro}T$ and $M^{pro}H$, S2 is lined by the side chains of His41, Thr47, Ile51, Leu164, and Pro188, despite for the residue Leu164 in $M^{pro}T$ being replaced by Ile. In $M^{pro}S$, S2 is lined by the side chains of His41, Asp48, Pro52, Met165, and Gln189. It indicates that S2 is not as conserved as S1 among these proteins. It is worthy of mentioning that the main chain of Leu164 of $M^{pro}T$ (or Ile164 of $M^{pro}H$ or Met165 of $M^{pro}S$) forms part of S1, while its side chain is involved in S2, indicating that these two subsites are somewhat influenced by each other towards substrate binding.

The analysis of ASAs of both S1 and S2 during the MD simulations indicates that both subsites are flexible enough to accommodate the substrates. The snapshots of both S1 and S2 for these proteins with the smallest and largest accessible surface areas (ASAs) sampled from the 200 ps MD simulations were illustrated in Fig. 5. It is interesting that the sizes and conformations of the smallest and the largest S1 pocket of $M^{pro}SH$ are very similar to those of $M^{pro}T$. The variation of

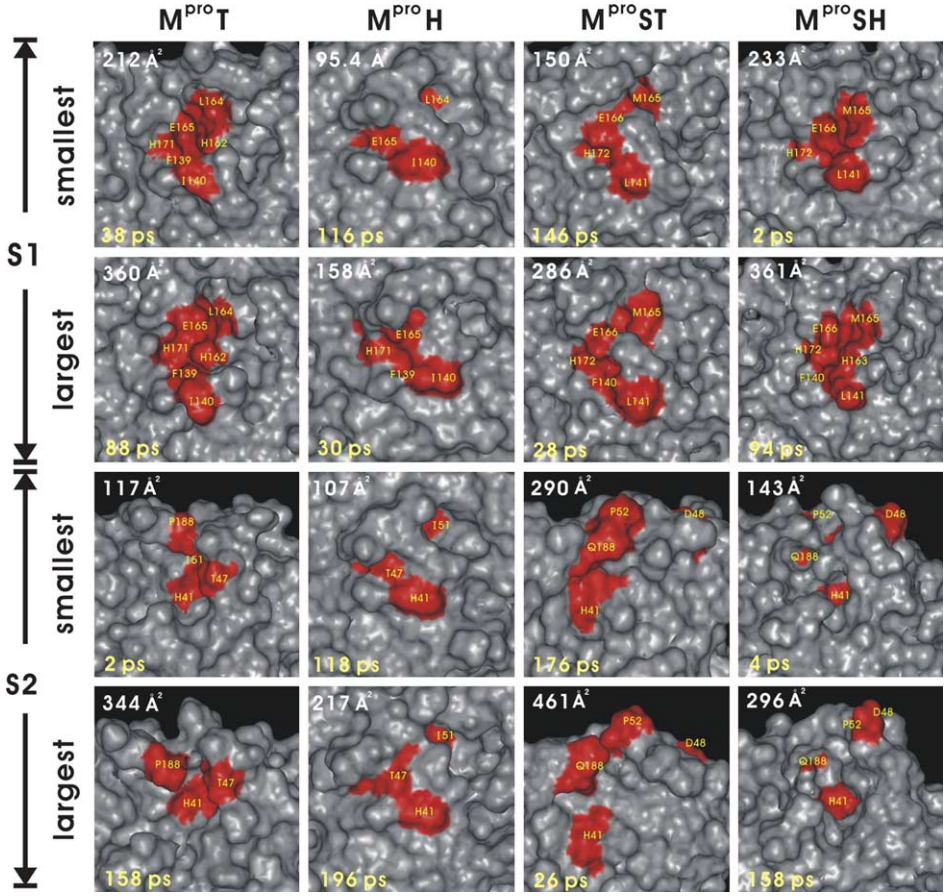

Fig. 5. Molecular surfaces of the substrate binding subsites S1 and S2 for $M^{pro}T$, $M^{pro}H$, $M^{pro}ST$, and $M^{pro}SH$ with the smallest and the largest ASAs during MD simulations. The residues forming these subsites are shown in red.

the size and conformation of S2 for these proteins is more significant than S1 during the MD simulations, probably because part of S2 is fully exposed to the solvent.

Acknowledgment

The authors gratefully acknowledge the financial support from the National Science Council of Taiwan (NSC-93-2214-E-027-001).

References

[1] D. Cavanagh, Arch. Virol. 142 (1997) 629.
[2] J.A. Den Boon, E.J. Snijder, E.D. Chirnside, A.A. de Vries, M.C. Horzinek, W.J. Spaan, J. Virol. 65 (1991) 2910.
[3] J. Herold, T. Raabe, B. Schelle-Prinz, S.G. Siddell, Virology 195 (1993) 680.
[4] V. Thiel, J. Herold, B. Schelle, S.G. Siddell, J. Virol. 75 (2001) 6676.
[5] J. Ziebuhr, J. Herold, S.G. Siddell, J. Virol. 69 (1995) 4331.
[6] J. Ziebuhr, G. Heusipp, S.G. Siddell, J. Virol. 71 (1997) 3992.
[7] A. Hegyi, A. Friebe, A.E. Gorbalenya, J. Ziebuhr, J. Gen. Virol. 83 (2002) 581.
[8] D.X. Liu, T.D. Brown, Virology 209 (1995) 420.
[9] A. Hegyi, J. Ziebuhr, J. Gen. Virol. 83 (2002) 595.
[10] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld, Science 300 (2003) 1763.
[11] P.A. Rota et al., Science 300 (2003) 1394.
[12] J.W. Ponder, F.M. Richards, J. Mol. Biol. 193 (1987) 775.
[13] B. Xiong et al., Acta Pharmacol. Sin. 24 (2003) 497.
[14] K. Anand, G.J. Palm, J.R. Mesters, S.G. Siddell, J. Ziebuhr, R. Hilgenfeld, EMBO J. 21 (2002) 3213.
[15] H.-L. Liu, W.-C. Wang, Chem. Phys. Lett. 366 (2002) 284.
[16] H.-L. Liu, Y. Ho, C.-M. Hsu, Chem. Phys. Lett. 372 (2003) 249.
[17] H.-L. Liu, C.-M. Hsu, Chem. Phys. Lett. 375 (2003) 119.
[18] H.-L. Liu, J.-C. Lin, Chem. Phys. Lett. 381 (2003) 592.
[19] M.-J. Hwang, X. Ni, M. Waldman, C.S. Ewig, A.T. Hagler, Biopolymers 45 (1998) 435.
[20] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, J. Comp. Phys. 81 (1984) 3684.
[21] M.A. Marra et al., Science 300 (2003) 1399.
[22] E.M. Bergmann, S.C. Mosimann, M.M. Chernaia, B.A. Malcolm, M.N. James, J. Virol. 72 (1997) 2436.
[23] S.C. Mosimann, M.M. Cherney, S. Sia, S. Plotch, M.N. James, J. Mol. Biol. 273 (1997) 1032.
[24] I.G. Kamphuis, K.H. Kalk, M.B. Swarte, J. Drenth, J. Mol. Biol. 179 (1984) 233.
[25] V.S. Lee et al., Sci. Asia 29 (2003) 181.