

Acknowledgements This work was supported by the Thai Graduate Institute of Science and Technology (TGIST) Funds (TG-22-11-845D) under the National Science and Technology Development Agency, Thailand. S.H. is grateful to the Thailand Research Fund (RTA5080005). The Postgraduate Education and Research Program in Petroleum Petrochemical Technology and Advanced Materials, the Commission on Higher Education (CHE), the Kasetsart University Research and Development Institute (KURDI), and the Graduate School Kasetsart University Scholarship are gratefully acknowledged for financial support. The authors would like to thank the Large-Scale Research Laboratory of the National Electronics and Computer Technology (NECTEC), the National Center for Genetic Engineering and Biotechnology (BIOTEC), LCAC, and the computing center of KU for computing research facilities. This work has been partially supported by the National Nanotechnology Center (NANOTEC), NSTDA, Ministry of Science and Technology, Thailand, through its Center of Excellence Network program. Medicines for Malaria Venture (MMV), Wellcome Trust, and UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) for Y.Y. and S.K. S.K. is an international research scholar of Howard Hughes Medical Institute (HHMI), USA. Finally, P.M. would like to thank Dr. Chak Sangma and Dr. Witcha Treesuwan for their AMBER supports.

References

- Prapunwattana P, Sirawaraporn W, Yuthavong Y, Santi DV (1996) *Mol Biochem Parasitol* 83:93
- Dasgupta T, Anderson KS (2008) *Biochemistry* 47(5):1336
- Yuthavong Y, Kamchonwongpaisan S, Leartsakulpanich U, Chitnumsub P (2006) *Future Microb* 1(1):113
- Yuthavong Y, Yuvaniyama J, Chitnumsub P, Vanichthanankul J, Chusaculttanachai S, Tarnchompoo B, Vilaivan T, Kamchonwongpaisan S (2005) *Parasitology* 130:249
- Yuvaniyama J, Chitnumsub P, Kamchonwongpaisan S, Vanichthanankul J, Sirawaraporn W, Taylor P, Walkinshaw MD, Yuthavong Y (2003) *Nat Struct Biol* 10:357
- Yuthavong Y (2002) *Microbes Infect* 4:175
- Nzila A (2006) *Drug Discov Today* 11(19–20):936
- Cody V, Schwalbe CH (2006) *Crystallog Rev* 12(4):301
- Martin S (2007) *ChemMedChem* 2(7):944
- Warhurst DC (2002) *Sci Prog* 85:89
- Hastings IM, Donnelly MJ (2005) *Drug Resist Updates* 8:43
- Nzila A (2006) *J Antimicrob Chemother* 57:1043
- Gregson A, Plowe CV (2005) *Pharmacol Rev* 57:117
- Uhlemann AC, Yuthavong Y, Fidock DA (2005) *Mol Approaches Malar*. ASM Press, Washington DC, pp 429–461
- Rastelli G, Sirawaraporn W, Sompornpisut P, Vilaivan T, Kamchonwongpaisan S, Quarrell R, Lowe G, Thebtaranonth Y, Yuthavong Y (2000) *Bioorg Med Chem* 8:1117
- Peterson DS, Milhous WK, Wellems TE (1990) *Proc Natl Acad Sci USA* 87:3018
- Sirawaraporn W, Yongkiettrakul S, Sirawaraporn R, Yuthavong Y, Santi DV (1997) *Exp Parasitol* 87:245
- Kamchonwongpaisan S, Quarrell R, Charoensetukul N, Ponsinet R, Vilaivan T, Vanichthanankul J, Tarnchompoo B, Sirawaraporn W, Lowe G, Yuthavong Y (2004) *J Med Chem* 47:673
- Maitarad P, Saparpakorn P, Hannongbua S, Kamchonwongpaisan S, Tarnchompoo B, Yuthavong Y (2008) Particular interaction between pyrimethamine derivatives and quadruple mutant type dihydrofolate reductase of *Plasmodium falciparum*: CoMFA and quantum chemical calculations studies. *J Enzym Inhib Med Chem*. doi:JEIMC07/1339
- Kubinyi H (1993) *QSAR: hansch analysis and related approaches*. VCH, Weinheim
- Hansch C, Leo A (1995) *Exploring QSAR. Fundamentals and applications in chemistry and biology*. American Chemical Society, Washington, DC
- Cramer RDIII, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959
- Kubinyi H, Flokers G, Martin YC (1998) *Perspect Drug Discov Des* 12–14:v–vii
- Cramer RDIII, Patterson DE, Bunce JD (1989) *Prog Clin Biol Res* 291:161
- SYBYL 7.0, Tripos Associates Inc., 1699 South Hanley Road, Suite 303, St Louis, Missouri 63144, USA
- Golbraikh A, Tropsha A (2002) *J Mol Graph Mod* 20:269
- Nilsson J, De Jong S, Smilde AK (1997) *J Chemom* 11:511
- Hannongbua S, Nivesanond K, Lawtrakul L, Pungpo P, Wolschann P (2001) *J Chem Inf Comput Sci* 41:848
- Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) *J Am Chem Soc* 115:9620
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Ciołowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) *Gaussian 03, revision B.05*. Gaussian, Inc., Pittsburgh
- AMBER parameter database. <http://www.pharmacy.manchester.ac.uk/bryce/amber/>
- Cummins PL, Ramnarayan K, Singh UC, Gready JE (1991) *J Am Chem Soc* 113(22):8247
- Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Matthews DH, Schafmeister C, Ross WS, Kollman PA (2006) *AMBER9.0*
- Boys SF, Bernardi F (1970) *Mol Phys* 19:553
- Saen-oon S, Kuno M, Hannongbua S (2005) *Proteins* 61:859
- Kuno M, Hongkrekrai R, Hannongbua S (2006) *Chem Phys Lett* 424:172
- Saen-oon S, Aruksakunwong O, Wittayanarakul K, Sompornpisut P, Hannongbua S (2007) *J Mol Graph Model* 26:720



Bridge water mediates nevirapine binding to wild type and Y181C HIV-1 reverse transcriptase—Evidence from molecular dynamics simulations and MM-PBSA calculations

Witcha Treesuwan^{a,b}, Supa Hannongbua^{a,b,*}

^a Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

^b Center of Nanotechnology, Kasetsart University, Bangkok 10900, Thailand

ARTICLE INFO

Article history:

Received 28 July 2008

Received in revised form 30 January 2009

Accepted 2 February 2009

Available online 9 March 2009

Keywords:

HIV-1 reverse transcriptase

Non-nucleoside reverse transcriptase inhibitors

Molecular dynamics simulations

Bridge water molecule

Binding energies

ABSTRACT

The important role of the bridge water molecule in the binding of HIV-1 reverse transcriptase (RT) inhibitor complex was elucidated by molecular dynamics (MD) simulations using an MM-PBSA approach. Binding free energies and thermodynamic property differences for nevirapine bound to wild type and Y181C HIV-1 reverse transcriptase were investigated, and the results were compared with available experimental data. MD simulations over 3 ns revealed that the bridge water formed three characteristic hydrogen bonds to nevirapine and two residues, His235 and Leu234, in the binding pocket. The energetic derived model, which was determined from the consecutive addition of a water molecule, confirmed that only the contribution from the bridge water was essential in the binding configuration. Including this bridge water in the MM-PBSA calculations reoriented the binding energies from −32.20 to −37.65 kcal/mol and −28.07 to −29.82 kcal/mol in the wild type and Y181C HIV-1 RT, respectively. From the attractive interactions via the bridge water, His235 and Leu234 became major contributions. We found that the bridge water is the key in stabilizing the bound complex; however, in the Y181C RT complex this bridge water showed weaker hydrogen bond formation, lack of attractive force to nevirapine and lack of binding efficiency, leading to the failure of nevirapine against the Y181C HIV-1 RT. Moreover, the dynamics of Val179, Tyr181Cys, Gly190 and Leu234 in the binding pocket showed additional attractive energetic contributions in helping nevirapine binding. These findings that the presence of a water molecule in the hydrophobic binding site plays an important role are a step towards a quantitative understanding of the character of bridge water in enzyme-inhibitor binding. This can be helpful in developing designs for novel non-nucleoside HIV-1 RT inhibitors active against the mutant enzyme.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Reverse transcriptase (RT) is a key enzyme in the replication cycle of the human immunodeficiency virus type 1 (HIV-1), catalyzing the conversion of virally encoded RNA into proviral DNA [1]. This essential step in the retroviral life cycle is targeted by a variety of drugs in clinical use to combat AIDS (acquired immune deficiency syndrome). RT is a heterodimeric enzyme with subunits of 66 and 51 kDa. The p66 subunit consists of fingers, palm and thumb subdomains (named for their resemblance to a right hand), as well as connection and ribonuclease H (RNase H) subdomains.

There are two main classes of RT inhibitors [2]. The first class consists of nucleoside analogues (nucleoside reverse transcriptase

inhibitors, or NRTIs) such as AZT (3'-azido-3'-deoxythymidine), ddI (dideoxymosine) and ddC (dideoxycytidine). These are competitive inhibitors of the nucleotide substrate, and bind to the polymerase active site upon metabolic activation. After incorporation in the DNA strand instead of dNTP they cause premature termination of the newly synthesized chain. In addition, NRTIs also act on other host DNA polymerases, which explains their toxicity [3]. The second class consists of non-nucleoside inhibitors (NNRTIs) such as HEPT (1-[(2-hydroxyethoxy)-methyl]-6-(phenylthio)thymine) [4], TIBO (tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepin-2(1H)-one) [5], nevirapine (dipyridodiazepinones) [6], and efavirenz ((-)-6-chloro-4-cyclopropyl ethynyl-4-trifluoromethyl-1,4-dihydro-2H-3,1-benzoxazin-2-one) [7]. These inhibitors are highly specific for HIV-1 RT and lock it into an inactive conformation by fitting into an allosteric site approximately 10 Å from the polymerase active site, causing a displacement of the catalytic aspartate residues. Furthermore, they show lower cellular toxicity than NRTIs.

* Corresponding author at: Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand. Fax: +66 2 562 5555x2140.

E-mail address: fscisph@ku.ac.th (S. Hannongbua).

Reported crystal structures of RT include those of unliganded RT [8], RT complexed to dsDNA [9], RT bound to several NNRTIs [10,11], and RT complexed to dsDNA and a deoxynucleoside triphosphate (RT/dsDNA/dNTP) [12]. The non-nucleoside binding pocket only exists in the structures of RT complexed to a NNRTI, with its formation probably being induced by the proximity of the inhibitor. In unliganded RT, the p66 thumb subdomain is folded into the DNA-binding cleft and lies over the palm subdomain, nearly touching the fingers subdomain in a “thumb down” configuration. As a consequence, the DNA-binding cleft is closed. Moreover, the conformational change effected by NNRTI binding reduces the catalytic efficiency of the enzyme [13].

The rapidity of the selection of drug-resistant HIV in patients was such that single-point mutations in the virus made first-generation NNRTIs such as nevirapine unusable in monotherapy [14]. Among the mutations in RT that were originally described for nevirapine resistance were those at Tyr181 and Tyr188, both of which gave rise to high-level resistance [15]. The mutation of tyrosine at position 181 has frequently occurred, not only when nevirapine was treated but also in many other NNRTIs, and the change is almost always to cysteine. In the case of the mutation of tyrosine at position 188, a variety of mutation was reported. Nevirapine and HEPT select the Tyr188Cys mutation, whereas TIBO or α -APA result in Tyr188His or Tyr188Leu mutations [16]. Experimental results show that the non-nucleoside inhibitors lose their inhibitory efficiency by 20- to 1,000-fold when the mutation occurs in the HIV-1 RT binding pocket [17].

Computational modeling studies on HIV-1 RT and nevirapine have been performed for a few decades to gain more understanding at the molecular level. The conformational analysis of nevirapine by quantum calculation revealed that the cyclopropyl group is rotatable, but the minima appeared at 218° [18]. The basic postulate Gaussian network model (GNM) revealed that the p66 thumb's mobility was extremely sensitive when HIV-1 RT was bound to nevirapine. Collective motions analysis showed the key residues, Leu100, Trp229 and Leu234, covered nevirapine inside the binding pocket [19]. High-level quantum mechanical (QM) theory, implemented through the ONIOM approach, revealed weak π – π and H– π attractive interaction between Tyr181 and nevirapine [20,21]. Although the QM studies provide accurate interaction data according to theory, the studies based on rigid structures still lack information about the dynamics of binding in the RT-nevirapine bound complex. Thus, the simulations approach has also been applied to investigate the properties, interactions and binding energy of the reverse transcriptase complex to non-nucleoside inhibitors, in order to handle larger systems than can be accommodated by QM techniques [22]. The conformational changes of the RT pocket have been investigated through molecular dynamics (MD) simulations. Specific demonstration of the dihedral angle rotation of N-CA-CB-CG in Y181 showed the induced effect from mutation [23]. However, the interrelationship of nevirapine and the Y181C mutation also needs some proof from energetic prediction to be linked with the experimental affinity. The combined Monte Carlo (MC) simulations with free energy perturbation (FEP) revealed a fold resistance energy of 3.88 ± 0.3 kcal/mol from nevirapine relatively to efavirenz [15], and the effectiveness of nevirapine against the Y181C mutated form of the enzyme versus the wild type [24]. Ways to calculate the absolute binding free energy between HIV-1 RT and nevirapine have been developing since first being introduced with the QSAR correlation [25]. One of the most useful methods is the Molecular Mechanics/Poisson–Boltzmann Surface Area (MM-PBSA) method [26] because it is based on a compromise between speed and accuracy in the calculations. The combined MD/MM-PBSA method could explain the lost activity of nevirapine when the Y181C mutation occurs from the change in free energy by -5.94 kcal/mol [27]. Currently,

numerous computational studies have shown quantitative agreement between calculated and experimental binding affinities in the HIV-1 RT system. However, the role of solvent molecules in the binding mechanism between the hydrophobic pocket and nevirapine is still of interest for further study.

Therefore, the main objective of this work was to study solvent molecules inside the hydrophobic binding pocket and introduce the key step in the MM-PBSA calculations when it was essential to include explicitly the solvent. The roles of the water molecule in HIV-1 RT/nevirapine complex have been reported previously by Rizzo and co-worker [28]. In this study we investigated a different water molecule, the bridge water molecule (WAT1067), which is located around the oxygen of nevirapine. A combination of molecular dynamics simulations and MM-PBSA calculations (MD/MM-PBSA) was applied to study the wild type and the Y181C HIV-1 RT/nevirapine complexes in order to obtain the theoretical binding energy, and observe the dynamics of the RT pocket. The motion and role of water inside the binding pocket were also investigated. The results will be useful in determining the molecular level of HIV-1 RT/nevirapine interaction in solution, and as guidance for other NNRTIs which have similar binding to nevirapine. This basic information can be used in the development of higher potency NNRTIs against mutant enzymes.

2. Computational methods

2.1. Molecular dynamics simulations

Model structures of the wild type and the Y181C HIV-1 RT complexes with nevirapine were constructed based on the crystallographic code 1VRT [29]. The enzymes were mutated at position 181 from tyrosine to cysteine by the SPdbV3.7 program [49]. After all missing residues were added, the mutant model was then minimized using 5,000 steps of steepest descent, and then switched to the conjugate gradient algorithm in the Sander module of the AMBER program package in order to remove bad steric interactions. Nevirapine was firstly optimized at the B3LYP/6-31G(d,p) level. Then the electrostatic potential that surrounds nevirapine was calculated by a single-point calculation at the HF/6-31G(d) level in GAUSSIAN98 [30]. The electrostatic potential was fitted into the partial atomic charge of nevirapine by the RESP [31,32] charge method. Preparation of the force field parameters of nevirapine was done by using the Antechamber module [33] of AMBER.

Molecular dynamics simulations were performed using an AMBER7 [34] program with an AMBER 1999 force field [35]. Each complex system was immersed in an octahedral box of 10 Å from the solute surface using TIP3P water [36]. A total of 42,170 solvent molecules were generated followed by 7 Cl[−] ions for neutralization, resulting in a total of 143,900 atoms in the system. Periodic boundary conditions and constant temperature and pressure were used. The non-bonded cutoff distance was set to 11.5 Å. The integration time step was 2 fs, with SHAKE [37] applied to constrain the bonds involving hydrogen atoms. The restraint force on the whole complex was slowly decreased from 4.0 kcal/mol to zero during the equilibration, and simulations were run for 1 ns at 300 K. Coordinates were saved every 1 ps. After the entire HIV-1 RT/nevirapine system had been simulated for 1 ns, the low energy structure at the equilibrium was selected as the next starting structure. In order to speed up the calculation, the structure of the enzyme complexes was reduced to a smaller model in which protein residues and water molecules outside 30 Å of the mass center of nevirapine were removed. Continuation of the simulations was performed by using the AMBER9 [38] program with the Duan et al. (2003) force field [39,40] which provided a better force field for proteins. The model was set up as follows: all water

molecules inside the allosteric binding site were kept in their positions, two magnesium ions were added to the active site region, and a 10 Å spherical water cap centered on nevirapine was added near the binding site [28,41]. The simulations were carried out for 3 ns at 300 K under constant volume periodic boundary conditions. Simulations were performed on a 2.4 GHz, 514 Mbyte system running on a Linux7.3 PC.

2.2. MM-PBSA analysis

All water molecules and ions were removed before the continuum solvent model was applied to calculate the MM-PBSA energies. The electrostatic contribution to the solvation free energy was calculated by the Poisson–Boltzmann method (PB). The hydrophobic contribution to the solvation free energy was determined using a solvent-accessible surface area (SA) dependent nonpolar solvation term. The description of the process to obtain the MM-PBSA energies derived from the thermodynamic cycle can be found in Wang et al. [33]. In this work, snapshots of 1.5–3 ns were sampled. In total, 500 snapshot structures were used in the MM-PBSA calculations and analysis. To investigate the effect of the explicit solvent water molecule, we included the closest water – one, two, and three – molecules to extend the MM-PBSA calculations.

2.3. Theoretical background

The experimental binding affinities of nevirapine with HIV-1 RT were obtained from enzymatic kinetic studies, which were represented variously by: K_d (the equilibrium dissociation constants); IC_{50} (50% inhibitory concentration of nevirapine against HIV-1 RT); or EC_{50} (50% effective concentration, or concentration required to protect cells against the HIV cytopathogenicity by 50%). These experimental data can be converted to the binding free energy between nevirapine and HIV-1 RT by using the thermodynamic equation ($\Delta G \sim RT \ln(\text{activity})$ in kcal/mol) [28].

To find the binding energies from the theoretical calculation, two strategies in the MM-PBSA calculations involving the explicit water molecule were derived. First, the MM-PBSA energies of individual solutes, including the explicit water molecule, were calculated. Then each individual energy term, which included the energy of the explicit water molecule, was subtracted from the free energy of the complex. The binding free energies ($\Delta G_{\text{binding}}$) were obtained as shown in Eq. (1):

$$\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - (\Delta G_{\text{enzyme}} + \Delta G_{\text{Nevirapine}} + \Delta G_{\text{WAT}}) \quad (1)$$

Secondly, the appearance of the explicit water molecule could be considered as a part of the receptor. This alternative approach kept the concept that the binding energies were calculated from two parts instead of the three components. The effect from the numbers of explicit water molecules were simultaneously embedded inside the enzyme. Thus, the binding free energies ($\Delta G_{\text{binding}}$) could be obtained from Eq. (2), which is the standard MM-PBSA approach for including selected solvent molecules:

$$\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - (\Delta G_{\text{enzyme-WAT}} + \Delta G_{\text{Nevirapine}}) \quad (2)$$

Entropy of the system was calculated through the normal mode analysis for a temperature of 298.15 K and 1.0 atm pressure. The total entropies for translation, rotation and vibration were obtained and converted to energy (kcal/mol) or heat capacity (cal/mol K). The $\Delta S_{\text{binding}}$ was obtained from the $\Delta S_{\text{complex}}$ in the same fashion as in Eq. (1) or Eq. (2).

An alternative way to compare the calculated result with the experimental data is to carry out relative measurements between the wild type and the mutant type enzymes. The

Table 1

Experimental binding affinities (μM) and energies (kcal/mol) from the wild type and Y181C HIV-1 RT complexes with nevirapine.

	Wild type HIV-1 RT/nevirapine		Y181C HIV-1 RT/nevirapine	
	Affinities (μM)	Energy (kcal/mol)	Affinities (μM)	Energy (kcal/mol)
1. K_d [45]	0.025 ± 0.010	−10.37	11.700 ± 4.30	−6.73
2. K_d [45]	0.019 ± 0.004	−10.53	2.500 ± 1.30	−7.64
3. IC_{50} [46]	0.060	−9.85	3.200	−7.50
4. IC_{50} [47]	0.032	−10.22	10.000	−6.82
5. EC_{50} [48]	0.044 ± 0.010	−10.04	3.040 ± 1.42	−7.53
6. IC_{50} [28]	0.084	−9.65	(Not reported)	

$\Delta G = RT \ln[\text{Activity}]$ in kcal/mol, where ΔG represents the binding free energy (kcal/mol) with activities expressed in units of M (molar), R represents the gas constant (1.988 cal/mol K), T represents the temperature (K) and [Activity] can be K_d , IC_{50} or EC_{50} .

relative activity of nevirapine against the wild type and the Y181C HIV-1 RT from experiment can be measured in terms of the fold resistance and the relative fold resistance energy ($\Delta \Delta G_{\text{fold resistance}} = \Delta G_{\text{mutant}} - \Delta G_{\text{wild type}}$) [42].

3. Results and discussion

3.1. Characteristic bridge water and MM-PBSA binding energies

The inhibitory activity of IC_{50} , representing affinities of the wild type and Y181C HIV-1 RT complexes with nevirapine, was investigated through kinetic studies. Since the IC_{50} was converted to the experimental binding energies, as reported in the literature, a list of nevirapine activities against wild type and Y181C HIV-1 RT is shown in Table 1. The major mutation at position 181 from tyrosine to cysteine clearly caused nevirapine to lose its efficiency, as indicated by the higher concentration of nevirapine usage. Also, variations of binding affinities depended on the bioactivity assay of each laboratory. The conversion of binding affinities into energies is the link from the experimental to the computational approach. Experimental binding energies are also used as a key to compare to the MM-PBSA calculations for each model in this work.

The MM-PBSA calculations were acquired by sampling 500 snapshots from the trajectory. In classical MM-PBSA calculations all solvent and ions are normally removed. Therefore, the continuum solvent model was used instead; but this failed to reproduce the experimental binding energies in the HIV-1 RT/nevirapine system, as the binding energies of −32.20 and −28.07 kcal/mol found in wild type and Y181C RT complexes, respectively. Among the various complex structures of HIV-1 RT/nevirapine available, the structure code 1VRT is the highest resolution from X-ray crystallographic structure. Plenty of crystal water is found in 1VRT, including bridge water which is only present in this complex structure. Since crystallographic water molecules in the complex code 1VRT.pdb were considered as an important part, it had been integrated in the binding pocket throughout the simulations. It was found that WAT1067, located close to oxygen atom of nevirapine.

The characteristics of WAT1067 were analyzed from the trajectory. We found that WAT1067 forms a moderate/weak hydrogen-bonded bridge between nevirapine and the binding pocket of both wild type and Y181C HIV-1 RT, as shown in Fig. 1. WAT1067 acts as a hydrogen donor to O1(nevirapine), O(Leu234) and O(His235). In addition, this water molecule is also a hydrogen acceptor from N-H(His235). The heteroatomic distances between O(WAT1067) and O1(nevirapine), O(Leu234), O(His235) and N(His235) were measured within the last 2 ns of the trajectories, as shown in Fig. 2. Average distances of 2.76, 3.60, 3.38 and 3.55 Å were found in the wild type complex, while average distances of

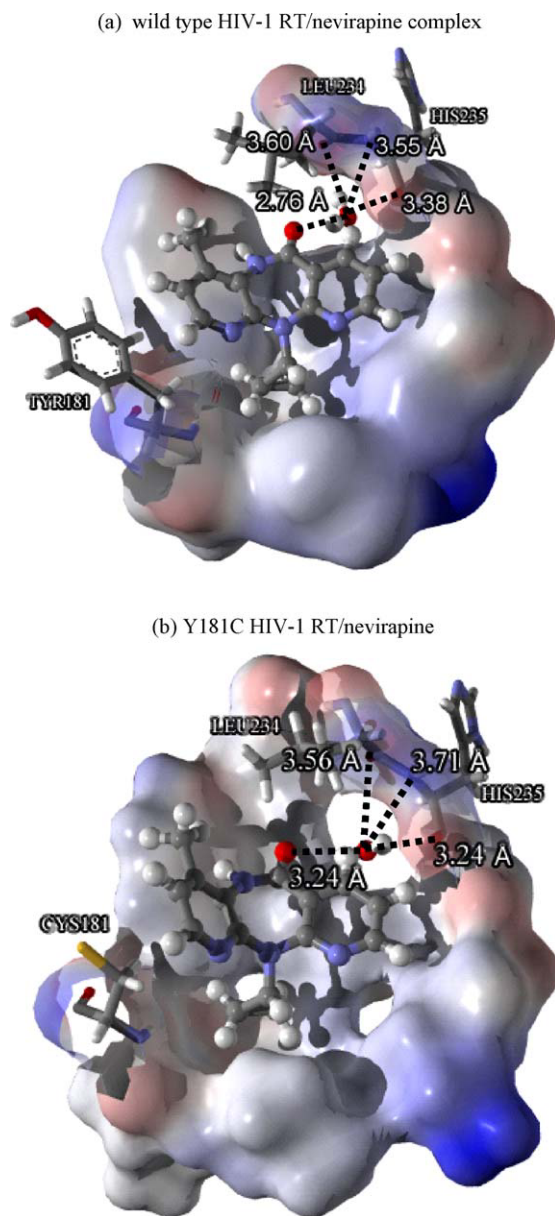


Fig. 1. Hydrogen bond network between the bridge water molecule, WAT1067, and residues in the binding pocket of HIV-1 RT present in the last snapshot of the simulations. Heteroatomic distances were measured in (a) wild type HIV-1 RT/nevirapine and (b) Y181C HIV-1 RT/nevirapine.

3.24, 3.56, 3.24 and 3.71 Å were found in the Y181C mutant type complex, corresponding to the distances between O(WAT1067) and O1(nevirapine), O(Leu234), O(His235) and N(His235), respectively. This suggests that the hydrogen bond formation between WAT1067 and nevirapine in the mutant RT complex was weaker than in the wild type RT. The space inside the binding pocket changed due to the Y181C mutation, allowing nevirapine and WAT1067 to move with more freedom. However, WAT1067 found new equilibrium and stability inside the pocket of Y181C RT/nevirapine, as shown in Fig. 3. A similar hydrogen bonded bridge position was also presented in the Monte Carlo simulations of the HIV-1 RT complex with nevirapine or MKC-442 [28]. Strong hydrogen bond interactions from the bridge water were also observed in another system, HIV-1 protease, as derived from quantum mechanical calculations [43]. Not only the bridge water, WAT1067, but three other water molecules – WAT1034, WAT1066 and WAT1183 – stay in contact with nevirapine at the beginning.

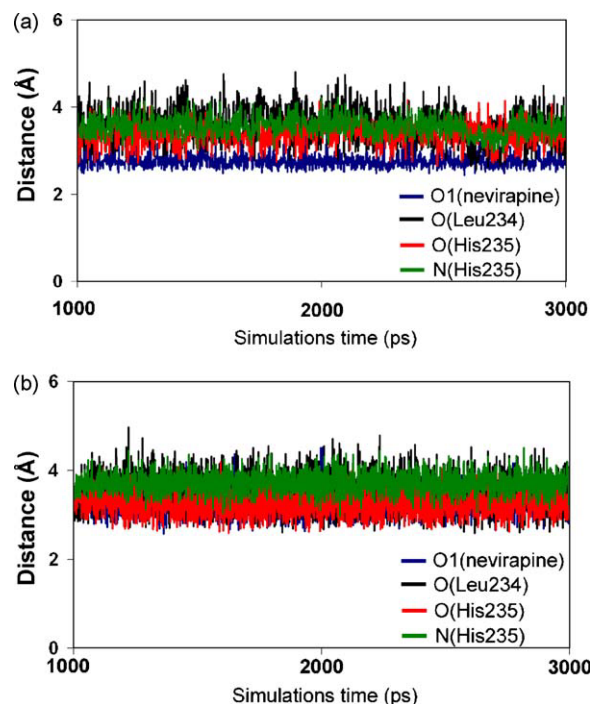


Fig. 2. Heteroatomic distances (Å) between WAT1067 and nevirapine, WAT1067 and selected residues, in the HIV-1 RT binding pocket were measured within the last 2 ns of the simulations. The distance between O(WAT1067) and O1(nevirapine) is represented in blue, O(WAT1067) and O(Leu234) in black, O(WAT1067) and O(His235) in red and O(WAT1067) and N(His235) in green. (a) Wild type HIV-1 RT/nevirapine and (b) Y181C HIV-1 RT/nevirapine.

During the whole RT complex simulations, only WAT1067 remained in both wild type and mutant type HIV-1 RT/nevirapine complexes. In the pocket of the wild type complex, WAT1034, WAT1066 and WAT1183 were absent. This is probably due to the hydrophobic properties of the RT pocket. On the other hand, the exchange of water found in the pocket of the Y181C RT complex when new water molecules, WAT1032, WAT1040 and WAT1057, came to substitute the former water position (see Figure 1 in supplementary data). Among the crystallographic structures of HIV-1 RT/nevirapine, there are eight structures that contain crystallographic water. However, bridge water is present only in 1VRT. The simulations suggest that water is able to move from place to place at all times. However, when comparing the water in the snapshot simulation structure with crystal structures (1JLF [50], 2HND [51], 2HNY [51], 1LWC [52], 1LWE [52], 1LWF [52], 1S1X [53] and 1VRT), the simulation waters showed good correspondence to the X-ray structure by locating in the solvent areas (see Fig. 4).

Due to the characteristics of the bridge water, WAT1067, as described above, the inclusion of only this water molecule changed the binding energies dramatically, as shown in Table 2. Shifts from –32.20 to –39.14 and –28.07 to –32.52 kcal/mol were found in the wild type and Y181C RT complexes, respectively. The major change of binding energy – from –32.20 to –39.14 kcal/mol in wild type HIV-1 RT/nevirapine, and from –28.07 to –32.53 kcal/mol in Y181C HIV-1 RT/nevirapine when WAT1067 was included in the MM-PBSA calculations – was evaluated by individual energy compositions. The main contribution comes from van der Waals interaction, whereas the bridge water significantly improves the electrostatic energy according to the binding process. This indicated that it was not only nevirapine and the RT residues but also one water molecule incorporated into the bound configuration that stabilized the complex (see Table 1 in supplementary data).

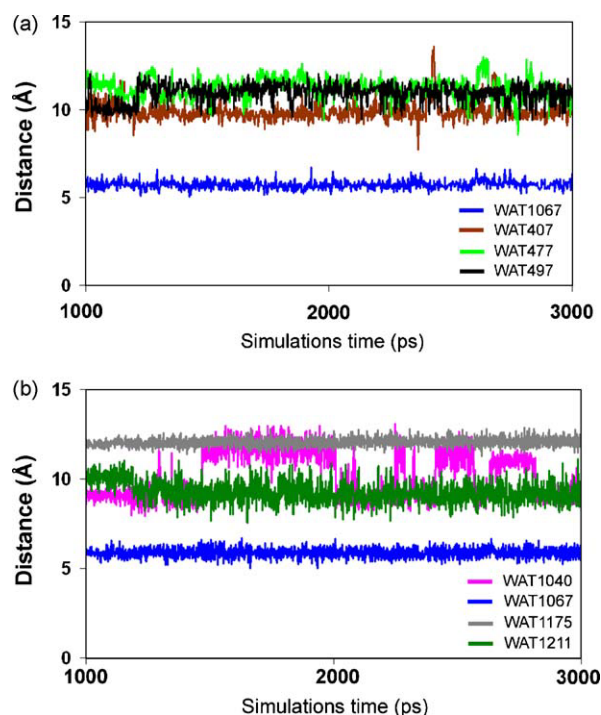


Fig. 3. Distances of the four closest water molecules to nevirapine measured by the center of mass (Å). (a) Wild type HIV-1 RT/nevirapine and (b) Y181C HIV-1 RT/nevirapine. Only WAT1067 presents in both the wild type and Y181C HIV-1 RT complexes (blue line).

While the appearance of WAT1067 proved to be a significant stabilizing force in the bonding between nevirapine and HIV-1 RT, especially in the mutant type complex, other water molecules were included to investigate their contribution as well. Each water molecule was added with respect to its distance to nevirapine. The distances, based on the center of mass from each water molecule to nevirapine in both the wild type and mutant type complexes, were investigated along with the simulations time. We found four water molecules closest to nevirapine: WAT1067, WAT407, WAT477 and WAT497 in the wild type RT/nevirapine; and WAT1040, WAT1067,

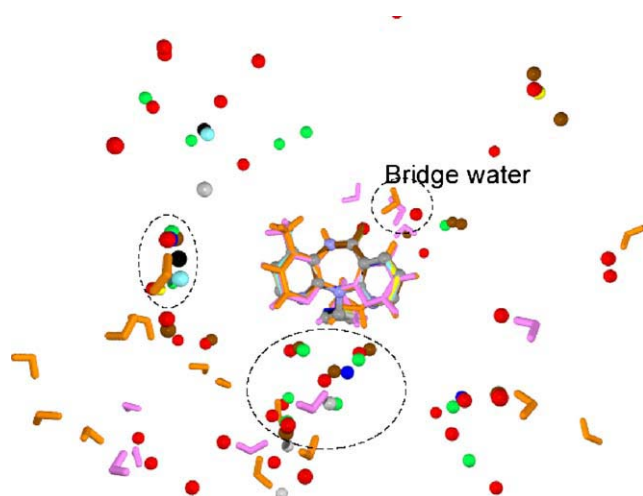


Fig. 4. Water accessible areas obtained by comparison of water position that surrounding nevirapine, while nevirapine in simulations (wild type HIV-1 RT (stick, pink) and Y181C HIV-1 RT (stick, orange)) and other crystal structures (2HND (brown), 2HNY (green), 1JLF (gray), 1LWC (blue), 1LWE (yellow), 1LWF (black) and 1S1X (light blue)) were superimposition referenced to nevirapine origin in structure of 1VRT (red).

Table 2

The MM-PBSA binding energies ($\Delta G^{\text{MM-PBSA}}$) of the wild type and Y181C HIV-1 RT/nevirapine complexes (kcal/mol).

	$\Delta G^{\text{MM-PBSA}}$ (kcal/mol)
1. Wild type HIV-1 RT + nevirapine	−32.20 (1.42)
2. Y181C HIV-1 RT + nevirapine	−28.07 (2.11)
3. Wild type HIV-1 RT + nevirapine + WAT1067	−39.14 (1.45)
4. Y181C HIV-1 RT + nevirapine + WAT1067	−32.52 (2.13)

Results obtained from classical MM-PBSA equation (i.e. model 1 and 2) and Eq. (2) (i.e. model 3 and 4).

Values in parenthesis show standard error in the calculations.

Table 3

Average distances of four water molecules closest to nevirapine measured by the center of mass from the wild type and Y181C HIV-1 RT complexes.

Wild type HIV-1 RT/nevirapine		Y181C HIV-1 RT/nevirapine	
WAT ID.	Average distance (Å)	WAT ID.	Average distance (Å)
1. WAT1067	5.7	1. WAT1067	5.9
2. WAT407	9.9	2. WAT1211	9.1
3. WAT497	11.0	3. WAT1040	10.5
4. WAT477	11.2	4. WAT1175	12.1

WAT1175 and WAT1211 in the Y181C complex. Water identification numbers higher than 1001 represented the crystal water from the starting crystallographic structure, while the others (WAT407, WAT477 and WAT497) were solvated water that had been generated by the leap subprogram in AMBER during the setup process. Average distances of these four water molecules to nevirapine, calculated over the last 2 ns trajectories, are listed in Table 3. The distances of each water molecule from nevirapine were ranked from the closest to the farthest. Water molecules in the wild type HIV-1 RT complex were ranked as WAT1067, WAT407, WAT497 and WAT477 with distances of 5.7, 9.9, 11.0 and 11.2 Å, respectively. On the other hand, water molecules in the Y181C HIV-1 RT complex were ranked as WAT1067, WAT1211, WAT1040 and WAT1175 with distances of 5.9, 9.1, 10.5 and 12.1 Å, respectively.

Distance fluctuation analysis of these four water molecules including the bridge water, as shown in Fig. 3, revealed the stabilities and deviations. Although all water molecules showed fluctuation, WAT1067 was the most stable throughout the simulations. Moreover, only WAT1067 was found to occupy the pocket without exchanges with other solvent molecules in both complexes (see Figure 1 in supplementary data). Even though the binding pocket was hydrophobic, one bridge water molecule was needed in the bound conformation. The thicker blue line in Fig. 3(b) shows that WAT1067 in the binding pocket of Y181C HIV-1 RT had a bit more freedom of space than in the pocket of wild type HIV-1 RT. Although this is a very minimal effect, it is a direct indication of the strength of hydrogen bond formation from this specific WAT1067. This stability of WAT1067 also corresponds to the previous hydrogen bonding distance analysis of the bridge water molecule. It was clear that the mutation at position 181 from tyrosine to cysteine affected the stability of the bridge water molecule. This is a new issue which explains the activity loss of nevirapine when the Y181C mutation occurs.

All four water molecules were considered in the MM-PBSA calculations, based on average distance analysis. However, a practical way to calculate the binding energies by including more than one water molecule could be done in two alternative ways. The calculated binding energies could be obtained either from the water molecule alone (following Eq. (1)) or kept with the enzyme receptor (following Eq. (2)). Energy calculations from each component based on the individual energy approach following Eq. (1) are listed in Table 4. In both cases of wild type and mutant

Table 4
MM-PBSA energies (kcal/mol) of the wild type and Y181C HIV-1 RT complexes with nevirapine calculated by including different numbers of solvent molecules based on individual energy.

Models	$\Delta G_{\text{Complex}}$	ΔG_{Enzyme}	ΔG_{Nev}	ΔG_{WAT}	$\Delta G^{\text{MM-PBSA}}$	Residual
RT ^{WT} . + Nev. + 1WAT	−11467.39 (0.69)	−11283.98 (0.69)	−137.22 (0.04)	−7.05 (0.03)	−39.14 (1.45)	–
RT ^{WT} . + Nev. + 2WAT	−11479.19 (0.69)	−11283.98 (0.69)	−137.22 (0.04)	−15.20 (0.02)	−42.79 (1.44)	−3.65
RT ^{WT} . + Nev. + 3WAT	−11490.39 (0.72)	−11283.98 (0.69)	−137.22 (0.04)	−23.25 (0.02)	−45.94 (1.47)	−3.15
RT ^{Y181C} . + Nev. + 1WAT	−11422.36 (1.04)	−11244.10 (1.02)	−138.69 (0.04)	−7.05 (0.02)	−32.52 (2.13)	–
RT ^{Y181C} . + Nev. + 2WAT	−11440.62 (1.03)	−11244.10 (1.02)	−138.69 (0.04)	−15.13 (0.02)	−42.70 (2.11)	−10.18
RT ^{Y181C} . + Nev. + 3WAT	−11454.34 (1.09)	−11244.10 (1.02)	−138.69 (0.04)	−23.09 (0.02)	−48.46 (2.17)	−5.76

RT^{WT}. = wild type HIV-1 RT.

RT^{Y181C}. = Y181C HIV-1 RT.

Nev. = nevirapine.

RT^{WT} 1WAT = WAT1067.

RT^{WT} 2WAT = WAT1067 + WAT407.

RT^{WT} 3WAT = WAT1067 + WAT407 + WAT497.

RT^{Y181C} 1WAT = WAT1067.

RT^{Y181C} 2WAT = WAT1067 + WAT1211.

RT^{Y181C} 3WAT = WAT1067 + WAT1211 + WAT1040.

Values in parenthesis show standard error in the calculations.

type complexes, the obtained binding energies became lower after each addition of a water molecule. The difference of binding energies after each addition was reported as the residual number to evaluate the methodology. In the wild type RT complex, residuals of −3.65 and −3.15 kcal/mol were obtained after addition of the second and third water molecules, respectively. In contrast, residuals of −10.18 and −5.76 kcal/mol were obtained, respectively, in the Y181C RT complex. The trend of slightly different binding energies in the wild type complex and fluctuating differences in the mutant complex suggested that the individual energy component approach used in the MM-PBSA calculations led to unstable binding energies. However, the binding energies from another calculation approach – by keeping the water molecule as part of the receptor – produced more reliable results, as shown in Table 5. Only the first addition – the important bridge water molecule – lowered the binding energy by a huge amount, because of its significant contribution to the electrostatic attraction (see Table 1 in supplementary data). Residuals of −0.06 and 0.00 kcal/mol were obtained after addition of the second and third water molecules, respectively, in the wild type complex. Likewise, residuals of −0.15 and −0.05 kcal/mol were obtained in the Y181C RT complex. The trend of small change in residuals – close to zero or not significantly different – suggested that the embedded water molecules provided good stability of the binding energies. Thus the appropriate methodology of including a water molecule into the receptor part would be used for the later binding energy

calculations. All these results strongly confirmed that only the bridge water molecule was essential in the binding mechanism between nevirapine and HIV-1 RT.

The most appropriate MM-PBSA binding energies from the wild type and Y181C HIV-1 RT complexes with nevirapine were −37.65 and −29.82 kcal/mol, respectively. Hence the improvement of estimating the binding free energy can be done by inclusion of a bridge water molecule in the binding site. This results in a closer match to the experimental data, as compared to the results of Zhou and co-worker (Predict I) [27] who found −54.78 and −48.48 kcal/mol for the wild type and Y181C RT/nevirapine, respectively, without including the bridge water molecule.

3.2. Fold resistance energies relative to the wild type HIV-1 RT

To determine the efficiency of nevirapine when the mutation occurred, the fold resistance value was calculated. The experimental data suggested that the fold resistance of nevirapine against Y181C HIV-1 RT could vary from about 50- to 500-fold, which was a rather wide range. The comparative results between experiment and modeling enabled the experimental fold resistance to be converted into the relative fold resistance energy, as shown in Table 6. The experimental relative fold resistance energies came from different laboratories, so the values varied from 2.35 to 3.64 kcal/mol. However, the previous computational prediction was 6.30 kcal/mol [27]. A similar number was found in this study when the MM-PBSA

Table 5
MM-PBSA energies (kcal/mol) of the wild type and Y181C HIV-1 RT complexes with nevirapine calculated by including different numbers of solvent molecules based on the embedded energy of WAT into the enzyme.

Models	$\Delta G_{\text{Complex}}$	$\Delta G_{\text{Enzyme-WAT}}$	ΔG_{Nev}	$\Delta G^{\text{MM-PBSA}}$	Residual
RT ^{WT} -1WAT. + Nev.	−11467.39 (0.69)	−11292.52 (0.69)	−137.22 (0.04)	−37.65 (1.42)	–
RT ^{WT} -2WAT. + Nev.	−11479.19 (0.69)	−11304.26 (0.69)	−137.22 (0.04)	−37.71 (1.42)	−0.06
RT ^{WT} -3WAT. + Nev.	−11490.39 (0.72)	−11315.46 (0.72)	−137.22 (0.04)	−37.71 (1.48)	0.00
RT ^{Y181C} -1WAT. + Nev.	−11422.36 (1.04)	−11253.85 (1.02)	−138.69 (0.04)	−29.82 (2.11)	–
RT ^{Y181C} -2WAT. + Nev.	−11440.62 (1.03)	−11271.96 (1.01)	−138.69 (0.04)	−29.97 (2.08)	−0.15
RT ^{Y181C} -3WAT. + Nev.	−11454.34 (1.09)	−11285.63 (1.07)	−138.69 (0.04)	−30.02 (2.20)	−0.05

RT^{WT}. = wild type HIV-1 RT.

RT^{Y181C}. = Y181C HIV-1 RT.

Nev. = nevirapine.

RT^{WT}-1WAT = wild type HIV-1 RT + WAT1067.

RT^{WT}-2WAT = wild type HIV-1 RT + WAT1067 + WAT407.

RT^{WT}-3WAT = wild type HIV-1 RT + WAT1067 + WAT407 + WAT497.

RT^{Y181C}-1WAT = Y181C HIV-1 RT + WAT1067.

RT^{Y181C}-2WAT = Y181C HIV-1 RT + WAT1067 + WAT1211.

RT^{Y181C}-3WAT = Y181C HIV-1 RT + WAT1067 + WAT1211 + WAT1040.

Values in parenthesis show standard error in the calculations.

Table 6

Relative fold resistance energies ($\Delta\Delta G$) in kcal/mol for Y181C HIV-1 RT/nevirapine complex normalized to wild type HIV-1 RT/nevirapine complex.

	$\Delta\Delta G$ (kcal/mol)
Expt. [45] fold resistance = 468	3.64
Expt. [45] fold resistance = 131.6	2.89
Expt. [46] fold resistance = 53.3	2.35
Expt. [47] fold resistance = 312.5	3.40
Expt. [48] fold resistance = 69.1	2.51
Predict I [27]	6.30
(this work)	7.83

Fold resistance = [Activity]_{mutant}/[Activity]_{wild type}.

Relative fold resistance energy, $\Delta\Delta G_{\text{fold resistance}} = \Delta G_{\text{mutant}} - \Delta G_{\text{wild type}} = RT \ln(\text{Resistance Fold})$.

(this work) acquired data of free energy of mutant minus wild type from Table 5 (including one water).

calculations including the bridge water molecule gave a relative fold resistance energy of 7.83 kcal/mol.

3.3. Interaction energies

After the binding free energies and the relative fold resistance energies had been obtained, the interaction energies between nevirapine and binding pocket residues of HIV-1 RT were further investigated to break down the main contributions. The individual interaction energies between each RT residue and nevirapine were calculated. This decomposition energy calculation was based on an MM-GBSA approach. The decomposition energies were calculated on a pairwise per-residue basis, in which the interactions included the contributions from side chains and the backbone of the pocket residues. Then the interactions between each residue and nevirapine were calculated. Decomposition energies of the wild type and Y181C HIV-1 RT compared to other theoretical calculations are shown in Table 7. The decomposition showed

Table 7

Interaction energies (kcal/mol) between nevirapine and the binding pocket residues of the wild type and Y181C HIV-1 RT.

Residues	Decomposition ^a		ONIOM ^b	MFCC approach ^c	
	Wild type	Y181C	Wild type	Wild type	Y181C
Pro95	−1.00 (0.10)	−1.29 (0.20)	−0.87	–	–
Leu100	−4.30 (0.43)	−4.24 (0.40)	−3.09	−0.58	2.81
Lys101	−0.71 (0.24)	−0.50 (0.17)	−1.10	−2.28	−3.08
Lys102	−0.26 (0.09)	−0.15 (0.06)	−0.37	−1.34	−1.25
Lys103	−1.92 (0.31)	−2.27 (0.40)	−1.20	−2.52	−2.38
Lys104	−0.09 (0.02)	−0.10 (0.02)	0.02	–	–
Ser105	−0.11 (0.03)	−0.08 (0.02)	−0.16	–	–
Val106	−2.99 (0.46)	−2.29 (0.30)	−0.08	–	–
Val179	−1.64 (0.26)	−1.89 (0.44)	1.47	–	–
Ile180	−0.72 (0.14)	−0.94 (0.11)	−0.50	–	–
Tyr181Cys	−3.12 (0.51)	−2.44 (0.34)	−2.79	1.34	7.63
Tyr188	−5.74 (0.40)	−4.94 (0.56)	−5.05	−2.07	−1.35
Val189	−1.11 (0.21)	−0.59 (0.13)	−0.64	–	–
Gly190	−0.84 (0.24)	−0.91 (0.13)	1.65	–	–
Phe227	−0.67 (0.27)	−0.72 (0.22)	−1.74	−2.23	−1.46
Leu228	−0.01 (0.00)	−0.02 (0.00)	−0.20	–	–
Trp229	−2.02 (0.30)	−1.52 (0.38)	−1.19	–	–
Leu234	−3.03 (0.40)	−2.43 (0.37)	1.22	–	–
His235	−1.19 (0.15)	−1.15 (0.20)	−2.39	−1.89	−0.64
Pro236	−1.30 (0.11)	−1.17 (0.17)	−1.63	−5.28	−2.41
Tyr318	−2.09 (0.40)	−2.17 (0.27)	−1.80	–	–
Glu138 ^d	−0.21 (0.08)	−0.12 (0.04)	−0.09	–	–

Values in parenthesis show standard error in the calculations.

^a Decomposition energies on a pairwise per-residue basis.

^b Interaction energies calculated at the MP2/6-31G(d,p) level with BSSE corrected from ONIOM3 (MP2/6-31G(d,p):B3LYP/6-31G(d,p):PM3) optimization [21].

^c Molecular fraction with conjugate caps approach [44].

^d Glu138 taken from p51 domain of RT.

that the main contributions were −4.30 and −5.74 kcal/mol from Leu100 and Tyr188, respectively, in the wild type RT complex. On the other hand, the main contributions of −4.24 and −4.94 kcal/mol were found from Leu100 and Tyr188, respectively, in the Y181C RT complex. Most of the decomposition energies for the wild type complex agree well with the interaction energies found from ONIOM (our own n-layered integrated molecular orbital and molecular mechanics) calculations [21]. Therefore, the decomposition revealed that the specific interactions changed from −3.12 to −2.44 kcal/mol at the position 181 mutation residue. This quantitative number explains clearly the loss of bound stability of nevirapine to the Y181C mutation in HIV-1 RT. Only a few interactions showed agreement between the decomposition and the MFCC (molecular fractionation with conjugate caps) approach [44]. Although the decomposition revealed that all of the interactions in the binding pocket were attractive forces, some repulsive energies were found in the ONIOM and the MFCC approaches. The repulsive forces were found at Leu100, Val179, Tyr181Cys, Gly190 and Leu234. These contradictory interactions at some amino acids occurred due to the geometry which had been used in the calculations. While the ONIOM and the MFCC approaches used a starting structure based on the crystallographic structure from the protein databank, the set of structures used in the decomposition energy calculations were taken by sampling from the MD trajectory. This introduced new information on how the dynamics of the binding pocket of HIV-1 RT – especially Leu100, Val179, Tyr181Cys, Gly190 and Leu234 – contributed to the nevirapine binding.

We further investigated the interaction between nevirapine and amino acids related to the bridge water molecule, since our analysis of MM-PBSA binding energies had proven the hydrogen bond bridge water molecule to be involved in nevirapine binding. The decomposition calculations via this bridge water molecule were investigated. The calculation of the decomposition energies from the bridge water molecule to nevirapine and the pocket residues, which refer to the hydrogen bond formation residues, are shown in Table 8. While the bridge water molecule formed similar levels of interaction to Leu234 and His235, significantly different interaction was found from nevirapine in the wild type and Y181C RT complexes (−2.36 and −0.56 kcal/mol, respectively). The lower attractive energy between WAT1067 and nevirapine in the mutant enzyme complex corresponded to the hydrogen bond distance analysis earlier. The larger freedom of WAT1067 in the Y181C RT complex caused the flexible distance between O(WAT1067) and O1(nevirapine); therefore, the attraction was broken. Losing interaction with the bridge water molecule had the important effect of making the nevirapine unstable inside the Y181C HIV-1 RT.

Moreover, if the decomposition energies via WAT1067 were taken into account, the attractive energies between Leu234, His235 and nevirapine (as shown in Table 7) had changed dramatically. Without the bridge water, nevirapine has attractive interactions of −3.03 and −1.19 kcal/mol to Leu234 and His235, respectively, in the wild type complex; and interactions of −2.43 and −1.15 kcal/mol to Leu234 and His235, respectively, in the mutant enzyme complex. When the attractive energies via the

Table 8

Decomposition energies (kcal/mol) from the bridge water (WAT1067) to nevirapine and the key amino acids in the HIV-1 RT binding site.

	Wild type	Y181C
Leu234	−0.60 (0.39)	−0.69 (0.30)
His235	−1.74 (0.45)	−1.68 (0.55)
Nevirapine	−2.36 (0.45)	−0.56 (0.61)

Values in parenthesis show standard error in the calculations.

bridge water (Table 8) concerning hydrogen bond formation had been taken into account, the additional attractive energies of nevirapine to the bridge water, and the bridge water to Leu234 and His235, were included in the interactions between nevirapine and pocket residues (Leu234 and His235). Thus the total attractive energy of nevirapine and Leu234 changed from -3.03 to -5.99 kcal/mol for the wild type RT, and from -2.43 to -3.68 kcal/mol for the Y181C mutant type RT. The attractive energy of nevirapine and His235 changed from -1.19 to -5.29 kcal/mol for the wild type RT, and from -1.15 to -3.39 kcal/mol for the Y181C mutant type RT. These results suggested that if the attractive energies via the bridging water molecule were included, the interaction energies from Leu234 and His235 would become the major contributions, as high as that found from the Tyr188. These significant interactions indicated clearly how WAT1067 plays an important role in the binding pocket.

4. Conclusion

The combined calculation of MD/MM-PBSA plus the bridge water molecule is the key to investigating the wild type and Y181C HIV-1 RT/nevirapine complexes. The binding energy from additional water molecules showed that only the bridge solvent needed to be involved as a receptor part in the MM-PBSA calculations. This revealed the significant contribution from bridge water in the binding of HIV-1 RT and nevirapine. The relative fold resistance energies are largely greater than predicted (i.e., experiment shows 2.5–3.64 kcal/mol, and theory shows 6.30–7.83 kcal/mol). The bridge water formed hydrogen bonding to nevirapine on one side, and amino acids inside the binding pocket on the other side. WAT1067 acts as a hydrogen donor to O1(nevirapine), O(Leu234) and O(His235), as well as a hydrogen acceptor from N-H(His235).

Interaction energies between nevirapine and residues in the binding pocket of HIV-1 RT were obtained from the decomposition energies calculation. Nevirapine showed attractive energies to all enzyme residues within 7 Å in both wild type and Y181C RT complexes. While most of the interaction energies from decomposition analysis agreed with ONIOM, MFCC calculations, some contradictory spots were found because of the structural dynamics of the binding site pocket. Normally, the major contribution came from Tyr188 and Leu100 which have the strong interactions of -5.74 and -4.30 kcal/mol, respectively, for the wild type HIV-1 RT complex, and -4.94 and -4.24 kcal/mol, respectively, for the Y181C HIV-1 RT complex. But when the interactions via the bridge water were taken into account, the contribution from Leu234 and His235 became prominent as well, due to the interactions of -5.99 and -5.29 kcal/mol, respectively, for the wild type HIV-1 RT complex, and -3.68 and -3.39 kcal/mol, respectively, for the Y181C HIV-1 RT complex. The key interaction was that the attractive energies depended on the bridging water molecule. Apart from losing the interaction at position 181 when tyrosine was mutated to cysteine, nevirapine also lost interaction with the bridge water molecule. The lower stability of nevirapine inside the Y181C HIV-1 RT pocket was proven from the lack of attractive force. Therefore, nevirapine lacks potential against Y181C HIV-1 RT. The significant role of the bridging water molecule and its dynamic properties provide key information to strengthen the fundamental understanding of the binding of non-nucleoside HIV-1 RT inhibitors.

Acknowledgements

This work was supported by the Thailand Research Fund through a TRF Research-Team Promotion Grant (RTA5080005). W.T. is grateful to the Royal Golden Jubilee Ph.D. Program (3C.KU/

47/B.1) for a scholarship. Partial support was also provided by the National Nanotechnology Center (NANOTEC), Ministry of Science and Technology, Thailand, through its program of Center of Excellence Network, and is gratefully acknowledged. Computing time and research facilities were made available through the generosity of the Postgraduate Education and Research Programs in Petroleum, Petrochemical Technology and Advanced Materials. Support from the Laboratory of Computational and Applied Chemistry (LCAC), the High Performance Computing and Networking Center (HPCNC) of Kasetsart University, and the Large Scale Research Laboratory at NECTEC are gratefully acknowledged. Thanks are due to Adrian Hillman and Christopher Salisbury for reading of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2009.02.007.

References

- [1] J.M. Whitcomb, S.H. Hughes, *Annu. Rev. Cell Biol.* 8 (1992) 275.
- [2] H. Mitsuya, R. Yarchoan, S. Hayashi, S. Broder, *J. Am. Acad. Dermatol.* 22 (1990) 1282.
- [3] E. De Clercq, *J. Med. Chem.* 38 (1995) 2491.
- [4] M. Baba, H. Tanaka, E. De Clercq, R. Pauwels, J. Balzarini, D. Schols, H. Nakashima, C.F. Perno, R.T. Walker, T. Miyasaka, *Biochem. Biophys. Res. Commun.* 165 (1989) 1375.
- [5] R. Pauwels, K. Andries, J. Desmyter, D. Schols, M.J. Kukla, H.J. Breslin, A. Raeymaeckers, J. Van Gelder, R. Woestenborghs, J. Heykants, *Nature* 343 (1990) 470.
- [6] V.J. Merluzzi, K.D. Hargrave, M. Labadia, K. Grozinger, M. Skoog, J.C. Wu, C.K. Shih, K. Eckner, S. Hattox, J. Adams, *Science* 250 (1990) 1411.
- [7] S.D. Young, S.F. Britcher, L.O. Tran, L.S. Payne, W.C. Lumma, T.A. Lyle, J.R. Huff, P.S. Anderson, D.B. Olsen, S.S. Carroll, *Antimicrob. Agents Chemother.* 39 (1995) 2602.
- [8] Y. Hsiou, J. Ding, K. Das, A.D. Clark Jr., S.H. Hughes, E. Arnold, *Structure* 4 (1996) 853.
- [9] J. Ding, K. Das, Y. Hsiou, S.G. Sarafianos, A.D. Clark Jr., A. Jacobo-Molina, C. Tantillo, S.H. Hughes, E. Arnold, *J. Mol. Biol.* 284 (1998) 1095.
- [10] A. Spallarossa, S. Cesarini, A. Ranise, M. Ponassi, T. Unge, M. Bolognesi, *Biochem. Biophys. Res. Commun.* 365 (2008) 764.
- [11] K. Das, J. Ding, Y. Hsiou, A.D. Clark Jr., H. Moereels, L. Koymans, K. Andries, R. Pauwels, P.A.J. Janssen, P.L. Boyer, P. Clark, R.H. Smith Jr., M.B.K. Smith, C.J. Michejda, S.H. Hughes, E. Arnold, *J. Mol. Biol.* 264 (1996) 1085.
- [12] H. Huang, R. Chopra, G.L. Verdine, S.C. Harrison, *Science* 282 (1998) 1669.
- [13] O. Kensch, T. Restle, B.M. Wohrl, R.S. Goody, H.-J. Steinhoff, *J. Mol. Biol.* 301 (2000) 1029.
- [14] M.S. Hirsch, F.O.B.-V. Zinet, B. Clotet, B. Conway, D.R. Kuritzkes, R.T. D'Aquila, L.M. Demeter, S.M. Hammer, V.A. Johnson, C. Loveday, J.W. Mellors, D.M. Jacobsen, D.D. Richman, *Clin. Infect. Dis.* 37 (2003) 113–128.
- [15] R.C. Rizzo, D.-P. Wang, J. Tirado-Rives, W.L. Jorgensen, *J. Am. Chem. Soc.* 122 (2000) 12898.
- [16] J. Ren, C. Nichols, L. Bird, P. Chamberlain, K. Weaver, S. Short, D.I. Stuart, D.K. Stammers, *J. Mol. Biol.* 312 (2001) 795.
- [17] L. Bacheler, S. Jeffrey, G. Hanna, R. D'Aquila, L. Wallace, K. Logue, B. Cordova, K. Hertogs, B. Larder, R. Buckery, D. Baker, K. Gallagher, H. Scarnati, R. Tritch, C. Rizzo, *J. Virol.* 75 (2001) 4999.
- [18] S. Hannongbua, S. Prasithichokekul, P. Pungpo, *J. Comput.-Aided Mol. Des.* 15 (2002) 997.
- [19] I. Bahar, B. Erman, R.L. Jernigan, A.R. Atilgan, D.G. Covell, *J. Mol. Biol.* 285 (1999) 1023.
- [20] M. Kuno, S. Hannongbua, K. Morokuma, *Chem. Phys. Lett.* 380 (2003) 456.
- [21] M. Kuno, R. Hongkrenkai, S. Hannongbua, *Chem. Phys. Lett.* 424 (2006) 172–177.
- [22] M.B.K. Smith, M.L. Lamb, J. Tirado-Rives, W.L. Jorgensen, C.J. Michejda, S.K. Rugs, R.H. Smith Jr., *Protein Eng.* 13 (2000) 413.
- [23] F.t. Rodríguez-Barrios, F. Gago, *J. Am. Chem. Soc.* 126 (2004) 15386.
- [24] M.B. Kroeger Smith, L.H. Rader, A.M. Franklin, E.V. Taylor, K.D. Smith, R.H. Smith Jr., J. Tirado-Rives, W.L. Jorgensen, *Bioorg. Med. Chem. Lett.* 18 (2008) 969.
- [25] R.C. Rizzo, M. Udier-Blagovic, D.-P. Wang, E.K. Watkins, M.B. Kroeger Smith, R.H. Smith Jr., J. Tirado-Rives, W.L. Jorgensen, *J. Med. Chem.* 45 (2002) 2970.
- [26] J. Srinivasan, T.E. Cheatham III, P. Cieplak, P.A. Kollman, D.A. Case, *J. Am. Chem. Soc.* 120 (1998) 9401–9409.
- [27] Z. Zhou, M. Madrid, J.D. Evanseck, J.D. Madura, *J. Am. Chem. Soc.* 127 (2005) 17253.
- [28] R.C. Rizzo, J. Tirado-Rives, W.L. Jorgensen, *J. Med. Chem.* 44 (2001) 145.
- [29] J. Ren, R. Esnouf, E. Garman, D. Somers, C. Ross, I. Kirby, J. Keeling, G. Darby, Y. Jones, et al., *Nat. Struct. Mol. Biol.* 2 (1995) 293.
- [30] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R.

- Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 98 user's reference, Gaussian Inc., Pittsburgh, 1998.
- [31] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, J. Phys. Chem. 97 (1993) 10269.
- [32] T. Fox, P.A. Kollman, J. Phys. Chem. B 102 (1998) 8070.
- [33] J. Wang, P. Morin, W. Wang, P.A. Kollman, J. Am. Chem. Soc. 123 (2001) 5221.
- [34] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, Comput. Phys. Commun. 91 (1995) 1.
- [35] J. Wang, P. Cieplak, P.A. Kollman, J. Comput. Chem. 21 (2000) 1049.
- [36] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
- [37] J.P. Ryckaert, G. Ciccoliti, H.J.C. Berendsen, J. Comput. Phys. 23 (1977) 327.
- [38] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, P.A. Kollman, AMBER 9, University of California, San Francisco, 2006.
- [39] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, J. Comput. Chem. 24 (2003) 1999.
- [40] M.C. Lee, Y. Duan, Proteins: Struct. Funct. Bioinform. 55 (2004) 620.
- [41] P. Weinzinger, S. Hannongbua, P. Wolschann, J. Enzyme Inhib. Med. Chem. 20 (2005) 129–134.
- [42] M. Udier-Blagovic, J. Tirado-Rives, W.L. Jorgensen, J. Med. Chem. 47 (2004) 2389.
- [43] L.L. Duan, Y. Tong, Y. Mei, Q.G. Zhang, J.Z.H. Zhang, J. Chem. Phys. 127 (2007), 145101/1.
- [44] X. He, Y. Mei, Y. Xiang, D.W. Zhang, J.Z.H. Zhang, Proteins: Struct. Funct. Bioinform. 61 (2005) 423–432.
- [45] R.A. Spence, K.S. Anderson, K.A. Johnson, Biochemistry 35 (1996) 1054.
- [46] P. Sarpapakorn, S. Hannongbua, D. Rognan, SAR QSAR Environ. Res. 17 (2006) 183.
- [47] D.W. Ludovici, M.J. Kukla, P.G. Grous, S. Krishnan, K. Andries, M.-P.d.B. Thune, H. Azijn, R. Pauwels, E.D. Clercq, E. Arnold, P.A.J. Janssens, Bioorg. Med. Chem. Lett. 11 (2001) 2225–2228.
- [48] A. Rao, J. Balzarini, A. Carbone, A. Chimiri, E.D. Clercq, A.M. Monforte, P. Monforte, C. Pannecouque, M. Zappalà, Antiviral Res. 63 (2004) 79–84.
- [49] N. Guex, M.C. Peitsch, Electrophoresis 18 (1997) 2714–2723.
- [50] J. Ren, C. Nichols, L. Bird, P. Chamberlain, K. Weaver, S. Short, D.I. Stuart, D.K. Stammers, J. Mol. Biol. 312 (2001) 795–805.
- [51] J. Ren, C.E. Nichols, A. Stamp, P.P. Chamberlain, R. Ferris, K.L. Weaver, S.A. Short, D.K. Stammers, Febs J. 273 (2006) 3850–3860.
- [52] P.P. Chamberlain, J. Ren, C.E. Nichols, L. Douglas, J. Lennerstrand, B.A. Larder, D.I. Stuart, D.K. Stammers, J. Virol. 76 (2002) 10015–10019.
- [53] J. Ren, C.E. Nichols, P.P. Chamberlain, K.L. Weaver, S.A. Short, D.K. Stammers, J. Mol. Biol. 336 (2004) 569–578.

Use of 3D QSAR to investigate the mode of binding of pyrazinones to HIV-1 RT

Patchreenart Saparpakorn · Ratsupa Thammaporn · Supa Hannongbua

Received: 15 September 2008 / Accepted: 22 October 2008 / Published online: 27 November 2008
© Springer-Verlag 2008

Abstract Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) based on the docked conformation were performed for 24 pyrazinone derivatives. All compounds were docked into the wild-type HIV-1 RT binding pocket and the lowest-energy docked configurations were used to construct the 3D QSAR models. The CoMFA and CoMSIA models enable good prediction of inhibition by the pyrazinones, with $r_{cv}^2 = 0.703$ and 0.735 . Results obtained from CoMFA and CoMSIA based on the docking conformation of the pyrazinones are, therefore, powerful means of elucidating the mode of binding of pyrazinones and suggesting the design of new potent NNRTIs.

Keywords HIV-1 · CoMFA · CoMSIA · Molecular docking · NNRTI

Introduction

Acquired immunodeficiency syndrome (AIDS) is caused by the human immunodeficiency virus (HIV). The reverse transcriptase (RT) of human immunodeficiency virus type-1 (HIV-1) is the essential enzyme converting the single-stranded viral RNA genome into double-stranded proviral DNA prior to its integration into the host genomic DNA.

HIV-1 RT is an asymmetric heterodimer consisting of a 560-residue chain (p66 subunit) and a 440-residue chain (p51 subunit). Nucleoside inhibitors (NRTIs) and non-nucleoside inhibitors (NNRTIs) are two types RT inhibitor widely used for treatment of AIDS. Nucleoside analog inhibitors, for example AZT, ddI, and ddC, are competitive inhibitors causing termination of the growing DNA chain. Although these nucleoside inhibitors are selective for HIV-1 RT, they are not highly specific and can affect cellular polymerases, with toxic effects [1, 2]. Non-nucleoside analog inhibitors, for example nevirapine, TIBO, and efavirenz, are non-competitive inhibitors binding in a hydrophobic pocket that is about 10 Å away from the enzyme's active site in the p66 subunit. Because of their use at low concentration and their high specificity, non-competitive inhibitors are of interest for developing novel potent inhibitors. Unfortunately, the major problem is the rapid development of NNRTI resistance, for example, Leu100Ile, Lys103Asn, and Tyr181Cys [3]. In order to improve the biological activity of NNRTIs, some classes of NNRTI, for example diaryltrialazine (DATA) [4], diarylpyrimidine (DAPY) [5], TMC-125 [6], indolyl aryl sulfone (IAS) [7], and TMC-278 [8], have been proposed and designed to enhance activity against wild-type and mutant type HIV-1 RT. A series of non-nucleoside reverse transcriptase inhibitors (NNRTIs), pyrazinones, have been found to be active against wild-type HIV-1 RT and some are also active against mutant HIV-1 RT [9]. Information about the specific binding mode of pyrazinones to HIV-1 RT is therefore needed to suggest designs of novel potent NNRTIs active against both wild and mutant HIV-1 RT.

Comparative molecular field analysis (CoMFA) [10] and comparative molecular similarity indices analysis (CoMSIA) [11] are three-dimensional quantitative structure–activity relationship (3D QSAR) methods widely used

P. Saparpakorn (✉) · R. Thammaporn · S. Hannongbua
Department of Chemistry, Faculty of Science,
Kasetsart University (KU), Chatuchak,
Bangkok 10903, Thailand
e-mail: fscipnks@ku.ac.th

P. Saparpakorn · R. Thammaporn · S. Hannongbua
Center of Nanotechnology, Kasetsart University (KU),
Bangkok, Thailand

to study and design new inhibitors. These methods have been used successfully for many classes of inhibitor by studying the relationship between their chemical structures and their biological activity. For NNRTIs, these methods are also able to reveal the dominant interaction of some types of NNRTI, for example TIBO [12], (aryl)bridged 2-aminobenzonitriles [13], and efavirenz [14]. In addition, the idea of a ligand's active conformation has been studied by using the docked conformation from the molecular docking method to construct 3D QSAR models. The docked conformation of some NNRTIs has been obtained by constructing and satisfying CoMFA and CoMSIA models [15, 16]. To avoid use of an incorrect conformation for construction of the 3D QSAR model, the docked conformation has been successfully used to select the bioactive conformation of the compounds [17]. Moreover, the 3D QSAR contour map obtained from the docked conformation can be superimposed on the active site for better explanation of the binding. In order to study the binding of pyrazinones to wild-type HIV-1 RT, the aim of this work was to derive their CoMFA and CoMSIA models based on their docked conformations.

Results and discussion

Molecular docking of pyrazinones

Because of the rotatable bonds connected to rings A and B of the pyrazinones, their optimized structures had different conformations from R185545, the conformation of which is known from X-ray studies. Therefore, the molecular docking method was used to propose bioactive conformations of the pyrazinones. The docked conformation was then used for constructing the 3D QSAR in the next step. To validate the AutoDock program for docking pyrazinones into the HIV-1 RT binding pocket, R185545 was docked back into its NNRTI binding pocket. From all 50 configurations, the results showed good ability to reproduce the same orientation as the X-ray position from RMSD values for the docked positions from the X-ray position of R185545 at about 1.03 to 1.38 Å. The final docked energies were about −61.92 to −63.81 kJ/mol. Therefore, these docking data were used to dock pyrazinones into the NNRTI binding pocket. From the docking results, each pyrazinone position was separated into several clusters but most of the positions (out of 50 configurations) were located in a binding orientation similar to that of R185545. Several positions have some parts of the compound located outside the binding pocket. For the final docked energies, most positions located in the binding orientation similar to that of R185545 had lower final docked energies than the other positions. Consequently,

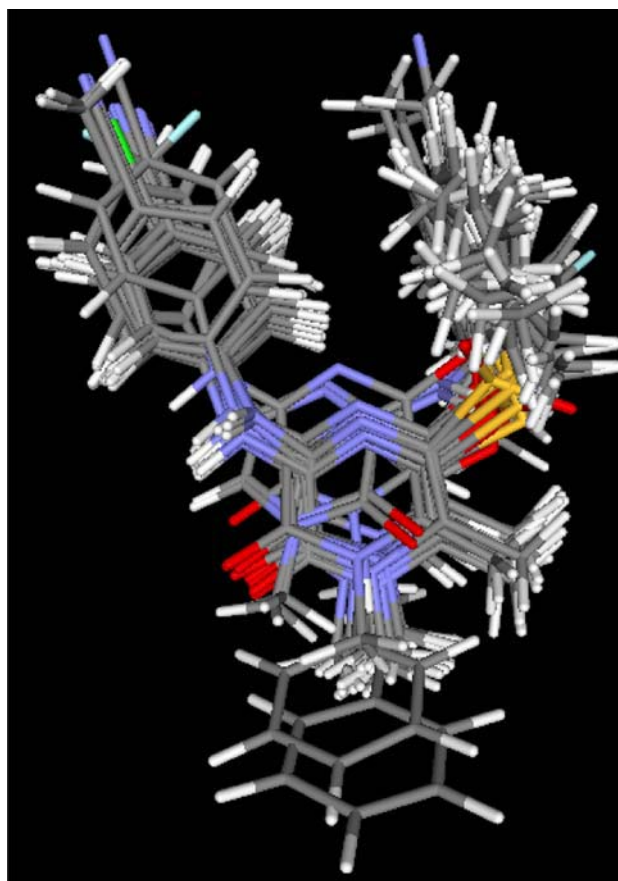


Fig. 1 Selected docked conformation of all the pyrazinones

each pyrazinone position with the lowest final docked energy having similar binding orientation to R185545 was selected as the orientation used to construct 3D QSAR models in the next step. The positions selected for the pyrazinones are shown in Fig. 1. It is clearly apparent that they adopt the same binding orientation, except for compounds **1** and **2**, both of which had the opposite position of the carbonyl group in the central pyrazinone ring. Because of the small atoms (H and F) attached at the R position of compounds **1** and **2**, the methyl group at the R¹ position formed an H-bond with His235 instead of the hydrogen atom at the R position. This caused an upwards shift of the orientation and the opposite position of the carbonyl group on the central pyrazinone ring of the docking positions of compounds **1** and **2**.

CoMFA and CoMSIA models

From PLS analysis, the results from the best CoMFA and CoMSIA models are shown in Table 1, where r_{cv}^2 , noc, s-press, and r^2 represent the cross-validation coefficient, the optimum number of components, the uncertainty of the prediction error sum of squares derived from the LOO method, and the regression coefficient. From Table 1,

Table 1 Statistical results for the CoMFA and CoMSIA models based on the docking conformation of the pyrazinones

	Model	
	CoMFA	CoMSIA
r_{cv}^2	0.703	0.735
noc	6	6
s-press	0.622	0.571
r^2	0.994	0.980
S	0.092	0.158
F	230.253	87.871
Field contribution (%)		
Steric	59.0	11.8
Electrostatic	41.0	15.7
H-acceptor	–	36.6
H-donor	–	2.9
Hydrophobic	–	33.0

CoMFA and CoMSIA models reveal the good predictive power with r_{cv}^2 of 0.703 and 0.735. The r^2 , s , and F values were 0.994, 0.092, and 230.253 for the CoMFA model and 0.980, 0.158, and 87.871 for the CoMSIA model. Outliers of the best CoMFA model were compounds **3**, **6**, and **13**. For the best CoMSIA model, the outlier was compound **3**. These outliers are because of the different substituent group at positions R and R⁴. The substituent groups at the R position of compounds **3** and **6** are the methyl and trifluoromethyl groups which are different from those of other compounds in the training set. In compound **13**, the R⁴ substituent is the benzyl group, which is different from the hydrogen or methyl group used in the training set. For the CoMFA model, the contributions of the steric and electrostatic interactions were 59.0 and 41.0%. The main contribution in the CoMFA model was steric interaction. The field contributions of the CoMSIA model were 11.8, 15.7, 36.6, 2.9, and 33.0% for steric, electrostatic,

H-acceptor, H-donor, and hydrophobic interactions, respectively. The H-acceptor and hydrophobic interactions were found to be the major contributions in the CoMSIA model. The 3D fields of the best CoMFA and CoMSIA models were represented as contour maps.

Steric and electrostatic contours of CoMFA and CoMSIA for compound **9** are shown in Figs. 2 and 3. Some steric contours of CoMFA and CoMSIA models (Fig. 2) are shown in the same region as the binding pocket. However, the steric contour map from the CoMSIA model was more localized than that from the CoMFA model. In both models, the green contours were localized near the bottom of ring A and substituent groups R and R¹. The green contour localized near the bottom of ring A was caused by the slightly different docked conformation of the ring A position. Ring A preferred the lower position, because in this position the ring formed a π – π interaction with Tyr318 and an H-bond with the backbone carbonyl group of Lys101. In addition, the favored steric area was found at the R³ and R⁴ substituted groups of the pyrazinone ring in the CoMSIA model. This favored steric area of the CoMSIA model also had the small yellow contour. This suggested that a group which was too bulky was not favored in this area. For the R³ substituted group, a group which was too bulky would cause steric interactions with the sidechain of Val179 and the C $_{\beta}$ of Glu138(B). Compared with compounds **9** and **20**, compounds with the hydrogen atom at the R³ position have greater activity than that with the methyl substituent. Because most of the R⁴ substituent groups in the training set were the methyl group, the exception being compound **22**, there was less effect at this substituent. Compound **22**, with the H substituent at the R⁴ position, had the same activity as compound **16**. The methyl group at the R³ position of compounds **16** and **22** caused slight movement of the structure compared with compound **9**. Therefore, the methyl group as the R⁴ substituent group of compound **16**

Fig. 2 Steric contour of the CoMFA (a) and CoMSIA (b) models. The green (G) color represents the favored steric area and the yellow (Y) color represents the disfavored steric area

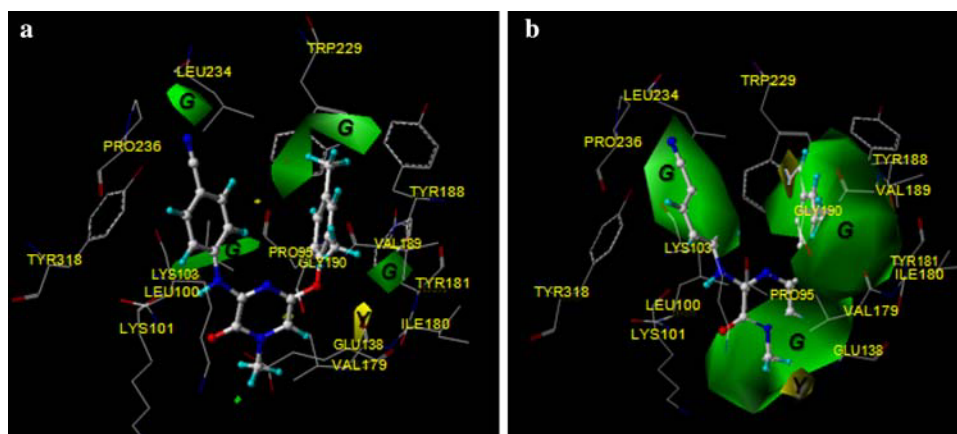
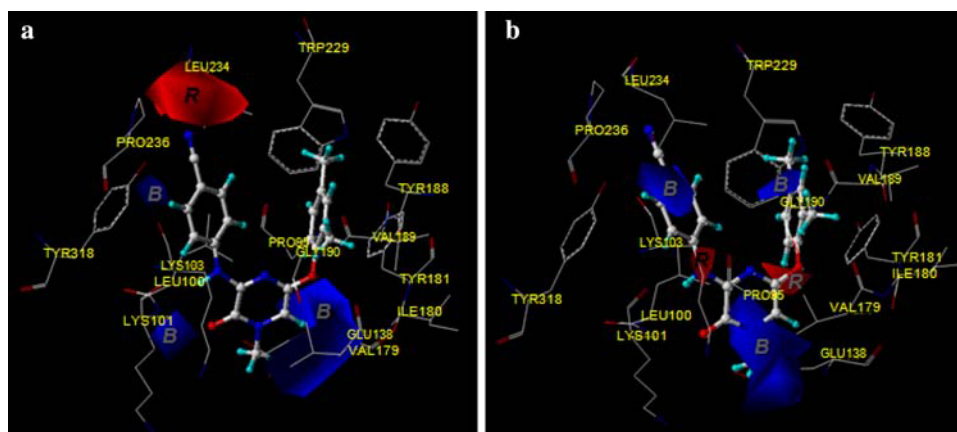


Fig. 3 Electrostatic contour of the CoMFA (a) and CoMSIA (b) models. The blue (B) color represents the favored positive electrostatic area and the red (R) color represents the favored negative electrostatic area



could not form an H-bonding interaction with Glu138(B). The green contours localized near the R and R¹ substituent groups suggest that increasing the bulkiness of groups attached at R and R¹ increase the activity but the increase of this bulky group should be electrostatic in character, as suggested by the following electrostatic contours. For the electrostatic contour in the CoMFA model, the favored negative electrostatic area was found at the R substituent groups. This may suggest the advantage of having an electron-withdrawing substituent group at the R position. This located contour was not found in the CoMSIA model. For compound **9**, substitution of a cyano group at the R position of ring A resulted in an attractive interaction with Leu234 and His235, and a methyl group at the R¹ position of ring B resulted in an H- π interaction with Tyr188 and Trp229. In both CoMFA and CoMSIA models, the favored positive electrostatic contours appeared in the same area. The contours were found near the R³ and R⁴ positions. The electrostatic CoMFA contour, excluding the common contour found in both CoMFA and CoMSIA models, can be explained by using the contours in the three additional field contributions of the CoMSIA model (i.e. hydrophobic, H-acceptor, and H-donor interactions).

The CoMSIA contours of three additional field contributions, i.e. hydrophobic, H-acceptor, and H-donor interactions, are shown in Fig. 4. One disfavored hydrophobic area is revealed at the R position. The results agree well with the favored negative electrostatic area in the CoMFA model. The other disfavored hydrophobic area appears at the bottom of the R⁴ position and a space between ring A and B. These basic structural interactions suggest the design of new inhibitors by adding a hydrophilic substituent group around these regions. From the docked conformation, the bottom of the R⁴ position had the space to interact with the sidechain of Lys101 and there was the space between ring A and B to interact with the sidechain of Leu234. For the H-acceptor contour, favored and disfavored H-acceptor areas were located at the R position. The favored H-acceptor area was

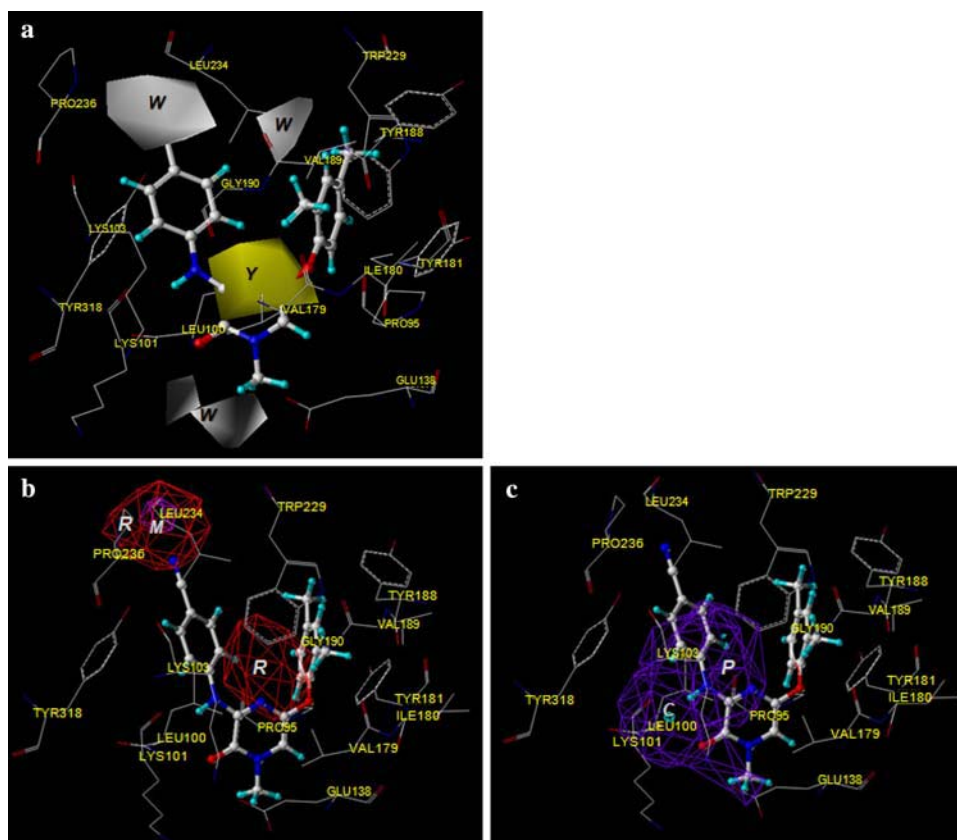
surrounded by the disfavored H-acceptor area, as shown in the line contour map in Fig. 4b. This suggests that the substituent group preferred at the R position is a H-acceptor group but that this substituted H-acceptor group should not be bulky. Considering the binding pocket, this indicated that this disfavored H-acceptor area was located near the nitrogen and the oxygen atoms of backbone His235 and Pro236. The disfavored H-donor area was found in the wide region from the bottom part of ring A to the carbonyl group of pyrazinone ring. Moreover, there was a small favored H-donor area inside the disfavored H-donor area, which agrees well with the position of amino group between ring A and the pyrazinone ring. From the docked conformation in the binding pocket, the amino group was important for these derivatives because H-bonding occurred between the hydrogen atom of an amino group and an oxygen atom of backbone Lys101.

To validate the best CoMFA and CoMSIA models, the biological activities of compounds in the test set were predicted by using the best CoMFA and CoMSIA models obtained from the training set. The predicted log(1/IC₅₀) and residuals for all the compounds are shown in Table 2. The residuals of the test set were within ± 0.9 log(1/IC₅₀) unit in both models. The r^2 of the test set are 0.7414 and 0.8659 for the CoMFA and CoMSIA models. The ability of both models to predict the biological activity of the pyrazinones is summarized in Fig. 5 as a plot of experimental activity [log(1/IC₅₀)] against activity predicted by the best CoMFA and CoMSIA models.

Conclusions

To investigate the important mode of binding of pyrazinones to HIV-1 RT, 3D QSAR methods were used with the docked conformations of the pyrazinones. Twenty-four pyrazinones were docked into the HIV-1 RT binding pocket and the orientations of the pyrazinones in the binding pocket were demonstrated. The docked conformation of each pyrazinone

Fig. 4 Hydrophobic (a), H-acceptor (b), and H-donor (c) contours of the CoMSIA models. The *yellow* (Y) color represents the favored hydrophobic area, the *white* (W) color represents the disfavored hydrophobic area, the *magenta* (M) color represents the favored H-acceptor area, the *red* (R) color represents the disfavored H-acceptor area, the *cyan* (C) color represents the favored H-donor area, and the *purple* (P) color represents the disfavored H-donor area



was used to construct CoMFA and CoMSIA models. These models had good predictive power, with r_{cv}^2 of 0.703 and 0.735. Steric and electrostatic contours of the CoMFA model are more or less similar to those of the CoMSIA model. The electrostatic contour found only in the CoMFA model agrees well with the contours from the three additional field contributions of the CoMSIA model (i.e. hydrophobic, H-acceptor, and H-donor interactions). Both models led to similar results and suggested that the preferred substituent at the R position should be a bulky, electron-withdrawing, H-acceptor, and that hydrophobic groups are not favored for forming an attractive interaction with Leu234 and His235. Therefore, the results obtained from the CoMFA model, and CoMSIA analyses based on the docked conformation of the pyrazinones, were powerful methods for describing the binding mode of pyrazinones and suggesting the design of new potent NNRTIs (Fig. 6).

Computational details

Data set and molecule preparation

Twenty-four pyrazinone structures and their experimental activities [$\log(1/IC_{50})$] used in this work are shown in Table 3. Starting geometries were optimized by quantum

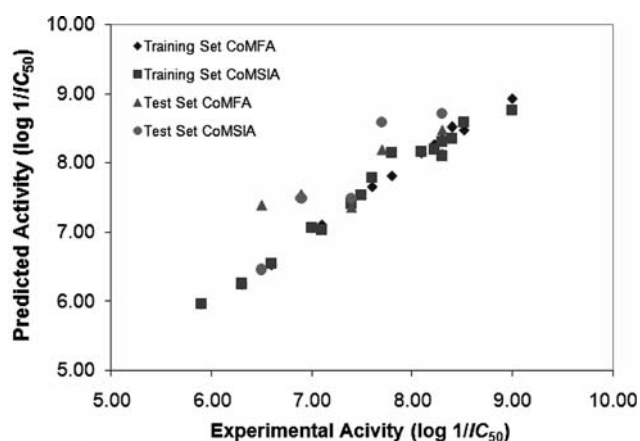
chemical calculations at the HF/3-21G level of theory using the Gaussian03 program [18]. Gasteiger–Hückel charges were added by using SYBYL software version 7.0 [19]. Because of the similar structures of the pyrazinones, the X-ray structure of HIV-1 RT complex with R185545 was selected from the Protein Databank (pdb code 1suq [20]) and used in this study. After removing R185545 from the complex structure, all-hydrogen atoms were added to the protein using standard SYBYL geometries, and Kollman charges were also added. For the R185545 structure, hydrogen atoms and Gasteiger–Hückel charges were added using SYBYL software.

Molecular docking of the pyrazinones

The pyrazinones were docked into the binding pocket of HIV-1 RT by using the software AutoDock3.05 [21] with the Lamarckian genetic algorithm. First, the validity of the docking method was verified by redocking R185545 back into the binding pocket of the protein structure (pdb code 1suq). The grid size and grid spacing were set to $60 \times 60 \times 60$ and 0.375 Å, respectively. The numbers of runs was set to 50. All other parameters were the default value. To select the configuration for setting up the 3D QSAR models, the lowest docked energy configuration having similar orientation of R185545 was used.

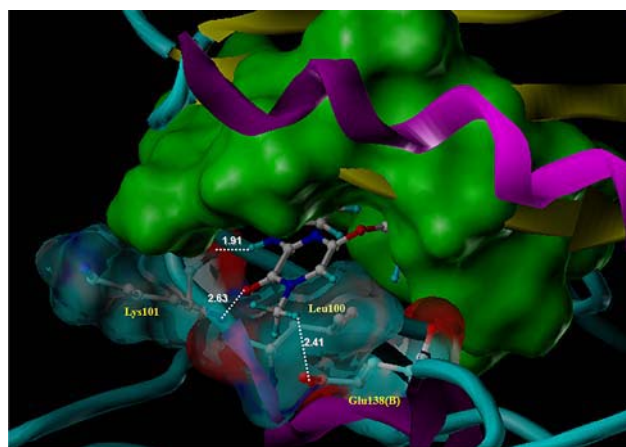
Table 2 Predicted $\log(1/IC_{50})$ and residuals for the CoMFA and CoMSIA models

Compounds	$\log(1/IC_{50})$	CoMFA		CoMSIA	
		Calc. $\log(1/IC_{50})$	Residual	Calc. $\log(1/IC_{50})$	Residual
1	6.60	6.52	0.08	6.54	0.06
2	5.90	5.96	-0.06	5.95	-0.05
3	4.00	6.44	-2.44	7.14	-3.14
4	7.60	7.65	-0.05	7.78	-0.18
5	7.10	7.10	0.00	7.02	0.08
6	6.30	8.85	-2.55	6.23	0.07
7	8.40	8.51	-0.11	8.35	0.05
8	7.70	8.19	-0.49	8.58	-0.88
9	9.00	8.92	0.08	8.76	0.24
10	8.30	8.34	-0.04	8.10	0.20
11	7.80	7.81	-0.01	8.14	-0.34
12	6.50	7.38	-0.88	6.45	0.05
13	7.49	8.80	-1.31	7.52	-0.03
14	7.00	7.05	-0.05	7.06	-0.06
15	8.30	8.13	0.17	8.29	0.01
16	7.40	7.36	0.04	7.48	-0.08
17	6.90	7.54	-0.64	7.48	-0.58
18	6.30	6.23	0.07	6.26	0.04
19	8.30	8.47	-0.17	8.70	-0.40
20	8.52	8.47	0.05	8.58	-0.06
21	8.22	8.26	-0.04	8.19	0.03
22	7.40	7.44	-0.04	7.40	0.00
23	8.10	8.14	-0.04	8.16	-0.06
24	8.30	8.31	-0.01	8.29	0.01

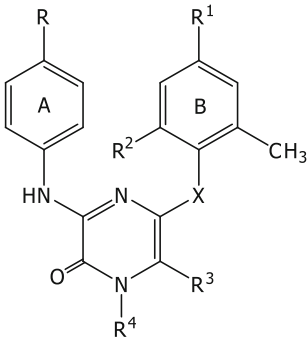
**Fig. 5** Plot of experimental activity [$\log(1/IC_{50})$] against activity predicted by the best CoMFA and CoMSIA models

CoMFA and CoMSIA models

The selected configurations of each compound from molecular docking were used to construct the 3D QSAR models. These docked structures were divided into the training set and the test set. Five compounds representing the diversity of the pyrazinone derivatives and the range of

**Fig. 6** Orientation of compound **9** (ball and stick model) in the binding pocket surface of HIV-1 RT

activities [$\log(1/IC_{50})$] served as the test set. These were compounds **8**, **12**, **16**, **17**, and **19**. The remaining compounds served as the training set for constructing 3D QSAR models. SYBYL software, version 7.0, was used to study the CoMFA and CoMSIA models. In CoMFA analysis, the steric and electrostatic fields were calculated using Lennard–Jones and Coulombic potentials. Default values for probes were

Table 3 Structures of the pyrazinones and antiviral activity against wild-type HIV-1 RT


Compound	R	X	R ¹	R ²	R ³	R ⁴	log(1/IC ₅₀)
1	H	O	CH ₃	H	H	CH ₃	6.60
2	F	O	CH ₃	H	H	CH ₃	5.90
3	CH ₃	O	CH ₃	H	H	CH ₃	4.00
4	Cl	O	CH ₃	H	H	CH ₃	7.60
5	Cl	S	CH ₃	CH ₃	H	CH ₃	7.10
6	CF ₃	O	CH ₃	H	H	CH ₃	6.30
7	CN	O	H	H	H	CH ₃	8.40
8	CN	S	H	H	H	CH ₃	7.70
9	CN	O	CH ₃	H	H	CH ₃	9.00
10	CN	S	CH ₃	H	H	CH ₃	8.30
11	CN	S	CH ₃	CH ₃	H	CH ₃	7.80
12	CN	O	H	H	H	CH ₂ Ph	6.50
13	CN	O	CH ₃	H	H	CH ₂ Ph	7.49
14	Cl	O	H	H	CH ₃	CH ₃	7.00
15	CN	O	H	H	CH ₃	CH ₃	8.30
16	CN	S	H	H	CH ₃	CH ₃	7.40
17	Cl	O	CH ₃	H	CH ₃	CH ₃	6.90
18	Cl	S	CH ₃	H	CH ₃	CH ₃	6.30
19	CN	O	CH ₃	CH ₃	CH ₃	CH ₃	8.30
20	CN	O	CH ₃	H	CH ₃	CH ₃	8.52
21	CN	O	CN	CH ₃	CH ₃	CH ₃	8.22
22	CN	S	H	H	CH ₃	H	7.40
23	CN	SO ₂	CH ₃	CH ₃	H	CH ₃	8.10
24	CN	SO ₂	CH ₃	CH ₃	CH ₃	CH ₃	8.30

positively charged sp^3 hybridized carbon, 2 Å of grid spacing, and energy cutoff values of 125.52 kJ/mol for steric and electrostatic fields. For CoMSIA analysis, five different similarity fields were studied: steric, electrostatic, hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor. The similarity indices of five fields at regularly spaced grid points were calculated. Probes with carbon atom with +1.0 charge and 1 Å radius were used. Partial least-squares (PLS) regression analysis was used to derive the CoMFA and CoMSIA models. The models represented the relationships between the independent variables (CoMFA and CoMSIA

fields) and the dependent variable (the experimental activities [$\log(1/IC_{50})$] for each compound). The maximum number of components was set to six and leave-one-out (LOO) was used to evaluate the predictive capability of the model and to obtain the optimum number of components. The cross-validated correlation coefficient (q^2 or r_{cv}^2) represented the quality of the model. The PLS procedure without cross-validation was performed to generate the CoMFA and CoMSIA models by using the chosen optimum number of components. The CoMFA and CoMSIA models were represented by use of contour maps. For better understanding of the pyrazinones, the contour maps were superimposed on to the NNRTI binding site.

Acknowledgments This investigation was supported by the Thailand Research Fund (RTA5080005). P.S. is grateful to the Thailand Research Fund (MRG5080267) and the Faculty of Science (ScRF), Kasetsart University for a research grant. The National Synchrotron Research Center, Thailand, is also acknowledged for R.T.'s scholarship. The Center of Nanotechnology at Kasetsart University, Laboratory for Computational and Applied Chemistry (LCAC), the Postgraduate Program in Education and Research on Petroleum, Petrochemical Technology and Advanced Materials, Ministry of Education, and the Large Scale Simulation Research Laboratory (LSR)/NECTEC are also gratefully acknowledged for research facilities and computing resources. Thanks are also due to Andrew Olson for reading the manuscript.

References

1. Tantillo C, Ding J, Jacobo-Molina A, Nanni RG, Boyer PL, Hughes SH, Pauwels R, Andries K, Janssen PA, Arnold E (1994) *J Mol Biol* 243:369
2. De Clercq E (1995) *Clin Microbiol Rev* 8:200
3. Menéndez-Arias L (2002) *Trends Pharmacol Sci* 23:381
4. Ludovici DW, Kavash RW, Kukla MJ, Ho CY, Ye H, De Corte BL, Andries K, De Béthune MP, Azijn H, Pauwels R, Moereels HEL, Heeres J, Koymans LMH, De Jonge MR, Van Aken KJA, Daeyaert FFD, Lewi PJ, Das K, Arnold E, Janssen PAJ (2001) *Bioorg Med Chem Lett* 11:2229
5. Ludovici DW, De Corte BL, Kukla MJ, Ye H, Ho CY, Lichtenstein MA, Kavash RW, Andries K, De Béthune MP, Azijn H, Pauwels R, Lewi PJ, Heeres J, Koymans LMH, De Jonge MR, Van Aken KJA, Daeyaert FFD, Das K, Arnold E, Janssen PAJ (2001) *Bioorg Med Chem Lett* 11:2235
6. Andries K, Azijn H, Thielemans T, Ludovici D, Kukla M, Heeres J, Janssen P, De Corte B, Vingerhoets J, Pauwels R, De Béthune MP (2004) *Antimicrob Agents Chemother* 48:4680
7. Ragno R, Coluccia A, La Regina G, De Martino G, Piscitelli F, Lavecchia A, Novellino E, Bergamini A, Ciapri C, Sinistro A, Maga G, Crespan E, Artico M, Silvestri RJ (2006) *J Med Chem* 49:3172
8. Mordant C, Schmitt B, Pasquier E, Demestre C, Queguiner L, Masungi C, Peeters A, Smeulders L, Bettens E, Hertogs K, Heeres J, Lewi P, Guillemont J (2007) *Eur J Med Chem* 42:567
9. Heeres J, De Jonge MR, Koymans LMH, Daeyaert FFD, Vinkers M, Van Aken KJA, Arnold E, Das K, Kilonda A, Hoornaert GJ, Compennolle F, Cegla M, Azzam RA, Andries K, De Béthune M, Azijn H, Pauwels R, Lewi PJ, Janssen PAJ (2005) *J Med Chem* 48:1910

10. Cramer RD III, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959
11. Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130
12. Barreca ML, Carotti A, Carrieri A, Chimirri A, Monforte AM, Calace MP, Rao A (1999) *Bioorg Med Chem* 7:2283
13. Sciabola S, Carosati E, Baroni M, Mannhold R (2005) *J Med Chem* 48:3756
14. Pungpo P, Saparpakorn P, Wolschann P, Hannongbua S (2006) *SAR QSAR Environ Res* 17:353
15. Zhou Z, Madura JD (2004) *J Chem Inf Comput Sci* 44:2167
16. Medina-Franco JL, Rodríguez-Morales S, Juárez-Gordiano C, Hernández-Campos A, Castillo RJ (2004) *J Comput Aided Mol Des* 18:345
17. AbdulHameed MDM, Hamza A, Liu J, Zhan C (2008) *J Chem Inf Model* 48:1760
18. Frisch MJ et al (2003) Gaussian03. Gaussian, Inc., Pittsburgh
19. SYBYL 7.0 TRIPOS, Assoc., Inc., St. Louis, MO
20. Das K, Clark AD Jr, Lewi PJ, Heeres J, De Jonge MR, Koymans LM, Vinkers HM, Daeyaert F, Ludovici DW, Kukla MJ, De Corte B, Kavash RW, Ho CY, Ye H, Lichtenstein MA, Andries K, Pauwles R, Debethune MP, Boyer PL, Clark P, Hughes SH, Janssen PA, Arnold E (2004) *J Med Chem* 47:2550
21. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) *J Comput Chem* 19:1639

Dipyridodiazepinone derivatives; synthesis and anti HIV-1 activity

Nisachon Khunnawutmanotham¹, Nitirat Chimnoi¹, Arunee Thitithanyanont², Patchreenart Saparpakorn³, Kiattawee Choowongkamon⁴, Pornpan Pungpo⁵, Supa Hannongbua³ and Supanna Techasakul^{*,1,3}

Full Research Paper

Open Access

Address:

¹Chulabhorn Research Institute, Vibhavadee-Rangsit Highway, Bangkok 10210, Thailand, ²Department of Microbiology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand, ³Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10903, Thailand, ⁴Department of Biochemistry, Faculty of Science, Kasetsart University, Bangkok 10903, Thailand and ⁵Department of Chemistry, Faculty of Science, Ubonratchathani University, Ubonratchathani 34190, Thailand

Email:

Supanna Techasakul* - supanna@cri.or.th

* Corresponding author

Keywords:

AIDS; anti HIV-1 RT; dipyridodiazepinone; nevirapine; synthesis

Beilstein Journal of Organic Chemistry 2009, 5, No. 36.

doi:10.3762/bjoc.5.36

Received: 09 April 2009

Accepted: 03 July 2009

Published: 22 July 2009

Editor-in-Chief: J. Clayden

© 2009 Khunnawutmanotham et al; licensee Beilstein-Institut.

License and terms: see end of document.

Abstract

Ten dipyridodiazepinone derivatives were synthesized and evaluated for their anti HIV-1 reverse transcriptase activity against wild-type and mutant type enzymes, K103N and Y181C. Two of them were found to be promising inhibitors for HIV-1 RT.

Introduction

Dipyridodiazepinone nevirapine (**1**) [1] (Figure 1) is a potent non-nucleoside inhibitor of human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT) and is approved as a therapeutic agent for the treatment of AIDS. In the clinic, nevirapine monotherapy results in relatively rapid drug resistance due to mutation of the RT enzyme. To develop a second-generation inhibitor with improved activity against the mutant RT enzyme, many efforts have been focused on the synthesis of dipyridodiazepinone derivatives [2-8]. On the basis of molecular modeling analysis on the wild-type (WT) and Y181C HIV-1 RT, it was found that the dipyridodiazepinone derivatives containing an unsubstituted lactam nitrogen and a 2-chloro-

8-arylthiomethyl moiety, when compared with **9** [4] (Figure 2) as reference, are effective inhibitors of this mutant enzyme. Some dipyridodiazepinone derivatives containing an *N*-methyl-

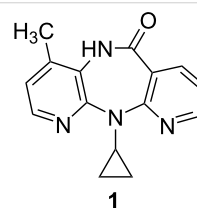
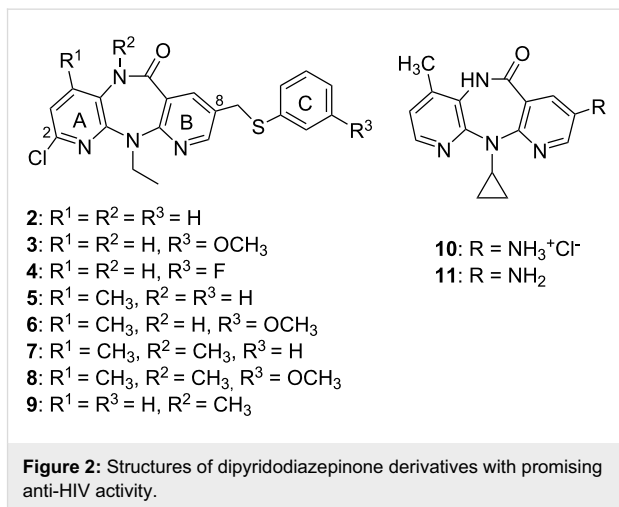


Figure 1: Structure of nevirapine (**1**).

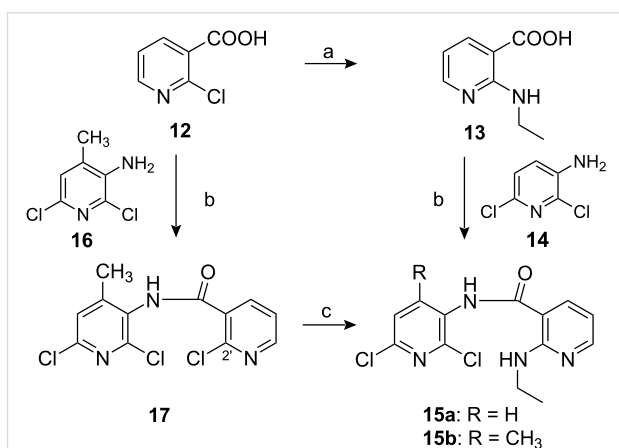


ated of lactam also exhibited good potency against the WT enzyme. The 8-amino derivative of nevirapine and its hydrochloride salt also provided interesting potency. The first two compounds, **2** and **3**, were synthesized and their virustatic and virucidal activities against HIV-1 subtype E were reported previously [9]. As part of our continuing efforts directed towards the development of potential HIV-1 RT inhibitors, we have extended the synthesis of dipyrindodiazepinone derivatives (Figure 2) and evaluation of their activity against wild-type RT and K103N and Y181C mutant RT enzymes.

Results and Discussion

Synthesis of compounds 2–9

Compounds **2–9** were synthesized *via* efficient routes as shown in Scheme 1 and Scheme 2. The aminopyridinecarboxamide **15a** ($R = H$) was prepared from 2-(ethylamino)-3-pyridinecarboxylic acid (**13**) and 3-amino-2,6-dichloropyridine (**14**)



Scheme 1: Reagents and conditions: (a) $EtNH_2$, 120 °C, 4 h, 99% (b) i) $(COCl)_2$, benzene, DMF, rt, 1 h; ii) amine **14** or **16**, dioxane, cyclohexane, pyridine, rt, 16 h, 80% (c) $EtNH_2$, dioxane, 100 °C, 20 min, 47%.

(Scheme 1) [9]. However, by using the same procedure to prepare aminopyridinecarboxamide **15b** ($R = CH_3$), only poor yields of **15b** were obtained. Therefore, **15b** ($R = CH_3$) was prepared by formation of pyridinecarboxamide **17**, obtained from **12** and **16** [10,11], followed by the displacement of 2'-chloro by the ethylamino group.

Afterwards, the aminopyridinecarboxamide intermediates **15** were treated as previously reported [4] to give 8-arylthiomethyl-dipyrindodiazepinones, as shown in Scheme 2.

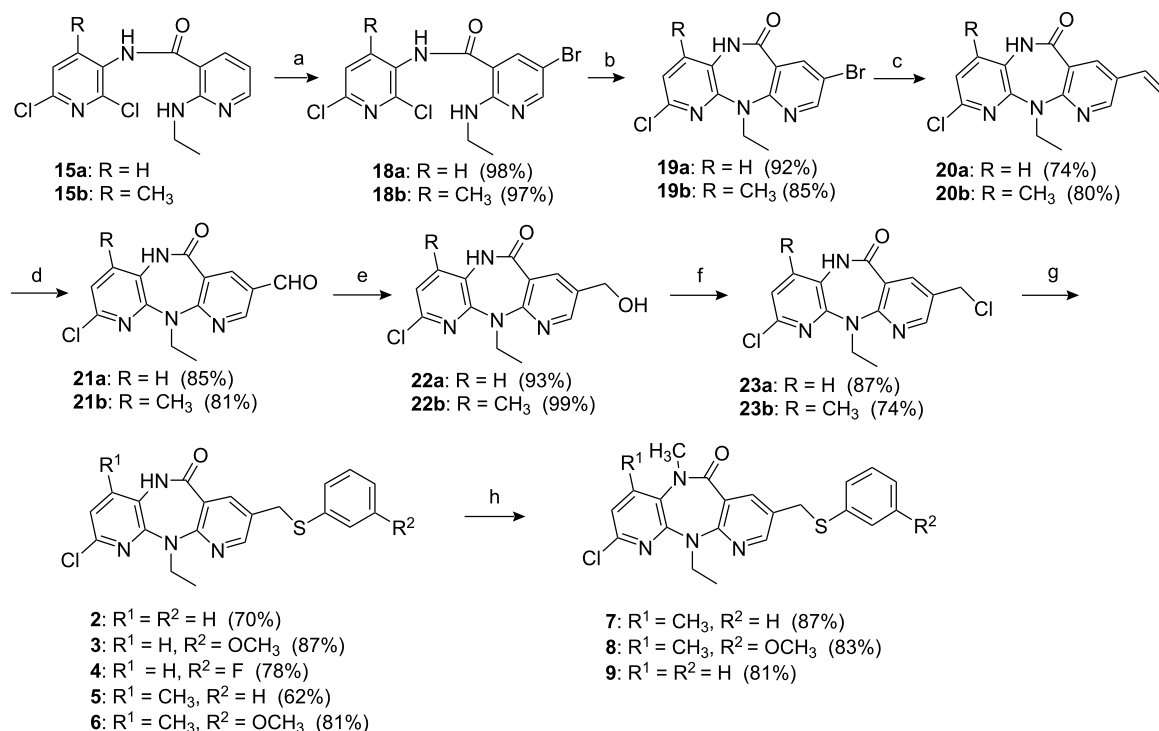
Aminopyridinecarboxamides **15** were regioselectively brominated to produce bromo compounds **18**. The diazepinone ring was formed by treatment with sodium hexamethyldisilazane in pyridine to yield tricyclic compounds **19**. Coupling of **19** with vinyltributyltin in the presence of tetrakis(triphenylphosphine) palladium(0) provided 8-vinyl compounds **20** which underwent ozonolysis to produce aldehydes **21** in good yields. The reduction of **21** with $NaBH_4$ produced alcohols **22**, which were converted to the corresponding chlorides **23** through treatment with thionyl chloride in dichloromethane. The reaction of **23a** with thiophenolate, 3-methoxythiophenolate, and 3-fluorothiophenolate in *N,N*-dimethylformamide yielded **2**, **3**, and **4**, respectively, whilst the reaction of **23b** with thiophenolate and 3-methoxythiophenolate yielded **5** and **6**, respectively. Methylation of the lactam nitrogen of **5** and **6** with methyl iodide provided **7** and **8**. Compound **9** was also prepared *via* methylation of **2**, which was used as the reference compound.

Synthesis of compounds 10 and 11

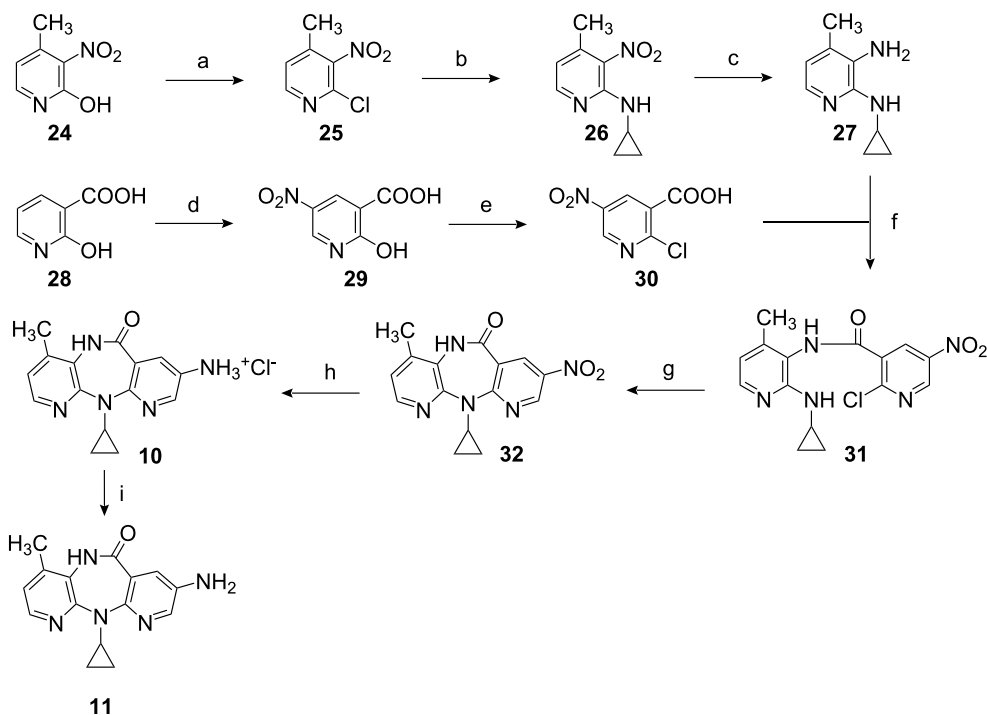
Compounds **10** and **11** were synthesized as shown in Scheme 3. The starting 3-amino-2-cyclopropylamino-4-methylpyridine (**27**) was prepared from commercially available 2-hydroxy-4-methyl-3-nitropyridine (**24**) through a sequence involving treatment with $POCl_3$, followed by chloro displacement from the resulting 2-chloro compound with the aminocyclopropyl group, and finally reducing the nitro to the amino group. 2-Chloro-5-nitronicotinic acid (**30**) was prepared by nitration of commercially available 2-hydroxynicotinic acid (**28**) followed by treatment with $POCl_3$. Then amine **27** and the acid **30** underwent coupling to produce carboxamide **31**. Diazepinone ring closure was performed by heating **31** in hexamethyldisilazane. Afterwards, the nitro group was reduced to produce the hydrochloride salt **10**. Treatment of **10** with 50% aqueous NaOH yielded its corresponding free amino compound **11**.

Biological testing against HIV-1 reverse transcriptase

The results from the biological testing of all compounds synthesized, compared with nevirapine (**1**) and **9**, against the wild-type RT together with K103N and Y181C mutant RT are



Scheme 2: Reagents and conditions: (a) Br₂, HOAc, KOAc, rt, 1 h; (b) NaHMDS, pyridine, 90 °C, 1 h; (c) CH₂=CH-SnBu₃, Pd(PPh₃)₄, DMF, 90 °C, 1 h; (d) O₃, CH₂Cl₂/MeOH, -78 °C then PPh₃, rt, 1 h; (e) NaBH₄, THF, H₂O, rt, 0.5 h; (f) SOCl₂, CH₂Cl₂, Et₃N, rt; (g) NaH, ArSH, DMF, rt, 1 h; (h) NaH, DMF, 50 °C, 0.5 h then MeI, rt, 0.5 h.



Scheme 3: Reagents and conditions: (a) POCl₃, 150 °C, 6 h, 85%; (b) cyclopropylamine, xylene, 105 °C, 4 h, 99%; (c) SnCl₂·2H₂O, conc. HCl, CH₃COOH, rt, 3 h, 83%; (d) 69% HNO₃, conc. H₂SO₄, 50 °C, 7 h, 79%; (e) POCl₃, reflux, 4 h, 78%; (f) i) **30**, (COCl)₂, benzene, DMF, rt, 1 h; ii) **27**, THF, DIPEA, rt, 5 h, 53%; (g) HMDS, 110 °C, 24 h, 90%; (h) SnCl₂·2H₂O, conc. HCl, CH₃COOH, rt, 3 h, 73%; (i) 50% aq. NaOH, rt, 1 h, 90%.

shown in Table 1. It was found that compounds **2–8** exhibited higher inhibitory activity against WT-RT and both mutant RTs compared to nevirapine. Interestingly, **5** and **6** were found to be about four times more potent against WT-RT than **9**, and they provided comparable activity against K103N mutant RT. Incorporation of a methyl group at the R¹ position and the free N of the amide seems to be responsible for this higher activity. However, **9** showed better potency against the Y181C mutant RT compared to the other two compounds. Compound **5**, without methoxy substituent, was found to be slightly more potent than **6** except for Y181C mutant RT. Additional *N*-methyl groups in **7** and **8** led to diminished activity relative to that of **5** and **6**. **10** and **11**, 8-amino analogues of nevirapine, were found to be ineffective inhibitors.

Table 1: Inhibitory activity of the synthesized compounds against HIV-1 reverse transcriptase.

Compounds	IC ₅₀ ^a (μM)		
	WT	K103N	Y181C
nevirapine (1)	1.070 ± 0.60 ^b	27.10 ± 5.20	228.5 ± 24.84
2	0.427 ± 0.31	6.23 ± 2.48	1.50 ± 0.34
3	0.757 ± 0.15	19.40 ± 2.80	2.90 ± 0.19
4	0.183 ± 0.10	13.90 ± 1.23	0.459 ± 0.29
5	0.0186 ± 0.002	0.224 ± 0.14	0.269 ± 0.08
6	0.0229 ± 0.01	0.428 ± 0.39	0.0593 ± 0.07
7	0.124 ± 0.03	4.37 ± 0.66	0.507 ± 0.36
8	0.0828 ± 0.03	4.59 ± 2.23	0.118 ± 0.06
9	0.0858 ± 0.00001	0.39 ± 0.23	0.00463 ± 0.0009
10	17.40 ± 2.11	62.10 ± 5.14	126.0 ± 28.03
11	6.05 ± 1.60	97.0 ± 16.8	61.0 ± 6.9

^aIC₅₀ is the concentration of inhibitor required for 50% inhibition of reverse transcriptase. ^bStandard errors obtained from duplicate experiments.

Molecular docking

To understand the binding mode of the new potent derivatives **5**, **6** and **9** were docked into the HIV-1 RT binding site by using the default parameters of the GOLD v3.2 program. The wild type HIV-1 RT structure (pdb code 1klm) was taken from the protein data bank. Additionally, two HIV-1 RT mutants, K103N and Y181C, were used and analyzed by mutating positions 103 and 181 of the wild-type structure through the use of the Sybyl 7.2 program. The docked conformations of **5**, **6**, **9** and nevirapine are shown in Figure 3, and their GoldScores are presented in Table 2. In the wild-type and K103N binding sites, the docked orientations of **5**, **6** and **9** are similar to that of nevirapine. In the Y181C binding site, except for **5**, the orientations of the others were similar to nevirapine orientation. The

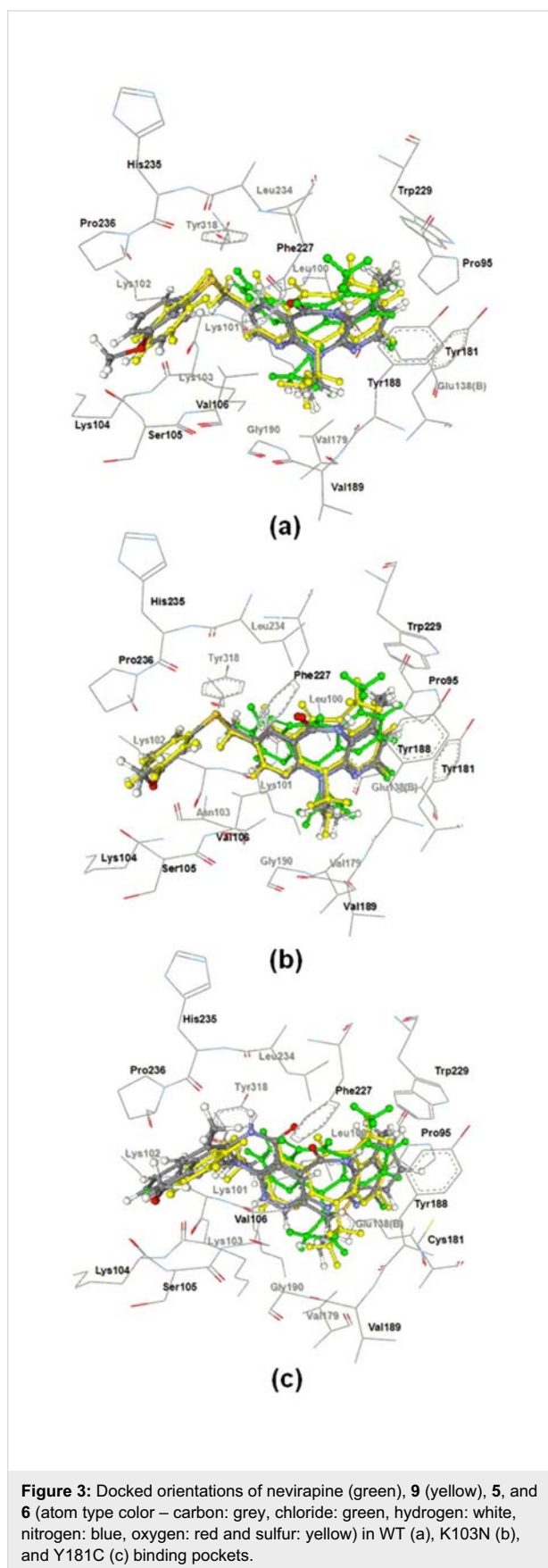


Figure 3: Docked orientations of nevirapine (green), **9** (yellow), **5**, and **6** (atom type color – carbon: grey, chloride: green, hydrogen: white, nitrogen: blue, oxygen: red and sulfur: yellow) in WT (a), K103N (b), and Y181C (c) binding pockets.

Table 2: GoldScores of nevirapine, **5**, **6** and **9** in wild-type (WT), K103N and Y181C HIV-1 RT.

Compounds	GoldScores		
	WT	K103N	Y181C
nevirapine	58.13 ($\pm$ 0.27) ^a	58.66 ($\pm$ 0.39)	56.92 ($\pm$ 0.04)
5	72.07 ($\pm$ 1.60)	79.23 ($\pm$ 1.42)	66.44 ($\pm$ 0.63)
6	79.19 ($\pm$ 0.83)	83.22 ($\pm$ 0.85)	71.92 ($\pm$ 1.69)
9	73.66 ($\pm$ 2.36)	76.27 ($\pm$ 1.00)	68.84 ($\pm$ 1.09)

^aIn parenthesis is the standard error of the GoldScore from the triplicate of docking calculations.

GoldScores of **5**, **6** and **9** were higher than those of nevirapine by 17.61–24.56, 9.52–15.00 and 15.33–21.06 in the WT, K103N, and Y181C binding pockets, respectively. In the wild-type binding pocket, the H-bond interaction with the backbone oxygen atom of Lys103 was found to contain **5**, **6** and **9** but no nevirapine was present. Compounds **5**, **6** and **9** also formed an H-bond interaction with the backbone nitrogen atom of Val106. Since there is a methyl group at the R¹ position of **5** and **6**, their docked conformations were slightly shifted below the binding pocket as compared to the docked conformation of **9**. This shift caused the formation of a stronger H-bond interaction of **5** and **6** with Lys101, Val179, Tyr188 and Val189 compared to **9**. The methyl group at R¹ position of **5** and **6** also formed stronger H– π interaction with Trp229. Moreover, the methoxy substituent of **6** revealed a strong attractive interaction with Lys104, but the movement of ring C in **6** caused a steric interaction with the side chain of Lys102.

In the case of the docked conformations of **5**, **6** and **9** in the K103N binding pocket, the H-bond interactions with the backbone atom of Asn103 were still detected, but their H-bond interactions with Val106 were lost. Furthermore, the docked conformation of **5** showed stronger H-bond interaction with Lys101 compared to **6** and **9**. The adjustment of the ethyl group also formed an important H-bond interaction with the oxygen atom of carbonyl Val179 at distances of 2.49, 2.74 and 2.68 Å for **5**, **6** and **9** respectively. The methyl group at the R¹ position of **5** and **6** presented the H– π interaction with side chain Trp229 closer than the hydrogen atom of **9**.

In the Y181C binding pocket, it was found that the docked conformation of **5** had a different orientation compared to **6** and **9**. Due to this orientation change of **5**, some attractive interactions found in the wild-type binding pocket were lost. It was observed that the docked conformations of **6** and **9** were aligned in the same orientation as nevirapine. For **9**, stronger attractive interactions with the backbone oxygen atoms of Val179 and

Tyr188 in the Y181C binding pocket were formed as compared with the WT binding pocket. For **6**, H-bond interaction between a hydrogen atom of the methoxy group of **6** and an oxygen atom of backbone Lys104 revealed a longer bond length (3.58 Å) in the Y181C binding pocket as compared to the WT binding pocket (2.14 Å).

Conclusion

The remarkable anti HIV-1 activity of dipyrrodozepinone derivatives, particularly compounds **5** and **6**, was presented in this study. A preliminary SAR study showed that the methyl group at the R¹ position and the free N of amide are crucial for potent activity. This is possibly because of the strong interaction with the amino acid residue in the RT enzyme. The secondary test of these two compounds was regarded to be valuable for future investigation.

Supporting Information

Supporting information provides details about the chemical methods, analytical data and biological testing.

Supporting Information File 1

Experimental part.

[<http://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-5-36-S1.doc>]

Acknowledgments

We would like to thank the Thailand Research Fund (DBG4780009 for S.T., RTA5080005 for S.H., DBG5180022 for P.P., and MRG5080267 for P.S.), National Synchrotron Research Center for K.C., and the Ministry of Education for their financial support. We also gratefully acknowledge the help of Dr. Carrie Dykes of the University of Rochester, New York for the mutant clones, K103N and Y181C; and Dr. Hiroaki Mitsuya of the HIV and AIDS Malignancy Branch, USA for the WT clone.

References

- Hargrave, K. D.; Proudfoot, J. R.; Grozinger, K. G.; Cullen, E.; Kapadia, S. R.; Patel, U. R.; Fuchs, V. U.; Mauldin, S. C.; Vitous, J.; Behnke, M. L.; Klunder, J. M.; Pal, K.; Skiles, J. W.; McNeil, D. W.; Rose, J. M.; Chow, G. C.; Skoog, M. T.; Wu, J. C.; Schmidt, G.; Engel, W. W.; Eberlein, W. G.; Saboe, T. D.; Campbell, S. J.; Rosenthal, A. S.; Adams, J. *J. Med. Chem.* **1991**, *34*, 2231–2241. doi:10.1021/jm00111a045
- O'Meara, J. A.; Yoakim, C.; Bonneau, P. R.; Bös, M.; Cordingley, M. G.; Déziel, R.; Doyon, L.; Duan, J.; Garneau, M.; Guse, I.; Landry, S.; Malenfant, E.; Naud, J.; Ogilvie, W. W.; Thavonekham, B.; Simoneau, B. *J. Med. Chem.* **2005**, *48*, 5580–5588. doi:10.1021/jm050255t

3. Yoakim, C.; Bonneau, P. R.; Déziel, R.; Doyon, L.; Duan, J.; Guse, I.; Landry, S.; Malenfant, E.; Naud, J.; Ogilvie, W. W.; O'Meara, J. A.; Plante, R.; Simoneau, B.; Thavonekham, B.; Bös, M.; Cordingley, M. G. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 739–742. doi:10.1016/j.bmcl.2003.11.049
4. Cywin, C. L.; Klunder, J. M.; Hoermann, M.; Brickwood, J. R.; David, E.; Grob, P. M.; Schwartz, R.; Pauletti, D.; Barringer, K. J.; Shih, C.-K.; Sorge, C. L.; Erickson, D. A.; Joseph, D. P.; Hattox, S. E. *J. Med. Chem.* **1998**, *41*, 2972–2984. doi:10.1021/jm9707030
5. Klunder, J. M.; Hoermann, M.; Cywin, C. L.; David, E.; Brickwood, J. R.; Schwartz, R.; Barringer, K. J.; Pauletti, D.; Shih, C.-K.; Erickson, D. A.; Sorge, C. L.; Joseph, D. P.; Hattox, S. E.; Adams, J.; Grob, P. M. *J. Med. Chem.* **1998**, *41*, 2960–2971. doi:10.1021/jm9707028
6. Kelly, T. A.; McNeil, D. W.; Rose, J. M.; David, E.; Shih, C.-K.; Grob, P. M. *J. Med. Chem.* **1997**, *40*, 2430–2433. doi:10.1021/jm960837y
7. Kelly, T. A.; Proudfoot, J. R.; McNeil, D. W.; Patel, U. R.; David, E.; Hargrave, K. D.; Grob, P. M.; Cardozo, M.; Agarwal, A.; Adams, J. *J. Med. Chem.* **1995**, *38*, 4839–4847. doi:10.1021/jm00024a011
8. Proudfoot, J. R.; Hargrave, K. D.; Kapadia, S. R.; Patel, U. R.; Grozinger, K. G.; McNeil, D. W.; Cullen, E.; Cardozo, M.; Tong, L.; Kelly, T. A.; Rose, J. M.; David, E.; Mauldin, S. C.; Fuchs, V. U.; Vitous, J.; Hoermann, M.; Klunder, J. M.; Raghavan, P.; Skiles, J. W.; Mui, P.; Richman, D. D.; Sullivan, J. L.; Shih, C.-K.; Grob, P. M.; Adams, J. *J. Med. Chem.* **1995**, *38*, 4830–4838. doi:10.1021/jm00024a010
9. Khunnawutmanotham, N.; Chimnoi, N.; Saparpakorn, P.; Pungpo, P.; Louisirirotchanaikul, S.; Hannongbua, S.; Techasakul, S. *Molecules* **2007**, *12*, 218–230. doi:10.3390/12020218
10. Grozinger, K. G.; Fuchs, V. U.; Hargrave, K. D.; Mauldin, S.; Vitous, J.; Campbell, S.; Adams, J. *J. Heterocycl. Chem.* **1995**, *32*, 259–263. doi:10.1002/jhet.5570320144
11. Bobbitt, J. M.; Scola, D. *J. Org. Chem.* **1960**, *25*, 560–564. doi:10.1021/jo01074a018

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the *Beilstein Journal of Organic Chemistry* terms and conditions: (<http://www.beilstein-journals.org/bjoc>)

The definitive version of this article is the electronic one which can be found at:
[doi:10.3762/bjoc.5.36](https://doi.org/10.3762/bjoc.5.36)

3D Pharmacophore Mapping Using 4D QSAR Analysis for the Cytotoxicity of Lamellarins Against Human Hormone-Dependent T47D Breast Cancer Cells

Poonsiri Thipnate,^{†,‡} Jianzhong Liu,^{§,||} Supa Hannongbua,^{*,†,‡} and A. J. Hopfinger^{§,||}

Department of Chemistry, Faculty of Science and Center of Nanotechnology, Kasetsart University, Chatuchak, Bangkok 10900, Thailand, College of Pharmacy, MSC09 5360, 1 University of New Mexico, Albuquerque, New Mexico 87131-000, and The Chem21 Group, Incorporated, 1780 Wilson Drive, Lake Forest, Illinois 60045

Received July 10, 2009

4D quantitative structure–activity relationship (QSAR) and 3D pharmacophore models were built and investigated for cytotoxicity using a training set of 25 lamellarins against human hormone dependent T47D breast cancer cells. Receptor-independent (RI) 4D QSAR models were first constructed from the exploration of eight possible receptor-binding alignments for the entire training set. Since the training set is small (25 compounds), the generality of the 4D QSAR paradigm was then exploited to devise a strategy to maximize the extraction of binding information from the training set and to also permit virtual screening of diverse lamellarin chemistry. 4D QSAR models were sought for only six of the most potent lamellarins of the training set as well as another subset composed of lamellarins with constrained ranges in molecular weight and lipophilicity. This overall modeling strategy has permitted maximizing 3D pharmacophore information from this small set of structurally complex lamellarins that can be used to drive future analog synthesis and the selection of alternate scaffolds. Overall, it was found that the formation of an intermolecular hydrogen bond and the hydrophobic interactions for substituents on the E ring most modulate the cytotoxicity against T47D breast cancer cells. Hydrophobic substitutions on the F-ring can also enhance cytotoxic potency. A complementary high-throughput virtual screen to the 3D pharmacophore models, a 4D fingerprint QSAR model, was constructed using absolute molecular similarity. This 4D fingerprint virtual high-throughput screen permits a larger range of chemistry diversity to be assayed than with the 4D QSAR models. The optimized 4D QSAR 3D pharmacophore model has a leave-one-out cross-correlation value of $xv - r^2 = 0.947$, while the optimized 4D fingerprint virtual screening model has a value of $xv - r^2 = 0.719$. This work reveals that it is possible to develop significant QSAR, 3D pharmacophore, and virtual screening models for a small set of lamellarins showing cytotoxic behavior in breast cancer screens that can guide future drug development based upon lamellarin chemistry.

INTRODUCTION

In 1985 the first four lamellarins, A–D were isolated from the marine prosobranch mollusk, *Lamellaria* sp. and their structures, and determined by X-ray crystallographic and ¹H NMR studies.¹ A family of more than 30 lamellarins which consist of three structural groups, such as an saturated D-ring fused (Figure 1a), a unsaturated D-ring fused (Figure 1b), and an unfused central pyrrole ring group (Figure 1c), have been isolated and investigated in terms of their biological activity profiles.^{2,3} These compounds, especially the fused central pyrrole ring lamellarins (Tables 1 and 2), have been found to be cytotoxic to a wide range of cancer cell lines. Lamellarins C and U demonstrate potent cytotoxicity against 10 human tumor cell lines (A549, HCT-116, LOX IMVI, MALME-3M, MCF-7, MOLT-4, OVCAR-3, PC-3, SF-295, UO-31),³ lamellarin D shows potent cytotoxic activity against human prostate cancer cells (DU-145, LNCaP) and leukemia

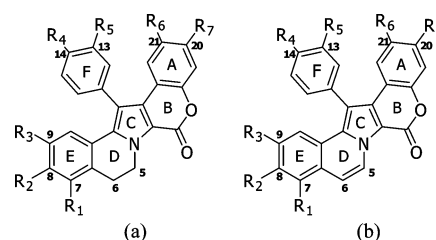


Figure 1. The scaffold groups forming the training set of lamellarins.

cells (K562),³ and lamellarins I, K, and L exhibit significant cytotoxicity against P388 and A549 cultured cancer cell lines.³ Additionally, lamellarin I and D have an effective cytotoxic activity against multidrug resistant reversal (MDR) cell lines by inhibiting P-glycoprotein (P-GP) mediated drug efflux.^{4,5} Some lamellarins have also been demonstrated to act on cancer cell mitochondria to induce apoptosis.^{6,7} Moreover, lamellarin D is an effective stabilizer of human topoisomerase I-DNA covalent complexes and, thus, is capable of stimulating DNA cleavage.^{2,8,9} Based on its biological actions, lamellarin D was identified as a novel lead candidate by Bailly and co-workers.^{2,3,6–11}

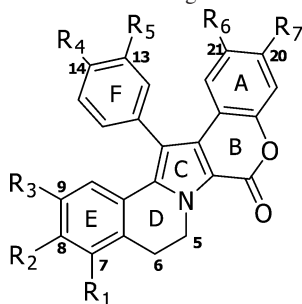
* Corresponding author. Telephone: +66 2 562 5555. Fax: +66 2 579 3955. E-mail: fscisph@ku.ac.th.

[†] Department of Chemistry, Faculty of Science, Kasetsart University.

[‡] Center of Nanotechnology, Kasetsart University.

[§] University of New Mexico.

^{||} The Chem21 Group, Incorporated.

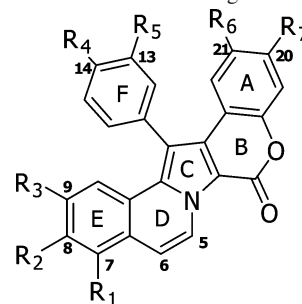
Table 1. Chemical Structures and Cytotoxic Activities, $-\log IC_{50}$, of Lamellarins with a Saturated D-Ring

lamellarin	MW	log P^a	substituent group							$-\log IC_{50}$
			R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	
C	545.6	3.4	OMe	OMe	OMe	OH	OMe	OMe	OH	5.11
E	531.5	3.4	OH	OMe	OMe	OMe	OH	OMe	OH	5.28
F	545.2	3.4	OH	OMe	OMe	OMe	OMe	OMe	OH	5.34
G	501.5	3.6	H	OH	OMe	OMe	OH	OH	OMe	5.07
I	559.6	3.4	OMe	OMe	OMe	OMe	OMe	OMe	OH	5.02
J	515.5	3.7	H	OH	OMe	OMe	OMe	OMe	OH	4.89
K ^b	531.51	3.7	OH	OMe	OMe	OH	OMe	OMe	OH	7.04
L	501.5	3.6	H	OH	OMe	OMe	OH	OMe	OH	5.36
T	545.6	3.4	OMe	OMe	OMe	OMe	OH	OMe	OH	4.88
U	515.5	3.7	H	OMe	OMe	OMe	OH	OMe	OH	4.99
Y	501.1	3.6	H	OMe	OH	OMe	OH	OMe	OH	5.14
χ	501.5	3.6	H	OH	OMe	OH	OMe	OMe	OH	5.42
χ triacetate	657.2	2.7	OAc	OMe	OMe	OAc	OMe	OMe	OAc	5.18
U diacetate	599.2	3.2	H	OMe	OMe	OMe	OAc	OMe	OAc	5.10
χ triacetate	627.2	3.0	H	OAc	OMe	OAc	OMe	OMe	OAc	5.54

^a Calculated by CS ChemDraw Ultra, version 5.0 (CambridgeSoft Corporation, Cambridge, MA). ^b This compound was not included in the original 25 lamellarin compounds training set used to build the 4D QSAR and 3D pharmacophore models. However, it shows unexpected high activity for the saturated D-ring series of compounds and, thus, is investigated as part of this study.

The first reported study on SARs of lamellarin was done by Ishibashi et al. in 2002.¹² It was reported that the hydroxyl groups at position C-8 and C-20 were important for cytotoxicity against a HeLa cell line, while the hydroxyl group at C-14 and the two methoxy groups at C-13 and C-21 were not essential for activity. The C5=C6 double bond in the D-ring or the planarity of the chromophore is necessary for activity.^{5,8,12} In more recent findings, Chittchang et al. not only substantiated the significant contributions of the C5=C6 olefin moiety, as well as the hydroxyl groups at C8 and C20, but also demonstrated the importance of the C-7 hydroxy group for the first time.¹³ These findings were also substantiated by carrying out three-dimensional quantitative structure–activity relationship (3D QSAR) analyses this past year.¹⁴

Treatment of the lamellarins data set is representative of a class of real world problems in drug discovery; namely, how to optimize the extraction of SAR information for, in turn, optimizing lead development efforts from a small number of structurally complex, hard to synthesize compounds that have been tested and observed to exhibit a wide-range of end point activity. Modeling such small data sets can be criticized on the basis of too little data to generate reliable and useful results to drive lead optimization. Yet doing nothing with the information resident in such data sets obviously contributes nothing to streamlining lead development efforts. A key to resolving this dilemma may reside in the type and level of sophistication of the modeling

Table 2. Chemical Structures and Cytotoxic Activities, $-\log IC_{50}$, of Lamellarins with an Unsaturated D-Ring

lamellarin	MW	log P^a	substituent group							$-\log IC_{50}$
			R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	
B	543.5	3.3	OMe	OMe	OMe	OH	OMe	OMe	OH	6.74
D	499.5	3.6	H	OH	OMe	OH	OMe	OMe	OH	10.10
M	529.5	3.3	OH	OMe	OMe	OH	OMe	OMe	OH	8.02
N	499.5	3.6	H	OH	OMe	OMe	OH	OMe	OH	9.22
W	543.5	3.3	OMe	OMe	OMe	OMe	OH	OMe	OH	5.37
α	513.5	3.6	H	OMe	OMe	OMe	OH	OMe	OH	6.23
X	529.5	3.3	OH	OMe	OMe	OMe	OH	OMe	OH	8.25
ε	543.5	3.3	OH	OMe	OMe	OMe	OMe	OMe	OH	8.26
ζ	557.6	3.4	OMe	OMe	OMe	OMe	OMe	OMe	OH	7.05
dehydro. J	513.5	3.6	H	OH	OMe	OMe	OMe	OMe	OH	10.01
dehydro. Y	499.5	3.6	H	OMe	OH	OMe	OH	OMe	OH	7.10

^a Calculated by CS ChemDraw Ultra, version 5.0 (CambridgeSoft Corporation, Cambridge, MA).

employed. A high-level modeling approach wherein detailed structural, thermodynamic, and electronic information about each complex compound of the data set may negate some of the drawbacks to the small size of the data set. In a sense, ‘quality’ is used to compensate for ‘quantity’.

Multiple complementary applications of the 4D QSAR paradigm¹⁵ may be a good way to extend our knowledge and understanding of the SARs of lamellarins using this ‘quality’ for ‘quantity’ argument. The fourth ‘dimension’ of the 4D QSAR paradigm is ensemble sampling the spatial features of the members of a training set.¹⁵ This sampling process, in turn, allows the construction of optimized dynamic spatial QSAR models in the form of 3D pharmacophores, which are dependent on conformation, alignment, and pharmacophore grouping.¹⁶ This method has been proven both useful and reliable for the construction of quantitative 3D pharmacophore models, especially for sets of flexible ligand analogues when the geometry of the corresponding receptor is not known.^{15–17}

Complementary to building 4D QSAR models that embed 3D pharmacophores is the construction of high-throughput 4D fingerprint models for virtual screening. The 4D fingerprints can be derived independently of any molecular alignment and are based upon an inductive approach to establish 4D molecular similarity measures across any collection of chemical compounds.¹⁸ The 4D QSAR paradigm has been successfully applied for a variety of chemical classes and biological end points including glucose, flavonoid, and propofol analogs, AHPBA and THP inhibitors of HIV-1 protease,¹⁸ human serum albumin (HSA),^{19,20} local lymph node assay (LLNA) database,^{21,22} skin penetration enhancers,²³ and HIV-1 integrase inhibitors.²⁴

The search for SARs and/or pharmacophores for natural and synthetic, lamellarins screened for cytotoxic activity

against 11 cancer cell lines is ongoing. However, the prominent activities observed in a few of the lamellarins screened against human hormone-dependent T47D breast cancer cells seemed to us to be best explored along with the corresponding SARs delineated and exploited by using the 4D QSAR methodology for the reasons cited above.

MATERIALS AND METHODS

Lamellarin Data Set and Cytotoxic Activity. Twenty six lamellarins were analyzed in this work. The chemical structures of all 26 lamellarins are given in Tables 1 and 2. These compounds were synthesized and purified by Ploypradith et al.²⁵ The cytotoxic activity ($-\log IC_{50}$) against human hormone-dependent T47D breast cancer cells have been recently reported and are included as parts of Tables 1 and 2.¹³

Receptor-Independent (RI) 4D QSAR Analysis Applied to the Lamellarin Data Set. Since the geometry of the receptor is not available in this study, the receptor-independent form of 4D QSAR analysis, referred to as RI-4D QSAR, has been employed. The 10 operational steps in RI 4D QSAR have been presented in detail previously¹⁵ and also given in the 4D QSAR software version 3.0 user guide.²⁶ Therefore, these 10 steps of RI-4D QSAR analysis are only summarized here as follows:

- Step 1 An initial 3D structure of each lamellarin was constructed in the neutral form using the HyperChem 7.5 software.²⁷ Partial atomic charges were computed using the semiempirical AM1 method. Each structure was then minimized with no geometric constraint. These energy-minimized structures were used as the initial structures in the conformational ensemble sampling of Step 3.
- Step 2 Atoms of each molecule were classified into seven types of interaction pharmacophore elements (IPEs). Each type is represented by a different number code from 0 to 6 as defined in Table 3.
- Step 3 Molecular dynamics simulation (MDS) was used to sample the conformational states available to each analogue and to generate its corresponding conformational ensemble profile (CEP). The MDSs were done using the MOLSIM package²⁸ and the MM2 force field.^{29,30} The temperature for the MDS is set at 300 K with a simulation sampling time of 40 ps with intervals of 0.001 ps for a total sampling of 40 000 conformations of each lamellarin compound. The atomic coordinates of each conformation and its intramolecular energy sampled during the MDS were recorded every 0.02 ps for a total of 2 000 "frames", or steps, in constructing the CEP of each compound.
- Step 4 The set of three-ordered atoms in trial alignments are defined in Table 4. In this study, eight alignments were explored across the overall lamellarin core structure.
- Step 5 Each conformation of a compound from its CEP was aligned in the grid cell lattice using the invariant coordinates of the three-ordered atom alignment. In this study, the size of the cubic grid cells of the lattice is 1 Å³, and the overall grid cell lattice size was chosen to fully enclose each compound of the training set. The normalized occupancy of each grid cell by each IPE atom type over the CEP for a given

alignment forms a unique set of QSAR descriptors referred to as grid cell occupancy descriptors or GCODs. The GCODs were computed and used as the trial descriptor pool in 4D QSAR analysis. Non-GCOD descriptors of the training set compounds can also be included in the trial basis set (descriptor pool). In this particular study, the logarithm of the 1-octanol/water partition coefficient ($\log P$) and the compound's molecular weight (MW) were selectively added to the trial basis set descriptors in some of the model building studies. The $\log P$ and the MW values of the training set compounds are reported in Table 1.

Step 6 A 4D QSAR analysis generates an enormous number of trial GCODs because of the large number of grid cells and the seven IPEs. Partial least-squares (PLS) regression analysis³¹ is used to perform a data reduction analysis between the observed dependent variable measures and the corresponding set of GCOD values.

Step 7 The most highly weighted PLS GCOD descriptors (currently the top 200) generated in Step 6 are used to form the trial descriptor pool for genetic algorithm (GA) model optimization. The specific GA currently used in the 4D QSAR software is a modification of the genetic function approximation (GFA).³² The GFA optimization is initiated using N (currently 300) randomly generated 4D QSAR models. Mutation probability over the crossover optimization cycle is set at 10%. The smoothing factor, a GFA operations variable, which controls the number of independent variables in the QSAR models is varied in order to determine the optimal number of descriptors for the 4D QSAR models. The diagnostic measures used to analyze the resultant 4D QSAR models generated by the GFA include: (i) descriptor usage as a function of crossover operation, (ii) linear cross-correlation among descriptors and/or dependent variables (biological activity measures), (iii) number of significant and independent 4D QSAR models, and (iv) indices of model significance including the correlation coefficient, r^2 , leave-one-out (LOO), cross-validation correlation coefficient, $xv - r^2$, and Friedman's lack-of-fit (LOF).³³ In this particular 4D QSAR application, the alignment similarity comparisons were limited to models having the same number of GCODs.

Step 8 Steps 4–7 are repeated until all trial alignments are included in the 4D QSAR analyses.

Step 9 The inspection and evaluation of the population of models are obtained from the set of trial alignments in this step. The goal of this step is to identify the best and distinct set of 4D QSAR models, which is referred to as the manifold model of the analysis.

Step 10 The "active" conformation of each compound is hypothesized at this step. This conformer is achieved by identifying all conformer states sampled for each compound that are within ΔE of the global minimum energy conformation of the CEP. Currently, ΔE is set at 2 kcal/mol. Each member of the resultant set of energy-filtered conformations is then individually evaluated in the best 4D QSAR model. The conformation within 2 kcal/mol of the apparent global

Table 3. The Set of Interaction Pharmacophore Elements (IPEs) Used in the RI 4D QSAR and 4D Fingerprint QSAR Analyses

IPE description	symbol	number code
all atoms in the molecule	any	0
nonpolar atoms	np	1
polar atoms of positive partial charge	p ⁺	2
polar atoms of negative partial charge	p ⁻	3
hydrogen-bond acceptor atoms	hba	4
hydrogen-bond donor atoms	hbd	5
aromatic atoms	aro	6
non-hydrogen atoms ^a	hs	7

^a Hydrogen suppressed use only in 4D fingerprint QSAR analysis.

minimum that predicts the highest activity in the best 4D QSAR model is defined as the active conformation.

4D Fingerprint Virtual Screening Analysis Applied to the Lamellarin Data Set. The theory and the corresponding methodology of the universal 4D fingerprints for constructing the main distance-dependent matrix (MDDM) and computing corresponding eigenvalues for each matrix, using 4D molecular similarity (MS), have been presented in detail in previous works.^{18,34} The types of atoms composing a molecule are currently defined as the IPEs shown in Table 3. A unique MDDM is constructed for each of the eight distinct and identical IPE pairs. The elements of the MDDM are defined as following:

$$E(v, d_{ij}) = e^{(-v \langle d_{ij} \rangle)} \quad (1)$$

The “universal constant (v)” in eq 1, which is equal to 0.25,³⁴ has been selected such that the difference in the sum of eigenvalues for any two arbitrary compounds with the same number, n , of a particular IPE type, m , is maximized. The term $\langle d_{ij} \rangle$ is the average distance between the atom pair ij of IPE type u and v :

$$\langle d_{ij} \rangle = \sum_k d_{ij}(k) p(k) \quad (2)$$

where $p(k)$ refer to the thermodynamic probability of the k^{th} conformer state sampled in the assessment of conformational flexibility, and $d_{ij}(k)$ is the corresponding distance between atom pair i and j of IPE types u and v for the k^{th} conformer state. Then, similarity eigenvalues are derived by the diagonalization of the MDDM. For same-term IPE pairs, such as $u = v$, the MDDM are square upper/lower triangular. These matrices can be directly diagonalized. The resulting eigenvalues determined from the MDDM are normalized and ranked in numerically descending order in their eigenvector representation. The n^{th} normalized eigenvalue for IPE type m of a compound α , $\epsilon_{mn}(\alpha)$, can be obtained by scaling the non-normalized eigenvalue $\epsilon_{mn}'(\alpha)$ relative to the rank of its MDDM:

$$\epsilon_{mn}(\alpha) = \epsilon_{mn}'(\alpha) / \text{rank}(\alpha)_m \quad (3)$$

Determination of eigenvalues of the MDDM for $u \neq v$, the so-called cross terms for IPE pairs that are not the same, requires a different strategy since these matrices may, or may not, be square. In the case of rectangular MDDM ($u \neq v$), the following square MDDM are constructed.

$$\text{MDDM}(u, u) = \text{MDDM}(n_u, n_u) \times \text{MDDM}(n_u, n_u)^T \quad (4)$$

$$\text{MDDM}(v, v) = \text{MDDM}(n_v, n_v) \times \text{MDDM}(n_v, n_v)^T \quad (5)$$

For $\text{MDDM}(u, u)$ and $\text{MDDM}(v, v)$ having the same rank and trace, both have the same set of eigenvalues. Hence, for each pair of IPE ($u \neq v$):

$$\epsilon(\alpha)_{u,v} = \{[\epsilon(\alpha)]_{\text{MDDM}(u,u)}\}^{1/2} \quad (6)$$

According to all possible combinations of the eight IPE types, there are 36 possible molecular similarity eigenvectors from the MDDM for each compound α . The similarity eigenvectors have been calculated for the set of compounds, and the estimation of molecular similarity for a pair of compounds α and β begins with a definition for molecular dissimilarity given by

$$D_{\alpha\beta} = \sum_i |\epsilon(\alpha)_i - \epsilon(\beta)_i| \quad (7)$$

where $i = i^{\text{th}}$ eigenvalue in the corresponding eigenvector of a specific IPE pair. Molecular similarity is then defined as

$$S_{\alpha\beta} = (1 - D_{\alpha\beta})(1 - \phi) \quad (8)$$

where $\phi = |\text{rank}(\alpha) - \text{rank}(\beta)| / (\text{rank}(\alpha) + \text{rank}(\beta))$. The rank of the matrices is essentially the number of atoms of a specific IPE type present. The ϕ term in eq 8 serves to reincorporate molecular size information. Similar to the measure for dissimilarity, the similarity measure is a value between 1 and 0, where a value closer to 1 refers to compounds that are more similar, and the value closer to 0 refers to compounds that are more dissimilar.

The descriptor set for α consists of all of the eigenvalues of all of the eigenvectors derived from all of the MDDM for the compound α . In this work, a threshold cutoff value, which is equal to 0.002, and those normalized eigenvalues below the threshold value are disregarded.

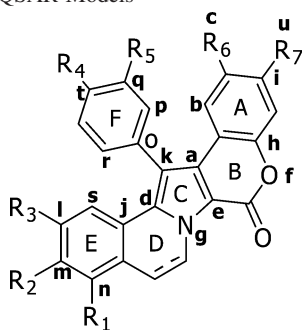
The maximum number of significant eigenvalues specific to that data set for a particular compound and a particular IPE type, m , is determined, $\epsilon_{m,\text{max}}$. All the eigenvectors for IPE type, m , for each molecule across a lamellarin data set are then assigned $\epsilon_{m,\text{max}}$ eigenvalues for IPE type m . Eigenvectors that otherwise contain less than $\epsilon_{m,\text{max}}$ elements have the “missing” eigenvalues set to zero.

The total set of descriptors, ϵ_{total} , for a compound in the data set will be the sum of the 36 eigenvalues of $\epsilon_{m,\text{max}}$ length, which can be a large number for the data set in this work.

Finally, the sets of 4D fingerprints across each of the molecules of the training set form the trial descriptor pool to build the 4D fingerprint virtual screens. The building procedure of these virtual screens is identical to that employed in constructing the RI 4D QSAR models. That is, Steps 6–9 given above for the RI 4D QSAR methodology are used.

RESULTS AND DISCUSSION

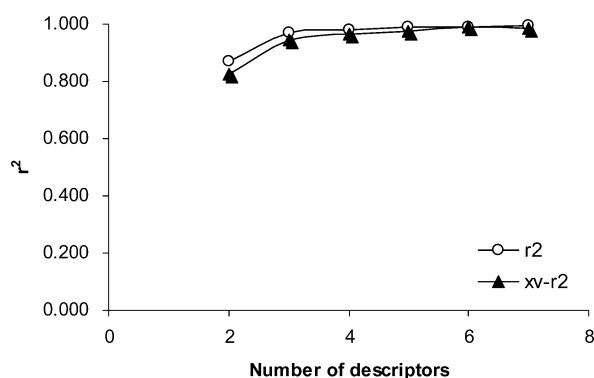
Receptor-Independent (RI) 4D QSAR Analysis. Optimized RI 4D QSAR models were constructed for each of the eight trial alignments listed in Table 4. Alignments 1, 2, 4, and 7 contain atoms from two rings (A and B), (B and C), (C and D), and (C and F), respectively. Alignment 5

Table 4. Set of Trial Alignment Used in Constructing the Best Five-Term RI 4D QSAR Models

alignment	first atom	second atom	third atom	r^2	$xv - r^2$
1	a	b	c	0.964	0.929
2	d	e	f	0.996	0.992
3	g	h	i	0.999	0.998
4	j	d	k	0.997	0.995
5	l	m	n	0.995	0.984
6	o	p	q	0.997	0.993
7	k	o	r	0.999	0.997
8	s	t	u	0.999	0.995

and 6 only contain atoms from ring (E and F), respectively. Only two alignments, 3 and 8, distribute the three-ordered atoms across three rings, namely, rings A, B, and C for alignment 3 and rings A, E, and F for alignment 8. The r^2 and $xv - r^2$ values from the best corresponding five-term RI 4D QSAR models of each alignment are given in Table 4. Five terms in a model corresponds to the largest model that can be built by allowing at least five observations [compounds] per model term for the training set. The optimized five-term model represents an initial upper-bound exploration of the type and a corresponding quality of an RI-4D QSAR model that can be expected from the structure–activity data set. Alignment 1 yields the poorest fits with $r^2 = 0.964$ and $xv - r^2 = 0.929$. The differences among r^2 and $xv - r^2$ of the remaining alignments are quite small, or the alignment of lamellarin is not significant to the 4D QSAR model. However, based on the greater r^2 (0.999) and $xv - r^2$ (0.998), alignment 3 appears to be the best alignment for 4D QSAR analysis of lamellarin data set.

The optimum number of descriptors in a model is determined by monitoring when $xv - r^2$ becomes effectively constant or decreases with an increasing model size. Figure 2 is a plot of the number of descriptor terms in an optimized alignment 3 model versus the corresponding r^2 and $xv - r^2$.

**Figure 2.** Plot of the number of RI 4D QSAR model descriptor terms versus r^2 and $xv - r^2$ for the complete training set.**Table 5.** The Cross-Correlation Matrix for the Top 10 Models of the 25 Lamellarins Training Set

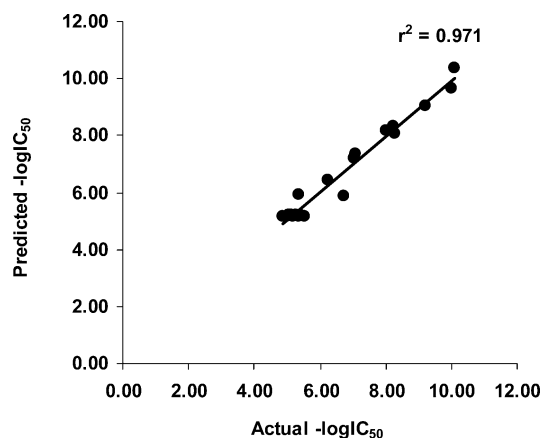
model no.	1	2	3	4	5	6	7	8	9	10
1	1.000									
2	0.518	1.000								
3	0.760	0.816	1.000							
4	0.578	0.872	0.903	1.000						
5	0.624	0.640	0.806	0.858	1.000					
6	0.525	0.405	0.635	0.631	0.683	1.000				
7	0.601	0.665	0.823	0.887	0.962	0.638	1.000			
8	0.618	0.689	0.837	0.902	0.987	0.654	0.979	1.000		
9	0.618	0.689	0.837	0.902	0.987	0.654	0.979	1.000	1.000	
10	0.608	0.674	0.828	0.892	0.963	0.644	0.998	0.981	0.981	1.000

An inspection of Figure 2 reveals that the maximum number of descriptor terms in the RI 4D QSAR model providing additional fit to the training set data is three. There is no meaningfully enhanced model fitting by including more than three descriptor terms. Thus, the optimized RI 4D QSAR model for the 25 lamellarins generated from alignment 3 is given by eq 9. Among top 10 4D QSAR models obtained from alignment 3, eq 9 (or model three) is the best 4D QSAR model since it has the highest $xv - r^2$, and all other top 10 models are basically the same as model three. This commonality to model three by the other top 10 models can be inferred from Table 5 by the high cross-correlations of their residuals of fit to those of eq 9:

$$\begin{aligned}
 -\log \text{IC}_{50} = & 5.14 + 16.90\text{GC1}(-5, 6, 2, \text{np}) - \\
 & 56.33\text{GC2}(-3, 4, -5, \text{any}) + \\
 & 64.62\text{GC3}(-1, 5, 0, \text{np}) \\
 n = 25, \quad r^2 = 0.971, \quad xv - r^2 = 0.947
 \end{aligned} \quad (9)$$

$\text{GC}_i(x, y, z, X)$ is the i^{th} GCOD descriptor term located at (x, y, z) in the reference grid cell and the alignment space and having the X type IPE as defined in Table 3. Figure 3 is a plot of the predicted, using eq 9, versus actual $-\log \text{IC}_{50}$ values. All of the predicted $-\log \text{IC}_{50}$ values are within ± 1 log unit of the corresponding observed values, and there are no outliers.

Two GCODS (GC1 and GC3) of eq 9 correspond to pharmacophore sites of nonpolar atom occupancy, both of which increase potency. These two GCODS both have positive regression coefficients with values of 16.90 and 64.62, respectively. GCOD GC2, having an ‘any’ IPE type,

**Figure 3.** Predicted $-\log \text{IC}_{50}$ values, using the RI 4D QSAR model for the 25 lamellarins data set versus the observed $-\log \text{IC}_{50}$ values.

has a negative regression coefficient with a value of -56.33 . Consequently, occupancy of the GC2 site by any type of atom will lead to a decrease in the potency of antitumor cancer activity of the corresponding lamellarin. From an analysis of eq 9, it is found that the any IPE type at $(-3, 4, -5)$ has about three times more of a negative effect upon $-\log IC_{50}$ than the positive effect of the nonpolar IPE type at $(-5, 6, 2)$, and about the same but opposite effect on $-\log IC_{50}$ as the nonpolar IPE type at $(-1, 5, 0)$. None of the best models from the GFA model optimization contain GCOD descriptors which deal with specific atom–atom interactions like hydrogen bonding.

In order to further search for pharmacophore sites, which are specifically associated with lamellarins, exhibiting high cytotoxic activity, an additional RI 4D QSAR analysis was carried out. The training set of this study was limited to the six lamellarins (D, M, N, X, ϵ , and dehydrolamellarin J of Table 2) that have the highest $-\log IC_{50}$ values and are not redundant in their structural features. The RI 4D QSAR models were constructed and optimized by using the same methodology and alignment used to build eq 9. The best RI 4D QSAR model from this small, high-activity data set of lamellarins is given by eq 10:

$$-\log IC_{50} = 10.31 - 50.52GC1(-1, 1, -6, any) + 1.58GC2(-1, 4, -6, np) \quad (10)$$

$$n = 6, \quad r^2 = 0.997, \quad x_{cv} - r^2 = 0.984$$

The regression coefficients of the descriptors of eq 10 suggest placing any type of atom at $(-1, 1, -6)$ has about 30 times more negative effect on $-\log IC_{50}$ than the positive gain by locating a nonpolar atom or group at $(-1, 4, -6)$. Certainly eq 10 is, or borders upon, being an overfit model. However, eq 10 and its 3D pharmacophore are only used as adjuncts to eq 9 and its 3D pharmacophore. That is, eq 10 is being used to provide a higher-resolution view of the SAR features most characteristic of the high-activity lamellarins of the training set. Equation 9 and its 3D pharmacophore are used outside that context.

The 3D pharmacophores defined by eqs 9 and 10 are shown in Figure 4a and b, respectively. The reference structure superimposed on each of the 3D pharmacophores in these two figures is the predicted active conformation of the most active compound (lamellarin D) using eq 9. The red spheres in Figures 4a and b represent those GCOD descriptor terms which have negative regression coefficients. Correspondingly, the blue spheres delineate GCOD descriptors having positive regression coefficients in the corresponding best RI 4D QSAR equation. From an inspection of Figure 4a, a red sphere near the R_2 and R_3 groups specifies a pharmacophore site where occupancy by any type of atom or group decreases the potency, since the corresponding regression coefficient is -56.33 . Two blue spheres are found near R_4 and R_5 , suggesting that substitution of nonpolar groups to occupy one or both sites is conducive to increasing the cytotoxic activity of the lamellarins.

The 3D pharmacophore of the high-activity model, eq 10, is represented by one red sphere (GCOD) located around R_1 and R_2 and a blue sphere (GCOD) positioned near R_2 and R_3 . The most active compounds of the potent lamellarins seemingly achieve most of their additional $-\log IC_{50}$

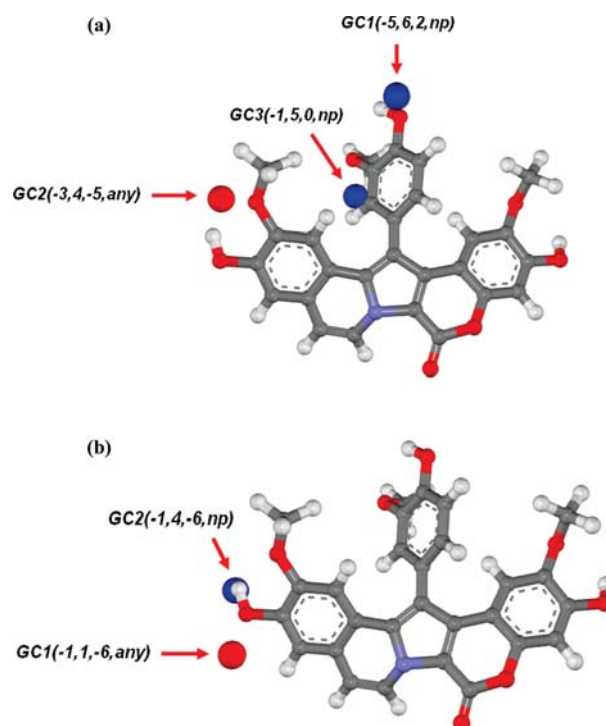


Figure 4. The 3D pharmacophores from (a) eq 9 based upon the full training set of 25 lamellarins, and (b) from the six high-activity compounds of the lamellarin training set. The 3D pharmacophores are shown relative to the predicted active conformation of the most active compound (lamellarin D). The red spheres refer to pharmacophore sites having negative regression coefficients in the 4D QSAR equation, and the blue spheres refer to pharmacophore sites having positive regression coefficients.

potency, as compared to the less potent lamellarins, by not having any atoms or groups at $(-1, 1, -6)$ in contrast to the increasing occupancy by nonpolar atoms or groups at the GCOD located at $(-1, 4, -6)$. The 30:1 ratio of not occupying the GCOD at $(-1, 1, -6)$ as compared to having a nonpolar atom or group at $(-1, 4, -6)$ is consistent with the relative binding energy contributions of an intermolecular hydrogen bond involving the OH near $(-1, 1, -6)$ as compared to a hydrophobic binding effect due to the methyl of the methoxy group near $(-1, 4, -6)$, as is shown in Figure 4b.

Overall, the highly active compounds are seemingly distinguished from one another in eq 10 by their ability to form an intermolecular hydrogen bond where the hydrogen-bond acceptor atom in the receptor is expected to be near $(-1, 1, -6)$. Some additional increase in $-\log IC_{50}$ can also be realized by having a hydrophobic substituent group of the ligand occupying the $(-1, 4, -6)$ site. The two GCODs of eq 10 may be a higher resolution representation of the single GC2 $(-3, 4, -5, any)$ GCOD found in eq 9.

In order to evaluate the possible roles of ligand MW on cytotoxic potency, $-\log IC_{50}$, this property was included as part of the trial basis set of descriptors in a GFA model optimization study. Unfortunately, no GFA model optimization could be realized. An inspection of the MW value of the training set compounds revealed that three lamellarins (lam K-triacetate, lam χ -triacetate, and lam U-diacetate) have very high MWs relative to the other training set compounds. These three lamellarins were removed to form a revised training set. Two lamellarins (lam F and K) were defined as

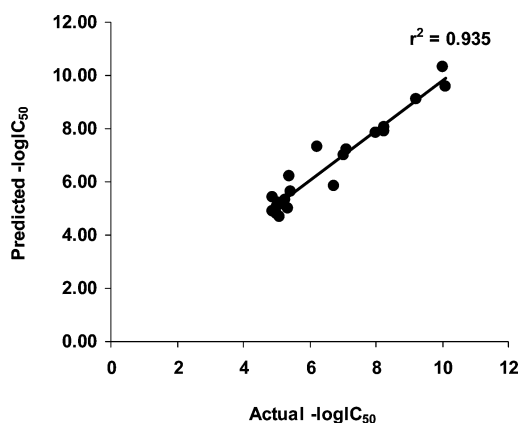
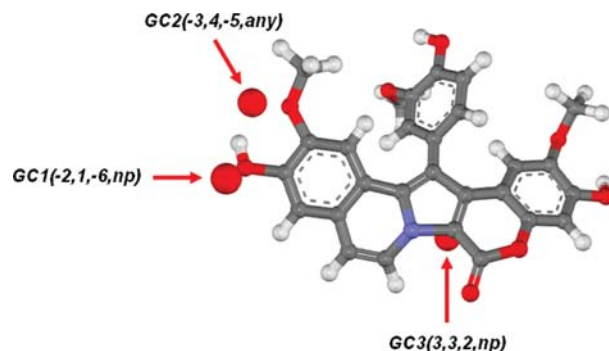
Table 6. The Cross-Correlation Matrix for the Top 10 Models of 21 Lamellarins Training Set

model no.	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.91	1.00								
3	0.83	0.97	1.00							
4	0.93	0.93	0.90	1.00						
5	0.92	0.99	0.97	0.93	1.00					
6	0.83	0.97	0.99	0.90	0.97	1.00				
7	0.93	0.93	0.90	0.99	0.93	0.90	1.00			
8	0.94	0.79	0.72	0.93	0.80	0.73	0.93	1.00		
9	0.91	0.98	0.95	0.91	0.98	0.95	0.91	0.80	1.00	
10	0.84	0.97	0.99	0.90	0.97	0.99	0.90	0.73	0.96	1.00

a test set. GFA model building and optimization was repeated for this 21 compound training set in the same manner as employed in developing eqs 9 and 10. The 10 best models were determined from the GFA optimization, and the residuals of fit cross-correlations between each pair of these models are given in Table 6. All pairs of the top 10 models have residuals of fit highly correlated to one another, with a value of at least 0.70, indicating these 10 models are all very nearly the same model. Therefore, the best of the 10 models was selected as the preferred RI 4D QSAR model for this training set and is given by eq 11:

$$\begin{aligned}
 -\log \text{IC}_{50} = & 10.31 - 4.77\text{GC1}(-2, 1, -6, \text{np}) - \\
 & 33.91\text{GC2}(-3, 4, -5, \text{any}) - \\
 & 8.12\text{GC3}(3, 3, 2, \text{np}) \quad (11) \\
 n = 21, \quad r^2 = 0.935, \quad xv - r^2 = 0.890
 \end{aligned}$$

Figure 5 is a plot of the observed versus the predicted $-\log \text{IC}_{50}$ values determined from using eq 11. The 3D pharmacophore embedded in the RI 4D QSAR model, given by eq 11, is shown in Figure 6 with lamellarin D again being the reference compound. All three GCOD descriptors of eq 11 correspond to pharmacophore sites where an increasing occupancy decreases activity. One pharmacophore site $(-3, 4, 5, \text{any})$ from eq 11 is identical to a site from eq 9, while the pharmacophore site at $(-2, 1, -6, \text{np})$ from eq 11 is very close to the pharmacophore site of eq 10 located at $(-1, 1, -6, \text{any})$, as can be seen by comparing Figure 6 to Figure 4a and b. The third pharmacophore site of eq 11 located at $(3, 3, 2)$, which predicts the occupancy of nonpolar groups to decrease $-\log \text{IC}_{50}$, is unique to this model as compared to

**Figure 5.** Predicted $-\log \text{IC}_{50}$ values, using the RI 4D QSAR model for the 21 lamellarins data set versus the observed $-\log \text{IC}_{50}$ values.**Figure 6.** The 3D pharmacophores from eq 11 based upon the 21 lamellarins training set. The 3D pharmacophores are shown relative to the predicted active conformation of the most active compound (lamellarin D). The red spheres refer to pharmacophore sites having negative regression coefficients in the 4D QSAR equation, and blue spheres refer to pharmacophore sites having positive regression coefficients.

eqs 9 and 10. This new GCOD descriptor term of eq 11 and the decrease in r^2 and $xv - r^2$ may be an indication of a significant pharmacophore site dependence on one or more of the four lamellarins eliminated from the training set used to build eq 11 and its corresponding 3D pharmacophore.

An attempt was made to further explore if $\log P$ plays a role in the SAR of the lamellarin training set by forcing an overfitting in the GFA model building and optimization process. The $\log P$ descriptor was the only non-GCOD descriptor added to the trial basis set (descriptor pool) at Step 5 of the 4D QSAR methodology. Overfit RI 4D QSAR models were permitted under the same methodology and alignment and for all lamellarins in training set as used to develop eq 9. None of the 10 most significant overfit four-term or five-term RI 4D QSAR models contained a $\log P$ descriptor term. Therefore, it was concluded that molecular lipophilicity is not a major contributing factor in the specification of the cytotoxic activity for the lamellarins studied in this analysis.

The predicted $-\log \text{IC}_{50}$ of lamellarin F calculated by using eq 11 is 5.74. This value is very close to actual $-\log \text{IC}_{50}$ value of 5.34. The RI 4D QSAR model obtained by removing the high MW compounds showed a good prediction of the lamellarin F activity. Lamellarin K was synthesized and tested after the 4D QSAR models, as reported in this paper, were constructed. However, lamellarin K has an unexpected high activity [$-\log \text{IC}_{50} = 7.04$] for the saturated D-ring series of compounds. Hence, it was thought important to see if this high activity could be predicted by the 4D QSAR models, or if this saturated D-ring analog had features outside those captured by the models. The predicted $-\log \text{IC}_{50}$ values of lamellarin K obtained from eqs 9–11 are 5.33, 9.77, and 6.00, respectively. Thus, the 4D QSAR models developed in this study cannot well predict the activity of lamellarin K, but their composite set of predicted activities brackets around the observed activity. Moreover, while the individual models did not adequately predict the experimental end point, it is to be noted that the average of these three predicted values is 7.03, which is a value very close to the experimental $-\log \text{IC}_{50}$ value of 7.04.

Lamellarin K has a unique three hydroxyl substituent pattern at R₁, R₄, and R₇. However, other analogs in Table 1 have three hydroxyl substituents, and some analogs without

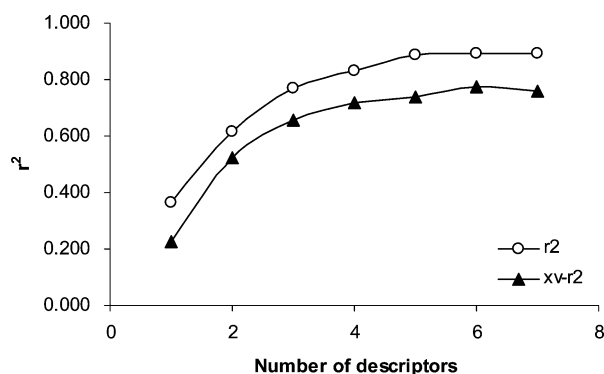


Figure 7. Plot of the number of 4D fingerprint model descriptor terms versus r^2 and $xv - r^2$ for the complete training set.

Table 7. The Linear Cross-Correlation Matrix of the Top 10 Models from the Four Descriptor Term 4D Fingerprint Models

model no.	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.93	1.00								
3	0.96	0.90	1.00							
4	0.93	0.94	0.93	1.00						
5	0.89	0.94	0.90	0.96	1.00					
6	0.91	0.86	0.94	0.88	0.85	1.00				
7	0.95	0.90	0.99	0.93	0.91	0.94	1.00			
8	0.92	0.91	0.92	0.91	0.89	0.92	0.93	1.00		
9	0.90	0.92	0.90	0.90	0.97	0.87	0.91	0.89	1.00	
10	0.91	0.92	0.92	0.90	0.96	0.89	0.94	0.90	0.98	1.00

hydroxyl substituents are more active than those with three hydroxyls, for example, compare lamellarin χ triacetate (5.54) to lamellarin E (5.28) in Table 1. All of the best 4D QSAR models, eqs 9–11 are rich in GCOD terms involving nonpolar IPE types. Polar- and/or hydrogen-bonding capabilities from hydroxyl groups are not explicitly present in the descriptor terms of the 4D QSAR models. All of these observations, in composite, suggest that the unique hydroxyl substituent pattern of lamellarin K make it the ‘magic bullet’ in terms of high inhibition potency relative to the other saturated D-ring analogs of Table 1.

4D Fingerprint Virtual Screens. 4D fingerprint virtual high-throughput screens permit a larger range of chemical diversity to be assayed more quickly than do RI 4D QSAR models. In this study, 4D fingerprints models were generated using all 25 of the lamellarins in the training set. Lamellarin K was used as a modest means to validate 4D fingerprints model as well as RI-4D QSAR models. Two types of 4D fingerprints can be constructed: 4D fingerprints explicitly dependent upon a particular alignment, and absolute 4D fingerprints which are alignment independent. Absolute 4D fingerprints were used in this analysis to maximize the range of lamellarin chemical diversity that could be reasonably screened. That is, a 4D fingerprint screening model built independent of alignment is more general than its corresponding alignment-dependent screen but at the cost of being somewhat less significant in its fit to the training set data.

The absolute 4D fingerprints were derived for each of the 25 training set lamellarins using the modeling methodology given above in the Methods Section. These 4D fingerprints formed the trial basis set for model building. No non-4D fingerprints were added to this trial descriptor pool. Model building and optimization in deriving the 4D fingerprint QSAR equations, which are the high-throughput virtual

Table 8. Frequency of Use and Corresponding Significance Ranking of Each Descriptor Term in 4D Fingerprint Virtual Screening Model

	$\epsilon_7(\text{any}, \text{np})$	$\epsilon_{11}(\text{any}, \text{hs})$	$\epsilon_3(\text{p}^+, \text{aro})$	$\epsilon_2(\text{np}, \text{hs})$
frequency	124	128	51	17
ranking	2	1	5	11

screens, were carried out in the identical fashion used to build the RI 4D QSAR models.

Figure 7 is a plot of number of descriptor terms in a 4D fingerprint model versus r^2 and $xv - r^2$. The $xv - r^2$ of the 4D fingerprints of the four- and five-term models are very nearly the same, and $xv - r^2$ behaves in something of an erratic fashion for models having five or more descriptor terms. The optimized four-descriptor term virtual screening model appears, on the basis of $xv - r^2$, to capture maximum fitting to the training set data without overfitting. Thus, the four-term QSAR model given by eq 12 was selected as the preferred absolute 4D-fingerprint virtual screen. Equation 12 is the best four-term model from the top 10 four-term models derived in the GFA optimization. Table 7 shows the linear cross-correlation matrix of the residual of fit for the top 10 four-term models. This table reveals that all pairs of models have highly correlated residuals of fit, greater than 0.85, to one another. Thus, eq 12 represents the best and only distinct fit to the training set data using absolute 4D fingerprints.

$$\begin{aligned}
 -\log \text{IC}_{50} = & -7.39 - 452.65\epsilon_7(\text{any}, \text{np}) + \\
 & 1357.10\epsilon_{11}(\text{any}, \text{hs}) + 9.58\epsilon_3(\text{p}^+, \text{aro}) - \\
 & 94.31\epsilon_2(\text{np}, \text{hs}) \quad (12) \\
 n = 25, \quad r^2 = 0.831, \quad xv - r^2 = 0.719
 \end{aligned}$$

For reference in defining the 4D fingerprints, $\epsilon_7(\text{any}, \text{np})$ represents the seventh largest eigenvalue from the MDDM of the IPEs $u = (\text{any})$ and $v = (\text{np})$ molecular similarity vector, capturing all pairs of atoms in each lamellarin assigned IPEs of any and nonpolar, respectively.

The relative significance and weight of each 4D fingerprint descriptor term in eq 12 was measured in terms of its frequency of use in the GFA model optimization process. The idea in monitoring frequency of use is that the more significant a descriptor is to establishing a fit to the training set data, the more often it will be used in the repetitive GFA optimization process. The frequencies of descriptor usage during GFA optimization are shown in Table 8. An inspection of Table 8 indicates that $\epsilon_{11}(\text{any}, \text{hs})$ and $\epsilon_7(\text{any}, \text{np})$ are the first and second important features governing the SAR of lamellarin cytotoxicity potency, respectively. Increased potency of the lamellarins arises from increasing the values of $\epsilon_{11}(\text{any}, \text{hs})$ and/or $\epsilon_3(\text{p}^+, \text{aro})$, while a decrease in lamellarin cytotoxicity accompanies an increase in the values of the $\epsilon_7(\text{any}, \text{np})$ and $\epsilon_2(\text{np}, \text{hs})$ 4D fingerprints. Figure 8 is a plot of $-\log \text{IC}_{50}$ values predicted using eq 12 versus the corresponding observed $-\log \text{IC}_{50}$ values.

The predicted activity of lamellarin K, the test compound, using eq 12 is 7.33, which differs from the observed activity of 7.04 by only 0.29 log units. Additional $-\log \text{IC}_{50}$ predictions using eq 12 were made for a small virtual library of eight lamellarin derivatives, see Table 9, generated by making substituent changes at R₁–R₅. These results indicate

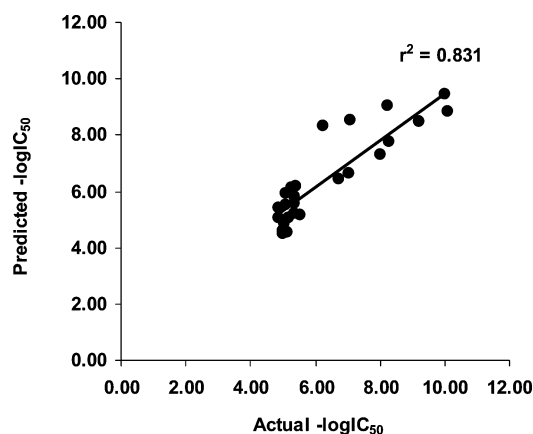
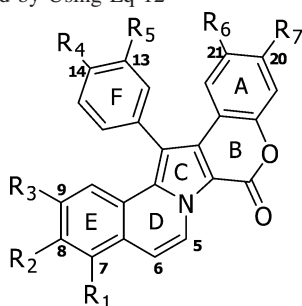


Figure 8. Predicted $-\log IC_{50}$ values, using the 4D-fingerprint model for the 25 lamellarins data set, versus the observed $-\log IC_{50}$ values.

Table 9. A Virtual Library of Lamellarins Built around Substituent Variations at R_1 – R_5 and the Corresponding Predicted $-\log IC_{50}$ Obtained by Using Eq 12



lamellarin	predicted $-\log IC_{50}$	substituent group						
		R_1	R_2	R_3	R_4	R_5	R_6	R_7
ML1	6.43	OH	OH	OMe	OH	OMe	OMe	OH
ML2	6.22	OMe	OH	OMe	OH	OMe	OMe	OH
ML3	7.93	H	OMe	OMe	OH	OMe	OMe	OH
ML4	8.74	H	H	OMe	OH	OMe	OMe	OH
ML5	10.05	H	OH	OH	OH	OMe	OMe	OH
ML6	3.03	H	OH	H	OH	OMe	OMe	OH
ML7	8.54	H	OH	OMe	H	OMe	OMe	OH
ML8	7.21	H	OH	OMe	OH	H	OMe	OH

Table 10. The Linear Cross-Correlation Matrix of the Predicted $-\log IC_{50}$ Values of the RI 4D QSAR Model (1), the 4D Fingerprint Model (2), and the Observed Cytotoxicity $-\log IC_{50}$ Values (3)

	1	2	3
1	1		
2	0.797	1	
3	0.972	0.823	1

that the 4D fingerprint model is responsive to predicting $-\log IC_{50}$ values over a wide $-\log IC_{50}$ potency range from nearly inactive values for ML6 to very potent activities for lamellarins ML4, ML5, and ML7. Equation 12 also has captured the SAR that both the number and positioning of $-\text{OH}$ on the E-ring is a critical factor to potency. In general, more hydroxyls on the ring are better. But the importance of hydroxyl positioning, particularly at R_3 , is dramatically shown for ML5, the most active analog in Table 9 [10.05] as compared ML6, the least active analog [3.03], which differs only from ML5 by having no hydroxyl at R_3 .

Comparison of the 4D Fingerprints QSAR Virtual Screening Model to the RI 4D QSAR Models. The RI 4D QSAR model given by eq 9 with three descriptor terms is a more significant fit to the training set data ($xv - r^2 = 0.947$ and $r^2 = 0.971$) than the four descriptor 4D-fingerprint model given by eq 12 ($xv - r^2 = 0.719$ and $r^2 = 0.831$). Presumably the inclusion of alignment information in eq 9 provides this boost in the overall fitting quality of this model as compared to eq 12. But eq 12 in not being dependent on alignment correspondingly permits a wider range of variations of lamellarin chemistry to be considered. Table 10 is the linear correlation matrix of the residuals of fit of eq 9, the RI 4D QSAR, to eq 12, the absolute 4D fingerprint virtual screen, as well as correlations of both models to the observed $-\log IC_{50}$ cytotoxicity values. The correlation coefficient of 0.797 between the residuals of fit for eqs 9 and 12 indicates that these two models are basically the same, but eq 9, owing to inclusion of alignment, fits the training set better overall than eq 12.

Comparison of the predicted inhibition potencies from the 4D fingerprints QSAR virtual screening model to the RI 4D QSAR model was also investigated using ML5 and ML6, the most and least potent compounds given in Table 9. The predicted $-\log IC_{50}$ values of ML5 obtained from the RI 4D QSAR models by eqs 9 and 11 are 5.39 and 9.20, respectively, and the $-\log IC_{50}$ values of ML6 obtained from the two equations are 7.44 and 10.23, respectively. It was found that there is an agreement in prediction only for ML5 between the RI 4D QSAR model (9.20 by eq 11) and the 4D fingerprint QSAR model (10.05 by eq 12).

CONCLUSION

This work puts forth a 'quality in place of quantity' strategy to handle small data sets composed of structurally complex and hard to synthesize compounds that can exhibit a wide range of end point activity. A high-level modeling approach providing detailed structural, thermodynamic, and electronic information about each complex compound of the data set is used to negate the lack-of-data drawbacks to the small size of the data set. In this study, the flexibility, yet high-level of modeling sophistication, of the 4D QSAR paradigm is used to explore different subpopulations of the data set in extracting the maximum SAR information from the data set in terms of a pseudo consensus RI 4D QSAR model and its corresponding 3D pharmacophore. The consensus aspect to the RI 4D QSAR modeling arises from the fact that the same methodology and parameters, including alignment, can be used in any manner across any subpopulations of the data set. As such, all resulting models are not only directly comparable, to an appreciable extent, but also can be combined to elucidate a high-resolution 3D pharmacophore. In addition, the 4D fingerprint formulation of the 4D QSAR paradigm permits alternate model generation, particularly useful in virtual screening. Still, the 4D fingerprint models are once again directly comparable to the RI 4D QSAR models so as to exact additional information from the data set as well as to evaluate the self-consistency across all the models constructed.

The consensus set of 4D QSAR models, expressed by eqs 9–12, suggests that the ability to form a ligand–receptor intermolecular hydrogen bond and a hydrophobic interaction

for substituents on the E-ring must modulate the cytotoxicity against T47D breast cancer cells. The optimization of this intermolecular hydrogen bond and, to a lesser extent, the hydrophobic interactions is coupled to the alignment freedom of a lamellarin owing, in turn, to other possible substitutions across the molecule and their possible interactions with sites on the receptor.

Hydrophobic substitutions on the F-ring can also enhance cytotoxic potency, but given that the 3D pharmacophore sites for these interactions arise for the entire data set and not the restricted high-activity data subset would indicate these are likely minor binding pharmacophore sites. Attempts to force the lipophilicity of the entire lamellarin into a 4D QSAR model were unsuccessful. Thus, the finding of 3D pharmacophore sites, where occupancy by nonpolar atoms and/or groups can modulate activity, likely reflect specific interactions at these sites and not global lipophilic features of the lamellarins.

Lamellarin K, synthesized and tested after the modeling studies reported here were carried out, likely has its very high activity relative to other saturated D-ring analogs because of its unique three hydroxyl group substituent pattern. The average predicted $-\log IC_{50}$ value developed in this study sufficiently predicts the activity of lamellarin K. This suggests that in order to get a high-resolution 4D QSAR model to distinguish some substituent patterns from others for the saturated D-ring lamellarins analogs, more lamellarins data sets are required.

The 4D fingerprint virtual screening model, eq 12, is highly consistent with the general RI 4D QSAR model given by eq 9. Consequently, eq 12 can be used to rapidly screen prospective compounds without concern for alignment, but with the expectation that the 3D pharmacophore of eq 9 will be relevant to helping understand findings from virtual screenings. A good test to evaluate how much SAR information is actually captured in eq 12 as a virtual screening tool, given it is based upon this relatively small training set of lamellarins, would be to make and test ML5 and ML 6 of Table 9. These two compounds are predicted to differ by 7 orders of magnitude in $-\log IC_{50}$ values, yet they differ by at their respective R_3 substituents. A large difference in measured $-\log IC_{50}$ values would help to validate eq 12, while a small difference would suggest that eq 12 has very limited resolution in correctly explaining small structural differences in the lamellarins.

ACKNOWLEDGMENT

The Thailand Research Fund (RTA5080005) and the Commission on Higher Education (CHE) are gratefully acknowledged for a research grant. We are grateful to the Faculty of Science and the Graduate School of Kasetsart University (KU), the Phetchaburi Rajabhat University, the Chulabhorn Research Institute (CRI), the Chulabhorn Graduate Institute (CGI), and the College of Pharmacy, University of New Mexico for research support. The National Center of Excellence in Petroleum, Petrochemical Technology, and Advanced Materials under the Ministry of Education, the KU Research and Development Institute, the Laboratory for Computational and Applied Chemistry, the NANOTEC Center of Excellence at KU, and the Large Scale Research Laboratory of the National Electronics and Computer Technology are also acknowledged for computing and

research facilities. This work was also supported by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1 R21 GM075775. Information on Novel Preclinical Tools for Predictive ADME-Toxicology can be found at <http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-04-023.html>. Links to nine initiatives are found here <http://nihroadmap.nih.gov/initiatives.asp>. This work was also supported, in part, by the Proter and Gamble Company. Resources at the Laboratory of Molecular Modeling and Design at University of New Mexico and the Chem21 Group, Incorporated, were used in performing these studies. We also thank the reviewers for the valuable comments.

REFERENCES AND NOTES

- Andersen, R. J.; John Faulkner, D.; Cun-heng, H.; Van Duyne, G. D.; Clardy, J. Metabolites of the marine prosobranch mollusc *Lamellaria* sp. *J. Am. Chem. Soc.* **1985**, *107*, 5492–5495.
- Bailly, C. Lamellarins, from A to Z: A family of anticancer marine pyrrole alkaloids. *Curr. Med. Chem. Anticancer Agents* **2004**, *4*, 363–378.
- Fan, H.; Peng, J.; Hamann, M. T.; Hu, J. F. Lamellarins and related pyrrole-derived alkaloids from marine organisms. *Chem. Rev.* **2008**, *108*, 264–287.
- Vanhuyse, M.; Kluza, J.; Tardy, C.; Otero, G.; Cuevas, C.; Bailly, C.; Lansiaux, A. Lamellarin D: A novel pro-apoptotic agent from marine origin insensitive to P-glycoprotein-mediated drug efflux. *Cancer Lett.* **2005**, *221*, 165–175.
- Quesada, A. R.; Garcia Gravalos, M. D.; Fernandez Puentes, J. L. Polyaromatic alkaloids from marine invertebrates as cytotoxic compounds and inhibitors of multidrug resistance caused by P-glycoprotein. *Br. J. Cancer* **1996**, *74*, 677–682.
- Gallego, M. A.; Ballot, C.; Kluza, J.; Hajji, N.; Martorati, A.; Castera, L.; Cuevas, C.; Formstecher, P.; Joseph, B.; Kroemer, G.; Bailly, C.; Marchetti, P. Overcoming chemoresistance of non-small cell lung carcinoma through restoration of an AIF-dependent apoptotic pathway. *Oncogene* **2008**, *27*, 1981–1992.
- Kluza, J.; Gallego, M. A.; Loyens, A.; Beauvillain, J. C.; Sousa-Faro, J. M. F.; Cuevas, C.; Marchetti, P.; Bailly, C. Cancer cell mitochondria are direct proapoptotic targets for the marine antitumor drug lamellarin D. *Cancer Res.* **2006**, *66*, 3177–3187.
- Facompre, M.; Tardy, C.; Bal-Mahieu, C.; Colson, P.; Perez, C.; Manzanares, I.; Cuevas, C.; Bailly, C. Lamellarin D: A Novel Potent Inhibitor of Topoisomerase I. *Cancer Res.* **2003**, *63*, 7392–7399.
- Marco, E.; Laine, W.; Tardy, C.; Lansiaux, A.; Iwao, M.; Ishibashi, F.; Bailly, C.; Gago, F. Molecular determinants of topoisomerase I poisoning by lamellarins: Comparison with camptothecin and structure-activity relationships. *J. Med. Chem.* **2005**, *48*, 3796–3807.
- Dias, N.; Vezin, H.; Lansiaux, A.; Bailly, C. Topoisomerase inhibitors of marine origin and their potential use as anticancer agents. *Top. Curr. Chem.* **2005**, *253*, 89–108.
- Tardy, C.; Facompre, M.; Laine, W.; Baldeyrou, B.; Garcia-Gravalos, D.; Francesch, A.; Mateo, C.; Pastor, A.; Jimenez, J. A.; Manzanares, I.; Cuevas, C.; Bailly, C. Topoisomerase I-mediated DNA cleavage as a guide to the development of antitumor agents derived from the marine alkaloid lamellarin D: Triester derivatives incorporating amino acid residues. *Bioorg. Med. Chem.* **2004**, *12*, 1697–1712.
- Ishibashi, F.; Tanabe, S.; Oda, T.; Iwao, M. Synthesis and structure-activity relationship study of lamellarin derivatives. *J. Nat. Prod.* **2002**, *65*, 500–504.
- Chittchang, M.; Batsomboon, P.; Ruchirawat, S.; Ploypradith, P. Cytotoxicities and structure-activity relationships of natural and unnatural lamellarins towards cancer cell lines. *ChemMedChem* **2009**, *4*, 457–65.
- Thipnate, P.; Chittchang, M.; Thasana, N.; Saparpakorn, P.; Ploypradith, P.; Hannongbua, S. 3D-QSAR analysis for cytotoxicity of lamellarins against human hormone-dependent T47D and hormone-independent MDA-MB-231 breast cancer cells, submitted.
- Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.
- Hopfinger, A. J.; Reaka, A.; Venkatarangan, P.; Duca, J. S.; Wang, S. Construction of a Virtual High Throughput Screen by 4D-QSAR Analysis: Application to a Combinatorial Library of Glucose Inhibitors of Glycogen Phosphorylase b. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1151–1160.
- Krasowski, M. D.; Hong, X.; Hopfinger, A. J.; Harrison, N. L. 4D QSAR analysis of a set of propofol analogues: Mapping binding sites

- for an anesthetic phenol on the GABA_A receptor. *J. Med. Chem.* **2002**, *45*, 3210–3221.
- (18) Senese, C. L.; Duca, J.; Pan, D.; Hopfinger, A. J.; Tseng, Y. J. 4D fingerprints, universal QSAR and QSPR descriptors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1526–1539.
 - (19) Liu, J.; Yang, L.; Li, Y.; Pan, D.; Hopfinger, A. J. Prediction of plasma protein binding of drugs using Kier-Hall valence connectivity indices and 4D-fingerprint molecular similarity analyses. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 567–583.
 - (20) Liu, J.; Yang, L.; Li, Y.; Pan, D.; Hopfinger, A. J. Constructing plasma protein binding model based on a combination of cluster analysis and 4D-fingerprint molecular similarity analyses. *Bioorg. Med. Chem.* **2006**, *14*, 611–621.
 - (21) Li, Y.; Pan, D.; Liu, J.; Kern, P. S.; Gerberick, G. F.; Hopfinger, A. J.; Tseng, Y. J. Categorical QSAR models for skin sensitization based upon local lymph node assay classification measures Part 2: 4D-Fingerprint three-State and Two-2-State logistic regression models. *Toxicol. Sci.* **2007**, *99*, 532–544.
 - (22) Li, Y.; Tseng, Y. J.; Pan, D.; Liu, J.; Kern, P. S.; Gerberick, G. F.; Hopfinger, A. J. 4D-fingerprint categorical QSAR models for skin sensitization based on the classification of local lymph node assay measures. *Chem. Res. Toxicol.* **2007**, *20*, 114–128.
 - (23) Iyer, M.; Zheng, T.; Hopfinger, A. J.; Tseng, Y. J. QSAR analyses of skin penetration enhancers. *J. Chem. Inf. Model.* **2007**, *47*, 1130–1149.
 - (24) Iyer, M.; Hopfinger, A. J. Treating chemical diversity in QSAR analysis: Modeling diverse HIV-I integrase inhibitors using 4D fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 1945–1960.
 - (25) Ploypradith, P.; Petchmanee, T.; Sahakitpichan, P.; Litvinas, N. D.; Ruchirawat, S. Total synthesis of natural and unnatural lamellarins with saturated and unsaturated D-rings. *J. Org. Chem.* **2006**, *71*, 9440–9448.
 - (26) *4D QSAR User's Manual*, The ChemBats21 Group, Inc.: Lake Forest, IL, 2003.
 - (27) *HyperChem Program Release 7.5 for Windows*; Hypercube, Inc.: Gainesville, FL.
 - (28) Doherty, D. C. M. U. S. G., The ChemBats21 Group, Inc.: Lake Forest, IL, 1997.
 - (29) Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.* **1997**, *99*, 8127–8134.
 - (30) Hopfinger, A. J. P., R. A. Molecular mechanics force-field parametrization procedures. *J. Comput. Chem.* **1984**, *5*, 486–492.
 - (31) Glen, W. G.; D, W. J.; Scott, D. R. Principal components analysis and partial least squares. *Tetrahedron Comput. Methods* **1989**, *2*, 349–354.
 - (32) Rogers, D.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
 - (33) Friedman, J. M. a. r. s. T. R. N. L. f. C. S., Department of Statistics, Stanford University: Stanford, CA, November 1988 (revised August 1990).
 - (34) Duca, J. S.; Hopfinger, A. J. Estimation of Molecular Similarity Based on 4D QSAR Analysis: Formalism and Validation. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1367–1387.

CI9002427

Roles of key residues specific to cyclooxygenase II: an ONIOM study

Darinee Sae-Tang · Prasat Kittakoop ·
Supa Hannongbua

Received: 21 January 2009 / Accepted: 17 September 2009 / Published online: 4 November 2009
© Springer-Verlag 2009

Abstract Binding energy calculations of flurbiprofen to the binding pocket of the cyclooxygenase (COX) enzyme were performed based on quantum chemical calculations. The interaction energies between flurbiprofen and two types of COX binding sites were studied. Quantum chemical calculations were used, based on the B3LYP hybrid functional and the MP2 method, with 6-31G(d) and 6-31G(d,p) basis sets. The results show that the main interaction between flurbiprofen and two COX isozymes (COX-1 and COX-2) is due to Arg120. In addition, selective SC558 COX-2 inhibitor was also compared. It was found that repulsive interaction plays a significant role in its inhibition of COX-2. ONIOM2(B3LYP/6-31G(d):PM3) calculations indicate that flurbiprofen interacts via moderate hydrogen bonding with Arg120 in the COX-2 binding site, while no hydrogen bond was detected with either Tyr355 or Val523. The ONIOM2 method can be used to describe the specific interaction of the inhibitor and is helpful in designing a specific COX inhibitor.

Keywords Cyclooxygenase · Flurbiprofen · Binding energy · Quantum chemical calculations · ONIOM

Introduction

Cyclooxygenase (COX) is a bifunctional enzyme catalyzing conversion of arachidonic acid into prostaglandins. This enzyme has been classified into two different isozymes, COX-1 and COX-2. COX-1 is a constituent of most tissues and creates prostaglandins used for basic house-keeping throughout the body, while COX-2 is more selective, and is induced by inflammation and pain [1–3]. Various side effects—such as ulcers, antiplatelet activity, gastrointestinal irritation, and suppression of renal function—can arise from the inhibition of both isozymes by non-steroidal anti-inflammatory drugs (NSAIDs) [4–6]. Traditional NSAIDs, such as aspirin, ibuprofen, indomethacin, and flurbiprofen [7–9], can inhibit prostaglandin synthesis and display anti-inflammatory action. However, these compounds show no selectivity and cause side effects by the inhibition of COX-1, whereas their anti-inflammatory activity results from the inhibition of the inducible COX-2. Therefore, specific COX-2 inhibitors—such as SC558, celecoxib (Celebrex), rofecoxib (Vioxx), valdecoxib, etoricoxib, and lumiracoxib [10–13]—provide a new class of anti-inflammatory drugs with significantly reduced side effects, although they pose an increased risk of cardiovascular events [14, 15].

COX-1 and COX-2 are bifunctional homodimer enzymes, in which each subunit consists of both cyclooxygenase and peroxidase active sites. Firstly, cyclooxygenase catalyzes conversion of arachidonic acid into prostaglandin G₂. Then, the peroxidase enzyme converts prostaglandin G₂ into prostaglandin H₂. The entrance of the COX active site is a long hydrophobic channel consisting of three main residues: Arg120, Tyr355, and Glu524. The active sites of both COX enzymes are very similar. Significantly different residues are found only at position 513 (His in COX-1 vs. Arg

D. Sae-Tang · S. Hannongbua (✉)
Department of Chemistry, Faculty of Science,
Kasetsart University, Bangkok 10900, Thailand
e-mail: fscisph@ku.ac.th

D. Sae-Tang · S. Hannongbua
Center of Nanotechnology,
Kasetsart University Research and Development Institute,
Kasetsart University, Bangkok 10900, Thailand

P. Kittakoop
Chulabhorn Research Institute, Bangkok 10900, Thailand

in COX-2) and position 523 (Ile in COX-1 vs. Val in COX-2). Deletion of a methylene group at position 523 in COX-2 allows access to an additional pocket by the side chains of selective COX-2 inhibitors. Therefore, the structural class of selective COX-2 inhibitors mostly consists of diaryl heterocyclic compounds, with methylsulfonyl (SO_2Me) or sulfonamide (SO_2NH_2) moieties substituted at the para-position of the benzene ring to fit with Arg at position 513, thus retaining COX-2 inhibitory selectivity.

Flurbiprofen (Fig. 1a), or (+/-)-2-(2-fluoro-4-biphenyl)propionic acid, is a chiral 2-arylpropionic acid. (*S*)-Flurbiprofen exhibits non-selective competitive COX inhibition. This compound produces side effects due to its inhibition of both COX isozymes. It inhibits human recombinant COX-1 and COX-2 with IC_{50} values of 0.04 and 0.51 nM [9]. The structure of flurbiprofen is both hydrophilic and hydrophobic at the carboxylate group and biphenyl moiety. Another class of inhibitor is SC558 (Fig. 1b), or 4-[5-(4-bromophenyl)-3-trifluoromethyl-1*H*-1-pyrazolyl]benzenesulfonamide. The structure of this inhibitor is composed of three important parts: a central five-membered ring, two aryl heterocyclic rings containing substituted bromine, and a substituted sulfonamide in the benzene ring. The sulfonamide side chain of SC558 is the most important moiety, specific to the COX-2 binding site.

In the present work, various theoretical methods were used to investigate COX/inhibitor interaction. Quantum chemical calculations and combined quantum mechanics, a method known as ONIOM (our own N-layered integrated molecular orbital and molecular mechanics) [16–18], were

applied to flurbiprofen in order to examine the particular interaction of the inhibitor with the COX-2 binding site. Use of the ONIOM method has recently been applied to large chemical and biological systems. This method is mostly applied to large biomolecules, organometallic complexes, and reaction mechanisms [19–25].

Calculations were based on the particular interaction of the flurbiprofen complex with COX-1 and COX-2. In addition, neutral and ionic forms of flurbiprofen, and some residues at the binding site, were also considered. Moreover, the interaction of SC558 with the COX-2 binding site was studied for comparison with flurbiprofen. The results obtained give greater insight into the specific and non-specific interactions in the binding pocket of cyclooxygenase. This structural information can be a guide for the design and development of a specific COX-2 inhibitor.

Results and discussion

Quantum chemical calculations of inhibitor and enzyme interaction—particular interaction of flurbiprofen with the COX-2 binding site

The interaction energies between flurbiprofen and each residue in the binding pockets of COX-1 and COX-2 were calculated by B3LYP/6-31G(d), B3LYP/6-31G(d,p), and MP2/6-31G(d) methods, as shown in Table 1. The results show that B3LYP/6-31G(d) and B3LYP/6-31G(d,p) calculations do not give significantly different interaction energies (less than 0.50 kJ/mol). This indicates that polarization of the hydrogen atom does not affect the system. The results obtained from the B3LYP and MP2 methods with a 6-31G(d) basis set clearly demonstrate that the key residues in the COX-2 binding site are Arg120 and Tyr355, which show important attractive (−16.62 kJ/mol) and repulsive (14.44 kJ/mol) interactions. According to the results, the MP2 method can exhibit attractive interaction for several residues, in contrast to the B3LYP method [23]. Therefore, in order to achieve a deeper understanding of COX/inhibitor interaction, the MP2 method was used in the next investigation.

The results obtained from MP2/6-31G(d) calculations indicate that attractive interaction plays a significant role in flurbiprofen binding. It is important to note that the main contribution is due to the interaction with Arg120, as indicated by an interaction energy of −16.62 kJ/mol. However, there are some repulsive interactions, especially with Tyr355, Val116, and Leu359 (14.44, 10.47, and 9.50 kJ/mol). Graphical representation of the attractive and repulsive interactions of flurbiprofen with the COX-2 binding site is shown in Fig. 2.

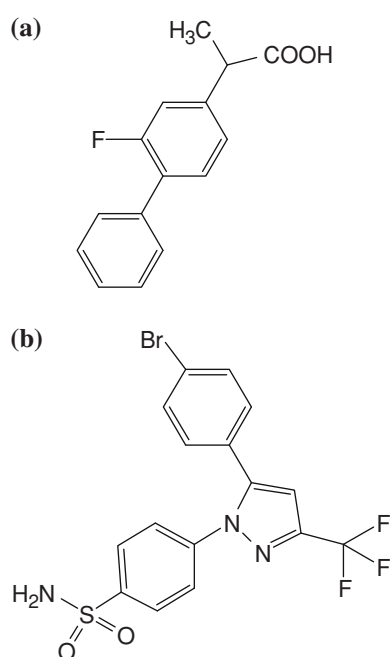
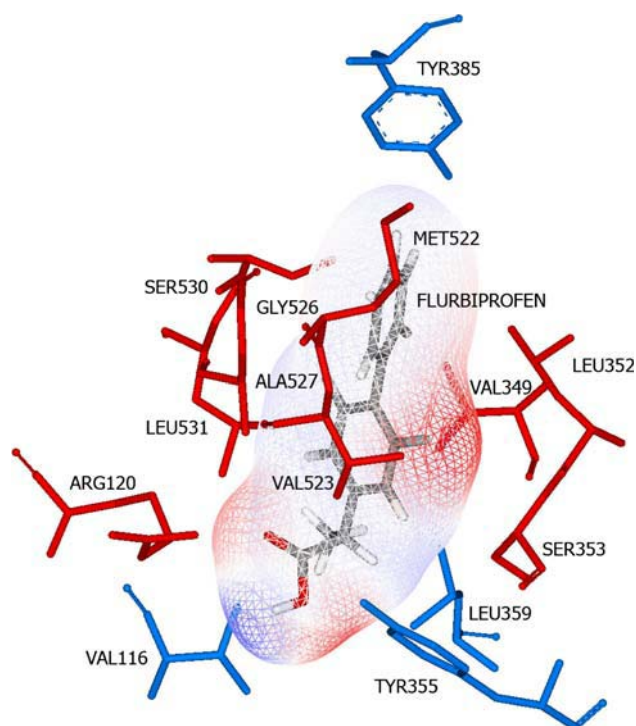


Fig. 1 Molecular structures of flurbiprofen (a) and SC558 (b)

Table 1 Interaction energies (*INT*) of flurbiprofen with individual amino acids (*X_i*) of the neutral system of binding site of COX isozymes (in kJ/mol), calculated at the B3LYP/6-31G(d), B3LYP/6-31G(d,p), and MP2/6-31G(d) levels of theory

	Interaction energies (kJ/mol)			
	COX-2		COX-1	
	B3YLP	MP2	MP2	
	6-31G(d)	6-31G(d,p)	6-31G(d)	6-31G(d)
Val116	13.10	12.77	10.47	7.74
Arg120	−11.30	−10.80	−16.62	−50.40
Val349	15.70	15.57	−0.38	−3.60
Leu352	2.22	2.22	−5.57	−5.15
Ser353	1.09	1.13	−4.02	−3.52
Tyr355	22.19	22.19	14.44	7.95
Leu359	12.85	12.35	9.50	−0.71
Tyr385	14.99	14.57	7.41	−8.12
Met522	−1.47	−1.34	−5.19	−6.24
Val/Ile523*	2.80	2.97	−6.03	−6.57
Gly526	0.75	0.75	−4.56	−7.16
Ala527	10.93	10.80	−8.20	−8.29
Ser530	1.34	1.42	−6.78	−0.04
Leu531	5.02	4.81	−5.78	−5.65

* Val523 in COX-2 and Ile523 in COX-1

**Fig. 2** Attractive (*red*) and repulsive (*blue*) interactions of flurbiprofen with individual residues in the binding site, calculated at the MP2/6-31G(d) level of theory

Comparison of interactions between flurbiprofen and COX-1 and COX-2

It is interesting to note that flurbiprofen can exhibit different interactions with the other isoform, COX-1. Therefore, similar calculations were performed; the results obtained from the MP2/6-31G(d) method are shown in Table 1 in which all residues were performed in a neutral form. There is a large difference in interaction energies between flurbiprofen and the COX-1 and COX-2 binding sites: about 34 kJ/mol. The results confirm that Arg120 is the key residue in both COX active sites. Flurbiprofen exhibits stronger interaction with Arg120 in COX-1 than in COX-2 because of two hydrogen bonds between the carboxylic moiety of flurbiprofen and the guanidine side chain of Arg120 (1.80 and 1.86 Å). In the case of Tyr385, there are opposite results in COX-1 and COX-2 binding. In COX-1, flurbiprofen demonstrates energetically favorable attractive interaction with this residue (−8.12 kJ/mol), whereas in COX-2, the inhibitor shows repulsive interaction (7.41 kJ/mol). Considering Tyr355 residue, the interaction with flurbiprofen in the COX-2 enzyme provides a higher repulsion (about 7 kJ/mol) than in COX-1. Another residue in both isozymes that shows different interaction energies is Ser530. Interaction between flurbiprofen and Ser530 in the COX-2 binding site gives higher attractive energy (about 6.7 kJ/mol) than that of the complex in COX-1. Van der Waals surfaces of these three residues in both enzymes binding with flurbiprofen are shown in Fig. 3. These results indicate that there is a different orientation of the hydroxyl group in the tyrosyl moiety of the inhibitor, which causes the different binding energies.

When considering the main amino acids at position 523 in the binding sites of COX-1 and COX-2, which are Ile and Val, similar interactions were found (−6.57 and −6.03 kJ/mol). This means that upon binding, flurbiprofen is specific to COX-1, and the main contribution comes from the interaction between the inhibitor and Arg120.

Analysis of different interactions of flurbiprofen and SC558 in binding with COX-2

The structures of flurbiprofen and SC558 are roughly equivalent in relation to the COX-2 binding site, based on the fact that the fluorophenyl ring of flurbiprofen overlaps with the bromophenyl ring of SC558. The sulfonamide moiety of SC558 attached to the pyrazole ring causes its selective inhibition of COX-2. For a comparison of the interactions at the same binding site, 23 amino acids in the binding site are shown in Table 2 in which basic and acidic chains were ionized at pH 7.0 based on their *pK_a* values. The calculated interaction energies between each inhibitor

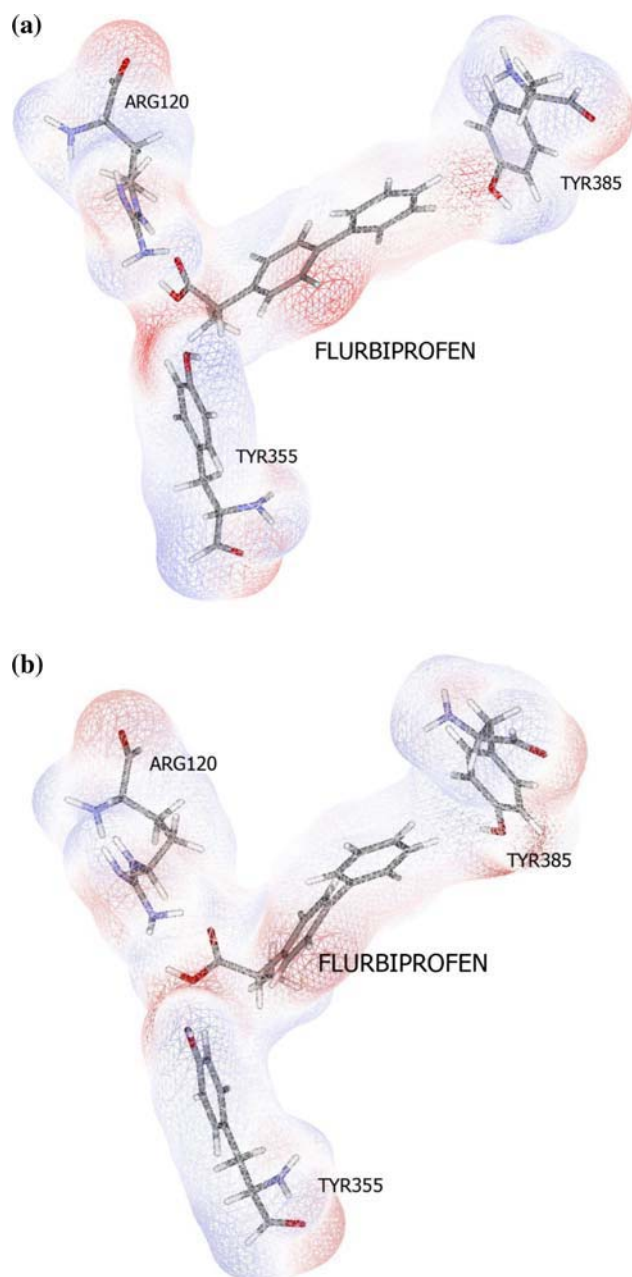


Fig. 3 Van der Waals surface of flurbiprofen with Arg120, Tyr355, and Tyr385 in COX-1 (a) and in COX-2 (b)

and individual residues are shown in this table. It is interesting to note that the general interaction within the binding site of flurbiprofen demonstrates attractive interaction with COX-2, whereas SC558 exhibits repulsive interaction. The results indicate that Arg120, Tyr355, Leu359, Arg513, Phe518, and Val523 residues show widely different interactions with flurbiprofen and SC558. There is a greater attractive interaction between flurbiprofen and Arg120 (−39.18 kJ/mol), while Tyr355 interacts more favorably with SC558 (−15.20 kJ/mol) than flurbiprofen (12.89 kJ/mol). The Val523 residue, which is the

Table 2 Interaction energies (*INT*) of flurbiprofen and SC558 with individual amino acids (X_i) (in kJ/mol) of the ionic system of the COX-2 binding pocket, calculated at the MP2/6-31G(d) level of theory

	Interaction energies (kJ/mol)	
	Flurbiprofen	SC558
His90	0.17	13.81
Val116	0.88	−2.26
Arg120	−39.18	−16.20
Gln192	−0.21	8.29
Val349	−0.59	−3.47
Leu352	−5.57	6.45
Ser353	−4.02	−9.08
Gly354	−0.42	−2.39
Tyr355	12.89	−15.20
Leu359	13.56	−2.13
Tyr385	7.12	11.13
Trp387	−3.18	−0.38
Arg513	−2.43	89.79
Ala516	−0.13	−5.90
Ile517	−0.29	−7.53
Phe518	−2.34	18.71
Gly519	0.13	−3.39
Met522	−5.11	−7.49
Val523	−5.90	6.57
Gly526	−3.47	−3.43
Ala527	−8.25	−2.80
Ser530	−6.57	−3.60
Leu531	−5.90	−2.01
Total	−58.81	67.48

key residue for discrimination of both isozymes, shows selectivity with the sulfonyl phenyl ring of SC558. These results clearly indicate that SC558 interacts with COX-2 better than flurbiprofen does, based on repulsive interaction—particularly in the case of Arg513.

ONIOM calculations

The main focus of our study is to investigate the specific interaction of flurbiprofen with key amino acids in the binding sites: Arg120, Tyr355, and Val 523. The binding energies obtained from ONIOM2 calculations are given in Table 3. The results show that the binding energies of flurbiprofen/Arg120 are higher than for flurbiprofen/Val523. This suggests that the hydrogen bonding between the carboxylic group of flurbiprofen and the guanidine group of Arg120 is stronger. Hydrogen bond distances are in the range of 1.99–2.28 Å (obtained by models A–D), whereas no hydrogen bonding is observed between Tyr355 and flurbiprofen.

Table 3 ONIOM2 binding energies with BSSE-CP correction of flurbiprofen bound into the COX-2 binding pocket, and key residues were set in the inner layer

ONIOM2 method	Binding energy (kJ/mol)			Hydrogen bonding distance (Å)	
	Arg120	Tyr355	Val523	N _{Arg120} –H–O _{FLP}	O _{Tyr355} –H–O _{FLP}
Model A	–39.73	–45.00	–39.06	2.24	4.17
Model B	–41.15	–36.04	–39.22	2.28	4.18
Model C	–54.29	–48.06	–43.45	1.97	4.10
Model D	–52.45	–57.06	–43.12	1.99	4.09

Hydrogen bonding distances are also indicated

Concerning the interaction between flurbiprofen and Tyr355, the bond distance between O_{TYR355}–H–O_{FLP} from model D is about 1.4 Å longer than that observed in the X-ray crystallographic structure (3.09 Å) (3PGH.pdb). The calculated results indicate no hydrogen bond formation between the thyroxyl side chain of Tyr355 and flurbiprofen (see Fig. 4). It is also known that a significant difference between COX-1 and COX-2 is due to the difference at position 523: Ile in COX-1 and Val in COX-2. Deletion of a methylene group at 523 in COX-2 allows access to an additional pocket. Van der Waals interaction between Val523 and flurbiprofen implies that the interaction might not be significant in binding with COX-2. Therefore, the selective COX-2 inhibitor exploits the additional pocket for enhanced binding through the sulfonamide or methylsulfone moiety; this clearly discriminates selective from non-selective COX inhibitors.

Interaction of key residues with flurbiprofen

In order to better understand the particular interaction of the key residues in the binding site with flurbiprofen, a comparison of their interactions was performed. In this case, two different states, neutral and ionic charge environments, were also considered. The ONIOM2(B3LYP/6-31G(d):PM3) method was used, and calculated results are shown in Table 4.

Effects of neutral and ionic charges on binding energy

Flurbiprofen and each key residue were included at a high level of the ONIOM2 model, while the other residues were set at a low level, for both neutral and ionic systems. It was clearly seen that in the neutral system, the binding energies of the three models are not significantly different. However, in ionic systems, binding energies between flurbiprofen and each key residue are in the following order: flurbiprofen—Arg120 ≫ Val523 > Tyr355. This means that ionic charges in the system play an important role in inhibitor-enzyme interaction, particularly the interaction between flurbiprofen and Arg120. However, too high-binding energies of the

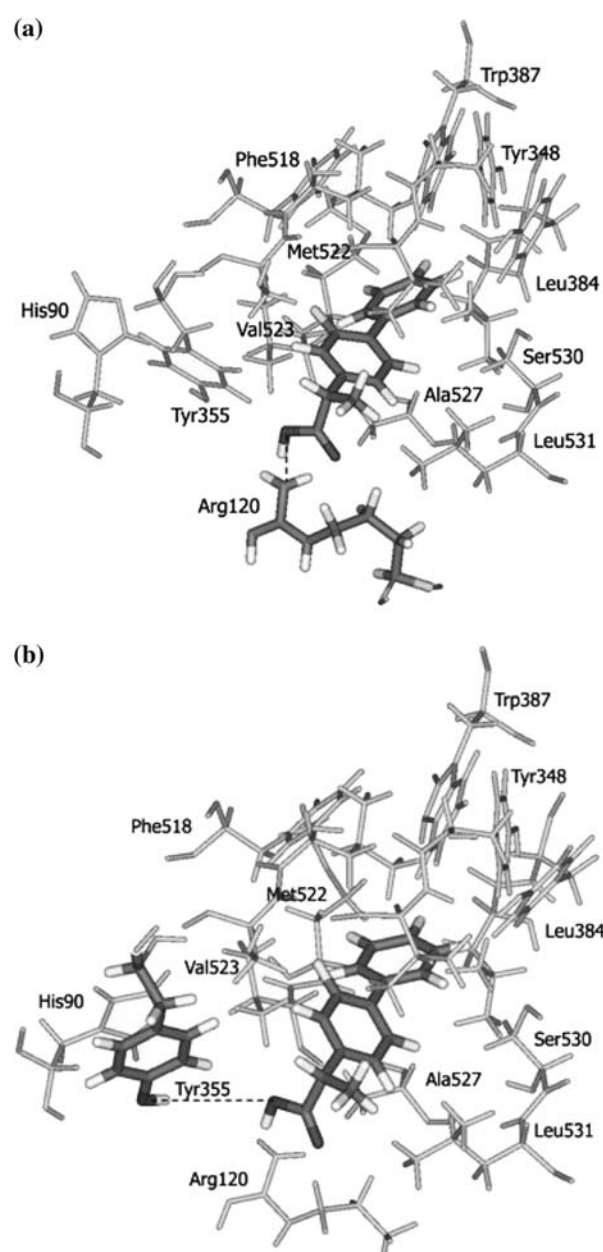
**Fig. 4** Hydrogen bonding between flurbiprofen/Arg120 (1.99 Å) (a) and flurbiprofen/Tyr355 (4.10 Å) (b), obtained from ONIOM2 (model D) calculations

Table 4 Binding energy of flurbiprofen bound into the COX-2 binding pocket, calculated by ONIOM2 (B3LYP/6-31G(d):PM3) method

	Binding energies (kJ/mol)		
	Flurbiprofen/ Arg120	Flurbiprofen/ Tyr355	Flurbiprofen/ Val523
Inner layer			
Outer layer	Rest residues	Rest residues	Rest residues
Neutral	−54.29	−48.06	−43.45
Ionic	−429.40	−145.34	−176.44

Flurbiprofen and each key residue (Arg120, Tyr355, and Val523) are set in different states (neutral and ionic) (in kJ/mol)

ionic system calculated from ONIOM calculations might come from electrostatic interaction between the carboxylate group of flurbiprofen and ionized residues such as Arg120. This point will be further investigated.

Effects of neutral and ionic charges on hydrogen bonding

Considering the hydrogen bonding interaction in the flurbiprofen-Arg120 ONIOM model, ionic charges produce a stronger hydrogen bond, with a distance of 1.68 (2.65) Å as compared to 1.97 (2.90) Å in the neutral system (values in parentheses are the intermolecular distance between the atom of the hydrogen bond donor and the atom of the hydrogen bond acceptor). X-ray crystallographic data provide a hydrogen bond distance of about 2.96 Å. Superimposition of the neutral and ionic charge structures, as well as the X-ray structure, are shown in Fig. 5.

In the flurbiprofen-Tyr355 model, the ionic system shows hydrogen bonding between carboxylate of the inhibitor and the hydroxyl group of the tyrosine residue of about 1.77 (2.76) Å, while the neutral system does not

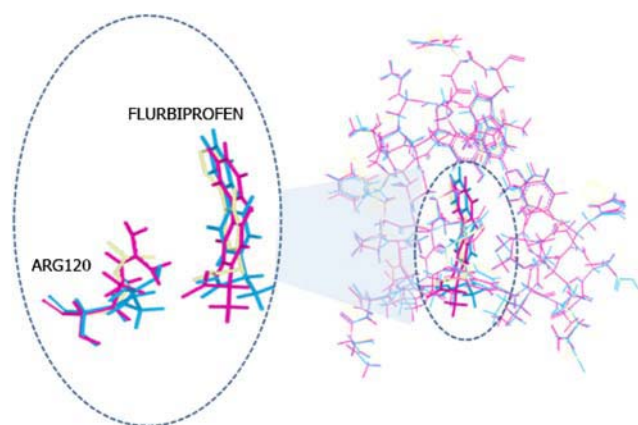


Fig. 5 Superimposed structures of neutral (blue) and ionic (red) systems obtained from ONIOM2 calculations, compared with X-ray structure (yellow). Arg120 and flurbiprofen are in inner layer, and the other residues are in outer layer

show hydrogen bonding. In contrast, in the flurbiprofen-Val523 model no hydrogen bonding is detected in either system. This confirms that most of the interaction energy of flurbiprofen/COX-2 binding comes from the main interaction with Arg120 (a result found in both inner and outer layers of model calculations).

Conclusions

The individual interaction energy of flurbiprofen with the COX-2 binding pocket was investigated. The results show that MP2/6-31G(d) is an appropriate method for studying enzyme-inhibitor interaction. A comparison of the interactions of flurbiprofen and SC558 in the COX-2 binding site clearly indicates that Arg120 strongly interacts with both inhibitors. However, in the case of SC558, the repulsive interaction with Arg513 in the COX-2 binding site plays a significant role in specific inhibition. ONIOM2 calculation was a useful method to investigate the interactions between flurbiprofen and key amino acids in the binding pocket. The results obtained indicate that flurbiprofen can produce moderate hydrogen bonding with Arg120, while no hydrogen bond is detected with Tyr355 or Val523.

Methods

Systems studied

Molecular structures of flurbiprofen bound with COX-1 and COX-2 were obtained from the 3.10 and 2.50 Å resolved crystal structures 1CQE.pdb and 3PGH.pdb [26, 27]. Based on these structures, only chain A was selected. The systems studied were defined for both COX enzymes consisting of 23 residues surrounding the flurbiprofen inhibitor in COXs' binding pockets with at least one atom interacting with any atom of flurbiprofen within the interatomic distance of 6 Å. In order to compare different enzyme/inhibitor interactions, SC558 was selected; the bound structure was obtained from the 3.0 Å resolved crystal structure (1CX2.pdb). Residues surrounding the SC558 within 6 Å were selected using the same criteria as in the before-mentioned system. Therefore, the residues consisting in 6 Å of both COXs/flurbiprofen and COX-2/SC558 are His90, Val116, Arg120, Val349, Leu352, Ser353, Gly354, Tyr355, Leu359, Tyr385, His/Arg513, Ser/Ala516, Ile517, Phe518, Gly519, Met522, Ile/Val523, Glu524, Gly526, Ala527, Ser530, and Leu531. Terminated cutting residues, the N- and C-terminal ends of cut residues were capped with an acetyl group (CH₃CO-) and a methyl amino group (-NHCH₃) from the adjacent residues in all systems. Hydrogen atoms were added to the

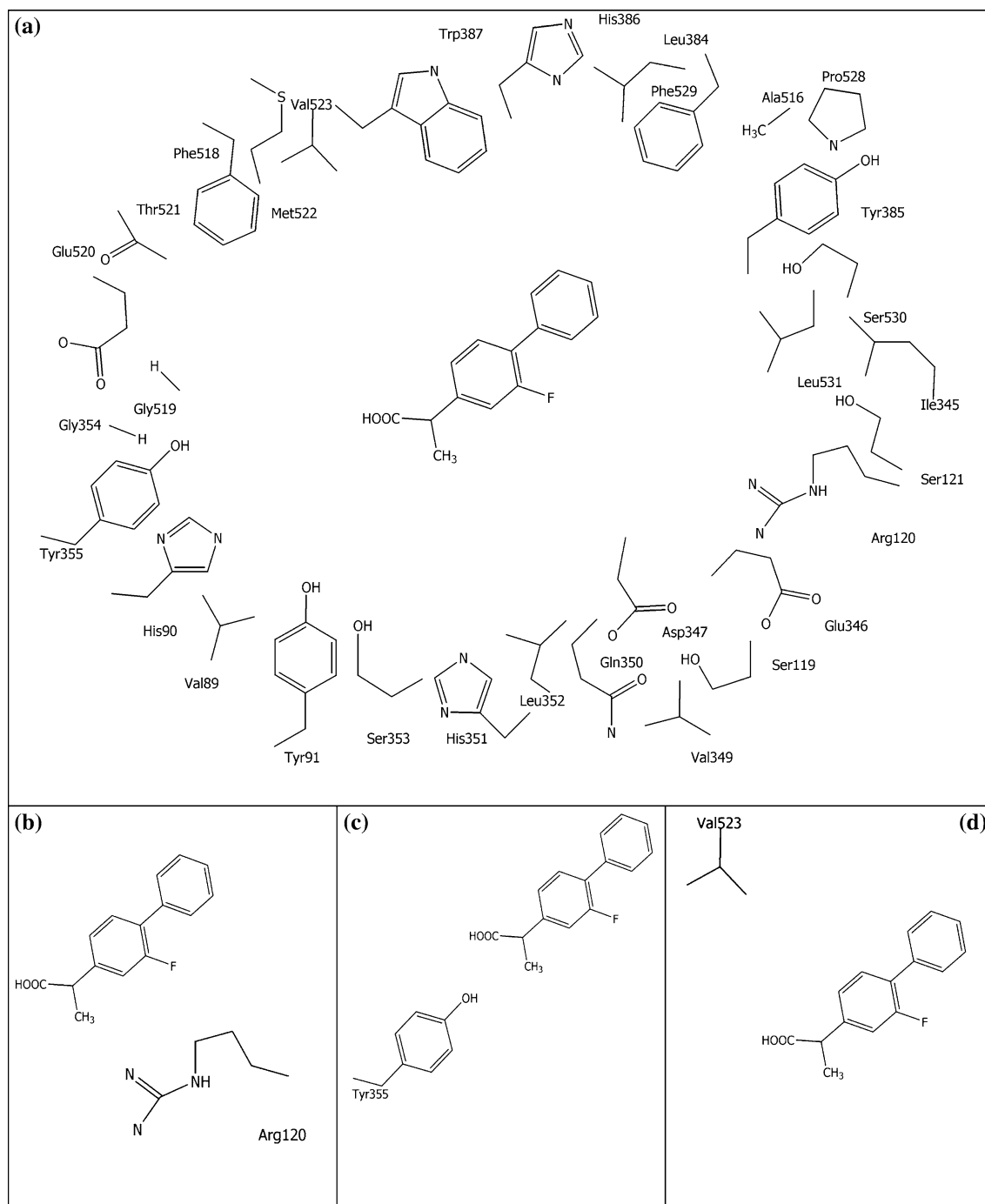


Fig. 6 Schematic representation of the flurbiprofen bound to the COX-2 binding site, consisting of 32 residues (a). For ONIOM2 calculations, each key residue (Arg120, Tyr355, and Val523) was added in the inner layer (b–d)

X-ray structure to generate a complete structure of the model using the SYBYL 7.0 program [28]. The structure was optimized with the semiempirical PM3 method [29], in which all the heavy atoms of the amino acids in the pocket were fixed in the X-ray structure. Therefore, only the geometry and position of the inhibitor molecules, as well as the positions of all the hydrogen atoms, were optimized.

These structures were used to generate the starting geometries for all calculations.

Inhibitor-individual residue interaction model setup

Interaction energies between the inhibitor (flurbiprofen or SC558) with the individual residue (defined as X_i) were

calculated using the optimized complex structures. For flurbiprofen, there are residues within 6 Å of the binding pocket consisting of 14 amino acids: Val116, Arg120, Val349, Leu352, Ser353, Tyr355, Leu359, Tyr385, Met522, Val523, Gly526, Ala527, Ser530, and Leu531. Seven amino acids within 6 Å of the binding pocket of the SC558 inhibitor were added due to the sulfonamide moiety. These residues are His90, Gln192, Arg513, Ala516, Ile517, Phe518, and Gly519. The interaction energy of each inhibitor-residue pair was calculated by three methods: B3LYP/6-31G(d), B3LYP/6-31G(d,p), and MP2/6-31G(d). The definition of the interaction energy (*INT*) is shown in Eq. 1:

$$INT_{(\text{Inhibitor}-X_i)} = E_{(\text{Inhibitor}-X_i)} - E_{(\text{Inhibitor})} - E_{(X_i)} \quad (1)$$

Calculations by the B3LYP/6-31G(d), B3LYP/6-31G(d,p), and MP2/6-31G(d) methods were used to search for the most appropriate method for this system. Then, interactions between the two COX isozymes and flurbiprofen were analyzed. In addition, interaction energies between the two inhibitors (flurbiprofen and SC558) and individual residues were compared. All calculations were performed using the GAUSSIAN 03 program package [30], implemented on a Linux PC-based multiprocessor (3.4 GHz).

Analysis of COX-2/flurbiprofen interaction using the ONIOM2 method

The ONIOM2 method was used to investigate the interaction between flurbiprofen and the residues in the COX-2 binding site (Fig. 6). In this study, we focused on the interactions between flurbiprofen and three important residues: Arg120, Tyr355, and Val523. Arg120 is a key amino acid of the active site of the substrate, in which the guanidinium group functions to stabilize the carboxylate of the inhibitor; while Tyr355 is a key amino acid for the flurbiprofen binding site. Both residues are located near the hydrophobic channel-binding region of the protein. Moreover, Val523 is considered to be the amino acid that differentiates the COX-1 and COX-2 isozymes. Therefore, these three residues were particularly investigated.

The ONIOM2 systems were set up by separation of the model into two layers. The inner layer, or the “interaction region,” was composed of flurbiprofen and the residues Arg120, Tyr355, and Val523, using a high level of calculation, while the outer layer (the “environmental region”) consisting of the remaining residues was treated at a lower level of calculation. Therefore, different ONIOM2 methods—*ab initio* and density functional theory combined with PM3—were applied to calculate the structural information and estimate interaction energies between flurbiprofen and

the COX-2 binding site. The calculation setup also included partitioning into a model layer and a real layer. Combinations of different levels of theory—HF/6-31G(d), HF/6-31G(d,p), B3LYP/6-31G(d), and B3LYP/6-31G(d,p)—with PM3 for ONIOM2 setup were performed, resulting in Models A, B, C, and D, as follows:

Model A : ONIOM2(HF/6-31G(d):PM3)
 Model B : ONIOM2(HF/6-31G(d,p):PM3)
 Model C : ONIOM2(B3LYP/6-31G(d):PM3)
 Model D : ONIOM2(B3LYP/6-31G(d,p):PM3)

In addition, the ionic system was also considered and calculated by the ONIOM2(B3LYP/6-31G(d):PM3) method. In this case, flurbiprofen was set as the ionic form, as well as ionic residues in the binding site. All ONIOM2 calculations were corrected by basis set superposition error using counterpoint (BSSE-CP) correction [23].

Acknowledgments This work was supported by the Thailand Research Fund (RTA5080005). DS is grateful to the Thailand Graduate Institute of Science and Technology (TG-B-11-22-11-733D) for a scholarship. The authors express their appreciation to: the National Center of Excellence for Petroleum, Petrochemicals, and Advanced Materials for its postgraduate education and research programs; the Commission on Higher Education; the Kasetsart University Research and Development Institute (KURDI); and the National Nanotechnology Center (through the NANOTEC Center of Excellence and Computational Nanoscience Consortium) and Large Scale Simulation Research Laboratory (LSR) at NECTEC for the use of its computing and research facilities. Thanks are also due to Christopher Salisbury for reading of a manuscript.

References

- Herschman HR (1996) *Biochim Biophys Acta* 1299:125
- Fu JY, Masferrer JL, Seibert K (1990) *J Biol Chem* 265:16737
- Kujubu DA, Fletcher BS, Varnum BC (1991) *J Biol Chem* 266:12866
- DeWitt DL (1999) *Mol Pharmacol* 4:625
- Marnett LJ, Kalgutkar AS (1998) *Curr Opin Chem Biol* 2:482
- Hawkey CJ (1999) *Lancet* 353:307
- Loll PJ, Picot D, Garavito M (1995) *Nat Struct Biol* 2:637
- Loll PJ, Picot D, Ekabo O, Garavito M (1996) *Biochemistry* 35:7330
- Barnett J, Chow J, Ives D (1994) *Biochim Biophys Acta* 1209:130
- Penning TD, Talley JJ, Bertenshaw SR, Carter JS, Collins PW, Docter S, Graneto MJ, Lee LF, Malecha JW, Miyashiro JM, Rogers RS, Rogier DJ, Yu SS, Anderson GD, Burton EG, Cogburn JN, Gregory SA, Koboldt CM, Perkins WE, Seibert K, Veenhuizen AW, Zhang YY, Isakson PC (1997) *J Med Chem* 40:1347
- Prasit P, Wang Z, Brideau C, Chan CC, Charleson S, Cromlish W, Ethier D, Evans JF, Ford-Hutchinson AW, Gauthier JY, Gordon R, Guay J, Gresser M, Kargman S, Kennedy B, Leblanc Y, Leger S, Mancini J, O'Neil GP, Ouellet M, Percival MD, Perrier H, Riendeau D, Rodger I, Tagari P, Therien M, Vickers P, Wong E, Xu LJ, Young RN, Zamboni R, Boyce S, Rupniak N, Forrest M, Visco D, Patrick D (1999) *Bioorg Med Chem Lett* 9:1773

12. Talley JJ, Brown DL, Carter JS, Graneto MJ, Koboldt CM, Masferrer JL, Perkins WE, Rogers RS, Shaffer AF, Zhang YY, Zweifel BS, Seibert KJ (2000) *J Med Chem* 43:775
13. Sorbera LA, Castaner RM, Silvestre J, Castaner J (2001) *Drugs Future* 26:346
14. Garcia-Rodriguez LA, Hernández-Díaz S (2003) *Epidemiology* 14:240
15. Szabo G, Fischer J, Kis-Varga A, Gyires K (2008) *J Med Chem* 51:142
16. Dapprich S, Komaromi I, Byun KS, Morokuma K, Frisch MJ (1999) *J Mol Struct (THEOCHEM)* 462:1
17. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) *J Phys Chem* 100:19357
18. Humbel S, Sieber S, Morokuma K (1996) *J Chem Phys* 105:1959
19. Maseras F, Morokuma K (1995) *J Comput Chem* 16:1170
20. Matsubara T, Sieber S, Morokuma K (1996) *Int J Quantum Chem* 60:1101
21. Vreven T, Morokuma M (2000) *J Comput Chem* 21:1419
22. Srivastava P, Hannongbua S (2008) *Chem Med Chem* 3:803
23. Kuno M, Hongkrenkai R, Hannongbua S (2006) *Chem Phys Lett* 424:172
24. Saen-oon S, Kuno M, Hannongbua S (2005) *Proteins Struct Funct Bioinform* 61:859
25. Nunriam P, Kuno M, Saen-oon S, Hannongbua S (2005) *Chem Phys Lett* 405:198
26. Kurumbai RG, Stevens AM, Gierse JK, McDonald JJ, Stegeman RA, Pak JY, Gildehaus D, Miyashiro JM, Penning TD, Seibert K, Isakson PC, Stallings WC (1996) *Nature* 384:644
27. Picot D, Loll PJ, Gaeavito RM (1994) *Nature* 367:243
28. SYBYL 7.0, Discovery Software Suite, Tripos Associates, Inc., St. Louis MO
29. Stewart JJP (1989) *J Comput Chem* 10:209
30. Frisch MJ, Trucks GW, Schlegel HB, Scuseria, GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) *Gaussian 03, Revision B.04*, Gaussian, Inc., Wallingford CT

This paper is published as part of a PCCP Themed Issue on:

Nucleic Acid Simulations

Guest Editors: Modesto Orozco and Charles Laughton



Editorial

Nucleic Acid Simulations

Modesto Orozco and Charles Laughton, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b921472k](https://doi.org/10.1039/b921472k)

Perspectives

Viral assembly: a molecular modeling perspective

Stephen C. Harvey, Anton S. Petrov, Batsal Devkota and Mustafa Burak Boz, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b912884k](https://doi.org/10.1039/b912884k)

Simulation of DNA catenanes

Alexander Vologodskii and Valentin V. Rybenkov, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b910812b](https://doi.org/10.1039/b910812b)

Papers

Stabilization of radical anion states of nucleobases in DNA

Alexander A. Voityuk, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b910690a](https://doi.org/10.1039/b910690a)

Effects of the biological backbone on DNA–protein stacking interactions

Cassandra D. M. Churchill, Lex Navarro-Whyte, Lesley R. Rutledge and Stacey D. Wetmore, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b910747a](https://doi.org/10.1039/b910747a)

The impact of monovalent ion force field model in nucleic acids simulations

Agnes Noy, Ignacio Soteras, F. Javier Luque and Modesto Orozco, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b912067j](https://doi.org/10.1039/b912067j)

Structural ultrafast dynamics of macromolecules: diffraction of free DNA and effect of hydration

Milo M. Lin, Dmitry Shorokhov and Ahmed H. Zewail, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b910794k](https://doi.org/10.1039/b910794k)

Simulation of DNA double-strand dissociation and formation during replica-exchange molecular dynamics simulations

Srinivasaraghavan Kannan and Martin Zacharias, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b910792b](https://doi.org/10.1039/b910792b)

Sensors for DNA detection: theoretical investigation of the conformational properties of immobilized single-strand DNA

Vincenzo Barone, Ivo Cacelli, Alessandro Ferretti, Susanna Monti and Giacomo Prampolini, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b914386f](https://doi.org/10.1039/b914386f)

Relaxation dynamics of nucleosomal DNA

Sergei Y. Ponomarev, Vakhtang Putkaradze and Thomas C. Bishop, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b910937b](https://doi.org/10.1039/b910937b)

Dynamics of a fluorophore attached to superhelical DNA: FCS experiments simulated by Brownian dynamics

Tomasz Wocjan, Jan Krieger, Oleg Krichinsky and Jörg Langowski, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b911857h](https://doi.org/10.1039/b911857h)

A detailed binding free energy study of 2 : 1 ligand–DNA complex formation by experiment and simulation

Witcha Treesuwan, Kitiyaporn Wittayanarakul, Nahoum G. Anthony, Guillaume Huchet, Hasan Alniss, Supa Hannongbua, Abedawn I. Khalaf, Colin J. Suckling, John A. Parkinson and Simon P. Mackay, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b910574c](https://doi.org/10.1039/b910574c)

Molecular simulation of conformational transitions in biomolecules using a combination of structure-based potential and empirical valence bond theory

Giuseppe de Marco and Péter Várnai, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b917109f](https://doi.org/10.1039/b917109f)

Dependence of A-RNA simulations on the choice of the force field and salt strength

Ivana Bešševová, Michal Otyepka, Kamila Réblová and Jiří Šponer, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b911169g](https://doi.org/10.1039/b911169g)

Protein–DNA binding specificity: a grid-enabled computational approach applied to single and multiple protein assemblies

Krystyna Zakrzewska, Benjamin Bouvier, Alexis Michon, Christophe Blanchet and Richard Lavery, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b910888m](https://doi.org/10.1039/b910888m)

Evaluation of molecular modelling methods to predict the sequence-selectivity of DNA minor groove binding ligands

Hao Wang and Charles A. Laughton, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b911702d](https://doi.org/10.1039/b911702d)

Mesoscale simulations of two nucleosome-repeat length oligonucleosomes

Tamar Schlick and Ognjen Perišić, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b918629h](https://doi.org/10.1039/b918629h)

On the parameterization of rigid base and basepair models of DNA from molecular dynamics simulations

F. Lankaš, O. Gonzalez, L. M. Heffler, G. Stoll, M. Moakher and J. H. Maddocks, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b919565n](https://doi.org/10.1039/b919565n)

Charge transfer equilibria of aqueous single stranded DNA

Marco D'Abramo, Massimiliano Aschi and Andrea Amadei, *Phys. Chem. Chem. Phys.*, 2009

DOI: [10.1039/b915312h](https://doi.org/10.1039/b915312h)

A detailed binding free energy study of 2 : 1 ligand–DNA complex formation by experiment and simulation†

Witcha Treesuwan,^{‡a} Kitiyaporn Wittayanarakul,^{‡b} Nahoum G. Anthony,^b Guillaume Huchet,^a Hasan Alniss,^b Supa Hannongbua,^a Abedawn I. Khalaf,^c Colin J. Suckling,^c John A. Parkinson^c and Simon P. Mackay^{*b}

Received 1st June 2009, Accepted 12th August 2009

First published as an Advance Article on the web 22nd September 2009

DOI: 10.1039/b910574c

In 2004, we used NMR to solve the structure of the minor groove binder thiazotropsin A bound in a 2 : 1 complex to the DNA duplex, d(CGACTAGTCG)₂. In this current work, we have combined theory and experiment to confirm the binding thermodynamics of this system. Molecular dynamics simulations that use polarizable or non-polarizable force fields with single and separate trajectory approaches have been used to explore complexation at the molecular level. We have shown that the binding process invokes large conformational changes in both the receptor and ligand, which is reflected by large adaptation energies. This is compensated for by the net binding free energy, which is enthalpy driven and entropically opposed. Such a conformational change upon binding directly impacts on how the process must be simulated in order to yield accurate results. Our MM-PBSA binding calculations from snapshots obtained from MD simulations of the polarizable force field using separate trajectories yield an absolute binding free energy ($-15.4 \text{ kcal mol}^{-1}$) very close to that determined by isothermal titration calorimetry ($-10.2 \text{ kcal mol}^{-1}$). Analysis of the major energy components reveals that favorable non-bonded van der Waals and electrostatic interactions contribute predominantly to the enthalpy term, whilst the unfavorable entropy appears to be driven by stabilization of the complex and the associated loss of conformational freedom. Our results have led to a deeper understanding of the nature of side-by-side minor groove ligand binding, which has significant implications for structure-based ligand development.

Introduction

Ligands interacting with DNA have the ability to regulate the gene machinery at the most fundamental level of expression.^{1–3} Through their selective association with a particular DNA sequence, such ligands may prevent protein binding and play a role in treating diseases that result from aberrant gene expression.^{4–7} The development of minor groove binders (MGBs) proceeded from the observation that two natural antibiotics, netropsin and distamycin, bind to A and T containing regions of the minor groove by a combination of hydrogen bonding with the bases on the groove floor facilitated by their natural curvature. Replacement of *N*-methyl pyrrole (Py) with *N*-methyl imidazole (Im) enabled the accommodation of the G–NH₂ by hydrogen bonding.^{8–11} A significant breakthrough in the field was the observation

that a number of MGBs could bind in the minor groove as a 2 : 1 complex, in a side-by-side fashion with the heterocyclic rings stacking against each other.¹² Since then, MGBs have been prepared that can discriminate not only GC from AT, but GC from CG and AT from TA base pairs.^{13–15} Whilst hydrogen bonding to the groove floor endows specificity for particular sequences, it is not the only driving force for association; lipophilic forces, particularly interactions with the sugar moieties that comprise the groove walls are also highly relevant.^{16,17} Furthermore, Haq *et al.*¹⁸ showed that the electrostatic interactions between Hoechst 33258 and DNA are not major components of the driving force for binding; they simply replace the electrostatic interactions that exist between the water and ions with the uncomplexed ligand and DNA species.¹⁸ This balance between enthalpic and entropic contributions to binding is the subject of extensive research, and appears to vary with MGB structure and the binding sequence of the DNA.¹⁹

When developing new DNA-binding ligands, it is therefore important to understand the factors that work both in favor of and counter to ligand binding. The molecular recognition event can be understood through detailed structural analyses using X-ray crystallography and NMR spectroscopy, but the intricacies of dynamic phenomena within ligand–DNA complexes need to be probed by other techniques such as molecular dynamics. The energetics associated with the

^a Chemistry Department and Center of Nanotechnology, Kasetsart University, Bangkok 10900, Thailand

^b Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 27 Taylor Street, Glasgow, UK G4 0NR. E-mail: simon.mackay@strath.ac.uk

^c WestCHEM Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, UK G1 1XL

† Electronic supplementary information (ESI) available: DNA step parameters: comparison between simulated and experimental structures. See DOI: 10.1039/b910574c

‡ Contributed equally to this work.

binding event can be elucidated through thermodynamic analysis using techniques such as isothermal titration calorimetry (ITC) and circular dichroism. Our approach to ligand design is conducted in cognizance of these issues. We have prepared a large library of MGBs made up from new heterocyclic and head–tail groups that seek to recognize both the hydrogen bonding capacity of the groove floor to achieve specificity, and to exploit the lipophilic nature of the groove walls to enhance affinity.^{20–23} We have found that increasing the size of the heterocyclic *N*- or *C*-alkyl groups, if placed towards the tail of the MGB, can offset side-by-side binding with the effect of extending the reading frame of the ligand from four to six base pairs. The first, well-characterized example of this effect was our detailed NMR study of the DNA duplex d(CGACTAGTCG)₂ with thiazotropsin A (formyl-Py-Py-ⁱPrTh-DMAP, where ⁱPrTh represents thiazole containing an isopropyl moiety and DMAP the dimethylamino-propyl tail, Fig. 1). Furthermore, the sulfur and isopropyl of the thiazole not only improved lipophilicity, but also ensured that the heterocyclic nitrogen was inward facing, which enhanced sequence reading by introducing a hydrogen bond between the exocyclic amine of G7 and the ⁱPrTh nitrogen.

To relate these structural examinations to binding energies, we have experimentally investigated by ITC the thermodynamics of thiazotropsin A interacting with the same DNA sequence. In parallel, we have used explicit solvent molecular dynamics (MD) simulations to relate the structural dynamics of ligand–DNA interactions explored by NMR with the experimental thermodynamic parameters to provide a unique insight into the processes that determine molecular association in this system.

Whilst MD simulations can provide predictions of ligand–DNA interactions through absolute binding free energy calculations, it is challenging to achieve accurate results.^{24–26} Thermodynamic integration (TI), free energy perturbation (FEP), linear interaction energy (LIE) or combined molecular mechanics Poisson–Boltzmann and surface area (MM-PBSA) methods constitute the four different approaches for free energy MD simulations which have been applied to, for example, netropsin, distamycin, Hoechst 33258 and DAPI.^{26–28} Although results have been rather encouraging using FEP methods,²⁸ the technique is not applicable to the general case. Accurate absolute binding affinities have been obtained, but only through extremely high computational cost.²⁹ Among the approximate methods, the MM-PBSA methodology has also been used to determine free energies in a number of nucleic acid systems,^{24,25,30–32} although results tend to overestimate the actual experimental value.³⁰ The advantage of the MM-PBSA approach is the speed, reproducibility, reliability and efficiency of the calculations,

compared with the resource requirements of FEP, LIE and TI. The MM-PBSA/GBSA approach was first introduced by Srinivasan³¹ and represents a post-processing method of evaluating binding free energies and absolute free energies of molecules in solution by the analysis of sets of structures collected by MD or Monte Carlo methods. Representative structures from the trajectory are post-processed with solvent and counterions removed to calculate the free energy (*G*) according to eqn (1):

$$\Delta G_{\text{binding}}^{\text{solvated}} = \langle G_{\text{complex}} \rangle - (\langle G_{\text{receptor}} \rangle + \langle G_{\text{ligand}} \rangle) \quad (1)$$

In previous studies aimed at analyzing ligand binding to DNA, eqn (1) has been applied to just the trajectory snapshots from the complex, which assumes that snapshots of the receptor and ligand taken from the complex trajectory are of comparable free energy to their separate trajectories. However, when the receptor and/or the ligand undergo significant conformational changes upon binding, this approach is less reliable and separate trajectories need to be considered to gain free energy values approaching experimental data, as demonstrated by Reyes and Kollman's study of RNA–protein associations.³³ Separate trajectories can only be considered if there are experimental structures available for the individual components, otherwise simulations extending beyond the nanosecond timescale may not yield accurate structures and ultimately produce imprecise energetic values for the separate systems. In this article, we have compared binding free energies derived using the single trajectory approach (where only one simulation of the receptor–ligand complex is performed and the co-ordinates of the receptor and ligand are subsequently extracted from the simulation of the complex for the purpose of free energy calculations) and the separate trajectory approach (where three distinct simulations are performed to obtain co-ordinates of the complex, receptor and ligand) to establish whether significant conformational changes impact upon our binding calculations.

According to the previous work we have alluded to,^{24,34} calculating binding free energies using the MM-PBSA method from MD simulations has employed non-polarizable force fields that tend to overestimate the experimentally determined values. Hence, our study also includes a comparison between calculated binding free energies obtained from simulations using the polarizable and non-polarizable AMBER force fields, which differ primarily in their treatment of electrostatic interactions. The ff03 is a non-polarizable force field which is itself a modification of the ff94³⁵ and ff99³⁶ force fields. AMBER ff02 is a polarizable force field that explicitly includes induction effects by the use of dipole polarizabilities on all atoms, and has the effect of significantly increasing computer time compared with ff03. The polarizable ff02 has been applied previously to explore protein–ligand interactions with limited success.³⁷ Although DNA itself has been simulated using ff02 and ff99 to compare the reorganization energies³⁸ and their ability to maintain its structural integrity,³⁹ binding free energy calculations for ligand–DNA interactions on trajectories generated using these force fields have not previously been performed in a comparative manner. We herein report the first attempt using MD simulations to obtain

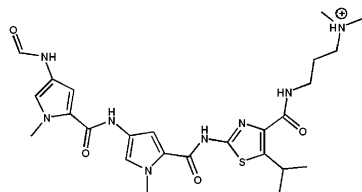


Fig. 1 Structure of thiazotropsin A.

absolute binding free energy data for 2 : 1 side-by-side binding of a ligand that evaluates the single and separate trajectory approaches using polarizable and non-polarizable force fields, all underpinned by structural and thermodynamic experimental assessment.

Methods

Isothermal titration calorimetry

Thiazotropsin A was prepared as before²⁰ and dissolved in degassed 0.01 M PIPES (piperazine-*N,N'*-bis(2-ethanesulfonic acid)), 0.2 M NaCl, 0.001 M EDTA which had been adjusted to pH 7.0. d(GCGACTAGTCGC)₂ was purchased from MWG-BIOTECH AG (Anzinger str. 7a, D-85560 Ebersberg, Germany) as HPLC-purified salt free oligonucleotides synthesized on the 1 μ mol scale. Ligand and DNA oligomer solutions were prepared in degassed buffer and the oligomer was heated to 90 °C for 12 minutes and allowed to anneal slowly over 12 hours. d(GCGACTAGTCGC)₂ was dissolved in 1 mL of the degassed PIPES buffer and the concentration of the resulting oligomer solution was determined spectroscopically at λ_{260} using the OD values supplied by the manufacturer. Aliquots were taken and diluted to achieve the concentration required for the ITC experiment ($\sim 15 \mu\text{M}$).

ITC was performed at 25 °C using a Microcal VP-ITC (Microcal Inc., Northampton, USA). The control units were interfaced to PCs equipped with the Origin software package for data manipulation and instrumental control. The DNA concentration in the sample cell was 15 μM . Mixing was carried out by stirring the sample cell at 329 revolutions per minute. A 250 μL rotating syringe with an impeller profiled needle was used to perform 25 repeat 10 μL injections of the ligand with a 300 s delay between the first five injections, a 600 s delay between the subsequent fourteen injections and a 300 s delay between the last six injections. To correct for the dilution heat of the ligand, control experiments were performed at the same temperature using similar conditions with buffer only. The heats of ligand dilution were subtracted from the subsequent heat obtained for the titration of d(GCGACTAGTCGC)₂ with the ligand, thereby yielding the heat of binding for the ligand–DNA complex. All experiments were performed in duplicate.

MD simulations

The NMR structures of the free DNA decamer duplex d(CGACTAGTCG)₂ (Protein Data Bank code 1RN9) and in a 1 : 2 complex with thiazotropsin A (Protein Data Bank code 1RMX) were used for all simulations. A terminal GC base pair was added to either end of the DNA duplex to generate the dodecameric d(GCGACTAGTCGC)₂ that was used to determine the binding free energy experimentally by ITC. The AMBER 2003 (ff03)⁴⁰ and AMBER 2002 (ff02)⁴¹ force fields that represent the non-polarizable and polarizable simulations, respectively, were applied to all DNA atoms. Since the RESP charges of thiazotropsin A are not available in the AMBER package, the RESP⁴² methodology was applied to this ligand for our study. We initially optimized the thiazotropsin A monomer at the B3LYP/6-31G(d, p) level

to adjust the structure obtained by NMR. The RESP fitting procedure was then used to obtain force field parameters for the ligand which was fitted by quantum mechanical electrostatic potentials at the HF/6-31G(d) level.

The separate trajectory approach used the co-ordinates for the free DNA solved experimentally (1RN9). The simulated systems were neutralized by the addition of 20 Na⁺, 22 Na⁺ and 2 Cl[−] counterions for the complex trajectory (using the default AMBER parameters for these ions), free d(GCGACTAGTCGC)₂ and free ligand, respectively. Each system was placed in a periodic octahedral box solvated with TIP3P and POL3 water for the non-polarizable and polarizable force fields, respectively, with outer edges approximately 10 Å in each direction from the closest solute atom. Periodic boundary conditions with a 15 Å cutoff for non-bonded interactions were applied, with the particle mesh Ewald (PME) method^{43,44} applied to account for the long-range electrostatic interactions.

Before the MD production phase, minimization and equilibration were carried out in three stages as follows: (i) the solvent and ions were minimized whilst the DNA and the ligand dimer were restrained by 10 kcal mol^{−1} Å^{−2} for 2000 steps, followed by heating using the NVT ensemble and Langevin dynamics with a collision frequency of 1 ps^{−1} from 100 to 300 K over 40 ps, followed by the NPT ensemble for 40 ps at a constant temperature of 300 K. (ii) Next, the restraints on the solvent and ions were applied at 10 kcal mol^{−1} Å^{−2} for the first 1000 steps and reduced to 5 kcal mol^{−1} Å^{−2} over 3000 steps of minimization using steepest descents and conjugate gradients, followed by equilibration from 100 to 300 K without pressure scaling. (iii) Minimizations were then performed again over a series of 1000 step intervals whilst restraints on the solute were gradually relaxed from 10, 5, 2, and 1 kcal mol^{−1} Å^{−2} and finally for 3000 steps without restraints. Further equilibration was applied to the completely unrestrained system using the NVT ensemble that involved heating from 100 to 300 K over 40 ps followed by NPT for 40 ps at a constant temperature of 300 K. The production phase involved the NPT ensemble at a constant temperature and pressure (300 K and 1 atm, respectively) for 5 ns using a timestep of 1 fs and the SHAKE⁴⁵ algorithm to constrain hydrogen to heavy atom bonds. The polarizable function was turned on throughout for systems that employed the polarizable force field. DNA structural parameters over the course of the trajectories were analyzed with CURVES.⁴⁶

Binding free energy calculations

In principle, the MM-PBSA approach calculates free energies based on eqn (2). We used 100 snapshots of the solute sampled regularly from the last ns of the MD trajectories, with the water and counterions stripped away. This method combines the enthalpic or molecular mechanics energies (E_{MM}) that represent the internal energies (bond, angle and dihedral; E_{BADH}) along with van der Waals (E_{vdW}) and electrostatic interactions (E_{elec}), with the solvation free energies (G_{sol}) calculated by the finite difference Poisson–Boltzmann (PB) model for polar solvation (G_{PB} or G_{polar})⁴⁷ and the non-polar

contribution ($G_{\text{non-polar}}$) as a function of the solvent-accessible surface area (SASA). All terms were computed from the MM-PBSA module in AMBER. The conformational entropy (S), was approximated by normal mode analysis of harmonic frequencies calculated at the molecular mechanics level (eqn (2)).

$$G = \langle E_{\text{MM}} \rangle + \langle G_{\text{sol}} \rangle - T\langle S \rangle \quad (2)$$

The G_{polar} contribution was calculated by applying a cubic lattice with 0.5 Å grid spacing and evaluating all pairwise interactions using an internal dielectric constant of 1.0 and an outside dielectric of 80. The $\Delta G_{\text{non-polar}}$ was determined as a function of the SASA estimated using eqn (3), where γ and b are empirical constants of 0.0054 kcal mol⁻¹ Å⁻² and 0.92 kcal mol⁻¹, respectively, for water.

$$\Delta G_{\text{non-polar}} = \gamma \text{SASA} + b \quad (3)$$

Solute entropic contributions were estimated from the sampled structures based on normal mode analysis using the *nmode* module in AMBER. Due to demanding computational times, configurations were selected every 100 ps (water molecules and ions removed) from the last ns of the trajectory. The selected structures were minimized using conjugate gradients for 9900 steps after 100 steps of steepest descents. Newton–Raphson algorithms were then used for 5000 steps with a distance-dependent dielectric of $1/r^2$ (with r being the distance between two atoms) and a dielectric constant of 4 for the electrostatic interactions until the root-mean-square of the elements of gradient vector was less than 10⁻⁴ kcal mol⁻¹ Å⁻¹. The frequencies of vibrational modes were computed at 300 K for these minimized structures using a harmonic approximation of the energies.

Binding free energies were determined by means of eqn (1) using snapshots from the last ns generated by the single and separate trajectory approach. For the former, co-ordinates for G_{receptor} and G_{ligand} were extracted from the G_{complex} trajectory. For the latter, snapshots were taken from the MD simulations performed separately on the solvated free DNA oligomer (using 1RN9 as the starting structure) and ligand dimer.

Results

Stability of the components in the single and separate trajectories

To assess the stability of the structures generated using the different protocols employed in our study for single (sg) and separate (sp) trajectories based on polarizable and non-polarizable force fields, RMSDs of the DNA oligomer–ligand complex (DNA_{Cpx}), the DNA in single (DNA_{sg}) and separate (DNA_{sp}) trajectories were calculated with respect to the initial MD (NMR) structure and are shown in Fig. 2a and b, respectively. For both force fields, the complex as a whole and the DNA oligomer within the complex (DNA_{sg}) displayed very stable trajectories, as indicated by the RMSD plots in Fig. 2. Greater, although acceptable, variations in RMSD were seen for the separate DNA oligomer trajectories (DNA_{sp}) when the co-ordinates of the free DNA solved by

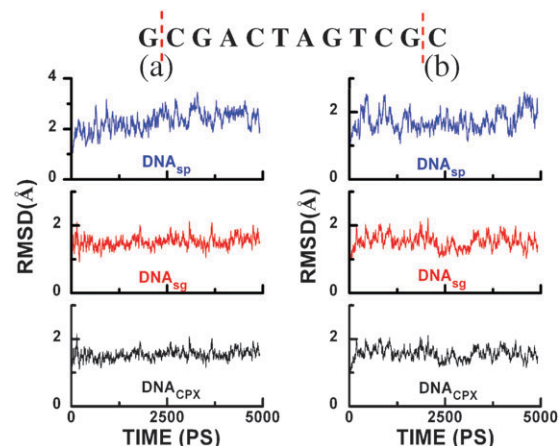


Fig. 2 All-atom root mean square deviation (RMSD) for simulations performed using the (a) polarizable and (b) non-polarizable AMBER force field. All-atom root mean square deviation of the complex (DNA_{Cpx}-black); DNA from the single trajectory (DNA_{sg}-red); and free DNA from the separate trajectory (DNA_{sp}-blue). All RMSD calculations excluded the terminal nucleotides. Average RMSDs for the polarizable system were 1.53, 1.52 and 2.31 Å for the DNA_{Cpx}, DNA_{sg}, and DNA_{sp}, respectively. Average RMSDs for the non-polarizable system were 1.55, 1.51 and 1.74 Å for the DNA_{Cpx}, DNA_{sg}, and DNA_{sp}, respectively.

NMR were input into the simulations. Our results reflect those of Babin *et al.* who reported that dodecameric B-form DNA oscillated around an RMSD of 2.9 Å when a polarizable force field was used.⁴⁸ These RMSD values applied to all the system components (except for the dimer of thiazotropsin A when simulated in the free form (2LIGsp) for both force fields; Fig. 3) and indicate that stability had been achieved in the production phases. Furthermore, the heavy atom RMSD for each nucleotide (Fig. 4) of d(GCGACTAGTCGC)₂ when present in the complex (DNA_{sg}), and when simulated separately (DNA_{sp}), ranged from 0.45 to 0.58 Å for both systems, and demonstrated that the nucleotide parameters for both forms of DNA exhibited small differences for both force fields when compared with the NMR structures (Fig. 4).

The RMSD for the dimeric form of thiazotropsin A when simulated in the solvated unbound form (2LIGsp-nopol)

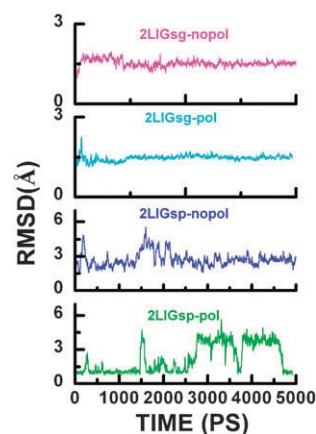


Fig. 3 RMSD for the dimeric form in the single (2LIGsg) and separate (2LIGsp) trajectories, when calculated for the polarizable (pol) and non-polarizable (nopol) force fields.

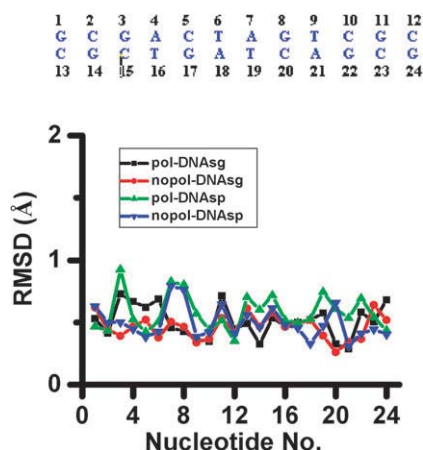


Fig. 4 RMSDs for the mean global heavy atoms of each nucleotide simulated using bound (DNAsg) or free (DNAsp) co-ordinates with the polarizable (pol) or non-polarizable (nopol) force field for 5 ns. Average RMSDs for the nucleotide heavy atoms in the polarizable system were 0.51 ± 0.1 and 0.58 ± 0.1 Å for the DNAsg, and DNAsp, respectively. Average RMSDs for the non-polarizable system were 0.46 ± 0.1 and 0.50 ± 0.1 Å for the DNAsg and DNAsp, respectively.

in Fig. 3) showed higher fluctuations using the non-polarizable force field than those for the polarizable force field for the first 2.5 ns. However, the opposite was observed for the last 2.5 ns. Furthermore, the free dimer favored two major conformations shown at an RMSD of about 4.0 and 1.0 Å from 2.5 ns to 5.0 ns for the polarizable force field. Whilst the solvated free form of the dimer does not directly impact on the structural integrity of our complex model, the presence of these two populations have implications for the free energy calculations (see Discussion).

Comparison of simulated structures with experimental data

To assess whether our MD simulations reproduced the experimental structural parameters we reported previously,²¹ we conducted a detailed comparative analysis of the different force fields in the model systems.

DNA structural parameters. In order to compare the simulated and experimental structures in more detail than RMSD, key structural parameters of the free and bound DNA dodecamer were investigated. The step parameters that characterize the relationship between contiguous base pairs (*rise*, *roll*, *shift*, *slide*, *tilt*, and *twist*) over 100 snapshots from the last 1 ns production phase shown as the average values of all base pairs are reported in Table 1 (individual base pair

parameters can be found in Fig. S1 and S2 in the ESI†). It appears that all step parameters for both force fields were consistent with the experimental data.

Minor and major grooves. The average groove width in the central -GACTAGTC- region for the bound DNA in all three systems were comparable (Table 2). A greater difference in groove width appeared between the simulated and experimental forms, which reflects the differences described by Laughton and Luisi,⁴⁹ who noted that groove features are affected by movement of the bases, which may enforce groove compression or expansion through a rolling action of the adjacent base pairs. These observations are also evident when comparing the step parameters of the free DNA forms (Table 1), particularly the *roll* parameter.

Ligand–DNA structural parameters

Hydrogen bonding. Hydrogen bond formation between the dodecamer and thiazotropsin A was determined based on a distance of 2.5 Å or less between the hydrogen bond donor (H) and acceptor (A) atom (H–A)—the empirical distance rule.⁵⁰ To investigate whether the simulated structures could reproduce the ten hydrogen bonds assigned to the experimental structure, the average distances in the polarizable and non-polarizable systems were measured (Table 3). In both systems, all hydrogen bonds detected by NMR were maintained, which demonstrated that our simulation protocols kept the integrity of the ligand–DNA complex within the parameters defined by experiment.

Inter-ligand distances. According to our experimental study, nuclear Overhauser effects (NOEs) indicated that there was little movement between the ligands not only relative to the DNA co-ordinates, but also with respect to one another. To demonstrate that the two ligands maintained this relationship within the simulated complexes, we showed that the distances between H4/C5 at the pyrrole methyl of LIG1 and C23/C24 of the isopropyl of LIG2 and *vice versa* for the polarizable and non-polarizable systems, respectively, remained consistent throughout the production phases (Fig. 5 and 6).

Free energy determinations using isothermal titration calorimetry

The thermodynamics of binding by thiazotropsin A to the dodecamer were examined using ITC to provide insight into the energetic basis for recognition and affinity by the ligand. The value of ΔH for a binding reaction is most reliably generated by calorimetry rather than indirectly from

Table 1 The average step parameters for d(GCGACTAGTCGC)₂ in the complex with thiazotropsin A (CPX) and in the unbound form (DNA) calculated from the last ns of the trajectories for the polarized (pol) and non-polarized (nopol) force fields. The corresponding average step parameters for the experimental structures solved by NMR were computed using the average for all sets of co-ordinate supplied in the 1RMX and 1RN9 pdb entries

Parameter	CPX-pol	CPX-nopol	NMR-cpx	DNAsg-pol	DNAsp-nopol	NMR-DNA
Rise/Å	3.3 ± 0.3	3.3 ± 0.3	2.8 ± 0.2	3.4 ± 0.3	3.3 ± 0.3	3.1 ± 0.8
Roll/°	6.2 ± 4.8	6.2 ± 4.7	7.8 ± 4.1	3.9 ± 5.3	4.8 ± 5.1	1.4 ± 8.5
Slide/Å	0.0 ± 0.4	0.0 ± 0.5	-0.1 ± 0.3	-0.2 ± 0.6	-0.6 ± 0.6	-0.1 ± 0.4
Shift/Å	0.0 ± 0.5	-0.1 ± 0.5	0.5 ± 0.3	0.1 ± 0.6	0.0 ± 0.7	-0.9 ± 0.5
Tilt/°	0.2 ± 4.9	0.1 ± 4.4	0.5 ± 2.3	0.3 ± 4.3	0.3 ± 4.3	-0.4 ± 5.6
Twist/°	33.3 ± 3.9	33.9 ± 3.9	31.4 ± 3.5	30.5 ± 5.0	31.9 ± 5.0	33.5 ± 6.2

Table 2 Comparison of the minor and major groove width for bound and free DNA with the experimental NMR structure in the polarized (pol) and non-polarized (nopol) systems

	Minor groove width/Å			Major groove width/Å		
	NMR	Pol	Nopol	NMR	Pol	Nopol
Bound	7.0	7.0	7.7	11.7	13.6	11.9
Free	4.0	6.3	6.7	14.5	13.5	12.8

Table 3 The average distance between the hydrogen bond donor and acceptor atoms identified by experiment between ligand and DNA, and reproduced for simulated structures based on the polarizable and non-polarizable force field. Deviations throughout the simulations are shown in parenthesis

Ligand atom	DNA atom	Distances/Å	
		DNA _{CPX-pol}	DNA _{CPX-nopol}
H2	T ⁵ O2	1.9(0.1)	1.9(0.1)
H9	A ⁶ N3	2.7(0.6)	2.5(0.3)
H16	G ⁷ N3	2.5(0.3)	2.4(0.2)
Thiazole N21	G ⁷ H22	2.2(0.2)	2.1(0.2)
H26	T ⁸ O2	2.6(0.4)	2.9(0.5)

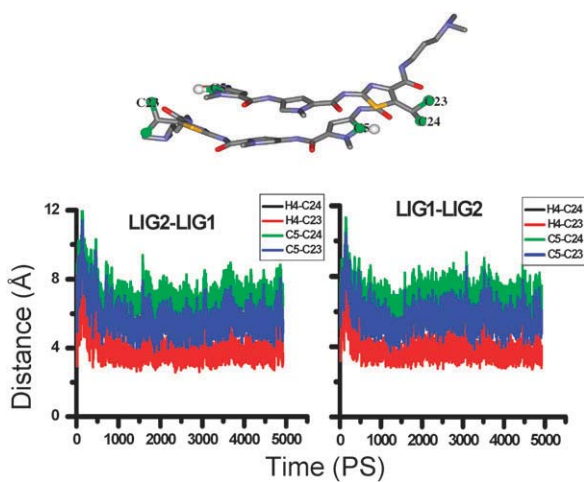


Fig. 5 Top: side-by-side anti-parallel conformation adopted by thiazotropsin A dimers with atoms involved in distance measurements annotated. Bottom: distances between atoms of the head group (H4, C5) and the isopropyl group (C23, C24) of the dimeric ligands (LIG1 and LIG2) in the bound complex (polarizable system).

van't Hoff determinations.^{51,52} Titration of thiazotropsin A with the dodecamer in PIPES buffer at 25 °C was clearly associated with an exothermic process (Fig. 7A). Dilution peaks (data not shown) were all endothermic, and their intensity decreased as more ligand was added, indicating that aggregation of the

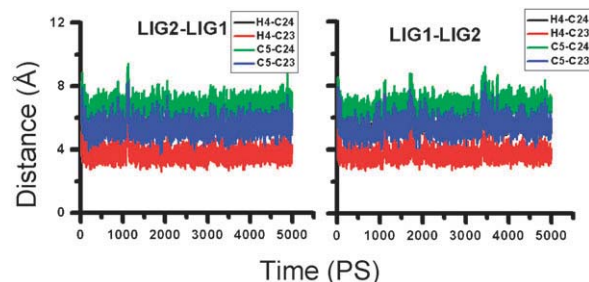


Fig. 6 The distances between atoms of the head group (H4, C5) and the isopropyl group (C23, C24) of the dimeric ligands (LIG1 and LIG2) in the bound complex (non-polarizable system).

thiazotropsin A occurs in buffered solution. The enthalpogram (Fig. 7B) was generated by integrating the raw data of the binding experiment and subtracting the heats of ligand dilution to yield binding enthalpy, ΔH , binding free energy, ΔG , and entropy changes, ΔS . Our analysis generated values for ΔG of $-10.2 \text{ kcal mol}^{-1}$, ΔH of $-12.9 \text{ kcal mol}^{-1}$, ΔS of $-9.1 \text{ cal mol}^{-1} \text{ K}^{-1}$ ($T\Delta S = -2.7 \text{ kcal mol}^{-1}$) and K of $3.0 \times 10^7 \text{ M}^{-1}$ and a binding stoichiometry of 2 : 1 (ligand to DNA). These results indicated that the binding interaction was enthalpically driven, and that there was an entropic penalty associated with the complexation process. Thiazotropsin A binding was measured at ligand : DNA ratios (r) that varied between 0 and 5.5 (Fig. 7), and analysis of the heat effects resulting from the binding process revealed that when $r \leq 2$, the binding enthalpy remained constant. A dodecamer can potentially provide two separate binding sites for a small molecule with dimensions similar to Hoechst 33258,²⁵ which raises the question: 'Does thiazotropsin A bind to two individual binding sites as a monomer or to one site as a dimer?' If monomeric binding occurred with relation to different base pair sequences of the two potential binding sites, a noticeable difference in the enthalpy of binding would be observed as the first and subsequently the second site was occupied. Since the measured ΔH values remained constant when $r \leq 2$, we propose that in this range, the ligand binds to the dodecameric DNA sequence exclusively in a non-stepwise dimeric 2 : 1 mode, which is consistent with our NMR study.

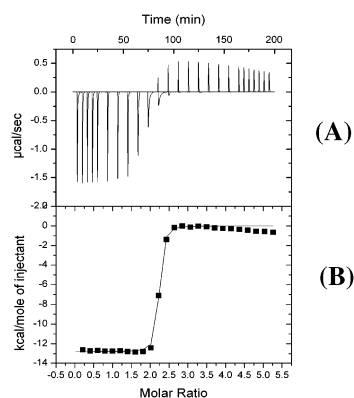


Fig. 7 (A) Raw data for titration of thiazotropsin A into the dodecamer in PIPES buffer at 25 °C (pH 7). (B) Enthalpogram retrieved from A, corrected for the heat of dilution, representing least-squares-fit to the single-site binding model.

Table 4 Adaptation energies of the d(GCGACTAGTCGC)₂ oligomer and the ligand dimer upon complexation using the polarizable and non-polarizable force fields. All values are in kcal mol⁻¹. Numbers in parentheses indicate the standard error. All terms are described in Methods

DNA/ligand	<i>E</i> (BADH)	<i>E</i> (vdW)	<i>E</i> (elec)	<i>E</i> (MM total)	<i>G</i> (PB)	<i>E</i> (elec) + <i>G</i> (PB)	<i>G</i> (non-polar)	<i>G</i> (total) + TDS
DNApol-bound	1111.7(18.9)	-188.5(10.1)	160.0(40.1)	1083.2(42.0)	-6244.8(37.3)	-6084.8(10.4)	27.0(0.2)	-4560.6
DNApol-free	1086.6(19.7)	-196.9(9.6)	153.3(56.3)	1043.0(56.5)	-6240.9(52.4)	-6087.6(10.6)	26.8(0.2)	-4602.6
Adaptation energy	25.1	8.4	6.7	40.2	-3.9	2.8	0.2	42.0
DNAopol-bound	1116.0(22.6)	202.2(9.8)	443.7(32.7)	1357.5(33.9)	-6299.7(29.1)	-5856.0(12.8)	27.3(0.2)	-4359.1
DNAopol-free	1102.0(19.2)	-206.1(10.6)	397.6(45.5)	1293.6(45.4)	-6255.6(41.8)	-5857.9(12.3)	27.2(0.2)	-4378.2
Adaptation energy	14.0	3.8	46.1	63.9	-44.2	1.9	0.1	19.1
LIGpol-bound	173.7(8.4)	-1.9(4.6)	-501.6(4.0)	-329.9(8.5)	-177.4(2.3)	-679.0(3.1)	8.3(0.1)	-355.0
LIGpol-free	182.0(10.1)	-9.8(3.8)	-514.5(10.3)	-342.3(12.5)	-175.6(8.8)	-690.1(3.7)	8.2(0.2)	-375.8
Adaptation energy	-8.4	7.9	12.8	12.4	-1.8	11.0	0.2	20.8
LIGopol-bound	172.8(8.5)	-2.6(4.3)	-501.6(4.2)	-331.8(8.4)	-176.9(2.4)	-678.9(3.1)	8.4(0.1)	-366.4
LIGopol-free	180.3(9.0)	-10.4(4.3)	499.5(11.5)	-329.6(14.2)	-187.2(10.7)	-686.7(3.4)	8.2(0.2)	-373.7
Adaptation energy	-7.5	7.9	-2.5	-2.3	10.3	7.8	0.2	7.3

Free energy calculations using MD simulations

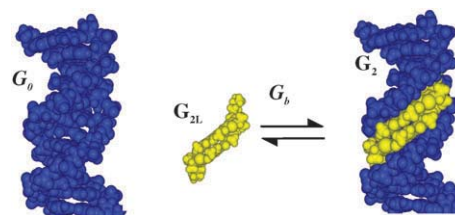
Adaptation energies of the bound and free forms of the complex components. Calculation of the adaptation free energy provides an indication of the conformational change and the energetic penalties involved when a system moves between a bound and free state. By this process, we can investigate whether a single or separate trajectory approach to determining the binding free energy is more appropriate. The single trajectory approach relies on minimal conformational rearrangement taking place when the free components associate into their bound complex and is the approach that has been predominantly used to evaluate ligands binding to DNA in the minor groove. When significant conformational adaptation within the DNA takes place, for example when ligands bind through intercalation or through an induced fit mechanism, the single trajectory approach fails to reproduce experimental binding free energies.^{24,53,54} The adaptation free energy associated with binding was calculated using 100 snapshots from the last ns of each trajectory (for components in the free and bound states) according to eqn (2), followed by subtracting the energy of the free state from the bound state (Table 4).

For the d(GCGACTAGTCGC)₂ dodecamer, the total average adaptation energies for the polarizable and non-polarizable systems were 42.0 and 19.1 kcal mol⁻¹, respectively, which implies that there is a significant penalty associated with conformational rearrangement from the free to the bound form using both force fields. Similarly, the ligand dimer undergoes a free energy penalty through conformational rearrangement on moving from the free dimer in solution to the bound dimer in the complex; the total average adaptation energies for the polarizable and non-polarizable force fields were 20.8 and 7.3 kcal mol⁻¹, respectively. Consequently, for ligand association with the DNA dodecamer to take place, the free energy of binding must offset these substantial conformational penalties.

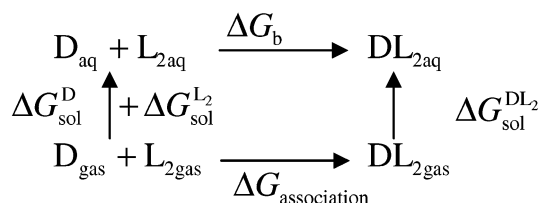
Binding free energies. Our NMR studies indicate the presence of only the 2 : 1 complex structure during titration with the ligand; indeed, all our experimental studies (circular dichroism; capillary electrophoresis;⁵⁵ data not shown) have thus far found no evidence for the formation of an initial 1 : 1

complex prior to 2 : 1 binding. Furthermore, we have shown that thiazotropsin type ligands behave like aggregates of head-to-tail dimers in solution, and monomers are not present at the ligand concentrations under investigation (data not shown, manuscript in preparation). We therefore suggest that the reaction scheme follows the equilibration process described in Scheme 1, and the thermodynamic cycle in Scheme 2 represents the parameters that can be calculated using the MM-PBSA treatment of the respective solvated trajectories (eqn 4–6), irrespective of whether the determination uses the single and separate trajectory approach.

The ΔG_b of the single trajectories were -52.1 and -76.6 kcal mol⁻¹ for the polarizable and non-polarizable force fields, respectively (Table 5). These clearly overestimate the experimental binding free energy of -10.2 ± 0.2 kcal mol⁻¹. Furthermore, for the separate trajectory approach using the non-polarizable force field, a ΔG_b value of -46.4 kcal mol⁻¹ is still an overestimation of the real value. For the separate trajectory approach (Table 5), the appearance of two populations for the ligand dimer when simulated alone with the polarizable force field (see Fig. 3) complicates the calculation and introduces two conformational groups that can inform the ΔG_b of this system. We can describe two ΔG_b energies accordingly: (i) -16.2 kcal mol⁻¹ over 4–4.6 ns. (ii) -14.5 kcal mol⁻¹ over 4.8–5 ns, both of which approach the experimental result of -10.2 kcal mol⁻¹ (the 4.6–4.8 ns interval is neglected because the conformations change dramatically over this period). The average ΔG_b (-15.4 kcal mol⁻¹) of these two conformations provides the closest approximation with the experimentally determined value.



Scheme 1 Schematic representation of the binding free energy of the duplex DNA d(GCGACTAGTCGC)₂ (blue) bound to the ligand dimer (yellow).



$$\Delta G_b = \Delta G_{\text{association}} + \Delta \Delta G_{\text{sol}} \quad (4)$$

$$\Delta G_{\text{association}} = \Delta H_{\text{association}} - T\Delta S_{\text{association}} ;$$

$$\Delta H_{\text{association}} = \Delta E_{\text{MM}}^{DL_2} ;$$

$$-T\Delta S_{\text{association}} = -T\Delta S^{DL_2} \quad (5)$$

$$\Delta \Delta G_{\text{sol}} = \Delta G_{\text{sol}}^{DL_2} - \Delta G_{\text{sol}}^D - \Delta G_{\text{sol}}^{L_2} \quad (6)$$

Scheme 2 Thermodynamic cycle for the association of the thiazotropsin A dimer (L_2) with d(GCGACTAGTCGC) $_2$ (D).

Discussion

Structural comparisons of simulated with experimental structures

A key element of any DNA simulation is the treatment of the long-range electrostatic interactions to ensure the stabilization of this highly polar and charged molecule. For biomolecules such as nucleic acids, a recognized source of inaccuracy in current force fields is the treatment of the electrostatic interactions, which has led to the development of polarizable force fields that are potentially more accurate than those based on point charges. However, these force fields have been largely unexplored with DNA, although extensive and detailed studies have started to emerge that examine their capabilities and limitations. In one such investigation, Babin *et al.*⁴⁸ compared polarizable and non-polarizable force fields based on the original Cornell *et al.*³⁵ description and found that the former tended to perform better at reproducing the structural features of DNA. Moreover, they were able to improve upon previous solution simulations,^{56–58} reducing RMSDs from around 4 Å to below 2 Å and represent the best results achieved to date that can be obtained with present force fields in solution phase studies. In this study, it appears that the DNA co-ordinates in the complex showed lower deviation than the free DNA, which suggested a stabilizing influence by the bound dimer for both force fields. This was reflected in the RMSD of the dimer itself: when associated with DNA in the complex it

exhibits minor deviations (Fig. 3; 2LIGsg-pol; 2LIGsg-nopol), which become more exaggerated when simulated in the free form (2LIGsp-pol; 2LIGsp-nopol).

To compare the simulated and experimental structures in more detail than RMSD, we examined several structural parameters of the free and bound DNA. Overall, both force fields were able to recreate stable trajectories that were structurally consistent with those we had solved previously by NMR and both performed equally for the simulations of the DNA–ligand complex. The RMSDs seen for the separate trajectories of the free DNA were more pronounced, but were still within the range reported in previous studies.⁵⁸ Unlike Babin *et al.*, greater oscillation was found for the free dodecamer in the polarizable system compared with the non-polarizable force field, although the average RMSD of 1.74 Å in the polarizable model generated here approximates to their value of 1.62 Å. The lower fluctuations that were observed throughout the bound complex trajectories replicate the known stiffening effect that minor groove binding ligands have on binding to nucleic acids,^{26,59} whilst the larger fluctuations for the free DNA reproduce the inherent flexibility observed in experimental systems. Having established that both force fields were able to describe this system reliably from a structural perspective, investigations were then sought to establish whether these simulations could be translated with any accuracy into the experimentally determined binding thermodynamics that were obtained from the ITC measurements.

Binding free energies: comparisons of simulated with experimental structures

The titration of thiazotropsin A with the DNA dodecamer was clearly associated with an exothermic process (Fig. 7A) together with an endothermic dilution process indicative of aggregation of the thiazotropsin A prior to DNA binding. A favorable enthalpy of interaction (ΔH of $-12.9 \text{ kcal mol}^{-1}$) is in agreement with the observation that exothermic interactions occur for the vast majority of ligands binding to DNA at room temperature.¹⁹ Analysis of our binding isotherm resulting from the titration of thiazotropsin A revealed that when $r \leq 2$, the ΔH values remained constant for the binding of thiazotropsin A with d(GCGACTAGTCGC) $_2$. It can be concluded that the ligand binds exclusively in a dimeric 2 : 1 mode, which is consistent with our previous NMR study, and agrees with the detailed comparative study of the

Table 5 The molecular mechanics ($\Delta E_{\text{MM}}^{DL_2}$), solvation (ΔG_{sol}), entropic ($-T\Delta S^{DL_2}$), and binding free energy (ΔG_b) terms from eqn (4) to (6) for thiazotropsin A binding to DNA using the polarizable and non-polarizable force fields. All values are in kcal mol^{-1} . Conf. 1 and Conf. 2 represent the two major conformational populations occupied by the free dimer (2LIGsp) using the polarizable force field (Fig. 3). All_conf. is the mean value for the two populations

Energy component	Single traj.			Separate		
	Pol	Nopol	All_conf.	Pol Conf. 1	Conf. 2	Nopol
$\Delta E_{\text{MM}}^{DL_2}$	−1281.1	−1287.2	−1229.8	−1226.3	−1233.2	−1225.0
ΔG_{sol}	1183.8	1179.6	1180.2	1178.2	1182.2	1146.0
$-T\Delta S^{DL_2}$	45.2	31.0	34.2	31.9	36.5	32.6
ΔG_b	−52.1	−76.6	−15.4	−16.2	−14.5	−46.4
ΔG^{exp}			−10.2			

thermodynamics of distamycin and netropsin binding performed by Lah and Vesnaver.⁶⁰ Moreover, Lah and Vesnaver's study⁶⁰ also revealed that distamycin binding in a 2 : 1 fashion is characterized by a strong enthalpy of similar magnitude to thiazotropsin A (e.g. $\Delta H = -12.5 \text{ kcal mol}^{-1}$) and is also accompanied by a substantial unfavorable entropy contribution (e.g. $T\Delta S = -2.0 \text{ kcal mol}^{-1}$). They also showed that binding free energy was dominated by a combination of non-covalent interactions such as hydrogen bond formation and van der Waals interactions, and through the hydrophobic transfer of the ligand from the surrounding solution to its binding site within the DNA minor groove.

Calculating absolute binding free energies poses a major challenge, particularly when considering practical problems such as inadequate sampling and the need to make approximations. This is particularly evident with the MM-PBSA approach, and the rigid-binding approximation that involves the single trajectory protocol has traditionally been used to overcome the incomplete sampling limitations associated with the separate trajectory approach. Rapidly improving computational architectures have helped reduce sampling limitations, and based on the relative success of the MM-PBSA method with rigid-binding approximations, researchers have been applying the methodology to tackle systems of higher flexibility where binding results in small conformational changes, with varying degrees of success.^{34,61–64} Reyes and Kollman,³³ when simulating an RNA–protein association, found that separate trajectories of the monomers, using their unbound experimental structures, provided the best agreement with the experimental binding free energies. Perhaps significantly, in the study by Spackova *et al.*²⁴ investigating the binding of the ligand DAPI to dodecameric DNA by the single and separate trajectory approach, the co-ordinates for the free DNA were taken from the simulated complex after removing the associated DAPI. These co-ordinates are appreciably different from those found in uncomplexed experimental free DNA, which could suggest why their flexible-binding approach was less successful than the rigid approach.

In order to establish whether significant conformational rearrangement took place when thiazotropsin A bound to the dodecamer, the adaptation energies of the nucleic acid and the ligand dimer upon binding were calculated as the difference between the free energy of the molecules from the simulation of the free state and the bound state.⁶⁵ Adaptation energy states that the receptor must undergo work to reorganize in order to accommodate the ligand, and the more positive the value, the greater the deformation required. By this process, it is possible to investigate whether a single or separate trajectory approach to determining the binding free energy is more appropriate (Table 4). For the DNA dodecamer, the adaptation energies of 42.0 and 19.1 kcal mol^{-1} , respectively, for the polarizable and non-polarizable force fields imply a significant penalty associated with binding. Overall, these values reflect the strain induced within the DNA structure by moving from the free to the bound state, which is only compensated in part by the solvation term (G_{PB}) that offers the only favorable drive. The cancelling of E_{Elec} with G_{PB} in both force fields appears to nullify the effects of

the charged contributions to the overall energetics, leaving the greatest difference (E_{BADH}) arising from the parameter set that is common to both force fields. The subtle differences in structure that arise through the influence of the non-bonded electrostatic term must therefore account for the differences between both force fields.

In summary, for ligand association with the DNA to take place, the free energy of binding must offset the substantial conformational penalties described by the adaptation energies and suggests that a flexible-binding approach using separate trajectories should be employed to simulate the association.

Both force fields significantly overestimate the binding free energy using the single trajectory protocol, which we relate to the neglect of conformational changes upon binding.²⁴ The positive adaptation free energies for all components in the system upon complexation (Table 4) indicate the magnitude of the conformational change that takes place, and the offset required by the contributions to binding free energy to overcome this effect. We found that the binding free energy obtained from the separate trajectory approach approximated and incorporated these conformational changes to a better degree, and returned values more consistent with our experimental studies, particularly using the polarizable force field.

Having established that structures generated by the separate trajectory approach described by the polarizable force field and post-processed using the MM-PBSA methodology furnished binding free energy values closest to experiment, evaluation of the contributions that promoted and countered ligand binding were sought. ITC had revealed that the binding process of thiazotropsin A to d(GCGACTAGTCGC)₂ was principally enthalpy-driven, although countered by an unfavorable entropy contribution. From our simulation studies, we parsed the contributions to the free binding energy described by eqn (4) into the individual components described by eqn (2), which were then broken down further into the bonded (BADH) and non-bonded (VDW; ELEC) terms and the polar (PB) and non-polar (SA) contributions. This was in order to establish where the driving force for binding at the molecular level arises from (Table 6). Like our experimental determination, the entropic term was unfavorable and opposed the binding process. In our simulation, the entropy contribution arises from a normal mode analysis, and reproduces the entropic penalty on complex formation through the loss in translational, rotational and vibrational degrees of freedom on moving from the free components to the complexed system.^{52,66,67} These calculations confirm our observations from the RMSD plots that thiazotropsin A stiffens DNA on binding, and therefore invokes the entropic penalty that the ITC studies indicate. The enthalpic term combines the molecular mechanics energy (E_{MM}) with the solvation energy (ΔG_{sol}) to give $-49.53 \text{ kcal mol}^{-1}$ and indicates the net exothermic drive for binding that our experimental studies demonstrated. At the molecular level, the stiffening of the macromolecule is reflected by a gain in steric or strain energy for the bonded component BADH, and confirms the total adaptation energy penalties for both ligand and DNA delineated in Table 6. Both the negative van der Waals and electrostatic terms suggest many favorable non-bonded interactions are formed between thiazotropsin A

Table 6 Parsing of the binding and adaptation free energies into the individual enthalpic, solvation and entropic components for the binding of thiazotropsin A with DNA using the separate trajectory approach with a polarizable force field. The binding energy components are the average values for both conformation populations (All_conf.) represented in Table 5. All values are in kcal mol⁻¹

Energy component	Total binding energy	Total adaptation energy of DNA	Total adaptation energy of ligand dimer
BADH	17.4	25.1	-8.4
VDW	-101.5	8.4	7.9
ELEC	-1145.7	6.7	12.8
PB	1189.3	-3.9	-1.8
SA	-9.1	0.2	0.2
ΔE_{MM}	-1229.8	40.2	12.4
ΔG_{sol}	1180.2	-3.7	-1.6
$-T\Delta S$	34.2	5.5	10.0

and the DNA dodecamer, a feature characteristic of minor groove binding ligands.¹⁹ Again, these values must be viewed in association with the total adaptation energies (Table 4); the non-bonded penalties associated with conformational rearrangements undertaken during the binding event are more than compensated for by the net formation of new interactions between the ligand and the macromolecule. We suggest that the magnitude of the van der Waals contributions arises in part through our premeditated incorporation of larger alkyl groups in the ligand to promote lipophilic interactions with the groove walls.^{20–23} The favorable electrostatics arise through a combination of the ten hydrogen bonds formed between the ligand dimer and the base pair edges on the groove floor (Table 3) and the two cationic DMAP tails interacting with the negative electrostatic potential of the DNA-phosphate backbone, which when combined with solvation show a clear net enthalpic drive towards complexation. Studies are currently underway with related analogues of thiazotropsin A binding to different oligodeoxynucleotide sequences to gain a fuller understanding of the key structural elements in both ligand and nucleic acid that drive association. We are performing these simulations using both explicit and implicit solvent representations in order to establish how to raise the throughput of the modelling without compromising accuracy (manuscript in preparation). Furthermore, extended simulation times will measure the impact of the parmbc force field parameters on the free energies we generate.⁶⁸

Conclusion

When developing new MGBs that bind selectively to DNA, it is essential to understand the factors that work both in favor of and counter to ligand binding. We have prepared a large library of MGBs that seek to recognize both the hydrogen bonding capacity of the groove floor to achieve specificity, and to exploit the lipophilic nature of the groove walls to enhance affinity. By way of this process, we found that the introduction of a C-alkyl into the thiazole ring of our ligand thiazotropsin A offset side-by-side binding and extended the reading frame of ligand from four to six base pairs. In 2004, we reported the first, well-characterized example of this effect in our detailed NMR study between the DNA duplex d(CGACTAGTCG)₂ with thiazotropsin A. To relate these structural examinations to binding energies, we have investigated the thermodynamics of thiazotropsin A interacting with the same sequence experimentally by ITC. Although thiazotropsin A has a greater

enthalpic drive, it suffers from an unfavorable entropy that is uncharacteristic of other MGBs and is more analogous with intercalative binding. The entropic cost associated with intercalation is most likely through the rigidification of the DNA helix, which is less evident for MGB binding, particularly in the 1 : 1 complexes. However, our MD simulations suggest that the 2 : 1 complex formed by thiazotropsin A is significantly stabilized, which is reflected by reduced RMSD fluctuations compared with the free DNA dodecamer and through the loss in translational, rotational and vibrational degrees of freedom on moving from the free components to the complexed system that our normal mode analysis revealed. With respect to these MD simulations, we have demonstrated that both the AMBER polarizable (ff02) and non-polarizable (ff03) force fields can reproduce the structural characteristics of thiazotropsin A binding with the target nucleic acid we previously solved by NMR. In terms of absolute binding free energies, the determination of adaption free energies clearly demonstrates that a significant energetic penalty is associated with binding that is a consequence of the conformational rearrangement of the DNA dodecamer to accommodate side-by-side binding by the ligand. As a result, using the MM-PBSA methodology in a single trajectory approach that assumes such conformational changes do not take place is not compatible with this system, which is reflected by the significant overestimations of binding free energies using both force fields. The separate trajectory approach, particularly for the polarizable force field, returned binding free energies that approximate well with our experimental data. Our simulations also reveal that non-bonded interactions feature strongly in the association that work to overcome the unfavorable adaptation energies and entropy that are a consequence of complex formation. Perhaps the incorporation of structural features that seek to maximize lipophilic interactions with the groove walls to enhance the enthalpic drive will always be counter-balanced by unfavorable entropies that arise from the helical rigidification of the DNA.

Our previous work has focused on designing drug-like molecules with high affinity for DNA while maintaining sufficient sequence selectivity to have useful therapeutic effects. With this objective in mind, we have increased the hydrophobicity of our ligands through the incorporation of alkyl groups larger than methyl to balance the hydrophilicity that the hydrogen bonding amides and cationic tails impart on the molecules. Our study demonstrates that such modifications can indeed improve the binding enthalpy. The challenge we

now face is to see whether the opposing entropic contribution is an inherent feature of these modifications, or whether we can use this knowledge to our advantage, and find structural adjustments that harness entropy in our favor. Further studies are underway towards this end.

Acknowledgements

We thank the Royal Golden Jubilee PhD Program (3C.KU/47/B.1) and the Thailand Research Fund for support of WT, EPSRC and the Scottish Funding Council for their funding of the Physical Organic Chemistry initiative at Strathclyde, and both Michael Feig, of Michigan State University and the National Grid Service for access to cluster facilities.

References

- 1 R. Burnett, C. Melander, J. W. Puckett, L. S. Son, R. D. Wells, P. B. Dervan and J. M. Gottesfeld, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11497–11502.
- 2 P. B. Dervan, A. T. Poulin-Kerstien, E. J. Fechter and B. S. Edelson, *DNA Binders and Related Subjects*, 2005, Springer, Berlin–Heidelberg, vol. 253, pp. 1–31.
- 3 N. G. Nickols and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 10418–10423.
- 4 P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 284–289.
- 5 J. W. Lown, *Drug Dev. Res.*, 1995, **34**, 145–183.
- 6 S. Neidle, *Nat. Prod. Rep.*, 2001, **18**, 291–309.
- 7 U. Pindur, M. Jansen and T. Lemster, *Curr. Med. Chem.*, 2005, **12**, 2805–2847.
- 8 M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 1376–1380.
- 9 M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 7586–7590.
- 10 D. J. Patel, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 6424–6428.
- 11 W. S. Wade, M. Mrksich and P. B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 8783–8794.
- 12 J. G. Pelton and D. E. Wemmer, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 5723–5727.
- 13 C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan and D. C. Rees, *Science*, 1998, **282**, 111–115.
- 14 S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird and P. B. Dervan, *Nature*, 1998, **391**, 468–471.
- 15 S. White, J. M. Turner, J. W. Szewczyk, E. E. Baird and P. B. Dervan, *J. Am. Chem. Soc.*, 1999, **121**, 260–261.
- 16 A. N. Lane and T. C. Jenkins, *Q. Rev. Biophys.*, 2000, **33**, 255–306.
- 17 V. K. Misra and B. Honig, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 4691–4695.
- 18 I. Haq, J. E. Ladbury, B. Z. Chowdhry, T. C. Jenkins and J. B. Chaires, *J. Mol. Biol.*, 1997, **271**, 244–257.
- 19 J. B. Chaires, *Arch. Biochem. Biophys.*, 2006, **453**, 26–31.
- 20 N. G. Anthony, K. R. Fox, B. F. Johnston, A. I. Khalaf, S. P. Mackay, I. S. McGroarty, J. A. Parkinson, G. G. Skellern, C. J. Suckling and R. D. Waigh, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 1353–1356.
- 21 N. G. Anthony, B. F. Johnston, A. I. Khalaf, S. P. MacKay, J. A. Parkinson, C. J. Suckling and R. D. Waigh, *J. Am. Chem. Soc.*, 2004, **126**, 11338–11349.
- 22 A. I. Khalaf, A. H. Ebrahimabadi, A. J. Drummond, N. G. Anthony, S. P. Mackay, C. J. Suckling and R. D. Waigh, *Org. Biomol. Chem.*, 2004, **2**, 3119–3127.
- 23 N. G. Anthony, D. Breen, J. Clarke, G. Donoghue, A. J. Drummond, E. M. Ellis, C. G. Gemmell, J. J. Helesbeux, I. S. Hunter, A. I. Khalaf, S. P. Mackay, J. A. Parkinson, C. J. Suckling and R. D. Waigh, *J. Med. Chem.*, 2007, **50**, 6116–6125.
- 24 N. Spackova, T. E. Cheatham, F. Ryjacek, F. Lankas, L. van Meervelt, P. Hobza and J. Sponer, *J. Am. Chem. Soc.*, 2003, **125**, 1759–1769.
- 25 S. A. Harris, E. Gavathiotis, M. S. Searle, M. Orozco and C. A. Laughton, *J. Am. Chem. Soc.*, 2001, **123**, 12658–12663.
- 26 H. Wang and C. A. Laughton, *Methods*, 2007, **42**, 196–203.
- 27 J. Dolenc, C. Oostenbrink, J. Koller and W. F. van Gunsteren, *Nucleic Acids Res.*, 2005, **33**, 725–733.
- 28 S. B. Singh, Ajay, D. E. Wemmer and P. A. Kollman, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 7673–7677.
- 29 H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin and V. S. Pande, *J. Chem. Phys.*, 2005, **123**, 084108.
- 30 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
- 31 J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401–9409.
- 32 V. Tsui and D. A. Case, *J. Phys. Chem. B*, 2001, **105**, 11314–11325.
- 33 C. M. Reyes and P. A. Kollman, *J. Mol. Biol.*, 2000, **297**, 1145–1158.
- 34 M. Lepsik, Z. Kriz and Z. Havlas, *Proteins: Struct., Funct., Bioinform.*, 2004, **57**, 279–293.
- 35 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
- 36 J. M. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.
- 37 A. Weis, K. Katebzadeh, P. Soderhjelm, I. Nilsson and U. Ryde, *J. Med. Chem.*, 2006, **49**, 6596–6606.
- 38 E. Vladimirov, A. Ivanova and N. Rösch, *J. Phys. Chem. B*, 2009, **113**, 4425–4434.
- 39 V. Babin, J. Baucom, T. A. Darden and C. Sagui, *Int. J. Quantum Chem.*, 2006, **106**, 3260–3269.
- 40 L. J. Yang, C. H. Tan, M. J. Hsieh, J. M. Wang, Y. Duan, P. Cieplak, J. Caldwell, P. A. Kollman and R. Luo, *J. Phys. Chem. B*, 2006, **110**, 13166–13176.
- 41 P. Cieplak, J. Caldwell and P. Kollman, *J. Comput. Chem.*, 2001, **22**, 1048–1057.
- 42 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145.
- 43 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 44 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
- 45 J.-P. Ryckaert, G. Ciccoti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 46 R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.*, 1989, **6**, 655–667.
- 47 K. Chin, K. A. Sharp, B. Honig and A. M. Pyle, *Nat. Struct. Biol.*, 1999, **6**, 1055–1061.
- 48 V. Babin, J. Baucom, T. A. Darden and C. Sagui, *J. Phys. Chem. B*, 2006, **110**, 11571–11581.
- 49 C. Laughton and B. Luisi, *J. Mol. Biol.*, 1999, **288**, 953–963.
- 50 I. Y. Torshin, I. T. Weber and R. W. Harrison, *Protein Eng.*, 2002, **15**, 359–363.
- 51 H. Naghibi, A. Tamura and J. M. Sturtevant, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 5597–5599.
- 52 J. B. Chaires, *Biophys. Chem.*, 1997, **64**, 15–23.
- 53 A. A. Gorfe and I. Jelesarov, *Biochemistry*, 2003, **42**, 11568–11576.
- 54 S. A. Shaikh, S. R. Ahmed and B. Jayaram, *Arch. Biochem. Biophys.*, 2004, **429**, 81–99.
- 55 I. Hamdan, G. G. Skellern and R. D. Waigh, *Nucleic Acids Res.*, 1998, **26**, 3053–3058.
- 56 D. R. Bevan, L. P. Li, L. G. Pedersen and T. A. Darden, *Biophys. J.*, 2000, **78**, 668–682.
- 57 T. E. Cheatham III and M. A. Young, *Biopolymers*, 2001, **56**, 232–256.
- 58 T. E. Cheatham and P. A. Kollman, *J. Mol. Biol.*, 1996, **259**, 434–444.
- 59 K. Boehncke, M. Nonella and K. Schulten, *Biochemistry*, 1991, **30**, 5465–5475.
- 60 J. Lah and G. Vesnaver, *J. Mol. Biol.*, 2004, **342**, 73–89.
- 61 H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891–913.

-
- 62 W. Wang, W. A. Lim, A. Jakalian, J. Wang, J. M. Wang, R. Luo, C. T. Bayly and P. A. Kollman, *J. Am. Chem. Soc.*, 2001, **123**, 3986–3994.
- 63 S. H. Huo, J. M. Wang, P. Cieplak, P. A. Kollman and I. D. Kuntz, *J. Med. Chem.*, 2002, **45**, 1412–1419.
- 64 H. Gohlke and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 238–250.
- 65 T. Lazaridis, *Curr. Org. Chem.*, 2002, **6**, 1319–1332.
- 66 R. S. Spolar and M. T. Record, *Science*, 1994, **263**, 777–784.
- 67 I. Haq, T. C. Jenkins, B. Z. Chowdhry, J. S. Ren and J. B. Chaires, *Methods Enzymol.*, 2000, **323**, 373–405.
- 68 A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.

Absorption and emission spectra of ultraviolet B blocking methoxy substituted cinnamates investigated using the symmetry-adapted cluster configuration interaction method

Malinee Promkatkaew,^{1,2} Songwut Suramitr,^{1,2} Thitinun Monhaphol Karpkird,^{1,3} Supawadee Namuangruk,⁴ Masahiro Ehara,^{5,a)} and Supa Hannongbua^{1,2,a)}

¹Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

²Center of Nanotechnology, Kasetsart University, Bangkok 10900, Thailand

³Functional Compounds Special Research Unit (FCSR), Kasetsart University, Bangkok 10900, Thailand

⁴National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency (NSTDA), 130 Paholyothin Rd., Klong 1, Klongluang, Pathumthani 12120, Thailand

⁵Institute for Molecular Science, 38 Myodaiji, Okazaki 444-8585, Japan

(Received 30 June 2009; accepted 27 October 2009; published online 8 December 2009)

The absorption and emission spectra of ultraviolet B (UVB) blocking cinnamate derivatives with five different substituted positions were investigated using the symmetry-adapted cluster configuration interaction (SAC-CI) method. This series included *cis*- and *trans*-isomers of *ortho*-, *meta*-, and *para*-monomethoxy substituted compounds and 2,4,5-(*ortho*-, *meta*-, *para*-) and 2,4,6-(*ortho*-, *para*-) trimethoxy substituted compounds. The ground and excited state geometries were obtained at the B3LYP/6-311G(d) and CIS/D95(d) levels of theory. All the compounds were stable as *cis*- and *trans*-isomers in the planar structure in both the S_0 and S_1 states, except the 2,4,6-trimethoxy substituted compound. The SAC-CI/D95(d) calculations reproduced the recently observed absorption and emission spectra satisfactorily. Three low-lying excited states were found to be relevant for the absorption in the UV blocking energy region. The calculated oscillator strengths of the *trans*-isomers were larger than the respective *cis*-isomers, which is in good agreement with the experimental data. In the *ortho*- and *meta*-monomethoxy compounds, the most intense peak was assigned as the transition from next highest occupied molecular orbital (next HOMO) to lowest unoccupied molecular orbital (LUMO), whereas in the *para*-monomethoxy compound, it was assigned to the HOMO to LUMO transition. This feature was interpreted as being from the variation of the molecular orbitals (MOs) due to the different substituted positions, and was used to explain the behavior of the excited states of the trimethoxy compounds. The emission from the local minimum in the planar structure was calculated for the *cis*- and *trans*-isomers of the five compounds. The relaxation paths which lead to the nonradiative decay were also investigated briefly. Our SAC-CI calculations provide reliable results and a useful insight into the optical properties of these molecules, and therefore, provide a useful tool for developing UVB blocking compounds with regard to the tuning of the photoabsorption. © 2009 American Institute of Physics. [doi:10.1063/1.3264569]

I. INTRODUCTION

Cinnamates have received much attention, as they are the most widely used ultraviolet B (UVB) blocking compounds among the various cosmetic sunscreen agents. Recently, 2-ethylhexyl-*para*-methoxy cinnamate, as well as other cinnamate derivatives, has been developed as a commercial product.^{1,2} Cinnamates achieve UVB blocking from a $\pi\pi^*$ absorption followed by a *cis-trans* isomerization at the propenyl double bond in the S_1 state and a relaxation to the ground state involving nonradiative decay.^{3,4} As a UVB blocking compound, the optical properties, in particular, the photoabsorption efficiency in the UVB energy region (290–320 nm) is an important factor. In order to achieve the favorable optical properties, molecular design using the variation

of the substituents has been extensively performed. The photochemistry of the process has also been studied, and the involvement of an intramolecular charge transfer (ICT) state resulting from a rotation at the C=C double bond is well recognized.^{5,6}

Several experimental and theoretical studies have been conducted to elucidate the optical properties and photochemistry of cinnamates. Time-dependent density functional theory (TD-DFT) has been utilized to investigate the photochemistry of *trans*-ethyl-*para*-(dimethylamino) cinnamate by examining the twist coordinates, corroborating the experimental observation that the formation of an ICT state is feasible in the excited state.⁵ Recently, the photophysical properties of methoxy substituted cinnamates, i.e., *ortho*-, *meta*-, and *para*-monomethoxy cinnamates and 2,4,5- and 2,4,6-trimethoxy 2-ethylhexyl-cinnamates, have been investigated experimentally to develop better UVB filter compounds.⁷ The excited states and spectroscopic properties of *para*-

^{a)}Authors to whom correspondence should be addressed. Electronic addresses: ehara@ims.ac.jp and fscisph@ku.ac.th.

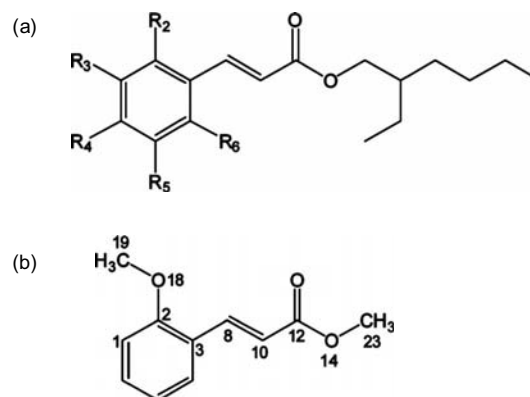
hydroxy cinnamate were investigated using TD-DFT and complete active space self-consistent field (CASSCF) methods.⁸ It was also observed that the UV absorption and fluorescence are affected by the solvent, with a shift to longer wavelengths in polar solvents.^{9,10} These studies have provided some useful insights into the optical properties of these molecules. However, reliable theoretical work is still necessary to understand the details of the optical properties of these molecules, such as the difference in the absorption and emission of the *trans*- and *cis*-isomers and the effect of the solvent.

The symmetry-adapted cluster configuration interaction (SAC-CI) method was developed by Nakatsuji^{11,12} to obtain a detailed interpretation and prediction of the molecular spectroscopy and photochemistry of the molecules. The SAC-CI method has been established as a reliable and useful method for investigating a wide variety of chemical phenomena through many successful applications.^{13,14} The method has been utilized for the accurate theoretical spectroscopy of many π -conjugated systems^{15–17} and has also been applied to the photochemistry of biological systems, such as porphyrins, photosynthetic reaction centers, retinal, and luciferin.^{18–21} Recently, the photophysical properties and excited state dynamics of fluorescent molecules, such as fluorine-thiophene oligomers²² and poly(*para*-phenylenevinylene) and poly(*para*-phenylene)²³ that are useful for organic light-emitting diodes were investigated and the absorption/emission spectra of these molecules were elucidated. These works confirm that the SAC-CI method is useful for investigating the electronic excitations and excited state dynamics of large π -conjugated systems.

In this work, the absorption and emission spectra of the *cis*- and *trans*-isomers of methoxy substituted cinnamates have been theoretically investigated using the SAC-CI method. The target molecules were *ortho*-(**1**), *meta*-(**2**), and *para*-(**3**) monomethoxy substituted cinnamates, and 2,4,5-(**4**) and 2,4,6-(**5**) trimethoxy cinnamates, as shown in Fig. 1, whose experimental spectra were reported recently.⁷ The vertical absorption and emission spectra were calculated at the theoretically optimized molecular geometries. The difference in electronic transitions among these molecules and the effect of methoxy substitution at the *ortho*-, *meta*-, and *para*-positions were analyzed. The change in geometry in the first excited state was qualitatively interpreted using electrostatic force (ESF) theory.^{24,25} The relaxation pathways that lead to the nonradiative decay were also briefly addressed.

II. COMPUTATIONAL DETAILS

The molecular structure of the *cis*- and *trans*-isomers of methoxy substituted 2-ethylhexyl-*ortho*-(**1**), *meta*-(**2**), and *para*-(**3**) monomethoxy cinnamates and 2,4,5-(**4**) and 2,4,6-(**5**) trimethoxy cinnamates is shown in Fig. 1(a). In this work, the photophysical properties of the model compounds shown in Fig. 1(b) were investigated to reduce the computational requirements without sacrificing the essence of the excitation, as the electronic structure relevant to the photophysical properties can be described using this model. The torsion angle (ω) along the $C_3-C_8=C_{10}-C_{12}$ chain was de-



Compounds	R ₂	R ₃	R ₄	R ₅	R ₆
1	OCH ₃	H	H	H	H
2	H	OCH ₃	H	H	H
3	H	H	OCH ₃	H	H
4	OCH ₃	H	OCH ₃	OCH ₃	H
5	OCH ₃	H	OCH ₃	H	OCH ₃

FIG. 1. Chemical structures of (a) 2-ethylhexyl-cinnamate derivatives and (b) the calculated model compounds and atom numbering of *trans*-isomers.

fined as $\omega=0^\circ$ for the *cis*-isomer and $\omega=180^\circ$ for the *trans*-isomer.

The ground state (S_0) geometries were fully optimized without restricting the symmetry using the B3LYP^{26,27}/6-311G(d)²⁸ method. The vertical excitation energies were calculated for the optimized geometries of the S_0 state using the SAC-CI method with the double-zeta basis set of the Huzinaga and Dunning plus polarization function [D95(d)].²⁹ To calculate the emission energy, the geometry optimization was performed for the first singlet excited (S_1) states using the CIS/D95(d) method with restricting the planar structure, except for the *cis*-**5** compound. The emission from this local minimum was observed in previous experimental work,⁷ although the global minimum exists in the nonplanar structure where the torsion angle (ω) of the $C_3-C_8=C_{10}-C_{12}$ chain is around 90° . The emission energies were calculated using the SAC-CI method with the D95(d) basis sets. The SAC-CI calculations based on the CIS optimized structure have been validated in many applications for photofunctional molecules and biological compounds.^{14,20–23} In addition, to investigate the energy barrier for the rotation of the methoxy group that affects the optical properties, the ground state potential energy curves along the rotation (θ) of the methoxy group with respect to the phenyl ring ($C_1-C_2-O_{18}-C_{19}$) were calculated using the B3LYP/6-311G(d) method.

In the SAC/SAC-CI calculations, the singles- and doubles-(SD)-R method with the direct calculation of the σ -vector, i.e., the direct SAC-CI approach³⁰ was used. The perturbation selection technique³¹ was used to reduce the computational cost and a LevelTwo accuracy was adopted. The threshold of the linked terms for the ground state was set to $\lambda_g=5.0 \times 10^{-6}$. All the product terms generated by the doubles were included in the SAC calculations. For the excited states, the threshold of the linked doubles was set to

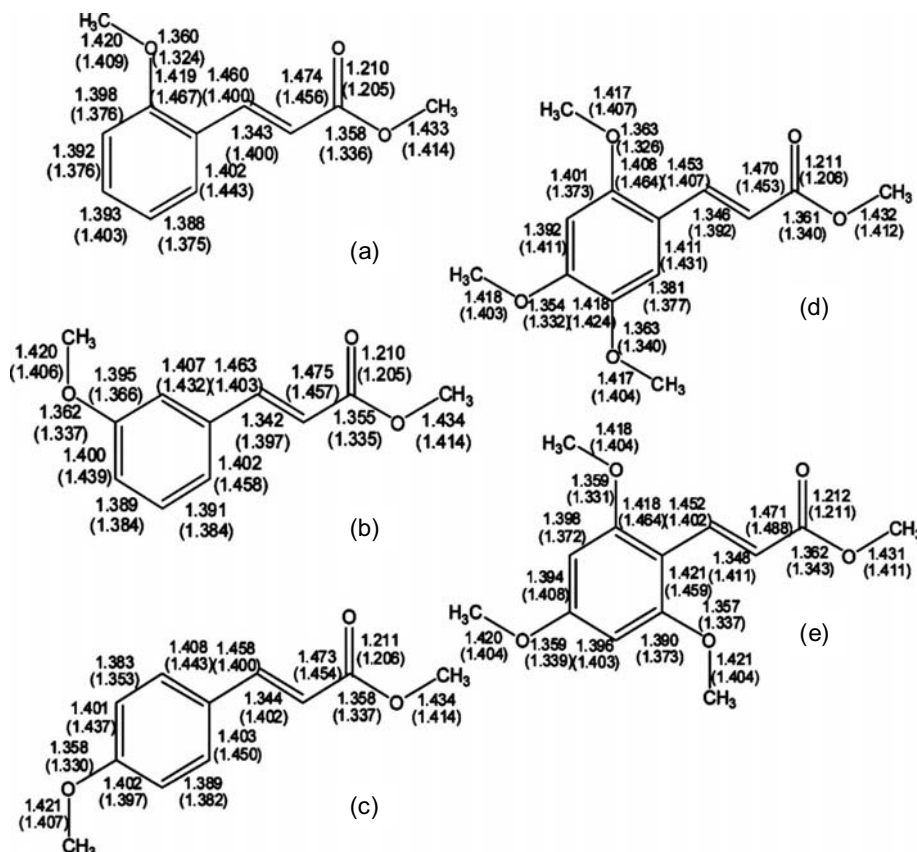


FIG. 2. A comparison of the changes in C–C and C–O bond lengths along the conjugation between the ground state and the first singlet excited state (in parentheses) of (a) *trans*-1, (b) *trans*-2, (c) *trans*-3, (d) *trans*-4, and (e) *trans*-5 calculated using the B3LYP/6-311G(d) and CIS/D95(d) methods, respectively.

$\lambda_e = 5.0 \times 10^{-7}$. All the product terms generated by the R_1S_2 and R_2S_2 operators were included in the SAC-CI calculations. The SAC/SAC-CI calculations were performed using the GAUSSIAN03 suite of programs³² with modifications for executing the direct SAC-CI method.³⁰

The results of LevelTwo are usually reliable for the direct SAC-CI calculations: In the present case, the deviations to those of LevelThree were within 0.05 eV for three excited states of *para*-(3) methoxy substitution.

III. RESULTS AND DISCUSSION

A. Ground state and excited state geometries

The ground and first excited state geometries of the *cis*- and *trans*-isomers of *ortho*-(1), *meta*-(2), *para*-(3) and the 2,4,5-(4) and 2,4,6-(5) methoxy substituted compounds were investigated. The results of the *trans*-isomers are shown in Fig. 2, and those of the *cis*-isomers are given in Fig. 3. All the compounds, except for the *cis*-5 compound, had local minima in the coplanar structure of the *cis*- and *trans*-isomers in both the S_0 and S_1 states due to π -conjugation. As shown later, in the S_0 state, the rotational energy barrier of the methoxy group is very low, although the planar structure is the most stable structure. In the S_1 state, the global minima are in the form of a nonplanar structure, but local minima exist in the planar structure. The emission was observed from this local minimum in the planar structure and, in particular, a strong emission was observed for compound 2. The life-

time was also measured for all the compounds with regard to this emission.⁷ A characteristic molecule is the *cis*-5 compound, which has a stable nonplanar structure with the torsion angle of $\omega = 7.6^\circ$ in the S_0 state because of steric effects. A local minimum of the *cis*-5 compound in the S_1 state could not be obtained.

A comparison of the C–C and C–O bond lengths along the conjugation between the S_0 and S_1 state geometries of the *trans*-isomers of the methoxy substituted compounds is shown in Fig. 2. It was found that the C_1 – C_2 , C_3 – C_8 , C_{10} – C_{12} , C_{12} – O_{14} , and O_{14} – C_{23} bond lengths decrease and the length of the other bonds increases. In the S_0 state, the carbon-carbon bond alternation exists for both the single and double bonds, but this bond alternation relaxes in the S_1 state. Since the excitation is relatively localized in the central unit, the prominent changes occur in the vinylenic unit. For example, in the *trans*-1 compound, the changes in bond length are $\Delta r = +0.048$, -0.060 , and $+0.057$ Å for $C_2=C_3$, C_3-C_8 , and $C_8=C_{10}$, respectively, while the other changes in bond length are within 0.022 Å.

The geometry change in the excited state could be qualitatively interpreted using ESF theory. We analyzed the geometry relaxation in the S_1 state of the *trans*-3 molecule. Based on ESF theory, the geometry change in the excited state is caused by the force acting on the nuclei due to the change in electron distribution. The molecular geometry is determined by balancing the atomic dipole, exchange (EC), and gross charge forces.^{24,25} The difference in the SAC/

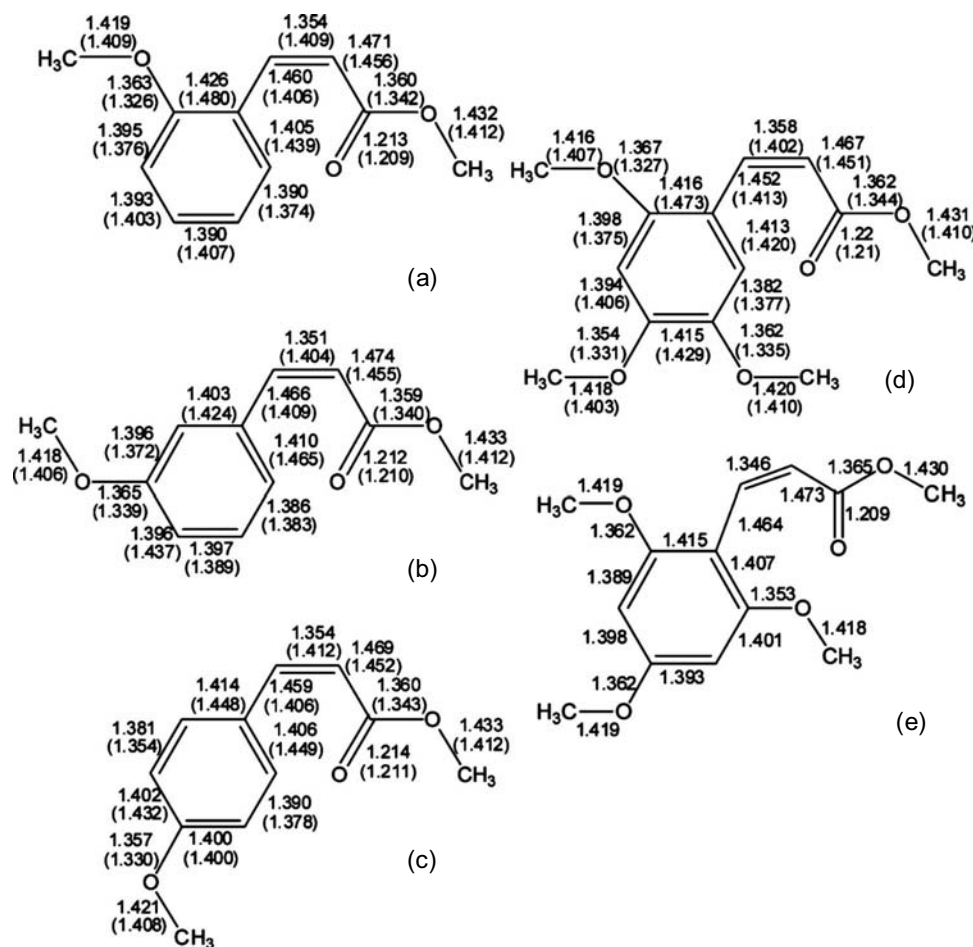


FIG. 3. A comparison of the changes in C–C and C–O bond lengths along the conjugation between the ground state and the first singlet excited state (in parentheses) of (a) *cis*-1, (b) *cis*-2, (c) *cis*-3, (d) *cis*-4, and (e) *cis*-5 calculated using the B3LYP/6-311G(d) and CIS/D95(d) methods, respectively.

SAC-CI electron density between the S_0 and S_1 states in the ground state geometry is shown in Fig. 4. The EC force of the C=C bond decreases because of the decrease in electron density, while the EC force enhances the C–C bond. For example, the C_1 – C_2 , C_3 – C_8 , C_{10} – C_{12} , C_{12} – O_{14} , and O_{14} – C_{23} bonds shrink where the electron density increases (blue) and the C_2 = C_3 and C_8 = C_{10} bonds become longer where the electron density decreases (yellow). Geometry relaxation of other molecules can also be explained in the same manner.

The rotation of the methoxy group relative to the plane of the phenyl ring was investigated. The *ortho*-(*trans*-1), *meta*-(*trans*-2), and the *para*-(*trans*-3) monomethoxy substi-

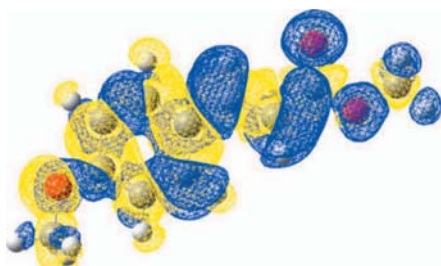


FIG. 4. Density difference maps of the *trans*-3 of methoxy substituted molecules with a positive diffuse (blue) and a negative diffuse (yellow).

tuted compound were examined for the C_1 – C_2 – O_{18} – C_{19} angle (θ) from 0° to 90° in steps of 15° . We calculated the potential energy curves with freezing other coordinates and estimated the energy barrier. The potential energy curves for this rotation are shown in Fig. 5, and these indicate that the most stable conformation was located at an angle of $\theta=0^\circ$ for the *trans*-1, *trans*-2, and *trans*-3 compounds. At the rotation angle of 90° , the energy barriers to the perpendicular confor-

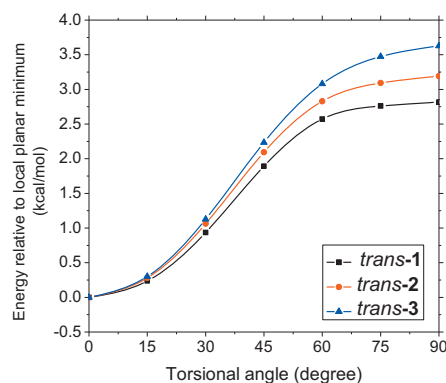


FIG. 5. Ground state potential energy curves along the rotation of the methoxy group (θ) for the *trans*-1, *trans*-2, and *trans*-3 compounds calculated using the B3LYP/6-311G(d) method.

mation of the methoxy group for the *trans*-**1**, *trans*-**2**, and *trans*-**3** compounds were 2.8, 3.2, and 3.6 kcal/mol, respectively. This shows that the π -conjugation of the *trans*-**3** compound is slightly more effective than that of the *trans*-**1** and *trans*-**2** compounds in the planar structure. The extent of π -conjugation between the methoxy oxygen and the phenyl ring could be reflected in the C–O bond length; when the bond is shorter, the conjugation is more effective. In the *trans* compounds, the CO bond lengths are 1.360, 1.362, and 1.358 Å, for *trans*-**1**, *trans*-**2**, and *trans*-**3** compounds, respectively. In *trans*-**1**, the intramolecular C–H $\cdots$ O hydrogen bond interaction between the oxygen atom of the methoxy group and the hydrogen atom in the vinylene group also exists, which causes a reverse effect on the energy barrier. This low rotational energy barrier of about 1.0 kcal/mol from $\theta=0^\circ$ – 30° indicates that a wide range of nonplanar conformations is possible at room temperature. The effect of the nonplanar structure on the absorption spectra will be discussed later.

B. Absorption spectra

The absorption spectra of the *cis*- and *trans*-isomers of the five methoxy substituted cinnamates were investigated. The vertical excitation energies were calculated using the SAC-CI/D95(d) method in their ground state optimized structure. The calculated excitation energies, oscillator strength, and dipole moments are summarized in Table I, along with the experimental values observed in hexane and methanol.⁷ The A' states were assigned to the $\pi\pi^*$ excited states with a large oscillator strength, while the A'' states were due to the $n\pi^*$ transition. Three low-lying excited states were found to be relevant in the energy region of the UV absorption. The higher excited states were also examined, up to around 7 eV, but as these states are located above 5 eV, they do not contribute to the important energy region for UVB absorption of 290–320 nm (4.28–3.87 eV). Compound **3** is the standard commercial product for UVB protection.

For the monomethoxy substituted compounds, the excited states that contribute to the absorption have a different character that depends on the methoxy substituted position. In the case of the *ortho*-(*cis*-**1** and *trans*-**1**) and *meta*-(*cis*-**2** and *trans*-**2**) compounds, the S_3 state has the highest transition probability and is characterized as being the transition from the next highest occupied molecular orbital (next HOMO) to lowest unoccupied molecular orbital (LUMO) (nH–L) transition, whereas the highest transition possibility of the *para*-(*cis*-**3** and *trans*-**3**) compounds was calculated for the $S_0 \rightarrow S_1$ transition, and the excitation character is a HOMO–LUMO (H–L) transition. The agreement with the experimental values was satisfactory. The deviations from the experimental values in peak position were within about 10 nm. For example, for the *cis*-**2** compound, the calculated values were 261 and 314 nm, compared to the experimental values of 274 (271) and 313 nm, respectively.⁷ For the transition probability, the observed absorption coefficients of the *trans*-**1** compound are 13 500 and 18 100 M^{−1} cm^{−1} for the

lower and higher peaks, respectively,⁷ and the calculated oscillator strengths are 0.26 and 0.38 for these peaks, respectively.

The deviation from the experimental data can be attributed to the solvent effect and/or our model neglecting the side chain. The polar solvent effect is relatively small for compounds **1** and **2**, as the difference in the experimental excitation energy between hexane and methanol solutions is within 4 nm, while it is large for compound **3** at 7–19 nm.⁷ The dipole moments of the ground and excited states were calculated to interpret the solvent effect qualitatively. The calculated dipole moments of the excited state were larger than that of the ground state, which causes a redshift of the peaks in a polar solvent. We also calculated the absolute values of the changes in the dipole moments from the ground state to excited states, $|\Delta\mu|=|\mu_{\text{ES}}-\mu_{\text{GS}}|$, which are shown in Table I, to examine the solvent effect before the solvent re-orientation due to the excitation of cinnamates. The *cis*-isomers of compounds **1** and **2** showed changes in their dipole moments between the ground and excited states of $|\Delta\mu|=0.27$ and 2.89 D, respectively, whereas the change in dipole moment of compound **3** was around $|\Delta\mu|=4.02$ D, which also explains the observed experimental trend. However, our calculations for compounds **4** and **5** could not explain the experimental data. The *cis*- and *trans*-isomers of compound **4** show large changes in their dipole moments, around $|\Delta\mu|=5.66$ and 4.38 D, compared to the energy shifts of only 4 and 1 nm, respectively. This may be explained by the direct interaction between methanol and cinnamates.

The trimethoxy substituted compounds at the *ortho*-, *meta*-, and *para*-positions (*cis*-**4** and *trans*-**4**) show two separate peaks, and the strong lowest absorption occurs around 350 nm. However, this lowest peak is below the UVB region of interest. The *trans*-**5** compound has the S_1 state at 307 nm with a large oscillator strength of 0.64. However, this value is lower than that of the *trans*-**3** compound of 0.76 (S_1+S_2). Experimental data also show the same trend. The absorption coefficients were 24 700 and 19 900 M^{−1} cm^{−1} in methanol for the *trans*-**3** and *trans*-**5** compounds, respectively.⁷ The $1A'$ and $2A'$ states were characterized as being H–L and nH–L transitions, respectively, for both compounds **4** and **5**. The *para*-methoxy substitution has an important effect in achieving a large oscillator strength of H–L transition in cinnamates. Methoxy substitution in the meta-position leads to a decrease in the oscillator strength of the H–L transition, as also seen in compound **2**. A comparison with the absorption spectrum of methyl hydroxyl cinnamate (pCA)⁸ in the gas phase shows that the substitution of methoxy for OH results in changes in the oscillator strengths of the S_1 and S_2 transitions. The oscillator strength of the H–nL transition increases in methoxy substituted compound: The TD-DFT calculation predicted that the S_2 state has smaller oscillator strength in hydroxy cinnamate.

Figure 6 shows the MOs relevant for the three low-lying excited states for the *trans*-**1**, *trans*-**2**, and *trans*-**3** compounds which correspond to the *cis*-**1**, *cis*-**2**, and *cis*-**3** compounds, whereas compounds **4** and **5** are shown in Fig. 7. The pattern of the MOs of the *cis*-forms is similar to that of the *trans*-ones. The HOMO, next HOMO, and LUMO are

TABLE I. Excitation energy (E_{ex}), absorption wavelength (λ_{max}), oscillator strength (f), excitation character, and dipole moment change ($|\Delta\mu|$) for the *cis*- and *trans*-isomers of methoxy substituted cinnamates calculated using the SAC-CI/D95(d) level of theory. Experimental values are cited from Ref. 7.

Molecule	State	SAC-CI					Experiment (nm)	
		E_{ex} (eV)	λ_{max} (nm)	f	Excitation character	$ \Delta\mu $ (D) ^a	MeOH	Hexane
<i>cis</i> -1	XA'					(2.44)		
	1A'	3.83	323	0.25	0.77(H→L)	4.27	313	
	1A''	4.49	276	0.00	0.78(H-3→L)	4.46		
	2A'	4.51	275	0.34	0.72(nH→L)	0.27	271	274
<i>trans</i> -1	XA'					(2.78)		
	1A'	3.87	321	0.26	0.77(H→L)	3.19	325	
	1A''	4.23	293	0.00	0.82(H-3→L)	5.96		
	2A'	4.64	267	0.38	0.71(nH→L)	0.25	276	272
<i>cis</i> -2	XA'					(1.09)		
	1A'	3.94	314	0.04	0.76(H→L)	5.56	313	
	1A''	4.26	291	0.00	0.82(H-3→L)	4.35		
	2A'	4.74	261	0.53	0.73(nH→L)	2.89	274	271
<i>trans</i> -2	XA'					(0.50)		
	1A'	4.06	306	0.12	0.69(H→L)	4.89	313	
	1A''	4.49	276	0.00	0.80(H-3→L)	5.91		
	2A'	4.61	269	0.47	0.62(nH→L)	2.49	278	274
<i>cis</i> -3	XA'					(0.80)		
	1A'	4.13	300	0.59	0.91(H→L)	4.02	303	296
	1A''	4.33	286	0.00	0.83(H-3→L)	4.26		
	2A'	4.41	281	0.04	0.71(H→nL)	0.51		
<i>trans</i> -3	XA'					(3.66)		
	1A'	4.17	297	0.44	0.79(H→L)	2.18	309	290
	2A'	4.44	279	0.32	0.65(H→nL)	1.98		
	1A''	4.54	273	0.00	0.79(H-3→L)	5.95		
<i>cis</i> -4	XA'					(3.02)		
	1A'	3.33	372	0.36	0.89(H→L)	5.66	345	349
	1A''	4.30	288	0.00	0.82(H-3→L)	4.52		
	2A'	4.34	286	0.24	0.77(nH→L)	1.70		
<i>trans</i> -4	XA'					(3.14)		
	1A'	3.45	360	0.41	0.88(H→L)	4.38	349	348
	2A'	4.48	277	0.30	0.73(nH→L)	1.95		
	1A''	4.52	274	0.00	0.79(H-3→L)	6.25		
<i>cis</i> -5	XA'					(2.26)		
	1A	4.09	303	0.24	0.83(H→L)	4.89	305	298
	2A	4.35	285	0.01	0.71(nH→L)	1.86		
	3A	4.56	272	0.11	0.70(H-4→L)	2.00		
<i>trans</i> -5	XA'					(5.04)		
	1A'	4.04	307	0.64	0.88(H→L)	0.73	320	312
	2A'	4.27	290	0.06	0.74(nH→L)	1.99		
	1A''	4.57	271	0.00	0.79(H-3→L)	5.88		

^aValues show the changes in the dipole moments from the ground to excited states ($|\Delta\mu|=|\mu_{\text{ES}}-\mu_{\text{GS}}|$) and values in parentheses are ground state dipole moment.

localized on the phenylene vinylene backbone. The methoxy substitution on the phenyl ring also has a small contribution to the π -conjugation. HOMO-3 contains the lone pairs of the C=O bond, and the $n\pi^*$ transition (1A'' state) causes a charge transfer from the C=O bond to the phenylene

vinylene. In the HOMO and next HOMO, the vinyl double bonds form bonding orbitals, and the single bonds linking the phenyl ring with the vinyl double bond are antibonding. In the LUMO, the vinyl double bonds are antibonding and the single bonds are bonding. On the other hand, the next

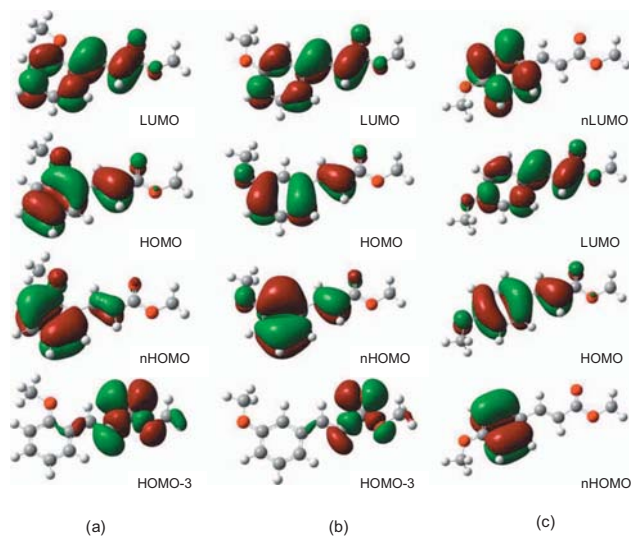


FIG. 6. MOs relevant to the low-lying excited states for (a) *ortho*-(*trans*-1), (b) *meta*-(*trans*-2), and (c) *para*-(*trans*-3) monomethoxy substituted molecules.

HOMO and next LUMO of the *trans*-3 compound are localized on the benzene ring and the amplitude on vinyl double bond diminishes. These characters of MOs also explain the geometry changes in the excited states as mentioned above. The HOMO of the *trans*-4 compound spreads over the oxygen lone pairs, which is the origin of the low-lying S_1 state in the *trans*-4 compound.

The SAC-CI excitation spectra for the monomethoxy substituted compounds are compared to the experimental ab-

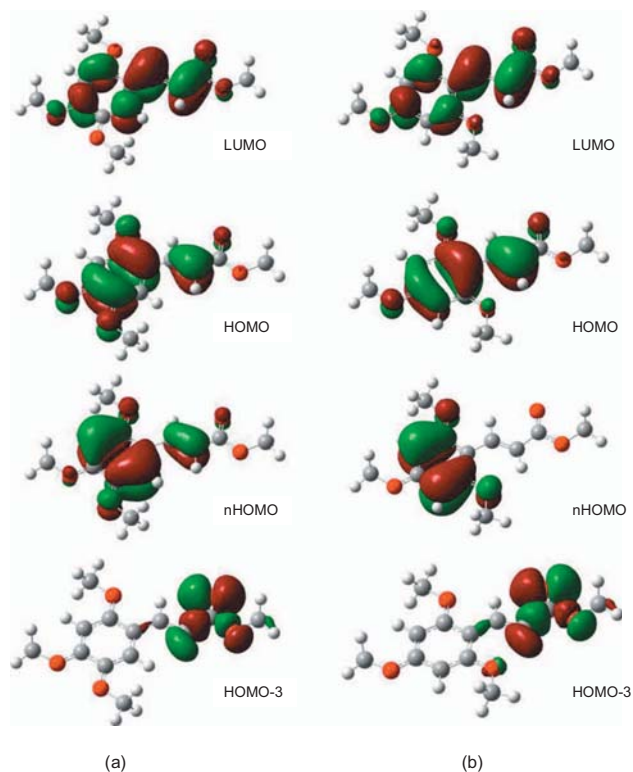


FIG. 7. MOs relevant to the low-lying excited states for (a) *trans*-4 and (b) *trans*-5 of the trimethoxy substituted molecules.

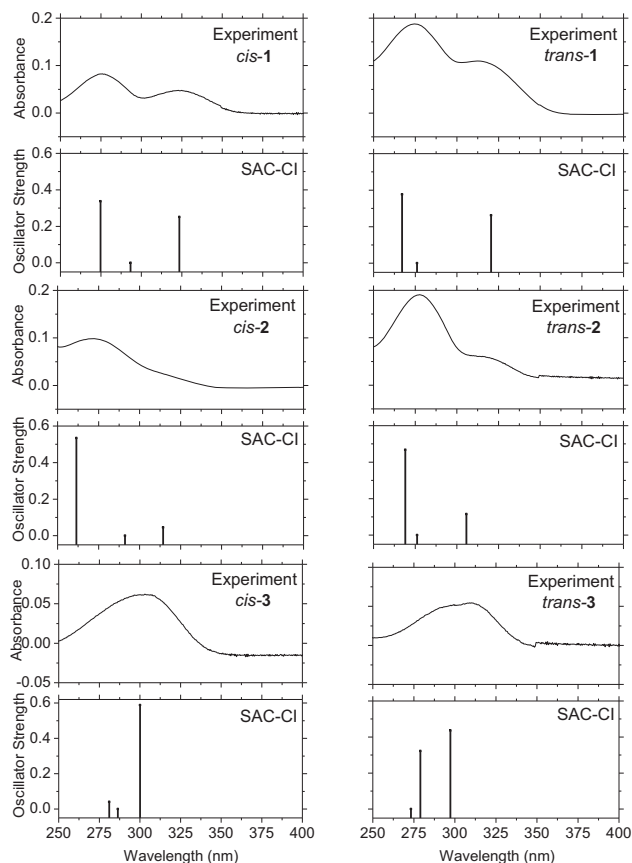


FIG. 8. SAC-CI absorption spectra of the *cis*- and *trans*-isomers of the monomethoxy substituted compounds compared with the experimental spectra in methanol (Ref. 7).

sorption spectra in Fig. 8. The experimental spectra were observed in methanol, and the polar solvent effect in terms of the energy shift was small for the *ortho*- and *meta*-substitution (less than 4 nm), and non-negligible for *para*-substitution (10–20 nm) from the experimental evidence.⁷ The SAC-CI absorption spectrum of the *cis*-3 substituted compound consists of a single absorption band, and that of the *trans*-3 compound consists of closely separated two peaks, which is in good agreement with the experimental spectra. The absorption spectra of the *ortho*-(*cis*-1 and *trans*-1) and *meta*-(*cis*-2 and *trans*-2) substituted compounds consist of two distinct absorption bands. Compound 1 has two peaks with a large oscillator strength, whereas the spectra of compound 2 show a single strong peak in the higher energy region, with a shoulder on the lower energy side. These trends were well reproduced by the present SAC-CI calculations.

The SAC-CI and experimental spectra of the trimethoxy substituted compounds are compared in Fig. 9. For the trimethoxy substituted compounds, the appearance of the SAC-CI absorption spectra of the *cis*-4 and *trans*-4 (*ortho*-, *meta*-, and *para*-substituted) compounds was similar to those of the *cis*-1 and *trans*-1 (*ortho*-substituted) compounds, showing two distinct bands but with the lower peak having a higher intensity. This is in good agreement with the experimental measurements. The absorption of compound 4 in the lower energy region may be useful for blocking of the UVA

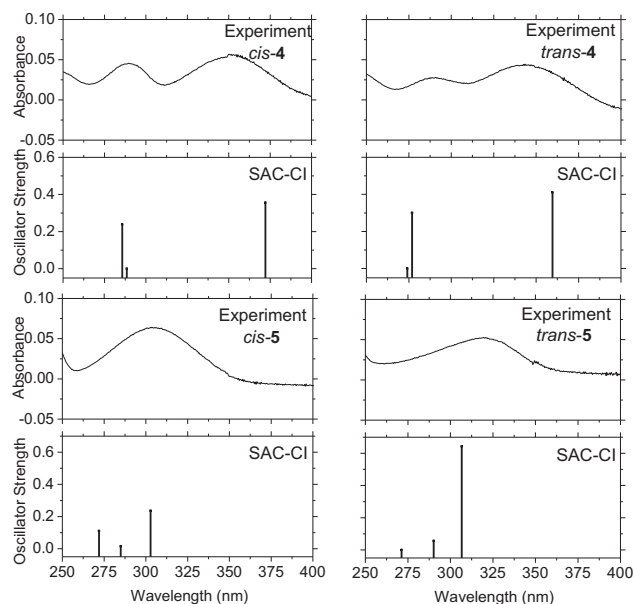


FIG. 9. SAC-CI absorption spectra of the *cis*- and *trans*-isomers of the trimethoxy substituted compounds compared to the experimental spectra in methanol (Ref. 7).

light. In contrast, the absorption spectra of the *cis*-5 and *trans*-5 (*ortho*- and *para*-substituted) compounds resemble those of the *cis*-3 and *trans*-3 (*para*-substituted) compounds that show a single band with a small shoulder in the higher energy region. Our theory also reproduced these experimental data.

The experimental spectra show that the absorption intensity of the *trans*-isomers is larger than that of the corresponding *cis*-isomers in all the compounds.⁷ This indicates that the *trans*-isomers have better absorption efficiencies than the *cis*-isomers do. This trend was also reproduced by our SAC-CI calculations. In general, the transition dipole moment is determined by the transition dipole integrals and configuration interaction. In the present case, although the oscillator strength was distributed over two $1A'$ and $2A'$ from the configuration interaction, the difference between the *cis*- and *trans*-isomers could be attributed to the transition dipole integrals. The HOMO and LUMO orbitals of the *trans*-isomer are spread more broadly than those of the *cis*-isomer, which leads to the difference in the oscillator strength.

The ground state potential energy curve is flat along the rotation angle $\theta=0^\circ$ – 30° , where the energy difference is less than 1.0 kcal/mol, and therefore, the nonplanar conformation contributes to the absorption spectra at room temperature. Therefore, we calculated the SAC-CI absorption spectra of the *trans*-1 and *trans*-3 compounds in the nonplanar structure. The calculated spectra with a rotation angles of $\theta=0^\circ$, 15° , and 30° are shown in Fig. 10. For the *trans*-1 compound, the lower energy peak shows a blueshift of about 10 nm from $\theta=0^\circ$ to 30° , and this contributes to the shoulder observed in the higher energy region. Since the HOMO of the *trans*-1 compound has an antiphase interaction between the benzene ring and the methoxy group, the HOMO stabilizes the nonplanar structure, and therefore, the blueshift occurs in the S_1 state of the H-L transition. The oscillator

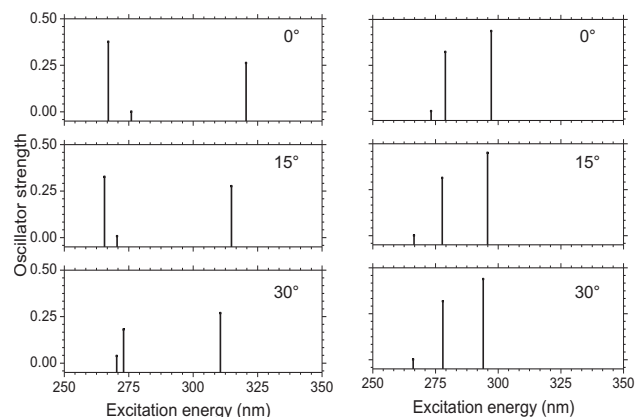


FIG. 10. Absorption spectra of the *trans*-1 (left) and *trans*-3 (right) compounds at a rotation angles $\theta=0^\circ$, 15° , and 30° calculated using the SAC-CI/D95(d) method.

strength of the *trans*-1 compound in the S_2 and S_3 states also interchanges along the rotation. The S_3 state is strong at $\theta=0^\circ$, while in the distorted structure, the oscillator strength of the S_2 state increases. On the other hand, the oscillator strength of the three excited states of the *trans*-3 compound does not change much.

C. Emission spectra

The emission energies of these molecules have also been calculated using the SAC-CI method. The stable geometries of the S_1 state were located using CIS followed by the SAC-CI calculations of the vertical emission energies. The calculated emission energies, oscillator strength, and Stokes shifts are compared to the experimental data in hexane and methanol solutions in Table II. The dipole moment and excitation character of each excited state are also given in the same table.

All the compounds, except for the *cis*-5 compound, were calculated to have the local minima in the planar structure. As discussed in Ref. 7, the global minimum of the excited state is the conical intersection where the torsion angle of the C_3 – C_8 = C_{10} – C_{12} chain is about 90° . Molecules excited to the S_1 state with sufficient energy relax through this conical intersection beyond an energy barrier to the ground state by nonradiative decay. Nonetheless, a weak emission was observed experimentally from this local minimum of the S_1 state in the coplanar structure, in particular, for compound 2.⁷ This emission from the local minimum of the S_1 state is discussed in this work. The bond distances in the S_1 state of these molecules are shown in Figs. 2 and 3.

In general, the SAC-CI calculations reproduced the experimental trends observed in hexane solution satisfactorily. The emission energies of molecules were in the order of compounds 2 and 3 (~ 350 nm) < compounds 1 and 5 (~ 360 nm) < compound 4 (~ 400 nm). The deviation from the experimental values was within 10 nm, except for the *cis*-4 compound. The transition was characterized as the HOMO-LUMO transition, and the SAC-CI coefficients are localized for this configuration. The calculated oscillator strengths were in the range of 0.41–0.68 for all the molecules, except for compound 2. As shown in the absorption

TABLE II. Excitation energy (E_{ex}), emission wavelength (λ_{max}), oscillator strength (f), excitation character, and dipole moment for the *cis*- and *trans*-isomers of methoxy substituted cinnamates calculated using the SAC-CI/D95(d) level of theory. Experimental values are cited from Ref. 7.

Molecule	State	SAC-CI						Experiment (nm)	
		E_{ex} (eV)	λ_{max} (nm)	f	Stokes shift (eV)	Excitation character	Dipole moments (D) ^a	MeOH	Hexane
<i>cis</i> - 1	1A'	3.43	361	0.41	0.40	0.83(H→L)	5.50(3.15)	409	361
<i>trans</i> - 1	1A'	3.37	367	0.44	0.50	0.85(H→L)	4.68(3.32)	405	359
<i>cis</i> - 2	1A'	3.53	351	0.10	0.41	0.76(H→L)	7.15(1.48)	410	351
<i>trans</i> - 2	1A'	3.62	343	0.29	0.44	0.84(H→L)	5.74(1.05)	409	350
<i>cis</i> - 3	1A'	3.55	349	0.63	0.58	0.91(H→L)	2.81(1.63)	468	354
<i>trans</i> - 3	1A'	3.52	352	0.61	0.65	0.90(H→L)	4.89(4.37)	462	351
<i>cis</i> - 4	1A'	2.92	424	0.44	0.41	0.90(H→L)	7.48(3.48)	461	397
<i>trans</i> - 4	1A'	3.04	408	0.51	0.41	0.90(H→L)	7.19(4.11)	461	398
<i>trans</i> - 5	1A'	3.43	362	0.68	0.61	0.90(H→L)	4.43(6.04)	463	358

^aValues in parentheses show the dipole moment of the ground state (XA').

data, the S_1 state of compound **2** also had a low oscillator strength in the ground state geometry, and the transition probability is distributed to the S_3 state. The calculated Stokes shifts were 0.40–0.50 eV for compounds **1**, **2**, and **4**, and these were slightly larger, 0.58–0.65 eV for compounds **3** and **5** where the para-position is substituted by a methoxy group. This means that the excited state geometry relaxations in the S_1 state of compounds **3** and **5** are larger than those for compounds **1**, **2**, and **4**.

A redshift was observed in methanol for all the compounds.⁷ To interpret this energy shift, the dipole moments of the S_0 and S_1 states calculated for the optimized geometry of the S_1 state are compared in Table II. The calculated dipole moments in the excited state are larger than those in the ground state, which explains the experimental redshift qualitatively. The redshift of the emission in methanol was marked in compounds **3** and **5** (about 100 nm), for which our calculations could not explain this effect.

The relaxation processes in the excited states are important for the UVB blocking molecules. We examined the S_1 potential energy surface of *trans*-**2** and *trans*-**3**, which mainly leads to decay by fluorescence and nonradiative decay through conical intersection, respectively.⁷ Since the SAC-CI method is based on the single-reference theory, it is difficult to obtain the potential energy surface around the conical intersection. The conical intersection should be examined by the multi-reference methods such as CASSCF or multi-reference configuration interaction (MRCI), but we failed the partial optimization using the state-averaged CASSCF (4,4), (4,6), and (4,8) calculations, although CASSCF is stable around conical intersection. Therefore, we performed the SAC-CI partial optimization of the S_1 state for the torsion angles from 0° to 75° with all the other coordinates being optimized. Because of the computational limitation, the small basis sets double-zeta valence (DZV) [3s2p/2s] and restricted active space (7 occupied and 33 unoccupied MOs) were used and the perturbation selection of the operator was not performed in the SAC-CI calculations. The potential energy curves along the minimum energy path of the S_1 state are shown for *trans*-**2** and *trans*-**3** in Fig. 11. In *trans*-**2**, the energy barrier to the conical intersection was calculated to be about 5.5 kcal/mol (0.24 eV), while *trans*-**3** has a low energy barrier about 0.6 kcal/mol (0.03 eV). These

calculations qualitatively explain the experimental fact that the activation energies based on the nonradiative deactive rate of *trans*-**2** and *trans*-**3** compounds were estimated to be 1.4 and 0.4 kcal/mol, respectively.⁷ The ground state surface of *trans*-**3** was too steep along the torsion, for which we think the errors come from the quasidegenerate character of the states and the restricted active space.

IV. CONCLUDING REMARKS

The electronic structures and optical properties of cinnamate derivatives at five different substituted positions were investigated using the SAC-CI method. Both the *cis*- and *trans*-isomers were examined for *ortho*-, *meta*-, and *para*-monomethoxy substituted compounds and 2,4,5-(*ortho*-, *meta*-, *para*-) and 2,4,6-(*ortho*-, *para*-) trimethoxy compounds.

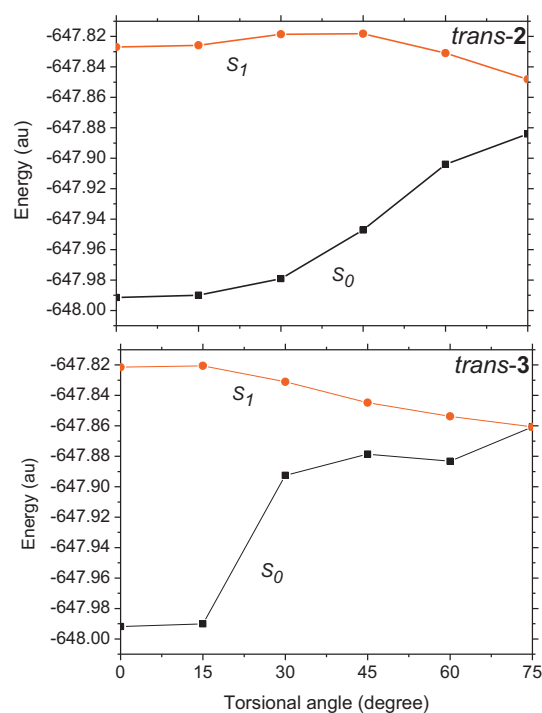


FIG. 11. The SAC-CI potential energy curves of the S_0 and S_1 states along the minimum energy path of the S_1 state are shown for *trans*-**2** and *trans*-**3**.

All compounds have a local minimum in the *cis*- and *trans*-isomers in the planar structure in both the S_0 and S_1 states, except for the 2,4,6-trimethoxy compound. The SAC-CI/D95(d) calculations reproduced the recently observed absorption and emission spectra satisfactorily, and allowed a detailed assignment and interpretation of the spectra. Three low-lying excited states were found to be relevant for the absorption in the UV blocking energy region. The calculated oscillator strengths of the *trans*-isomers were larger than all the *cis*-isomers, which is in good agreement with experimental data. In the monomethoxy compounds at the *ortho*- and *meta*-positions, the most intense peak was assigned to the transition from the nH-L, whereas in the *para*-compound it was assigned to the H-L transition. This feature was interpreted from the variation of the MOs due to the different substituted positions, and was used to explain the behavior of the excited states of the trimethoxy substituted compounds.

Our SAC-CI calculations have provided reliable results and a useful insight into the optical properties of these molecules and, therefore provide a useful tool for developing UVB blocking compounds with regard to the tuning of the photoabsorption. High absorbance, broad absorption peak with small fluorescence quantum yield, and low radiative rate are expected for superior UVB sunscreen. Nonradiative decay back to the initial ground state is also relevant. Thus, the theoretical study of the relaxation process is important to design the superior UVB blocking molecules. In the present case, both *trans*- and *cis*-forms can be generated in the course of the relaxation at around the conical intersection and *cis*-form also has absorption in the UVB region.

ACKNOWLEDGMENTS

We would like to thank Asst. Professor Fukuda and Professor Nakatsuji for the direct SAC-CI program code. We also thank the reviewers for the valuable comments. This work was supported by the Thailand Research Fund and the Commission on Higher Education (CHE) (Grant No. RTA5080005 for S.H, Grant No. MRG5180287 for S.S., and Grant No. MRG4980160 for T.K.). M.P. is grateful to the Thailand Graduate Institute of Science and Technology (TGIST) for a scholarship. M.E. was supported by the grant from the JST-CREST, Scientific Research in Priority Areas "Molecular Theory for Real Systems" from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), the Next Generation Supercomputing Project, and the Molecular-Based New Computational Science Program, NINS. The Postgraduate Education and Research Program in Petroleum, Petrochemical Technology and Advance Materials (NCE-PPAM), Center of Nanotechnology Kasetsart University, Kasetsart University Research and Development

Institute (KURDI), National Nanotechnology Center (NANOTEC), Laboratory of Computational and Applied Chemistry (LCAC), Bilateral Research Cooperation (BRC), and Institute for Molecular Science (IMS) are gratefully acknowledged for partial support and research facilities. The parts of the computations were performed using Research Center for Computational Science, Okazaki, Japan.

- ¹T. Monhaphol, B. Albinsson, and S. Pattanaargson, *J. Pharm. Pharmacol.* **59**, 279 (2007).
- ²L. Oriol, M. Pinol, J. L. Serrano, and R. M. Tejedor, *J. Photochem. Photobiol., A* **155**, 37 (2003).
- ³S. Pattanaargson and P. Limphong, *Int. J. Cosmet. Sci.* **23**, 153 (2001).
- ⁴H. B. Wang, B. C. Zhai, W. J. Tang, J. Y. Yu, and Q. H. Song, *Chem. Phys.* **333**, 229 (2007).
- ⁵T. S. Singh, S. Mitra, A. K. Chandra, N. Tamai, and S. Kar, *J. Photochem. Photobiol., A* **197**, 295 (2008).
- ⁶A. Chakraborty, S. Kar, and N. Guchhait, *Chem. Phys.* **320**, 75 (2006).
- ⁷T. M. Karpkird, S. Pattanaargson, and B. Albinsson, *Photochem. Photobiol. Sci.* **8**, 1455 (2009).
- ⁸M. de Groot, E. V. Gromov, H. Köppel, and W. J. Buma, *J. Phys. Chem. B* **112**, 4427 (2008).
- ⁹S. Pattanaargson, T. Munhapol, N. Hirunpachot, and P. Luangthongaram, *J. Photochem. Photobiol., A* **161**, 269 (2004).
- ¹⁰G. J. Smith and I. J. Miller, *J. Photochem. Photobiol., A* **118**, 93 (1998).
- ¹¹H. Nakatsuji, *Chem. Phys. Lett.* **59**, 362 (1978).
- ¹²H. Nakatsuji, *Chem. Phys. Lett.* **67**, 329 (1979).
- ¹³H. Nakatsuji, *Acta Chim. Hung.* **129**, 719 (1992).
- ¹⁴M. Ehara, J. Hasegawa, and H. Nakatsuji, in *Theory and Applications of Computational Chemistry: The First 40 Years*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Oxford, 2005), p. 1099–1141.
- ¹⁵H. Nakatsuji, O. Kitao, and T. Yonezawa, *J. Chem. Phys.* **83**, 723 (1985).
- ¹⁶J. Wan, M. Ehara, M. Hada, and H. Nakatsuji, *J. Chem. Phys.* **113**, 5245 (2000).
- ¹⁷J. Wan, J. Meller, M. Hada, M. Ehara, and H. Nakatsuji, *J. Chem. Phys.* **113**, 7853 (2000).
- ¹⁸H. Nakatsuji, J. Hasegawa, and M. Hada, *J. Chem. Phys.* **104**, 2321 (1996).
- ¹⁹H. Nakatsuji, J. Hasegawa, and K. Ohkawa, *Chem. Phys. Lett.* **296**, 499 (1998).
- ²⁰K. Fujimoto, J. Y. Hasegawa, S. Hayashi, S. Kato, and H. Nakatsuji, *Chem. Phys. Lett.* **414**, 239 (2005).
- ²¹N. Nakatani, J. Y. Hasegawa, and H. Nakatsuji, *J. Am. Chem. Soc.* **129**, 8756 (2007).
- ²²P. Poolmee, M. Ehara, S. Hannongbua, and H. Nakatsuji, *Polymer* **46**, 6474 (2005).
- ²³B. Saha, M. Ehara, and H. Nakatsuji, *J. Phys. Chem. A* **111**, 5473 (2007).
- ²⁴H. Nakatsuji, *J. Am. Chem. Soc.* **95**, 345 (1973).
- ²⁵H. Nakatsuji and T. Koga, in *The Force Concept in Chemistry*, edited by B. M. Deb (Van Nostrand Reinhold, New York, 1981), p. 137.
- ²⁶A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- ²⁷C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ²⁸R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *J. Chem. Phys.* **72**, 650 (1980).
- ²⁹T. H. Dunning, Jr. and P. J. Hay, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer III (Plenum, New York, 1976), Vol. 3, p. 1.
- ³⁰R. Fukuda and H. Nakatsuji, *J. Chem. Phys.* **128**, 094105 (2008).
- ³¹H. Nakatsuji, *Chem. Phys.* **75**, 425 (1983).
- ³²M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, Revision B.01, Gaussian, Inc., Pittsburgh, PA, 2003.

Article

Molecular Docking Studies and Anti-Tyrosinase Activity of Thai Mango Seed Kernel Extract

Saruth Nithitanakool ¹, Pimolpan Pithayanukul ^{1,*}, Rapepol Bavovada ² and Patchreenart Saparpakorn ³

¹ Department of Pharmacy, Faculty of Pharmacy, Mahidol University, Bangkok 10400, Thailand; saruth_pipe@yahoo.com (S. N.)

² Department of Pharmaceutical Botany, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10300, Thailand; E-mail: brapepol2@hotmail.com (R. B.)

³ Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand; E-mail: patchareenart_s@yahoo.com (P. S.)

* Author to whom correspondence should be addressed; E-mail: pypph@mahidol.ac.th; pypph@hotmail.com; Tel.: +662 6448694; Fax: +662 6448694.

Received: 16 November 2008; in revised form: 18 December 2008 / Accepted: 4 January 2009 / Published: 7 January 2009

Abstract: The alcoholic extract from seed kernels of Thai mango (*Mangifera indica* L. cv. ‘Fahlan’) (Anacardiaceae) and its major phenolic principle (pentagalloylglucopyranose) exhibited potent, dose-dependent inhibitory effects on tyrosinase with respect to L-DOPA. Molecular docking studies revealed that the binding orientations of the phenolic principles were in the tyrosinase binding pocket and their orientations were located in the hydrophobic binding pocket surrounding the binuclear copper active site. The results indicated a possible mechanism for their anti-tyrosinase activity which may involve an ability to chelate the copper atoms which are required for the catalytic activity of tyrosinase.

Keywords: *Mangifera indica* L; Molecular docking study; Polyphenols; Pentagalloylglucopyranose; Tyrosinase inhibitor.

Introduction

Tyrosinase (E.C. 1.14.18.1) is a multifunctional Cu-containing enzyme which is widely distributed in nature. The enzyme is mainly involved in the first two steps of melanin biosynthesis, which consists of the hydroxylation of L-tyrosine (monophenolase activity) and the oxidation of the product of this reaction, the L-DOPA (diphenolase activity), to the corresponding *o*-quinone [1]. Tyrosinase, which is contained in vegetables, fruits and mushrooms, is a key enzyme in the browning which occurs upon brushing or long term storage [2]. In mammals, this enzyme is responsible for skin pigmentation abnormalities such as flecks and defects [3]. Tyrosinase is also linked to Parkinson's and other neurodegenerative diseases [4, 5], oxidizing excess dopamine to produce DOPA quinones, highly reactive compounds which induce neuronal damage and cell death.

The crystallographic structure of tyrosinase has been established recently [6], enabling a close look at its three-dimensional structure and a better understanding of its mechanism of action. This three-dimensional structure revealed the presence of a hydrophobic protein pocket adjoining the binuclear copper active site. Each of these sites, CuA and CuB, are coordinated by three nitrogen donor atoms from histidine residues: CuA binds to the ϵ -nitrogen atoms of His³⁸, His⁵⁴, and His⁶³, and CuB binds to those of His¹⁹⁰, His¹⁹⁴, and His²¹⁶.

As plants are a rich source of bioactive chemicals, which are mostly free from harmful side-effects, there is an increased interest to identify natural tyrosinase inhibitors from plants. Tannins are diverse compounds with a wide variation in structure and concentration within and among plant species. Therefore, biomedical research on the health benefits of these compounds is of great interest. Mangos (*Mangifera indica* L.), which belong to the family *Anacardiaceae*, grow in tropical and subtropical regions, and their components are commonly used in folk medicine to produce a wide variety of remedies [7, 8, 9]. Among the edible portions, the mango seed kernel has been shown to exhibit potent antioxidant activity and to have a relatively high phenolic content [10]. The total antioxidant capacity and the amount of polyphenolic constituents present vary considerably from one variety of plant to another [11, 12]. The ethanolic extract of Thai mango seed kernel (MSKE) cultivar 'Fahlun' was found to contain 61.28% pentagalloylglucopyranose (PGG), 0.68% methyl gallate (MG) and 0.44% gallic acid (GA); the MSKE and its isolates exhibited potent free radical scavenging, antioxidant and anti-inflammatory activity as well as anti-hepatotoxicity against liver damage induced by carbon tetrachloride in rats [13]. In this study, the MSKE and its isolated phenolic principles were evaluated for their inhibitory effects on L-DOPA oxidation by mushroom tyrosinase. We also performed molecular docking studies using the Gold v3.2 program with the aim of explaining the differences in activity of the plant polyphenols isolated from MSKE. Understanding how the tannin principles in the extract interact with tyrosinase may explain how they inhibit tyrosinase and melanin formation.

Results and Discussion

Effect of MSKE and its phenolic principles on the activity of mushroom tyrosinase

Figure 1 shows the dose-response curves of MSKE and its phenolic principles (GA, MG and PGG) and a well-known tyrosinase inhibitor, kojic acid (KA) on the inhibition of L-DOPA oxidation by mushroom tyrosinase. It can be seen that MSKE and its phenolic principles clearly showed a

concentration-dependent inhibitory activity against tyrosinase. The order of potency as judged by the half-inhibition concentration (IC_{50}) was KA ($2.21 \pm 0.05 \mu\text{g/mL}$) > PGG ($42.65 \pm 1.85 \mu\text{g/mL}$) > MG ($62.50 \pm 0.50 \mu\text{g/mL}$) > MSKE ($98.63 \pm 1.62 \mu\text{g/mL}$) > GA ($644.00 \pm 14.00 \mu\text{g/mL}$) (Table 1). These results indicated that the anti-tyrosinase potency of MSKE may be attributed to its major principle (PGG) and other unidentified constituents, since PGG exerted its effect at the lowest IC_{50} value and had the highest percentage content (61.28%) within the MSKE compared with GA and MG. Although the anti-tyrosinase potency of MG was close to PGG, its percentage content within the MSKE was very low (0.68%), therefore MG may only have a negligible effect, similar to GA (0.44%). The results for MSKE were in accordance with that of Kim *et al.* [14] who found that *Galla rhois* which contains PGG as its major constituent, exhibited potent anti-tyrosinase activity. Since MSKE, GA, MG and PGG have been demonstrated to possess chelating activity when compared to EDTA [13]; a possible mechanism for their anti-tyrosinase activity may involve the chelation of copper atoms which are required for the catalytic activity of tyrosinase [14, 15]. Therefore, we performed docking calculations to explain the possible mechanism of the phenolic principles isolated from MSKE towards the binding site of tyrosinase.

Figure 1. Dose-dependent inhibition of mushroom tyrosinase by MSKE, GA, MG, PGG and the positive reference (KA). Tyrosinase activity was measured using L-DOPA as the substrate. Each value represents mean $\pm$ SEM (n = 2).

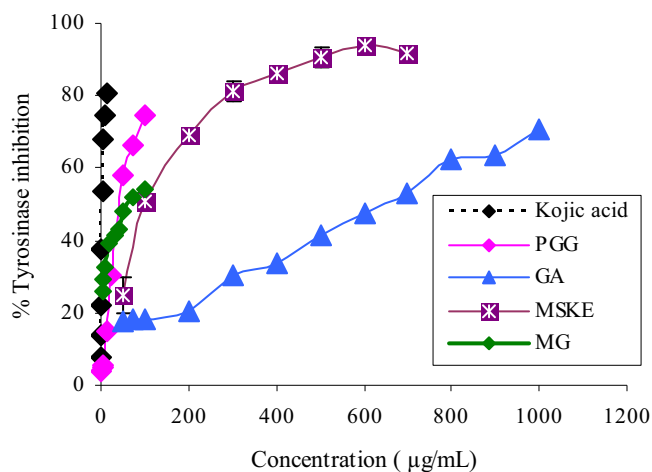


Table 1. Tyrosinase inhibitory activity of MSKE and its phenolic principles compared with the positive reference, KA.

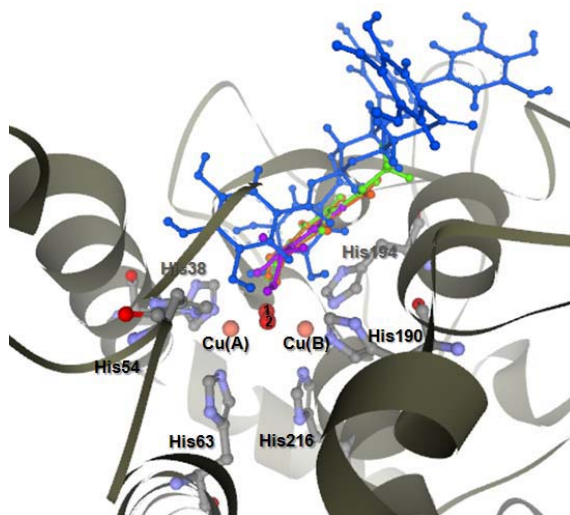
Test compounds	Anti-tyrosinase $IC_{50} \pm \text{SEM} (\mu\text{g/mL})$
MSKE	98.63 ± 1.62
Phenolic principles of MSKE [13]:	
GA	644.00 ± 14.00
MG	62.50 ± 0.50
PGG	42.65 ± 1.85
Positive reference: KA	2.21 ± 0.05

Molecular docking study

The selected docked conformations of KA, GA, MG and PGG in the tyrosinase binding site are shown in Figure 2. The docked conformations revealed that all ligands were located in the hydrophobic binding pocket surrounding the binuclear copper active site. The location of the docked compounds agreed well with that of docked 2,4-resorcinol derivatives which would contribute to their tyrosinase inhibitory potency [16].

In this study, all docked ligands were found to have some interaction between an oxygen atom of the ligands and CuB within 4 Å. Moreover, these docked conformations also formed an H-bonding interaction (<3.00 Å) with peroxide in the active site. In the binding pocket, common H-bonding interactions were formed between all docked ligands and Ile42, Asn191, Thr203 and Ser206. The specific H-bonding interaction with Phe59 was only found in the docked conformation of KA. In order to explain the binding of these compounds, the H-bonding interactions with the other surrounding residues in the hydrophobic binding pocket were also investigated. In Figure 3A, strong H-bonding interactions between the hydroxyl group of KA and an oxygen atom of Glu182 and Met201 were formed. H-bonding interactions were also formed with His38, His54, His190, and His194 which are important residues coordinated with the two copper ions in the active site. In addition, a π - π interaction between KA and His194 was found. The docked PGG, MG, and GA are shown in Figure 3B-3D. Additional similar H-bonding interactions with His38, His190, Val195, and Ala202 and a π - π interaction with His194 were found in their docked conformations. In the case of docked PGG, H-bonds with Arg55, Ser146, Val147, Glu182, Gly183, Trp184, Arg185, Asn188, and Gly204 were also formed with PGG but these H-bonding interactions were not found with MG and GA. However, a strong H-bond with Met201 was found in the docked MG and GA. Docked MG and GA revealed similar binding modes with the exception of the methyl group in the MG structure. The methyl group in the MG structure showed an extra H-bond with an oxygen atom on the backbone Asn191. The docking results agreed well with the observed *in vitro* data, which showed that the tyrosinase inhibitory activity of PGG was higher than those of MG and GA, respectively.

Figure 2. Docked conformation of ligand structures in the binding site of tyrosinase. KA (purple), PGG (blue), MG (green), GA-ionize (orange).



Experimental

Chemicals

Mushroom tyrosinase (E.C. 1.14.18.1) and L-3,4-dihydroxyphenylalanine (L-DOPA; $\geq 98\%$) were obtained from Sigma Chemical Co. (St. Louis, MO, USA). Gallic acid (GA; $\geq 98\%$), kojic acid (KA; $\geq 98\%$) and methyl gallate (MG; $\geq 98\%$) were purchased from Fluka (Buchs, Switzerland). Pentagalloylglucopyranose (PGG; $> 95\%$) was obtained from Endotherm GmbH (Germany). All reagents used in the *in vitro* experiment were obtained from commercial sources and were of analytical grade.

Plant materials

Fully grown unripened Thai mango fruits (*Mangifera indica* L. cv. 'Fahlun') (Anacardiaceae) were purchased from a local market. The voucher specimen (RB 20007) was deposited at the Herbarium of the Faculty of Pharmaceutical Sciences, Chulalongkorn University, Thailand. Fresh seeds were homogenized in a blender, using hot ethanol (80°C). After removal of the solvent under reduced pressure, the extract was defatted with hexane, evaporated, and then freeze-dried to afford a crude mango seed kernel extract (MSKE) with a yield of 8.66% (w/w).

Standardization

GA, MG and PGG were used as chemical markers and the calibration curves for each compound were obtained by densitometric scanning of different quantities of the chemical marker bands on developed chromatographic plates. An aliquot of the crude extract (MSKE) (8 μ L, 25 mg/mL) was applied along with serial amounts of the chemical marker stock solution. The thin-layer chromatographic (TLC) plates were developed in a pre-saturated twin trough glass tank using $\text{CHCl}_3/\text{MeOH}/\text{EtOAc}/\text{ethyl methyl ketone}$ (6:1.6:2:2) with five drops of formic acid as the mobile phase for GA and MG and $\text{CHCl}_3/\text{EtOH}/\text{formic acid}$ (3:5:1) for PGG. The developed TLC plates were scanned at 286 nm and the amount of each compound (GA 4.4, MG 6.8 and PGG 612.8 mg/g dry weight) in MSKE was calculated from the calibration curves [13].

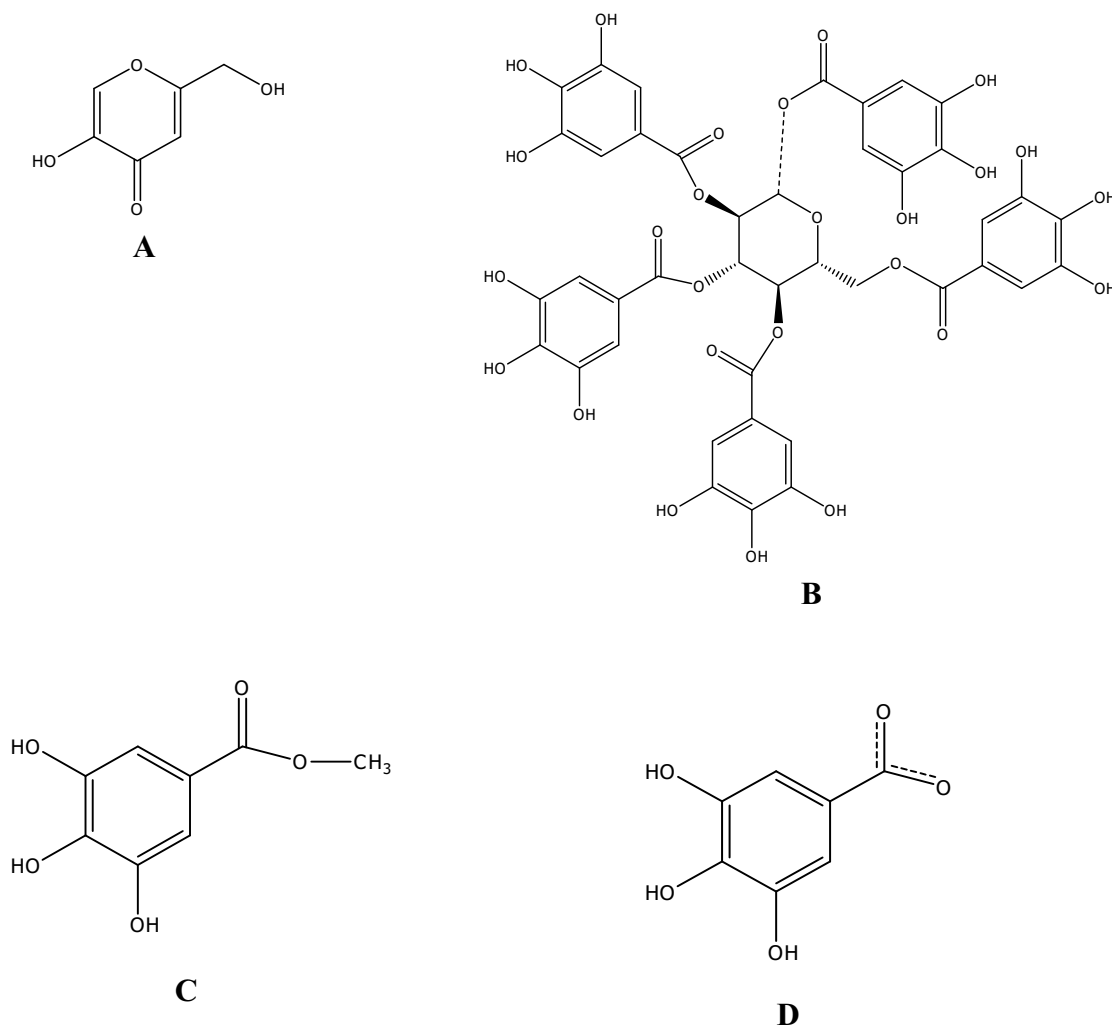
Determination of mushroom tyrosinase inhibition *in vitro*

The dopachrome method described by Iida *et al.* [17] was followed with slight modifications. Briefly, 20 mM phosphate buffer (pH 6.8, 120 μ L), 48 U/ml mushroom tyrosinase (40 μ L) and various sample concentrations (20 μ L) with or without enzyme were placed in the wells of a 96-well microplate. After pre-incubation at 25°C for 10 min, 0.85 mM L-DOPA (20 μ L) was added and subsequently incubated at 25°C for 20 min. The amount of dopachrome was measured at 492 nm in an ELISA reader (Anthos 2010, Wals, Austria) and KA was used as a positive tyrosinase inhibitor control. The extent of inhibition by the test samples was expressed as the percentage of concentration necessary to achieve 50% inhibition (IC_{50}).

Molecular modeling

Docking studies of PGG, MG and GA and the positive reference (KA) were performed. The structures of these compounds (Figure 4) were constructed and optimized at the HF/3-21G level of theory using a Gaussian 03 program [18]. To prepare the tyrosinase structure, the crystal structure of the oxy form of tyrosinase was taken from the Protein Data Bank (PDB code 1wx2) [6]. The caddy protein (ORF378) and water molecules were removed. Hydrogen atoms were added to the enzyme using the SYBYL version 7.2 program (TRIPOS Assoc., Inc., St. Louis, MO, USA). The molecular docking method was performed using the Gold version 3.2 program [19] to study the binding orientation of PGG, MG, GA (ionize) and KA into the tyrosinase structure. The radius of the binding site was set to 10 Å. The default parameters of the automatic settings were used to set the genetic algorithm parameters. The docked conformation which had the highest GoldScore was selected to analyze the mode of binding.

Figure 4. Structures of the studied compounds. (A) kojic acid, (B) 1,2,3,4,6-penta-O-galloyl- β -D-glucopyranose, (C) methyl gallate and (D) ionized gallic acid.



Statistical analysis

The sample concentration which provided 50% inhibition (IC_{50}) was calculated from the graph which plotted inhibition percentage against sample concentration. The data were expressed as mean $\pm$ standard error of the mean (SEM). All statistical analyses were carried out using Minitab[®] Release 14 for Windows (Minitab Inc., State College, PA). Analysis of variance was performed by ANOVA procedures. Significant differences between the means were determined by Tukey's pairwise comparison test at a level of $P < 0.05$.

Acknowledgements

Financial support from the Thailand Research Fund (TRF) through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0009/2549) is gratefully acknowledged. The authors are grateful to Faculty of Science, Kasetsart University for computing resources.

References

1. Kim, Y.-J.; Uyama, H. Tyrosinase inhibitors from natural and synthetic sources: structure, inhibition mechanism and perspective for the future. *Cell. Mol. Life Sci.* **2005**, *62*, 1707-1723.
2. Nerya, O.; Ben-Arie, R.; Luzzatto, T.; Musa, R.; Khatib, S.; Vaya, J. Prevention of *Agaricus bisporus* postharvest browning with tyrosinase inhibitors. *Postharv. Biol. Technol.* **2006**, *39*, 272-277.
3. Seo, S.Y.; Sharma, V.K.; Sharma, N. Mushroom tyrosinase: recent prospects. *J. Agric. Food Chem.* **2003**, *51*, 2837-2853.
4. Asanuma, M.; Miyazaki, I.; Ogawa, N. Dopamine- or L-DOPA induced neurotoxicity: the role of dopamine quinone formation and tyrosinase in a model of Parkinson's disease. *Neurotox. Res.* **2003**, *5*, 165-176.
5. Xu, Y.; Stokes, A.H.; Roskoski, R., Jr.; Vrana, K.E. Dopamine, in the presence of tyrosinase, covalently modifies and inactivates tyrosine hydroxylase. *J. Neurosci. Res.* **1998**, *54*, 691-697.
6. Matoba, Y.; Kumagai, T.; Yamamoto, A.; Yoshitsu, H.; Sugiyama, M. Crystallographic evidence that the dinuclear copper center of tyrosinase is flexible during catalysis. *J. Biol. Chem.* **2006**, *281*, 8981-8990.
7. Coe, F.G.; Anderson, G.J. Screening of medicinal plants used by the Garífuna of Eastern Nicaragua for bioactive compounds. *J. Ethnopharmacol.* **1996**, *53*, 29-50.
8. Singh, Y.N. Traditional medicine in Fiji. Some herbal folk cures used by Fiji Indians. *J. Ethnopharmacol.* **1986**, *15*, 57-88.
9. Sharma, L.D.; Bahga, H.S.; Srivastava, P.S. *In vitro* anthelmintic screening of indigenous medicinal plants against *Haemonchus contortus* (Rudolphi, 1803) cobbold, 1898 of sheep and goats. *Indian J. Anim. Res.* **1971**, *5*, 33-38.
10. Soong, Y.Y.; Barlow, P.J. Antioxidant activity and phenolic content of selected fruit seeds. *Food Chem.* **2004**, *88*, 411-417.

11. Soong, Y.Y.; Barlow, P.J. Quantification of gallic acid and ellagic acid from longa (*Dimocarpus longan* Lour.) seed and mango (*Mangifera indica* L.) kernel and their effects on antioxidant activity. *Food Chem.* **2001**, *97*, 524-530.
12. Abdalla, A.E.M.; Darwish, S.M.; Ayad, E.H.E.; El-Hamam, R.M. Egyptian mango by-product 1. Compositional quality of mango seed kernel. *Food Chem.* **2007**, *103*, 1134-1140.
13. Nithitanakool, S.; Pithayanukul, P.; Bavovada, R. Antioxidant and hepatoprotective activities of Thai mango seed kernel extract. *Planta Med.* (Submitted).
14. Kim, J.H.; Sapers, G.M.; Choi, S.W. Isolation and identification of tyrosinase inhibitor from *Galla rhois*. *Food Sci. Biotechnol.* **1998**, *7*, 56-59.
15. Duckworth, H.W.; Coleman, J.E. Physicochemical and kinetic properties of mushroom tyrosinase. *J. Biol. Chem.* **1970**, *245*, 1613-1625.
16. Khatib, S.; Nerya, O.; Musa, R.; Tamir, S.; Peter, T.; Vaya, J. Enhanced substituted resorcinol hydrophobicity augments tyrosinase inhibition potency. *J. Med. Chem.* **2007**, *50*, 2676-2681.
17. Iida, K.; Hase, K.; Shimomura, K.; Sudo, S.; Kadota, S.; Namba, T. Potent inhibitors of tyrosinase activity and melanin biosynthesis from *Rheum officinale*. *Planta Med.* **1995**, *161*, 425-428.
18. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R. *et al.* Gaussian03, Gaussian, Inc: Pittsburgh, PA, USA, 2003.
19. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727-748.

Sample Availability: Samples are available from the authors.

© 2009 by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Understanding on absorption and fluorescence electronic transitions of carbazole-based conducting polymers: TD-DFT approaches

Songwut Suramitr · Wichanee Meeto ·
Peter Wolschann · Supa Hannongbua

Received: 7 May 2009 / Accepted: 5 October 2009 / Published online: 30 October 2009
© Springer-Verlag 2009

Abstract The electronic excitation transitions of carbazole-based oligomers, $(\text{Cz-co-Cz})_N$, $(\text{Cz-co-Fl})_N$ and $(\text{Cz-co-Th})_N$ ($N = 2-4$) were investigated using density functional theory (DFT) and time-dependent (TD) DFT methods. Our results show that the calculated ground state geometries favor a more aromatic, planer structure, while the electronically excited geometries favor a quinoidic type structure. Absorption and fluorescence energies have been obtained from TD-B3LYP/SVP calculations performed on the S_1 optimized geometries and are in excellent agreement with experimental data. The experimental fluorescence excitation energies for $(\text{Cz-co-Cz})_4$, $(\text{Cz-co-Fl})_4$ and $(\text{Cz-co-Th})_4$ (2.76, 2.63, and 2.25 eV, respectively) correspond closely with the predicted S_1 transitions (2.84, 3.91 and 2.43 eV, respectively). We also report the predicted radiative lifetimes 0.52, 0.47, and 0.99 ns for $(\text{Cz-co-Cz})_N$, $(\text{Cz-co-Fl})_N$ and $(\text{Cz-co-Th})_N$, discuss the origin of the small stoke shift of the carbazole based oligomers and the magnitude of bathochromic shifts. We conclude by discussing the benefits of theoretical calculations, which can provide critical structural and electronic understanding of

excitation–relaxation phenomena that can be exploited in design of novel optical materials.

Keywords Carbazole-based · Density functional theory · TDDFT · Radiative lifetimes

1 Introduction

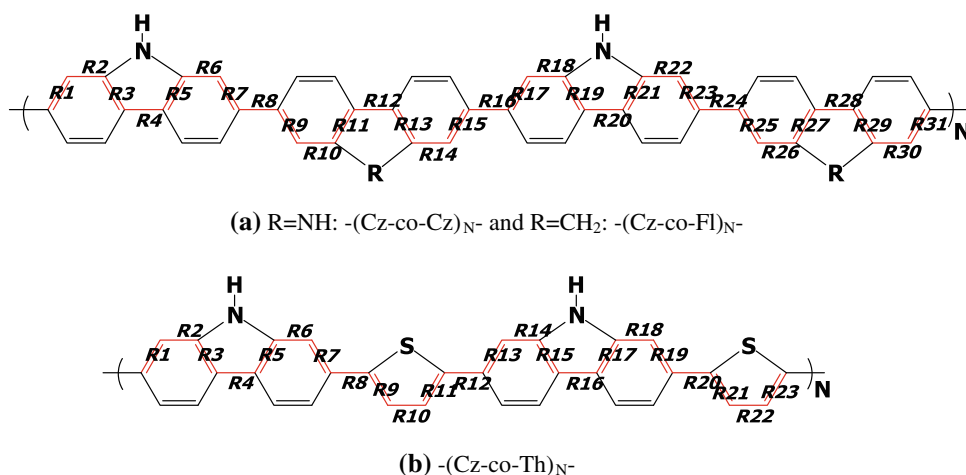
Conducting polymers as light-emitting diodes, field effect transistors, charge storage devices, photodiodes, sensors, etc. [1, 2] are currently of interest. In the last year, novel well-defined 2,7-carbazole-based (Cz) polymers were synthesized by Leclerc et al. [2–7]. 2,7-Carbazole-based polymers and derivatives with thiophene, pyrrole, phenylene, and fluorene subunits have been synthesised [8–10], are currently of both industrial and academic interest because of their wide-ranging potential in electronic devices. These novel polymeric materials are stable in air and soluble in many usual organic solvents. Interestingly, the absorption and fluorescence spectra of these materials exhibit significant differences [11–13]. It is found that the existence of multi-components in the fluorescence decay profiles of such polymers in the solid state is caused by several distinct intermolecular $\pi-\pi^*$ interactions. However, these interactions are not strong enough to provoke the appearance of distinct fluorescence bands or even to increase the bandwidths of the emission bands in solution. Fundamental understanding on structural and energetic properties of this kind of copolymers could lead to beneficial knowledge for the design of novel copolymers. Therefore, it is of interest to compare the absorption and fluorescence transitions of 2,7-carbazole-based polymers and its dependence on the structural and the electronic properties. We are also interested in exploring the

S. Suramitr · W. Meeto · S. Hannongbua
Department of Chemistry, Faculty of Science,
Kasetsart University, Bangkok 10900, Thailand
e-mail: fscisph@ku.ac.th

S. Suramitr (✉) · W. Meeto · S. Hannongbua
The Center of Nanoscience, Kasetsart University,
Bangkok, 10900, Thailand
e-mail: fsciswsm@ku.ac.th

P. Wolschann
Institute for Theoretical Chemistry, University of Vienna,
Währinger Straße 17, 1090 Vienna, Austria

Fig. 1 Structures and numbering schemes of **a** (Cz-co-Cz)_N and (Cz-co-Fl)_N and **b** (Cz-co-Th)_N oligomers



excitation mechanism since this will affect the pattern of the emission bands.

In previous studies of carbazole-based molecules [11–13], calculations were performed using DFT as well as (TD-DFT) with the B3LYP functional and three basis sets: 6-31G, 6-31G(*d*) and 6-311G(*d,p*). However, only ground state conformational analysis and calculations of the vertical excitation energies were carried out. From these calculations, it was reported that the optimized ground state geometries of oligomers with six-membered heterocyclic rings copolymers are nonplanar, whereas planar copolymer structures were found for with five-membered heterocyclic ring a results of the subtle balance between minimizing steric repulsion and maximizing electronic conjugation [14].

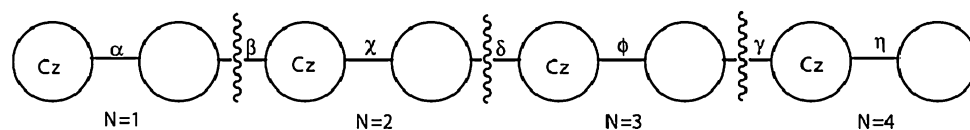
To increase our understanding of this important polymer class, we have performed calculations on the excited state properties of carbazole-based homopolymers and copolymers with fluorene and thiophene substituents (formulas are given in Fig. 1). In this study, we put particular emphasis on understanding the ground and low-lying excited states of the carbazole-homopolymer (Cz-co-Cz)_N, carbazole-co-fluorene (Cz-co-Fl)_N and carbazole-co-thiophene (Cz-co-Th)_N oligomers, which are explored by theoretical studies. The transitions associated with the absorption and fluorescence spectra of carbazole-based oligomers are explored. The fluorescence energies and radiative lifetimes are also analyzed in an attempt to gain better, more general understanding of the behavior of such systems.

2 Computational details

All QM calculations were performed using TURBOMOLE version 5.7 [15]. All geometry optimizations were performed using TURBOMOLE's JOBEX program with generalized internal coordinates and the corresponding

STATPT module [16]. The ground state and the lowest singlet excited-state geometries of the carbazole-based oligomers were optimized by DFT and TDDFT, respectively, using the B3LYP [17–21] functional for the carbazole homopolymer (Cz-co-Cz)_N, the carbazole-co-fluorene (Cz-co-Fl)_N and the carbazole-co-thiophene (Cz-co-Th)_N oligomers (Fig. 1). The default *m*³ numerical quadrature grid [22] was employed in all DFT calculations. The geometries of carbazole-based analogs were optimized without symmetry constraints using redundant internal coordinates and were considered converged if the gradient was less than 10^{−4} au. In all optimizations, the criterion for convergence was set to 10^{−8} for the energy and 10^{−7} for the density. To calculate excitation energies and analytic excited-state gradients, the TD-DFT method was used. Modules DSCF [22], GRAD, and ESCF [23] have been used. Sufficiently converged results were obtained with the valence-double-zeta quality with polarization functions SVP basis sets in close agreement with many previous studies; the conclusions are thus not expected to vary upon further basis set extensions [11–13]. In order to consider solvent effects on excitation energies, we adopted the conductor-like screening model (COSMO) [24] with a dielectric constant of $\epsilon = 4.8$ to simulate the chloroform solvent and optimized atomic radii (C, 2.00 Å; N, 1.83 Å; O, 1.72 Å; H, 1.30 Å) for the construction of the molecular cavity were used for the calculations of all molecules.

For each oligomer in Fig. 1, the chain lengths studied varied from dimers to tetramers (*N* = 1, 2, 3 and 4). The first five singlet–singlet electronic transitions (*S*₀ → *S*_{*n*}) were calculated for (Cz-co-Cz)_N, (Cz-co-Fl)_N, and (Cz-co-Th)_N oligomers using the B3LYP/SVP and TD-B3LYP/SVP methods, respectively. Based on the optimized geometries of the oligomers, the electronic absorption and fluorescence spectra were calculated at the DFT and TD-DFT levels. The absorption and fluorescence excitation energies were obtained from the ground state and the

Table 1 Bond torsional angles of oligomers in ground (S_0) and lowest excited state (S_1) (in brackets) for the (Cz-co-Cz) $_N$, (Cz-co-Fl) $_N$ and (Cz-co-Th) $_N$ molecules optimized using B3LYP/SVP and TD-B3LYP/SVP (in parenthesis) methods

Oligomers	Bond torsional angles (°)						
	α	β	χ	δ	ϕ	γ	η
(Cz-co-Cz) $_N$							
$N = 1$	140.51 (170.24)						
$N = 2$	141.54 (155.15)	141.15 (163.88)	141.56 (155.03)				
$N = 3$	141.94 (147.13)	142.79 (155.54)	142.80 (160.24)	142.76 (155.39)	142.00 (146.68)		
$N = 4$	141.99 (146.06)	141.58 (152.82)	142.25 (159.91)	143.44 (154.50)	141.88 (155.69)	141.77 (148.60)	141.91 (143.88)
(Cz-co-Fl) $_N$							
$N = 1$	142.14 (170.71)						
$N = 2$	142.20 (156.46)	142.19 (167.58)	142.30 (155.39)				
$N = 3$	142.53 (146.86)	143.38 (157.03)	142.40 (164.52)	143.21 (157.34)	142.42 (148.24)		
$N = 4$	142.69 (156.46)	142.60 (167.58)	142.44 (155.39)	141.78 (167.58)	142.66 (156.46)	142.59 (167.58)	143.04 (155.39)
(Cz-co-Th) $_N$							
$N = 1$	21.06 (0.01)						
$N = 2$	27.23 (0.02)	27.41 (0.00)	27.00 (0.00)				
$N = 3$	24.19 (0.02)	24.52 (0.04)	25.67 (0.02)	24.72 (0.00)	25.61 (0.03)		
$N = 4$	26.82 (0.00)	25.78 (0.01)	27.81 (0.01)	26.14 (0.00)	27.53 (0.01)	25.49 (0.00)	26.04 (0.02)

All torsional angles are given in degrees. N represents the number of the oligomer units

lowest singlet excited-state optimized geometries. The fluorescence electronic transitions were calculated as the vertical de-excitation based on the optimized geometry of the lowest excited state.

3 Results and discussion

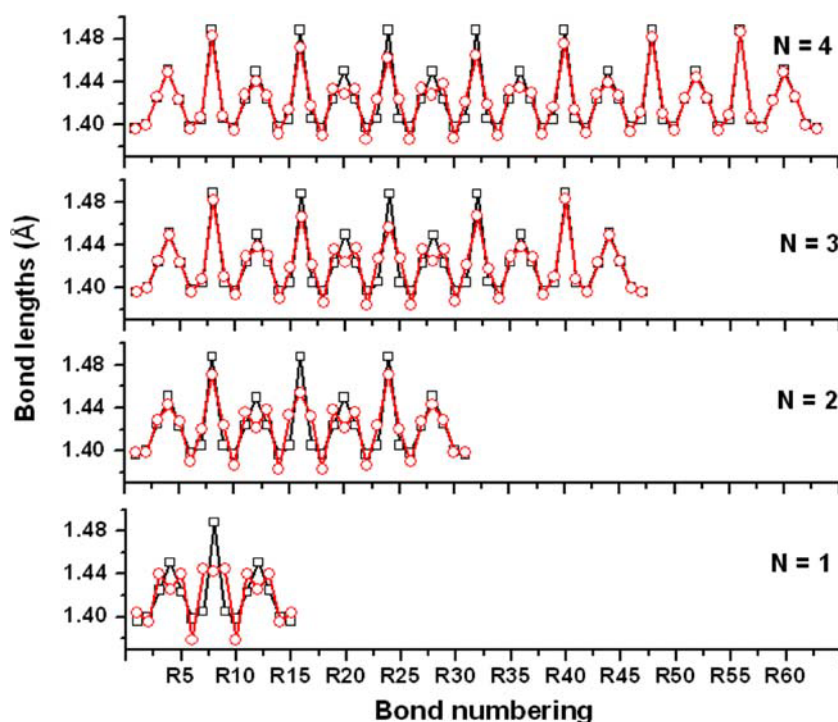
3.1 Ground and excited states structural properties

The chain length dependence on the oligomer torsional angles (θ), both at ground and excited state has been

investigated for the carbazole-based oligomers using our ground state optimized geometries. The torsion angles of the oligomers are given in Table 1 for the ground (S_0) and the lowest excited state (S_1) for the (Cz-co-Cz) $_N$, (Cz-co-Fl) $_N$ and (Cz-co-Th) $_N$ molecules.

The optimized structures of carbazole-based oligomers in the ground states are generally more distorted than the structures in excited states. We found twisted conformations with torsional angles (α , β , χ , δ , ϕ , γ and η) around 140° for both Cz-co-Cz-co-Cz and Cz-co-Fl-co-Cz. For Cz-co-Th-co-Cz, the torsional angles are around 30° , which means that the *syn* or *cis* structure of the copolymerized

Fig. 2 Computed bond lengths of ground state and excited state of homocarbazole oligomers $(\text{Cz-co-Cz})_N$. The *open square* denote electronic ground state and the *open circle* indicate the first excited state. Calculated at the B3LYP/SVP level of theory



heterocycles ring (thiophene) is energetically more favorable than the “trans” conformations [11, 12].

These ground state torsional angles appear independent the number of oligomer subunits used in the calculations. Only in the case of $(\text{Cz-co-Th})_N$ is the monomer more planar than the oligomers. On the contrary, the excited state-torsion angles are more planar than the ground state geometries with torsional angles around 180° or 0° , respectively. The differences in the bond torsional angles between the ground and lowest singlet excited state can be explained by considering the bond length changes. The structures of the ground state and the lowest singlet excited-state optimized oligomers geometries are given in Figs. 2 and 3, where the changes bond lengths for copolymer derivatives and carbazole-based oligomer can be compared. The bond numbering schemes of $(\text{Cz-co-Cz})_N$, $(\text{Cz-co-Fl})_N$ and $(\text{Cz-co-Th})_N$ oligomers are depicted in Fig. 1. In Fig. 2, the conjugated carbon–carbon bonds are illustrated for the ground state (S_0) and singlet excitations (S_1) for $(\text{Cz-co-Cz})_N$ oligomers, dimer to octamer.

From Fig. 2, it is found that the carbon–carbon bonds which lie parallel to the polymer chain (i.e. bond numbering R2, R4, R6, R8, R10, R12 and R14) become shorter while those that lie at angles other than 180° become longer. The carbon–carbon bonds are found to alternate in length between a single and double bond. In the lowest excited states, the double bond lengths increase, whereas the single bond length decreases with the changes being localized toward the centre of the oligomers. For example, the geometric changes in $(\text{Cz-co-Cz})_1$, $(\text{Cz-co-Cz})_2$,

$(\text{Cz-co-Cz})_3$ and $(\text{Cz-co-Cz})_4$ due to excitation affect the central units only. It should be noted that the spatial extent of the geometry deformations is not constant; the deformations continue to extend over the entire chain when the length increases, at least up to $(\text{Cz-co-Cz})_{N=4}$ and this can be understood in terms of the degree of charge delocalization. For longer chains, the amount of charge per monomer unit is lower and thus the geometry change is smaller. The geometry change, as obtained from DFT calculations, are found in all the $(\text{Cz-co-Cz})_N$ oligomer chains. This suggests that the center rings in the larger carbazole-based oligomers have more quinoidic character than the terminal rings [12, 13]. Therefore, the central rings in the larger oligomers were selected to confirm the trends of the C–C bond alternation along the backbone of the $(\text{Cz-co-Cz})_3$, $(\text{Cz-co-Fl})_3$ and $(\text{Cz-co-Th})_3$ as shown in Fig. 3. We monitor the changes in bond length of these aromatic systems using the bond length alternation (BLA) [25]. The BLA values for selected molecular fragment can be defined as the differences in lengths between single bonds (d_{single}) and double bonds (d_{double}) of carbon–carbon atoms (Eq. 1). A positive BLA value indicates that the molecular unit has an aromatic (quinoidic) character [26–28].

$$BLA = \sum \frac{(d_{\text{single}} - d_{\text{double}})}{N}, \quad (1)$$

The BLA associated with the carbon–carbon conjugated bond for the ground and singlet excitations states for the central rings of $(\text{Cz-co-Cz})_3$, $(\text{Cz-co-Fl})_3$ and $(\text{Cz-co-Th})_3$ oligomer was estimated. We find that the BLA changes

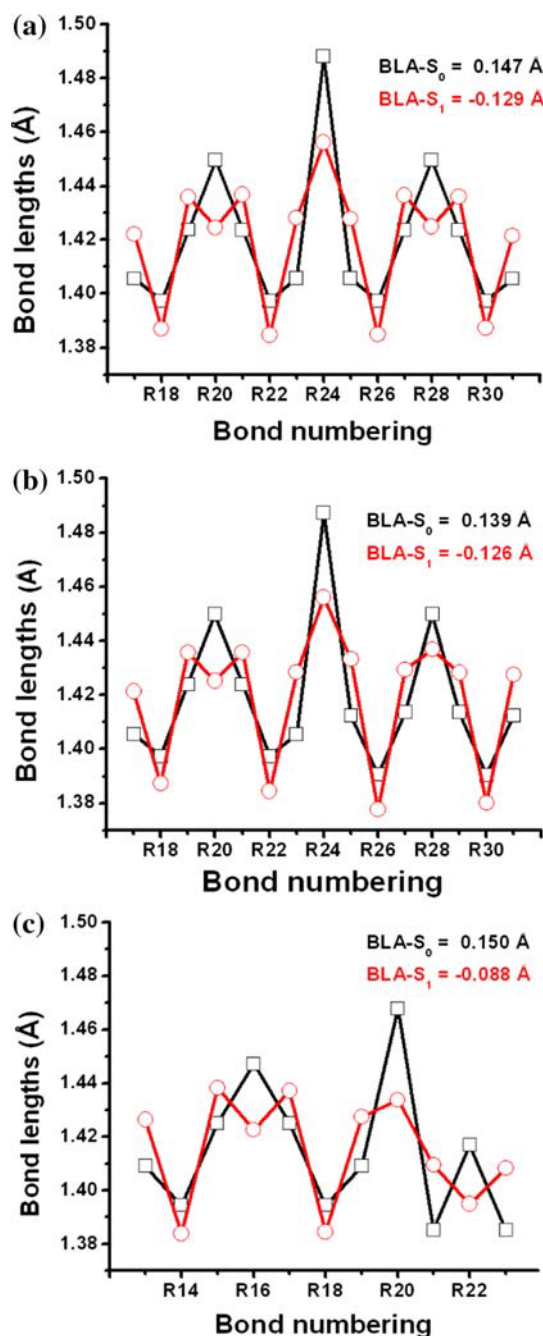


Fig. 3 The bond length alternating (BLA) of the central rings of the oligomers of (a) (Cz-co-Cz)₃, (b) (Cz-co-Fl)₃ and (c) (Cz-co-Th)₃ as shown. The open square denote electronic ground state and the open circle indicate the first excited state. Calculated using TD-DFT at the B3LYP/SVP level of theory

significantly in the excited states compared to the ground states, in the case of (Cz-co-Cz)₃ and (Cz-co-Fl)₃, decreasing from 0.147, 0.139 Å in S₀ to −0.129, −0.125 Å in S₁, respectively. In both, the ground and singlet states, the BLA differences are between 0.014 and 0.019 Å, whereas in the case of (Cz-co-Th)₃, the BLA decreases from 0.150 Å in S₀ to −0.088 Å in S₁. These

results indicate that the thiophene unit leads to a decrease in the BLA value by 0.063 Å with respect to the (Cz-co-Cz)₃ and (Cz-co-Fl)₃ oligomers. In addition, the center of the quinoidic structures is located at the linking bonds between the copolymer units such as R8, as seen clearly in Fig. 1. Chidthong et al. [29] and Wichanee et al. [13] suggest that the elongation of the molecular chain leads to minor changes in the inter-ring distances of oligomers, the largest change localized at the terminal ring. In addition to, it was found that the inter-ring bond distances do not display appreciable variations with the oligomer size. Moreover, the bond-changing pattern is varied systematically when the molecular chain is elongated. These behaviors have been also found in the case of carbazole-based homopolymer and copolymer oligomers.

3.2 Absorption and fluorescence transitions

The absorption and fluorescence excitation energies calculated by the TD-B3LYP/SVP method are reported in Table 2. The excitations energies with highest oscillator strength (π – π^* transition) of each polymer calculated by TD-B3LYP/SVP method and were extrapolated by linear regression. There is a good linear relation ($r^2 = 0.99$) between the lowest excitations and the inverse chain length. A comparison the extrapolated energy of the absorption and fluorescence excitation with the experimental results and other computed values is shown in Table 2. From these results, it was found that the excitation energies of these materials are lower than the experimental data, 0.34, 0.31, and 0.18 eV (absorption) and 0.33, 0.51, and 0.69 eV (fluorescence) for its carbazole-based, (Cz-co-Cz)_N, (Cz-co-Fl)_N and (Cz-co-Th)_N, respectively. Cornil et al. [30, 31] have shown that this can result from the overestimation of long-range electron correlation effects in the TD-DFT methods. Previous works [30–33] have shown that the high accuracy of DFT functionals such as TPSS functional is more suitable for calculation of conjugated oligomers. However, for small molecules, the impact of inductive and/or mesomeric effects induced by substituents appears to be well reproduced; however, the agreement with the correspondingly experimental values deteriorates when the chain size is increased. Our results confirm previous reports that a proper extrapolation procedure recommends the use of a rather large number of oligomers to improve the accuracy of the fit, and an accurate fitting function [30, 31, 34–37]. In fact, in order to obtain more accurate excitation energy for an infinite oligomer, one needs to use higher order polynomials. Jansson et al. studied the chain length dependence of singlet and triplet excited states of oligofluorene and used an empirical relationship proposed by Meier et al. [35, 36]. They discussed in detail the concept of “effective conjugation length”

Table 2 The calculation absorption (E_{abs}), fluorescence energies (E_{flu}) and fluorescence lifetimes of carbazole-based polymers

Oligomers	Absorption	Fluorescence	
	E_{abs} (eV)	E_{flu} (eV)	Lifetime (ns)
(Cz-co-Cz)_N			
$N = 1.0$	3.84	3.28 (1.484)	1.44
$N = 1.5$	3.51	2.91 (2.482)	1.10
$N = 2.0$	3.36	2.83 (3.222)	0.89
$N = 2.5$	3.26	2.77 (3.933)	0.76
$N = 3.0$	3.21	2.76 (4.561)	0.66
$N = 4.0$	3.17	2.76 (5.814)	0.52
$N = \infty$	2.91	2.51	
Expt.	3.25 ^a	2.84 ^a	
(Cz-co-Fl)_N			
$N = 1.0$	3.80	3.23 (1.517)	1.46
$N = 2.0$	3.32	2.77 (3.242)	0.93
$N = 3.0$	3.18	2.68 (4.537)	0.71
$N = 4.0$	3.14	2.63 (7.076)	0.47
$N = \infty$	2.89	2.40	
Expt.	3.20 ^a	2.91 ^a	
(Cz-co-Th)_N			
$N = 1.0$	3.87	3.59 (0.862)	2.07
$N = 2.0$	3.14	2.57 (2.318)	1.50
$N = 3.0$	2.87	2.34 (3.517)	1.20
$N = 4.0$	2.81	2.25 (4.598)	0.99
$N = \infty$	2.42	1.74	
Expt.	2.60 ^b	2.43 ^b	

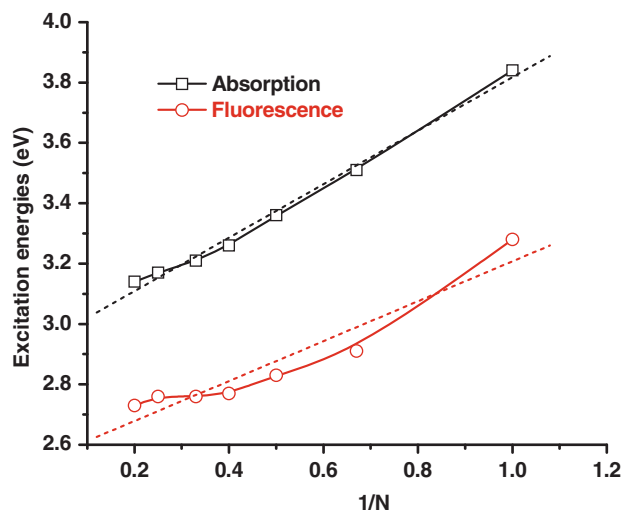
Values in parentheses are oscillator strengths

^a Ref. [6]

^b Ref. [7]

(ECL), defined as the conjugation length at which the wavelength of the absorption maximum in the series of oligomers is not more than 1 nm above the lower limit, which is given by the infinitely long polymer chain [30, 35–37]. From this definition, the ECL largely differs for the various oligomer series. Given these issues, we focus on ECL for estimate the excitation energies of these systems in the following section.

We estimate the ECL for all selected oligomers based on the convergence of the calculated excited energies of the first dipole-allowed excited states with the increasing chain length (Fig. 4). The ECL was estimated by the convergence of excitation energies with the chain length within a threshold of 0.05 eV, based on the obtained linearity between the excitation energy and reciprocal chain length. Apart from the selected carbazole-based oligomers, well-studied oligomers, such as (Cz-co-Cz)_N, (Cz-co-Fl)_N and (Cz-co-Th)_N oligomers are also reexamined with TDDFT for further validating the theory and for comparison. These results will be discussed only briefly.

**Fig. 4** First singlet excitation energies as calculated by TDDFT for absorption and fluorescence energies in poly(2,7-carbazole) oligomers, as a function of $1/N$

We present their absorption and fluorescence energies in Table 2 by applying the ECLs for the absorption and fluorescence energies for compounds (Cz-co-Cz)_N, (Cz-co-Fl)_N and (Cz-co-Th)_N oligomers. The ECL value at which a convergence of the optical properties, absorption and emission properties is reached corresponds to $N = 4$ for (Cz-co-Cz)_N, (Cz-co-Fl)_N and (Cz-co-Th)_N oligomers. Comparison of the results between experiment and absorption spectrum calculations of are shown in Fig. 5 and indicate that the estimated excitation energies, at ECLs $N = 4$, are in good agreement with the optical properties. We therefore conclude that this procedure can be used to reliably estimate the excitation energies of such polymers.

We next look at the details of the electronic transitions of each carbazole-based tetramer ($N = 4$) at TD-B3LYP/SVP level (Table 3) which can be used to describe the possible excitations of all carbazole-based molecules. From the absorption transitions, it was found that for (Cz-co-Cz)₄, (Cz-co-Fl)₄ and (Cz-co-Th)₄ molecules, the $S_0 \rightarrow S_1$ excitation primarily corresponds to the promotion of an electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO) ($H \rightarrow L$) (in Fig. 6) as indicated by large oscillator strengths (f) of : 6.112, 6.169 and 4.277, respectively. The isosurface plot of the HOMO and LUMO (Fig. 6) indicate an exchange of the double and single bonds as is typical for a $\pi \rightarrow \pi^*$ transition in conjugated polymers. On the other hand, the S_2 , S_3 and S_4 electronic transitions of each compound possess very small oscillator strengths. The fluorescence energies and the radiative lifetimes of (Cz-co-Cz)₄, (Cz-co-Fl)₄ and (Cz-co-Th)₄ computed with the TD-B3LYP/SVP method using S_1 state optimized geometries are collected in Table 3. The fluorescence energies were

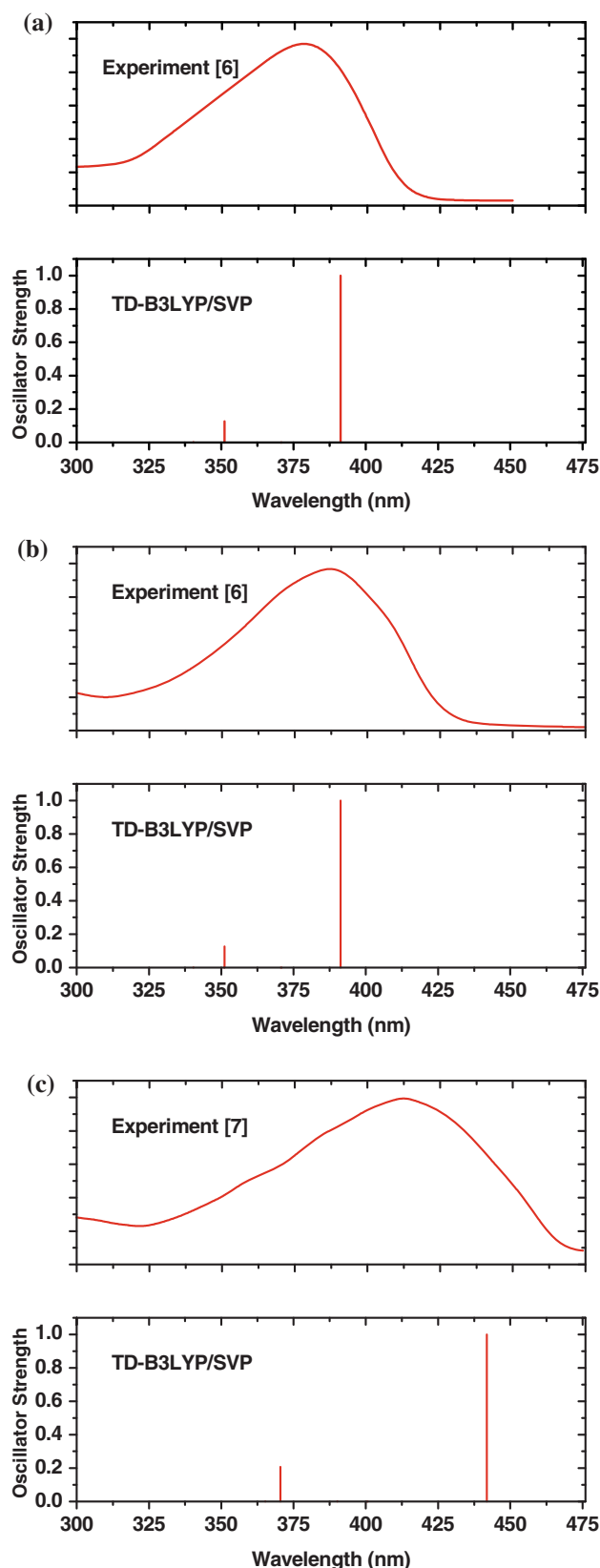


Fig. 5 Absorption spectra [6, 7] and TD-B3LYP/SVP calculations of $(\text{Cz-co-Cz})_4$, $(\text{Cz-co-Fl})_4$ and $(\text{Cz-co-Th})_4$ oligomers

also investigated. From Table 2, it is clear that the fluorescence energies of $(\text{Cz-co-Th})_N$ and $(\text{Cz-co-Fl})_N$ molecules are red-shifted from the excitation energies of $(\text{Cz-co-Cz})_N$ with values of 2.76, 2.63 and 2.25 eV for $(\text{Cz-co-Cz})_N$, $(\text{Cz-co-Fl})_N$ and $(\text{Cz-co-Th})_N$, respectively. These energies are in good agreement with experimental values [6, 7].

The differences in absorption and fluorescence energies should also lead to different Stokes shifts. We therefore evaluate the Stokes-shift as the differences $\Delta E = E_{\text{abs}} - E_{\text{flu}}$. The TD-B3LYP/SVP values exhibit Stokes shift of about 0.3 eV for $(\text{Cz-co-Fl})_N$ and are lower than those for $(\text{Cz-co-Cz})_N$ and $(\text{Cz-co-Th})_N$ by about 0.6 eV. This result demonstrates that the $(\text{Cz-co-Fl})_N$ structure is more relaxed than those of $(\text{Cz-co-Cz})_N$ and $(\text{Cz-co-Th})_N$ upon excitation. These results also show that the electronic excitation leads to the formation of a quinoid-type structure.

Using the computed structures, we can also relate the differences in the bond lengths between the ground (GS) and lowest singlet excited state (ES) to the molecular orbital nodal patterns. Because the lowest singlet state corresponds to an excitation from the HOMO to the LUMO in all of the oligomers considered here (Fig. 5), the bond-length variations were explored further in terms of the changes to the HOMO and LUMO. By comparing Fig. 5, we can see that the HOMO has nodes across the *R1*, *R3*, *R5*, *R7*, *R9*, *R11* and *R13* bonds in all molecules, but the LUMO is bonding in these regions. Therefore, one would expect a contraction of these bonds. The data reported in Fig. 2 and Fig. 3 do in fact show this given the bonds are in fact considerably shorter in the excited state. However, the bond length will increase when the bonding changes to antibonding. The dihedral angle (Table 1) between the two adjacent units shortened from 140° to 170° in $(\text{Cz-co-Cz})_N$ and $(\text{Cz-co-Fl})_N$ molecules. Whereas the dihedral angle of $(\text{Cz-co-Th})_N$ shortened from 28° to nearly 0° . It is obvious that the excited structure has a strong coplanar tendency in all molecules. It is indicated that is, the conjugation is better in the excited structure. In this result, it can see that geometry of excited state is more planar than ground state.

Finally, to investigate the effects of the structural relaxation upon excitation, radiative lifetimes were investigated. Based on the fluorescence energy and oscillator strength, the radiative lifetimes have been computed for spontaneous emission using the Einstein transition probabilities according to the formula (in au) [29, 38, 39].

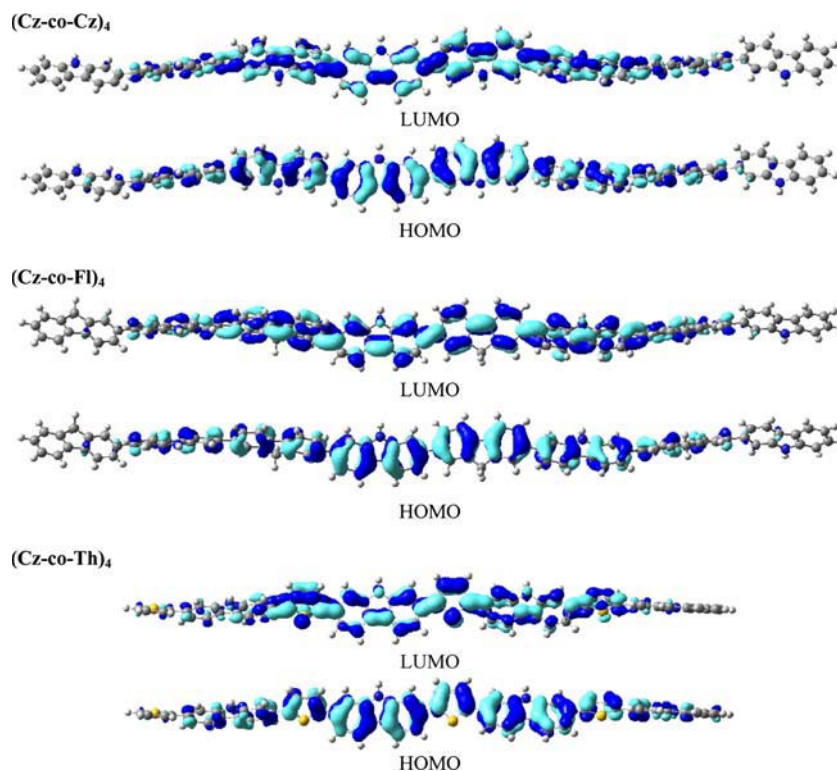
$$\tau = \frac{c^3}{2(E_{\text{Flu}})^2 f} \quad (2)$$

In Eq. 2, c is the velocity of light, E_{Flu} is the transition energy and f is the oscillator strength. The computed lifetimes, τ , for the carbazole-based oligomers are depicted

Table 3 Excitation energies (E_{ex} (eV)), oscillator strengths (f), and wave function compositions for the lowest singlet electronic states of (Cz-co-Cz)₄, (Cz-co-Fl)₄ and (Cz-co-Th)₄ molecules computed by TD-B3LYP/SVP

Electronic transitions	E_{ex}	f	Wave function composition
(Cz-co-Cz) ₄			
Absorption			
$S_0 \rightarrow S_1$	3.17	6.112	H $\rightarrow$ L(61.8%), H-1 $\rightarrow$ L+1(25.1%)
$S_0 \rightarrow S_2$	3.34	0.002	H-1 $\rightarrow$ L(45.8%), H $\rightarrow$ L+1(45.8%)
$S_0 \rightarrow S_3$	3.51	0.773	H-1 $\rightarrow$ L+1(46.5%), H $\rightarrow$ L+2(31.8%)
$S_0 \rightarrow S_4$	3.54	0.008	H-1 $\rightarrow$ L+1(47.4%), H $\rightarrow$ L+1(46.8%)
Fluorescence			
$S_1 \rightarrow S_0$	2.76	5.814	H $\rightarrow$ L(65.3%), H-1 $\rightarrow$ L+1(17.8%)
(Cz-co-Fl) ₄			
Absorption			
$S_0 \rightarrow S_1$	3.14	6.169	H $\rightarrow$ L(61.4%), H-1 $\rightarrow$ L+1(25.8%)
$S_0 \rightarrow S_2$	3.31	0.011	H-1 $\rightarrow$ L(46.4%), H $\rightarrow$ L+1(45.4%)
$S_0 \rightarrow S_3$	3.50	0.039	H $\rightarrow$ L+1(46.9%), H-1 $\rightarrow$ L(46.3%)
$S_0 \rightarrow S_4$	3.51	0.731	H $\rightarrow$ L+1(47.1%), H-2 $\rightarrow$ L(30.6%)
Fluorescence			
$S_1 \rightarrow S_0$	2.63	7.076	H $\rightarrow$ L(59.9%), H + 1 $\rightarrow$ L + 1(28.7%)
(Cz-co-Th) ₄			
Absorption			
$S_0 \rightarrow S_1$	2.81	4.277	H $\rightarrow$ L(65.0%), H-1 $\rightarrow$ L+1(18.4%)
$S_0 \rightarrow S_2$	3.10	0.000	H $\rightarrow$ L+1(47.9%), H-1 $\rightarrow$ L (45.3%)
$S_0 \rightarrow S_3$	3.18	0.012	H-1 $\rightarrow$ L+1(50.0%), H $\rightarrow$ L+1(47.0%)
$S_3 \rightarrow S_4$	3.35	0.886	H-1 $\rightarrow$ L+1(63.3%), H $\rightarrow$ L (19.1%)
Fluorescence			
$S_1 \rightarrow S_0$	2.25	4.598	H $\rightarrow$ L(65.7%), H-1 $\rightarrow$ L+1(13.3%)

Fig. 6 HOMO and LUMO of (Cz-co-Cz)₄, (Cz-co-Fl)₄ and (Cz-co-Th)₄ oligomers. Depicted are two isosurfaces of equal values but opposite sign



in Table 2. The lifetime of carbazole-based oligomers at $N = 4$ amounts to 0.52, 0.47, and 0.99 ns for (Cz-co-Cz) $_N$, (Cz-co-Fl) $_N$ and (Cz-co-Th) $_N$, respectively. Among the carbazole-based molecules, the (Cz-co-Fl) $_N$ shows the lowest lifetime, which is close to that of (Cz-co-Cz) $_N$. For the purpose of comparison, the results of a chemically similar system were used. The radiative lifetimes of poly(*N*-octyl-2,7-carbazole) and poly(*N*-octyl-2,7-carbazole-alt-9,9-dioctyl-2,7-fluorene) in THF solution are 0.51 and 0.45 ns [28] which is in good agreement with the predicted lifetimes of (Cz-co-Cz) $_N$ (0.52 ns) and (Cz-co-Fl) $_N$ (0.47 ns), respectively. Similar results have been reported for several other polymers (thin films) [40–44] and it can be concluded that their low radiative lifetime can produce useful fluorescent emission [40–44].

4 Conclusions

Absorption and fluorescence properties of (Cz-co-Cz) $_N$, (Cz-co-Fl) $_N$ and (Cz-co-Th) $_N$, are presented herein. The optimized ground state and the first singlet excited electronic state have been obtained using B3LYP and TD-B3LYP, methods, respectively, in conjunction with the SVP basis set. A chloroform solvent effect on excitation has been assessed using the COSMO implicit solvent model.

The estimated excitation energies of the absorption and fluorescence excitation based on ECLs $N = 4$ are in good agreement with the optical properties of (Cz-co-Cz) $_N$, (Cz-co-Fl) $_N$ and (Cz-co-Th) $_N$ polymers, 3.17, 3.14 and 2.81 eV (absorption) and 2.76, 2.63 and 2.25 eV (fluorescence), respectively. Compared to experimental fluorescence excitation energies available for (Cz-co-Cz) $_N$, (Cz-co-Fl) $_N$ and (Cz-co-Th) $_N$, it can be seen that TD-B3LYP/SVP calculations give good predictions of the excitation energies for the S_1 transition (2.84, 2.91 and 2.43 eV, respectively). This suggests that the procedure used herein is reliable method for the estimation of excitation energies of such polymers.

Moreover, we find that the geometry of excited states is more planar than that of the ground states. The excitation to the S_1 state causes significant changes in the predicted geometry which is in agreement with the small Stokes shifts observed experimentally. Furthermore, the radiative lifetime of carbazole-based oligomers at $N = 4$ amounts to 0.52, 0.47 and 0.99 ns for (Cz-co-Cz) $_N$, (Cz-co-Fl) $_N$ and (Cz-co-Th) $_N$, respectively, which is in agreement with the experiment lifetimes of (Cz-co-Cz) $_N$ (0.51 ns) and (Cz-co-Fl) $_N$ (0.45 ns), respectively. It is shown that the existence of multi-components in the fluorescence decay profiles of polymers is caused by several distinct intermolecular π – π^* interactions. We therefore conclude that homopolymers and copolymers derived from *N*-substituted-2,7-carbazoles

appear to be very promising materials for the future development of light-emitting diodes, electrochromic windows, photovoltaic cells, photorefractive materials.

Acknowledgments Supporting from the Thailand Research Fund (RTA5080005 to SH and MRG5180287 to SS), and the National Center of Excellence in Petroleum, Petrochemical Technology, Center of Nanotechnology Kasetsart University, Kasetsart University Research and Development Institute (KURDI), National Nanotechnology Center (NANOTEC) and Laboratory of Computational and Applied Chemistry (LCAC) are gratefully acknowledged. The calculations were performed in part on the Schrödinger III cluster of the University of Vienna. Thanks are also due to Matthew Paul Gleeson for helpful comments and reading of the manuscript.

References

- Burroughes JH, Bradley DDC, Brown AR, Marks RN, Friend RH, Burn PL, Holmes AB (1990) Light-emitting diodes based on conjugated polymers. *Nature* 347:339–341
- Morin JF, Beaupre S, Leclerc M, Levesque I, D'Iorio M (2002) Blue light-emitting devices from new conjugated poly(*N*-substituted-2,7-carbazole) derivatives. *Appl Phys Lett* 80(3):341–343
- Morin JF, Leclerc M (2001) Syntheses of conjugated polymers derived from *N*-alkyl-2,7-carbazoles. *Macromolecules* 34(14):4680–4682
- Zotti G, Schiavon G, Zecchin S, Morin JF, Leclerc M (2002) Electrochemical, conductive, and magnetic properties of 2,7-carbazole-based conjugated polymers. *Macromolecules* 35(6):2122–2128
- Morin JF, Boudreault PL, Leclerc M (2003) Blue-light-emitting conjugated polymers derived from 2,7-carbazoles. *Macromol Rapid Commun* 23:1032–1036
- Bouchard J, Belletete M, Durocher G, Leclerc M (2003) Solvatochromic properties of 2,7-carbazole-based conjugated polymers. *Macromolecules* 36:4624–4630
- Morin JF, Leclerc M (2002) 2,7-Carbazole-based conjugated polymers for blue, green, and red light emission. *Macromolecules* 35(22):8413–8417
- Belletete M, Bedard M, Leclerc M, Durocher G (2004) Absorption and emission properties of carbazole-based dyads studied from experimental and theoretical investigations. *Synth Met* 146(1):99–108
- Belletete M, Bedard M, Leclerc M, Durocher G (2004) Ground and excited state properties of carbazole-based dyads: correlation with their respective absorption and fluorescence spectra. *J Mol Struct (THEOCHEM)* 679(1–2):9–15
- Belletete M, Bedard M, Bouchard J, Leclerc M, Durocher G (2004) Spectroscopic and photophysical properties of carbazole-based triads. *J Can J Chem* 82:1280–1288
- Suramitr S, Hannongbua S, Wolschann P (2007) Conformational analysis and electronic transition of carbazole-based oligomers as explained by density functional theory. *J Mol Struct (THEOCHEM)* 807(1–3):109–119
- Marcon V, Van der Vegt N, Wegner G, Raos G (2006) Modeling of molecular packing and conformation in oligofluorenes. *J Phys Chem B* 110(11):5253–5261
- Meeto W, Suramitr S, Vannarat S, Hannongbua S (2008) Structural and electronic properties of poly(fluorine-vinylene) copolymer and its derivatives: Time-dependent density functional theory investigation. *Chem Phys* 349(1–3):1–8
- Jespersen KG, Beenken WJD, Zaushtsyn Y, Yartsev A, Andersson M, Pullerits T, Sundström V (2004) The electronic states

- of polyfluorene copolymers with alternating donor–acceptor units. *J Chem Phys* 121:12613–12617
15. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Electronic structure calculations on workstation computers: the program system turbomole. *Chem Phys Lett* 162(3):165–169
 16. von Arnim M, Ahlrichs R (1999) Geometry optimization in generalized natural internal coordinates. *J Chem Phys* 111(20):9183–9190
 17. Dirac PAM (1929) Quantum mechanics of many-electron systems. *Proc R Soc Lond A* 123:714–733
 18. Slater JC (1951) A simplification of the Hartree–Fock method. *Phys Rev* 81:385–390
 19. Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can J Phys* 58(8):1200–1211
 20. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 38(6):3098–3100
 21. Kohn W, Becke AD, Parr RG (1996) Density functional theory of electronic structure. *J Chem* 100(31):12974–12980
 22. Treutler O, Ahlrichs R (1995) Efficient molecular numerical integration schemes. *J Chem Phys* 102(1):346–354
 23. Bauernschmitt R, Ahlrichs R (1996) Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory. *Chem Phys Lett* 256(4–5):454–464
 24. Klamt A, Schüürmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc Perkin Trans* 2:799–805
 25. Jacquemin D, Perpète EA, Chermette H, Ciofini I, Adamo C (2007) Comparison of theoretical approaches for computing the bond length alternation of polymethineimine. *Chem Phys* 332(1):79–85
 26. Liu Q, Liu W, Yao B, Tian H, Xie Z, Geng Y, Wang F (2007) Synthesis and chain-length dependent properties of monodisperse oligo(9, 9-di-n-octylfluorene-2,7-vinylene)s. *Macromolecules* 40(6):1851–1857
 27. Yang L, Ren AM, Feng JK, Wang JF (2005) Theoretical investigation of optical and electronic property modulations of pi-conjugated polymers based on the electron-rich 3, 6-dimethoxy-fluorene unit. *J Org Chem* 70(8):3009–3020
 28. Wang JF, Feng JK, Ren AM, Liu XD, Ma YG, Lu P, Zhang HX (2004) Theoretical studies of the absorption and emission properties of the fluorene-based conjugated polymers. *Macromolecules* 37(9):3451–3458
 29. Chidthong R, Hannongbua S, Aquino A, Wolschann P, Lischka H (2007) Excited state properties, fluorescence energies, and lifetime of a poly(fluorene-pyridine) copolymer, based on TD-DFT investigation. *J Comput Chem* 28(10):1735–1742
 30. Cornil J, Gueli I, Dkhissi A, Sancho-Garcia JC, Hennebicq E, Calbert JP, Lemaire V, Beljonne D (2003) Electronic and optical properties of polyfluorene and fluorene-based copolymers: a quantum-chemical characterization. *J Chem Phys* 118:6615–6623
 31. Gierschner J, Cornil J, Egelhaaf HJ (2007) Optical bandgaps of π -conjugated organic materials at the polymer limit: experiment and theory. *Adv Mater* 19:173–191
 32. Sancho-García JC (2006) Assessing a new nonempirical density functional: difficulties in treating π -conjugation effects. *J Chem Phys* 124:124112/1–124112/10
 33. Sancho-García JC (2007) Treatment of singlet–triplet splitting of a set of phenylene ethylenes organic molecules by TD-DFT. *Chem Phys Lett* 439:236–242
 34. Jansson E, Chandra Jha P, Ågren H (2007) Chain length dependence of singlet and triplet excited states of oligofluorenes: a density functional study. *Chem Phys* 336(2–3):91–98
 35. Meier H, Stalmach U, Kolshorn H (1997) Effective conjugation length and UV/vis spectra of oligomers. *Acta Polymer* 48:379–384
 36. Oelkrug D, Tompert A, Egelhaaf HJ, Hanack M, Steinhuber E, Hohloch M, Meier H, Stalmach U (1996) Towards highly luminescent phenylene vinylene films. *Synth Met* 83(3):231–237
 37. Peach GMJ, Tellgren EI, Saek P, Helgaker T, Tozer DJ (2007) Structural and electronic properties of polyacetylene and polyyne from hybrid and coulomb-attenuated density functionals. *J Phys Chem A* 111(46):11930–11935
 38. Luke V, Aquino A, Lischka H, Kauffmann HF (2007) Dependence of optical properties of oligo-*para*-phenylenes on torsional modes and chain length. *J Phys Chem B* 111(28):7954–7962
 39. Beenken WJD, Lischka H (2005) Spectral broadening and diffusion by torsional motion in biphenyl. *J Chem Phys* 123:144311–144319
 40. Tirapattur S, Belletete M, Drolet N, Leclerc M, Durocher G (2003) Steady-state and time-resolved studies of 2,7-carbazole-based conjugated polymers in solution and as thin films: determination of their solid state fluorescence quantum efficiencies. *Chem Phys Lett* 370(5–6):799–804
 41. Jenekhe SA, Lu L, Alam MM (2001) New conjugated polymers with donor–acceptor architectures: Synthesis and photophysics of carbazole-quinoline and phenothiazine-quinoline copolymers and oligomers exhibiting large intramolecular charge transfer. *Macromolecules* 34(21):7315–7324
 42. Schenning APH, Tsipis AC, Meskers SCJ, Beljonne D, Meijer EW, Brédas JL (2002) Electronic structure and optical properties of mixed phenylene vinylene/phenylene ethynylene conjugated oligomers. *Chem Mater* 14(3):1362–1368
 43. Pålsson LO, Wang C, Russell DL, Monkman AP, Bryce MR, Rumbles G, Samuel DW (2002) Photophysics of a fluorene copolymer in solution and films. *Chem Phys* 279(2–3):229–237
 44. Sun RG, Wang YZ, Wang DK, Zheng QB, Kylo EM, Gustafson TL, Wang F, Epstein AJ (2000) High PL quantum efficiency of poly(phenylene vinylene) systems through exciton confinement. *Synth Met* 111–112:595–602



QM methods in structure based design: Utility in probing protein–ligand interactions

M. Paul Gleeson^{a,*}, Supa Hannongbua^{a,b}, Duangkamol Gleeson^c

^a Department of Chemistry, Faculty of Science, Kasetsart University, 50 Phaholyothin Rd, Chatuchak, Bangkok 10900, Thailand

^b Center of Nanotechnology KU, Kasetsart University, 50 Phaholyothin Rd, Chatuchak, Bangkok 10900, Thailand

^c Department of Chemistry, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

ARTICLE INFO

Article history:

Received 8 June 2010

Received in revised form

27 September 2010

Accepted 30 September 2010

Available online 8 October 2010

Keywords:

Hydrogen bond strength

Kinase inhibitors

QM

QM/MM

Structure based design

ABSTRACT

Small changes in ligand structure can lead to large unexpected changes in activity yet it is often not possible to rationalize these effects using empirical modeling techniques, suggesting more effective methods are required. In this study we investigate the use of high level QM methods to study the interactions found within protein–ligand complexes as improved understanding of these could help in the design of new, more active molecules.

We study aspects of ligand binding in a set of protein ligand complexes containing ligand efficient, fragment-like inhibitors as these structures are often challenging to determine experimentally. To assess the reliability of our theoretical models we compare the MP2/6-31+G** QM results to the original X-ray coordinates and to QM/MM B3LYP/6-31G**/UFF results which we have previously reported. We also contrast these results with data obtained from an analysis of the distribution of comparable interactions found in (a) high resolution kinase complexes (≤ 1.8 Å) from the PDB and (b) more generic, small molecule crystal structures from the CSD.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Structure based design (SBD) involves the structurally driven, iterative alteration of a molecule to increase its affinity for a given receptor by modulating the fine balance between attractive and repulsive interactions present in the protein–ligand complex. It is known that even small changes in ligand structure can often have very dramatic effects on activity confirming subtle interactions are often at play [1,2]. An ability to more effectively understand and predict these subtleties is one of the next challenges in SBD which will allow now routinely determined 3D crystallographic coordinates to be used in a more quantitative fashion to guide the structural modification of a lead series [3,4]. As a result of this requirement, an increasing number of research articles are being reported in the literature that focus on improving our fundamental understanding of the interactions between molecules including pi–pi stacking [5–8], halogen bond interactions [9,10] and the conformational preferences of functional groups [11,12] amongst others. These studies have realized valuable insights through both the mining of experimental structural databases [3,4,11–14] and quantum mechanical (QM) calculations [5–11,15,16].

Kinases represent an ideal target class to investigate the subtleties of protein–ligand binding since their implication in a wide variety of medical conditions has resulted in a huge amount of high quality structural information being amassed on a wide variety of sub-families and inhibitors [17]. The majority of kinase inhibitors bind to the ATP binding site (Type 1) by forming between 1 and 3 hydrogen bonds (H-bonds) with the 3 amino acids that constitute the hinge region (Fig. 1). These inhibitors are ATP mimics, consisting of either a single or multiple aromatic rings which inhibit the protein through a combination of polar interactions with the so called hinge and non-specific binding due to their relatively high lipophilicity. The interactions formed at the hinge are exemplified by the crystal structure of CDK2 (1WCC, Fig. 1). The inhibitor makes a single H-bond interaction to the central hinge acceptor, and as a result of the conformation of the hinge backbone, makes rather short CH mediated interaction with one of the hinge donors [18,19]. Given the ligand has only 350 μ M affinity for CDK2 and the structure is of resolution 2.2 Å it might be appropriate to assess the binding interactions with those of comparable structures in the PDB, CSD or to QM calculations. This might also be pertinent in light of research that show the ligand density associated with relatively weak binding inhibitors can be less distinct, even for protein structures of good overall resolution [20], which can complicate their application to SBD [21]. Indeed classical molecular mechanical (MM) [22,23] and QM based methods [24,25] have been specifically developed to help improve the active site placement of ligands and co-factors.

* Corresponding author. Tel.: +66 86 5242120; fax: +66 2 5793955.
E-mail address: paul.gleeson@ku.ac.th (M.P. Gleeson).

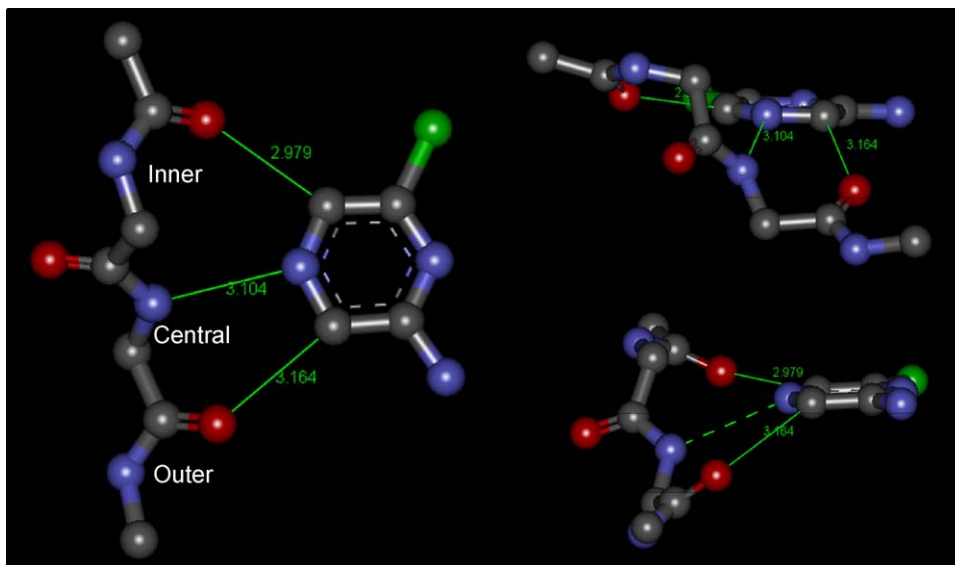


Fig. 1. Interactions formed between a small molecules inhibitor of CDK2 and the hinge backbone (PDB ID 1WCC, 2.2 Å). Clockwise from top right, the images show the binding pocket from a position (1) behind the hinge, (2) looking in from solvent, and (3) looking down from the N-lobe. Note the apparently ideal H-bond distances, angles and dihedrals formed between the inner hinge C=O and H–C of the ligand.

MM techniques are widely used in the pharmaceutical industry as they are both fast and the results are easily interpretable due to the relatively simple terms used in the derivation of their forcefields. They have proved very effective in a wide variety of applications yet it is becoming increasingly clear that these methods may lack the required accuracy to describe the subtleties associated with protein–ligand complexes due to limitations in empirical forcefield themselves [8,9,26–29]. QM calculations are a more accurate alternative to study chemical phenomenon in proteins including aspects of protein–ligand interactions [15,30–35] under investigation here, as well as their better known role in assessing chemical reactivity [36]. QM methods are typically employed on a small representation of the active site, consisting of the key residues for computational efficiency, which means the effects of potentially important longer range interactions will be lacking. This has prompted interest in the so-called QM/MM methods [37–39] which overcome some of the deficiencies associated with both QM and MM methods separately. In this type of calculation, the residues involved in the ligand binding event are treated using QM while the remainder of the protein is modeled using MM. In this way, the critical interactions between the ligand and a receptor are treated QM, the long range electrostatics can be included explicitly and medium range VDW terms accounted for classically.

QM calculations offer certain advantages over QM/MM methods including their ease of setup and the ability to use much higher levels of theory. Thus, the neglect of protein environment is balanced by the fact that we can more accurately model the local interaction between the protein and ligand.

In this study, we assess the utility of QM methods to probe aspects of protein ligand interactions using the experimental X-ray derived atomic coordinates of eight different kinase protein–ligand complexes. High level MP2/6-31+G** gas-phase models are employed to assess the ability of the method to describe the structural properties observed in the active sites. We specifically focus on small kinase inhibitors which interact solely with the 3 amino acid residues of the hinge region so that we can use the highest possible levels of QM theory.

In such cases a small gasphase model might be considered a more reasonable surrogate for the overall system since the protein environment is expected to play a much more marginal role in binding due to the nature of the ligands (at least compared to

popular QM methods to assess H-bond strength based on simple probes [15]). Even so, the lack of the protein environment provides us with a useful opportunity to assess how ideal the ligands in question interact with the key polar moieties in their respective active sites. By neglecting the nonpolar residues in our QM calculations (i.e. residues defining the ATP non-polar cavity), we expect that an inhibitor which optimally interacts with the hinge region will show a gasphase binding mode that will differ only slightly from that in the protein. This situation would suggest that the association between the inhibitor and the polar moiety is optimal (i.e. not unfavorably hindered due to the steric constraint imposed by the protein environment). Furthermore, an inhibitor whose binding mode alters substantially on gasphase optimization must make sub-optimal interactions to the polar active site feature (s) due to steric constraints arising from the protein. In the latter case, an obvious way of increasing inhibitor binding would be to maximize the interaction of the inhibitor with the polar active site features by minimizing the unfavorable steric clash that prevents the interaction from being optimal.

We assess the utility of the gasphase QM models of the 8 kinase complexes (Table 1) by comparing the predicted H-bond interaction characteristics (distances, angles and ligand RMSDs) to the original experimental results and those derived using QM/MM based models previously reported by us elsewhere [26,27] (for 4 out of 8 structures). We also contrast the results to comparable H-bond interactions values obtained from a search of the PDB (kinase type-1 inhibitors with a resolution <1.8 Å) and interactions derived from high resolution small molecule structures from the Cambridge Structural Database (CSD) [40].

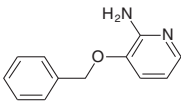
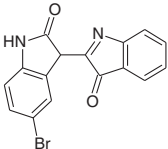
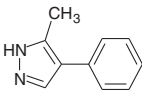
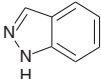
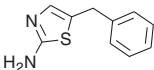
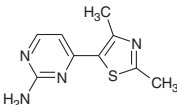
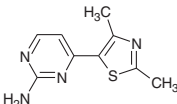
2. Computational procedures

Crystal structures of the 8 protein-kinases listed in Table 1 were downloaded from the RCSB protein databank [41]. These structures were obtained by searching the PDB for small molecule inhibitors that mediate their interaction with the hinge only (i.e. no water molecules). They are of resolutions typically employed in SBD applications.

QM models were created by extracting only the key protein residues that interact with the ligand, and the ligand itself. The QM representation used in this study is exemplified in Fig. 1. It has been

Table 1

PDB structures used in the QM analyses. Reported are the PDB ID, small molecule inhibitor, resolution, kinase target and a description of the H-bonds mediated with the hinge. Outer (O), central (C) and inner (I) H-bonds correspond to those defined in Fig. 1. CH refers to a short interaction distance between a carbonyl group of the hinge and a CH hydrogen atom of the inhibitor.

PDB ID	Inhibitor	Resolution (Å)	Target	H bonds
1W7H		2.2	P38	O(CH), C, I
2BHE ^a		1.9	CDK2	O, C, I
2UVX		2.0	PKA-B	O(CH), C, I
2UW3		2.2	PKA-B	C, I
2VTA		2.0	CDK2	O, C, I(CH)
3DND		2.3	CDK2	O(CH), C, I
2C5O ^a		2.1	CDK2(active)	O, C, I(CH)
1PXJ		2.3	CDK2(inactive)	O(CH), C, I

^a No density available to assess the X-ray atomic coordinates.

used as the core region in QM/MM calculations by both us [26,27] and others [10] to elucidate aspects of non-bonded interactions in kinase-inhibitor complexes. This suggests it is a sufficiently large representation of the key active site interactions for use in these gasphase calculations. Geometry optimization were performed in Gaussian 03 [42] at the MP2/6-31+G** level of theory to accurately account for the non-bonded interactions observed within the model systems. The C_α atoms of the truncated amino acids were frozen during geometry optimization.

QM/MM calculations were also used to compare and contrast the QM results to those of the original X-ray structure. These calculations not only employ the same QM representation but also include the effect of protein environment using an MM potential. The QM/MM results have been obtained from two distinctly different studies reported elsewhere: (a) based on a QM/MM optimization of the ligand in its native protein conformation [27] and (b) based on a QM/MM optimization of the ligand in a non-native conformation of the same protein, where the docked conformation was derived from the 3D alignment of both conformations [26]. Briefly, all QM/MM were performed using the ONIOM methodology developed by Morokuma and co-workers [43,44] and have been reported elsewhere in detail [26,27]. The QM regions were described using the B3LYP/6-31G** method and the MM using the Universal Force Field (UFF) [45]. Hydrogen link atoms were used

to satisfy atoms at the QM and MM interface. Atoms in the QM region were optimized using the electrical embedding scheme. The hydrogen link atoms and MM atoms remained fixed during optimization. Note, all atoms that directly interact with the inhibitors are treated both QM and flexibly. As noted by Bizzanti et al., in such studies it is advisable to keep the protein rigid with the exception of the residues that are known to be flexible [3]. This is because repulsive interactions, either real or due to sub-optimal refinement, can be unrealistically dissipated into the wider protein by very small changes in the overall protein conformation following MM optimization. This could arise due to the limitations of MM force-fields [8,9,24–26] or the coupling between QM and MM regions in QM/MM calculations [46].

We have also extracted information on the H-bond interactions formed between type-1 kinase inhibitors from the PDB with X-ray resolutions ≤1.8 Å to compare with the QM, QM/MM and X-ray structural models. We have also extracted protein non-specific information for comparable H-bonds formed between high resolution small molecules crystal structures in the ISOSTAR database [47]. For all models we determine the distance between HBA and HBD, the corresponding heavy atoms, and the bond angle. As X-ray structures extracted from the PDB lack hydrogens, these were added using the AMBER [48]/GAFF [49] forcefield prior to determining the HBA–HBD distance. RMSDs (in Å) of the QM opti-

Table 2

Characteristics of kinase-inhibitor H-bonds interactions with the hinge region for (a) 57 high resolution kinase-inhibitor PDB complexes, (b) 8 fragment-like kinase-inhibitor PDB complexes and (c) the equivalent QM MP2/6-31+G** optimized coordinates of the 8 PDB complexes. Also reported are the characteristics of comparable interactions found in small molecule crystal structures contained within the ISOSTAR database. In all cases the mean distances (Å), angles (degrees) and number of cases (N) are reported, along with the corresponding standard deviations in parenthesis.

	NH...N(ar)	C=O...H-N(ar)	C=O...H-C(ar)		
A. PDB ≤ 1.8 Å	Central H-Bond	outer H-Bond	inner H-Bond	outer H-Bond	inner H-Bond
X–HY distance	2.09 (0.18)	1.97 (0.14)	1.92 (0.14)	2.60 (0.31)	2.19 (0.18)
X...Y distance	3.06 (0.17)	2.86 (0.12)	2.87 (0.12)	3.29 (0.29)	3.19 (0.14)
Angle	123.4 (3.5)	102.9 (8.7)	133.5 (5.1)	91.7 (14.2)	143.8 (3.9)
N	57	30	25	27	32
B. Isostar	Generic HN...N(ar)		Generic C=O...H-N(ar)		Generic C=O...H-C(ar)
X–HY distance	1.95 (0.16)		1.92 (0.12)		2.40 (0.16)
X...Y distance	2.91 (0.32)		2.87 (0.08)		3.15 (0.08)
Angle	117.6 (12.1)		130.3 (27.7)		118.5 (16.8)
N	737		3991		1841
C. PDB (N=8)	Central H-bond	Outer H-bond	Inner H-bond	Outer H-bond	Inner H-bond
X–HY distance	1.99 (0.35)	1.97 (0.11)	1.93 (0.10)	3.04 (0.67)	2.46 (0.71)
X...Y distance	3.04 (0.35)	2.86 (0.24)	2.88 (0.07)	3.69 (0.48)	3.23 (0.37)
Angle	118.9 (6.4)	117.6 (22.8)	135.8 (5.8)	86.2 (6.7)	146.8 (89.0)
N	8	2	6	5	2
D. MP2/6-31+G** (N=8)	Central H-bond	Outer H-bond	Inner H-bond	Outer H-bond	Inner H-bond
X–HY distance	1.89 (0.04)	2.00 (0.21)	1.90 (0.07)	2.66 (0.62)	2.47 (0.07)
X...Y distance	2.91 (0.04)	2.96 (0.11)	2.91 (0.06)	3.58 (0.55)	3.25 (0.11)
Angle	121.1 (4.5)	117.0 (23.5)	134.2 (4.3)	77.7 (8.0)	150.2 (3.9)
N	2	8	6	5	2

mized coordinates were determined by optimally realigning the molecules to the original X-ray coordinates using ROCS [50]. The RMSDs were calculated using non-hydrogen atoms only.

3. Results and discussion

The theoretical results from the QM and QM/MM models for the complexes in Table 1 can be compared to their original experimental coordinates and to searches of comparable interactions from the PDB and CSD (Table 2). To understand the utility of QM calculations in SBD we must first assess their ability to describe the interactions found within the set of test protein–ligand complexes. For example, in the 2.2 Å CDK2 structure shown in Fig. 1, the ligand makes a single traditional H-bond interaction with the hinge, formed between its aromatic nitrogen acceptor and the hinge central nitrogen donor (3.1 Å). The bond angle of 111° deviates from the idealized value of ~120° that would be expected. It is also apparent that a relatively strong interaction is made between the hinge carbonyl acceptors and aromatic CH atoms ($C_{ar}H$) of the pyrazine ring. The heavy atom distances correspond to 3.0 Å and 3.2 Å for the inner and outer hinge interactions, respectively. These correspond to $C=O...HC_{ar}$ H-bond distances of 2.0 Å and 2.4 Å, respectively when hydrogen atoms are added according to the AMBER forcefield. Surprisingly, the former CH interaction distance is essentially equivalent to a conventional $C=O...HN$ H-bond formed with the central hinge donor [18]. This suggests the $C_{ar}H$ mediated interaction is either abnormally strong interaction or corresponds to a rather subtle error introduced in the PDB refinement process.

Any structural deviation could of course be due to issues with either the QM representation (i.e. level of theory, inadequacies in the gasphase model) or the X-ray derived atomic solution (i.e. the X-ray coordinates are in reality an empirically derived theoretical model fitted to the experimental electron density, solved using classical methods) [20,21,51]. To put these results in better context, and to better understand the characteristics of kinase inhibitors more generally, we have also performed 2 separate database analyses. An analysis of high resolution X-ray structures (≤ 1.8 Å) of

kinase type-1 inhibitors reported in the PDB will give us information about the frequency of interactions made with the 3 possible H-bonds of the hinge (are they equally probable?) and their relative strength. Additionally, an analysis of comparable interactions found in small molecules crystal structures from the CSD will tell us how optimal the kinase based interactions are compared to those formed in a more ideal, less sterically constricted situations.

3.1. Crystallographic analysis of high resolution kinase-inhibitor complexes

We first discuss the results obtained from the PDB and CSD database searching exercises. In Table 2 and Fig. 2, we report the distance between heavy atoms involved in the interaction, the angle associated with the interaction and the actual H-bond distance which was determined by adding hydrogen atoms to the experimental coordinates. Also reported is the number of times that the interaction is classified as being made.

Fifty-seven high resolution type-1 kinase inhibitor–protein structures were extracted from the PDB and the characteristics of their hinge–ligand interactions are summarised in Table 2a. The inhibitors contain at least 1 aromatic heterocyclic ring which is involved in the formation of a strong H-bond to the hinge H-bond donor (central donor). The average distance observed between the corresponding heavy atoms is 3.06 Å and the distribution displays a low standard deviation of just 0.17 Å. The bond angle is only slightly larger (by ~3°) than the “ideal” value of 120°. However, the corresponding average H-bond distance is 2.1 Å, which is nearly 0.15 Å longer than the average distance observed from comparable interactions in the CSD ($N=737$) (Table 2b). We also find that the heavy atom distance is 0.15 Å shorter suggesting it is not an artefact of AMBER (H atoms are resolved in the CSD). Furthermore, the bond angles display comparable values to those of the 57 kinase complexes, being ~3° from the ideal value. This suggests the interaction formed between the inhibitors and the central H-bond donor may be sub-optimal due to steric factors arising from the close proximity of the outer and inner hinge acceptor features. Interestingly, this

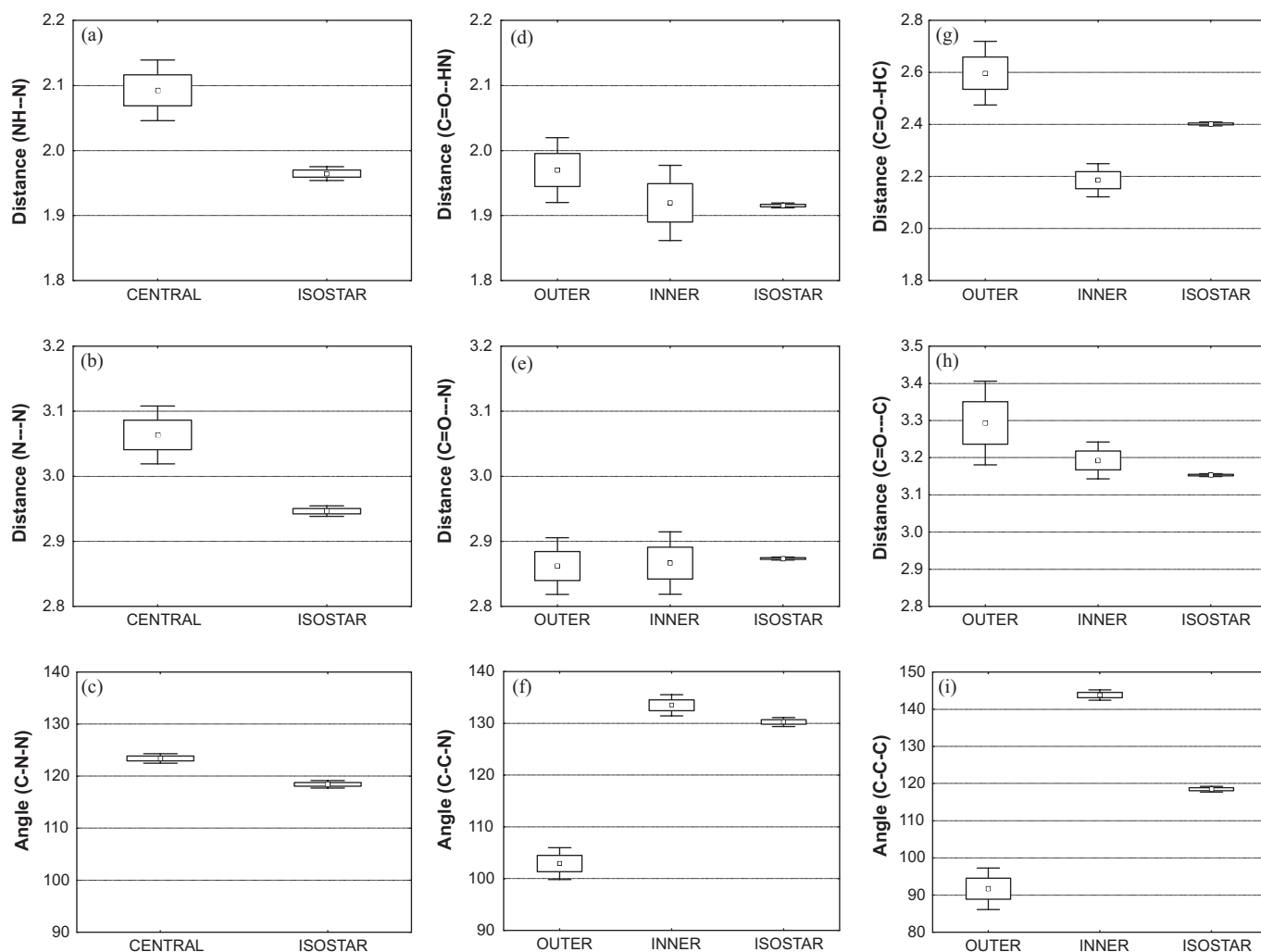


Fig. 2. The characteristics of the central and outer/inner H-bond interactions for 57 kinase X-ray structures with resolutions <1.8 Å. Also shown are the comparable ISOSTAR derived distances. All involve nitrogen containing aromatic heterocycles that interact with an amide donor. Squares denote the 95% confidence in the mean and bars the 99% confidence. Panels A–C correspond to (N–H...N) interactions, D–F correspond to (C=O...HN) interactions and G–I non-classical (C=O...HC) interactions. All distances are given in Å.

analysis suggests that the central H-bond interaction to inhibitors is sub-optimal when compared to a comparable interaction found in the CSD due to the requirement that additional interactions be made with the adjacent hinge acceptors. This can be appreciated visually from Fig. 2 where the differences in the mean values, and the corresponding errors in the mean are given.

Traditional N donor mediated H-bonds are formed to the outer carbonyl acceptor by 30 of the 57 PDB complexes while 25 are formed by the inner acceptor. It is unclear whether this means the outer H-bond distance is preferred given the magnitude of the different populations and the relatively small sample size. The average heavy atom distance in both cases is almost identical at 2.9 Å with bond angles of $\sim 103^\circ$ and 134° . Analysis of the comparable interactions formed by generic small molecules reported in the CSD ($N=3991$) reveals that the mean heavy atom distance obtained from the high resolution PDB set are almost identical (2.87 Å) (Fig. 2). They also suggest that the inner H-bond appears more optimal from an angle perspective. Additionally, a similar preference is noted when looking at the H-bond distances of 1.97 Å and 1.92 Å to the outer and inner donors, respectively. An unpaired *t*-test reveals that the differences in mean distances and angles are significant at the $>90\%$ and $>99\%$ confidence levels respectively ($>95\%$ being considered statistically significant). It is difficult there-

fore to conclude whether one hinge acceptor interaction is truly preferred over the other given that the outer interaction is more frequently made in the set of 57 PDB structures, but when made, the inner interaction appears to be stronger.

In the majority of cases, kinase inhibitors do not make all 3 H-bonds to the hinge that are possible. In the case of 1WCC, for example (Fig. 1), the inhibitor orients two carbon bonded hydrogen (CH) atoms towards the hinge acceptors. CH atoms are of course not strong hydrogen bond donors so this interaction will be viewed as sterically repulsive in SBD terms. From an analysis of the 57 high resolution PDB structures we can see that 32 (56%) of the kinase form CH interactions with the inner hinge donor while 27 form them with the outer hinge donor (47%). Interestingly, while the heavy atom distance between donor and acceptor is comparable (3.2–3.3 Å), because of the angle difference, the interaction distance to the inner H-bond is dramatically shorter (0.4 Å) with a mean value of 2.2 Å. The difference in the mean values is significant above the 99% confidence level suggesting it is a real effect. Analysis of similar H-bond interactions found in small molecule crystal structures reported in the CSD ($N=1841$) reveals that these values are intermediate to the interactions to the outer and inner acceptors found in the PDB set. While these interactions are unquestionably weak [18,19], there may be a benefit in optimizing such weak

interactions, particularly when viewed alongside halogen mediated interactions which are now considered important from a SBD perspective [9,10]. Additionally, this observation raises concerns for the application of MM methods to study such subtle interactions as it is known, for example, that relatively weak, non-standard interactions are not well described using standard MM methods [9].

3.2. QM analysis of fragment-like kinase complexes

The affinity of a ligand for a given receptor depends on a subtle balance of attractive and repulsive forces that define the extent of partitioning between itself and solution. For a good discussion, see Ref. [3]. The free energy of binding to a receptor can essentially be altered by (a) maximising the strength of polar interactions (b) reducing the conformational energy penalty associated with ligand binding to the receptor to a level approaching its energy in solution, or (c) increasing ligand lipophilicity (i.e. the hydrophobic effect), in particular at positions that result in the expulsion of high energy, labile water molecules from the active site.

QM calculations can be employed to analyse certain aspects of this process. For example, QM model calculations could be used to assess how ideal an X-ray binding mode is through the use of a model consisting of the ligand and the key polar residues it interacts with in the active site. Optimization of the QM model in the gasphase will lead to either a large change in the ligand binding conformation and an increase in interaction with the active site model, or no significant change at all. Both results provide us with useful information about the binding mode. Should gasphase optimization lead to a dramatic change in the ligand binding mode we might conclude that steric aspects associated with the protein environment prevent the ligand from forming optimal interactions with the key active site residues. Alternatively, where optimization leads to a negligible change in ligand binding, we might conclude that the interactions present are essentially optimal. This information is important since it allows us to direct synthetic efforts more precisely on finding activity gains by (a) maximizing the key non-bonded interactions in the active site, or alternatively (b) focusing elsewhere in the molecule by searching for additional, more distant polar interactions, or hydrophobic pockets to support targeted lipophilicity increases.

Prediction of the *relative* strength of interactions between simple models of a ligand and its corresponding receptor can be useful to help rationalize structure activity relationships (SAR). This relatively simple approximation has led to the development of in-silico tools using small probe molecules to estimate H-bond strength which can prove very useful in lead optimization efforts [30–35]. However, simplifications involved in these models might mask SAR in cases where the generic probes are not fully representative of the receptor, or the effect of the local active site conformation is important. Kinase ATP site inhibitors are a case in point since up to 3 H-bonds can be formed between the ligand and the important hinge region. Thus an assessment of H-bond strength would ideally be assessed using all the residues involved in the binding process, using the precise hinge conformation from the protein itself. Such target specific models could then be used to assess the strength of interaction resulting from different substituents on the core template or even modifications to the core template.

We now report the results obtained on the target specific QM model calculations of 8 kinase protein–ligand complexes. Complexes were obtained by searching the PDB for kinases containing small molecule inhibitors <350 da, with reasonable resolutions (~ 2.0 Å), where the interactions with the receptor were mediated only through H-bonding with the hinge itself (Table 1). The PDB structures have been reported as part of fragment based screening efforts and are particularly applicable to QM analysis due to their relatively small size. In addition, given their activity is pri-

marily governed by the H-bond interactions with the hinge region, these relatively small QM calculations should be particularly representative. We assess this by comparing results from the QM and QM/MM protein based models. For these structures we also look at the atomic solution derived from the electron density in greater detail, putting particular emphasis on the accuracy on the ligand placement with respect to the ligand electron density as well as conformational aspects of the ligand itself.

Analysis of the average structural parameters for the 8 ligands listed in Table 2c shows that the N-based H-bond interactions are comparable to those from the high resolution PDB set, suggesting they are quite representative set overall. However in the analysis of the often neglected CH based interactions formed with the outer and inner hinge acceptors, we can see the mean distances are ~ 0.4 Å longer than the high resolution PDB set suggesting either (a) the interactions are much more repulsive for these fragment-like

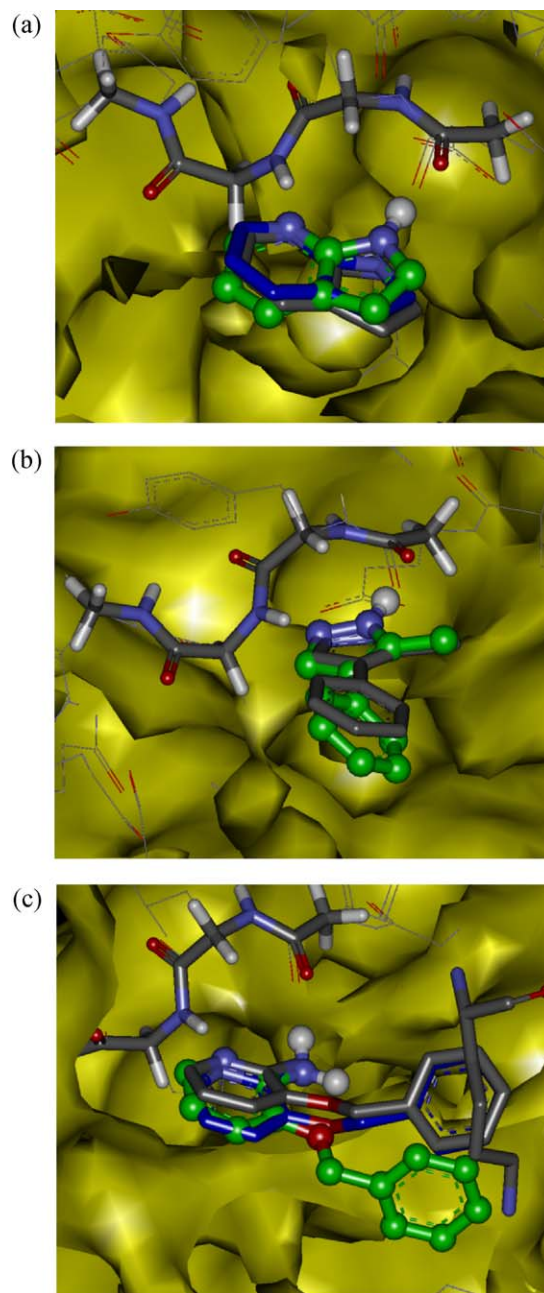


Fig. 3. Superimposition of the QM (green) and QM/MM (blue) and X-ray (grey) models for (a) 2UVX, (b) 2UW3 and (c) 1W7H.

Table 3

H-bond characteristics of 8 kinase-inhibitor complexes derived from X-ray, MP2/6-31+G** QM and B3LYP/6-31G**//UFF QM/MM models. Parameters reported are the same as those in Table 2. Distances are given in Å and angles in degrees and RMSDs in Å. QM/MM results are taken from Refs. [26,27].

ID	Model	RMSD	Outer H-bond X	C=O...H-X			Central H-bond NH...N			Inner H-bond X	C=O...X			C-C-X
				C=O...H-X	C=O...X	C-C-X	NH...N	N...N	C-N-N		C=O...H-X	C=O...X	C-C-X	
1W7H	XRAY	–	C	2.59	3.31	87.4	2.00	3.00	127.1	N	2.00	2.80	133.6	
	MP2/6-31+G**	1.6		2.42	3.37	78.3	1.86	2.88	124.6		1.96	2.97	127.9	
	B3LYP/6-31G**//UFF ^a	0.2		2.69	3.56	84.6	2.04	3.05	126.2		1.81	2.81	131.7	
2BHE	XRAY	0.0	N	2.10	3.06	106.4	1.45	2.50	110.6	N	1.95	2.98	141.1	
	MP2/6-31+G**	0.2		2.23	3.06	110.5	1.81	2.83	113.3		1.84	2.86	136.3	
2UVX	XRAY	–	C	2.57	3.48	80.8	1.92	2.98	122.6	N	1.88	2.88	129.0	
	MP2/6-31+G**	0.5		2.39	3.45	73.9	1.87	2.88	127.4		1.84	2.85	133.3	
	B3LYP/6-31G**//UFF ^a	0.2		2.53	3.50	80.1	1.97	1.99	122.7		1.80	2.79	127.9	
2UW3	XRAY	–	C	3.57	4.10	77.8	1.86	2.91	117.7	N	1.81	2.82	143.9	
	MP2/6-31+G**	0.7		3.77	4.55	66.2	1.87	2.89	120.3		1.82	2.84	139.5	
	B3LYP/6-31G**//UFF ^b	0.4		3.30	3.97	76.9	1.97	2.97	122.3		2.01	2.98	145.8	
2VTA	XRAY	–	N	1.89	2.92	143.9	1.96	3.02	113.5	C	2.96	3.49	153.1	
	MP2/6-31+G**	0.3		1.83	2.85	143.1	1.95	2.96	123.2		2.52	3.32	152.9	
3DND	XRAY	–	C	3.94	4.29	92.6	2.66	3.70	127.5	N	2.09	2.93	136.2	
	MP2/6-31+G**	2.5		2.80	3.28	84.3	1.88	2.91	123.8		1.94	2.95	131.0	
2C5O	XRAY	–	N	1.93	2.59	102.6	1.84	2.88	118.2	C	1.96	2.96	140.4	
	MP2/6-31+G**	0.9		1.95	2.97	97.4	1.91	2.93	117.5		2.42	3.17	147.4	
	B3LYP/6-31G**//UFF ^b	1.1		1.76	2.70	124.4	1.88	2.90	124.4		2.20	3.25	152.4	
1PXJ	XRAY	–	C	2.52	3.25	92.2	2.23	3.30	113.7	N	1.85	2.86	130.9	
	MP2/6-31+G**	0.2		2.33	3.26	85.8	1.93	2.96	118.9		1.97	2.96	137.2	

^a Obtained using the native protein conformation.

^b Obtained using a non-native protein conformation.

inhibitors or (b) the structures themselves are sub-optimal due to their only moderate resolutions.

Looking in more detail at the individual complexes we can see that the interaction characteristics displayed by 1W7H, 2UVX, 2UW3 and 2VTA lie within the expected values obtained from an analysis of the high resolution PDB structures. However, for the other four complexes, 2BHE, 3DND 1PXJ and 2C5O, analysis of their intermolecular interactions reveals uncharacteristically long or short interactions made with the hinge which in some case appear not to be supported by the electron density. We discuss these issues further based on a consideration of the original electron density as well as computationally quite rigorous MP2/6-31+G** QM calculations (Tables 2c and 3).

3.3. Validity of the QM and QM/MM Models

To assess the utility of the QM based models to probe the binding interactions we first focus on those QM models where the structures were in good agreement with the original X-ray coordinates, as well as the interaction characteristics from high resolution kinase structures from the PDB, and CSD experimental H-bonds characteristics. This will allow us to more objectively assess the performance of the QM based models. Additionally, for PDBs 2UVX and 1W7H we have previously reported QM/MM results for the native protein conformation [27] which can be compared to the QM results. In the case of 2UW3 and 2C5O we have also reported the QM/MM results obtained in a non-native protein conformation, where the binding mode was determined by aligning the protein C_α atoms [26] with subsequent optimization. The former QM/MM models provide a useful assessment of how subtle changes in the protein conformation (obtained for a different inhibitor) affect the inhibitor binding mode.

In Fig. 3, the QM, QM/MM and X-ray conformations of 2UVX, 2UW3 and 1W7H have been superimposed using their common QM atoms. The relatively small, inflexible molecules found in 2UVX and 2UW3 are both well modeled using both the QM and QM/MM approach as can be seen from a qualitative assessment of Fig. 3. Moreover, from Table 3, we can see that the QM models display RMSDs to the original X-ray of 0.5 Å and 0.7 Å, respectively, which are marginally larger than those obtained with the QM/MM models (0.2 Å and 0.4 Å, respectively). To put

these values into perspective, it is expected that the intrinsic error in the coordinates of a structure at 2 Å resolution is ~0.2. The relatively large RMSD of 0.7 Å observed for 2UW3 is a result of the subtle movement of its phenyl substituent rather than that of the pyrazole which interacts with the hinge (Fig. 3b). In fact, the interactions between the ligand and hinge are in very good agreement with the QM model results for both PDBs (Fig. 4).

In the case of 1W7H, a molecule with 2 rotatable bonds, the gas-phase representation is not accurate since it lacks the important steric effects of an active site lysine side-chain (Fig. 3c). Nonetheless, these results are still interesting from an SBD perspective since they suggest that this ligand is bound in a relatively high energy conformation and when given sufficient freedom to relax (in the gas-phase), the benzyl substituent conformation significantly alters, improving the interaction with the hinge in the process (Fig. 4). This suggests optimization of the benzyl portion of this chemotype may lead to greater dividends, potentially reducing unfavorable steric interactions in the active site, while simultaneously allowing a more optimal interaction of its pyrazole portion with the hinge. In the case of 2VTA, no QM/MM results were previously reported so we cannot make the comparison, however the predicted H-bond distances are in good agreement with the X-ray structure and the RMSD is very low at 0.3 Å. The only sizeable difference occurs for the C=O...HC interactions distance which is 2.96 Å in the crystal structure but 2.52 Å in the QM model. For comparison, the mean value from the high resolution PDB set is 2.46 Å (Table 2).

Analysis of the theoretical results for 2BHE, 3DND, 1PXJ and 2C5O shows that they are not in as good agreement compared to the other four complexes studied here. Nevertheless, the QM results do reproduce the same overall binding mode, it is only the strength of the interactions between the ligand and the protein that differ. Analysis of the intermolecular interactions found in the X-ray structures reveals significant deviations from the expected parameters derived from the PDB dataset analysis and ISOSTAR. Furthermore, the atomic solutions to the electron density are in some cases not necessarily consistent with the reported density as can be seen in Fig. 5. In each case, 1 or more interaction between the ligand and the hinge shows uncharacteristically short, or long interactions (Table 3), which are highlighted. The average resolution of the four structures is 2.1 Å with none greater than 2.3 Å.

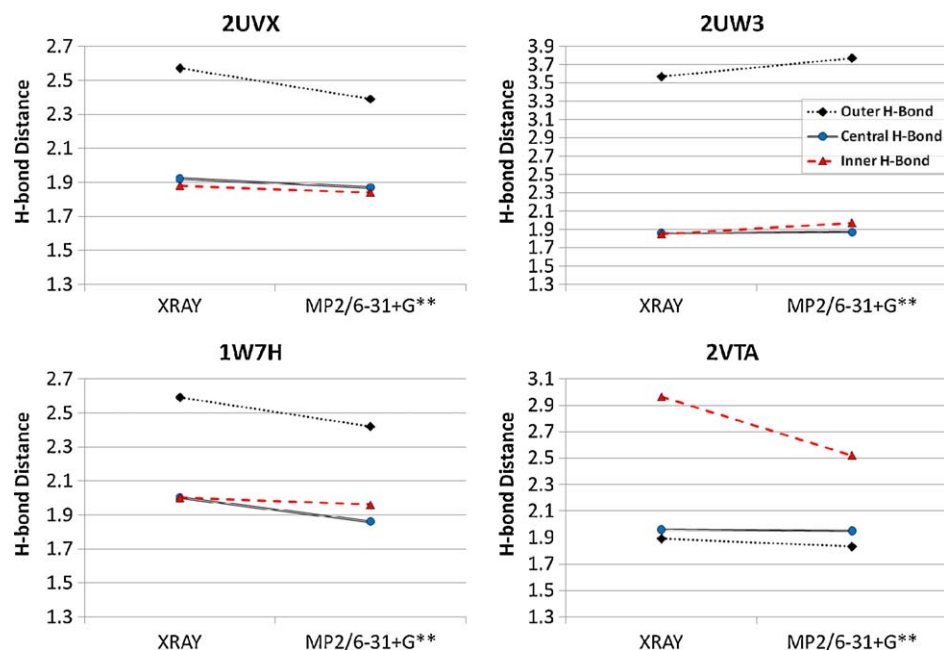


Fig. 4. A comparison of the X-ray and QM model H-bond distances for cases where good agreement is observed. Distances >2.1 Å correspond to CH mediated interactions. Circles with solid lines denote the central H-bond, triangles with dashed line denote the inner H-bond and diamonds with a dotted line denote the outer H-bond.

2BHE: The 2BHE structure is the best resolved here (1.9 Å) yet displays the clearest flaw of all PDB structures under consideration. The outer and inner C=O...HN interaction distances lie within those expected from the 57 high resolution kinase structures how-

ever the central N...N mediated H-bond is a major outlier. The N-N distance is found to be 2.5 Å, or just 1.5 Å after hydrogen atoms are added according to the AMBER forcefield. This is nearly 3 standard deviations (SD) shorter than the comparable values from the high

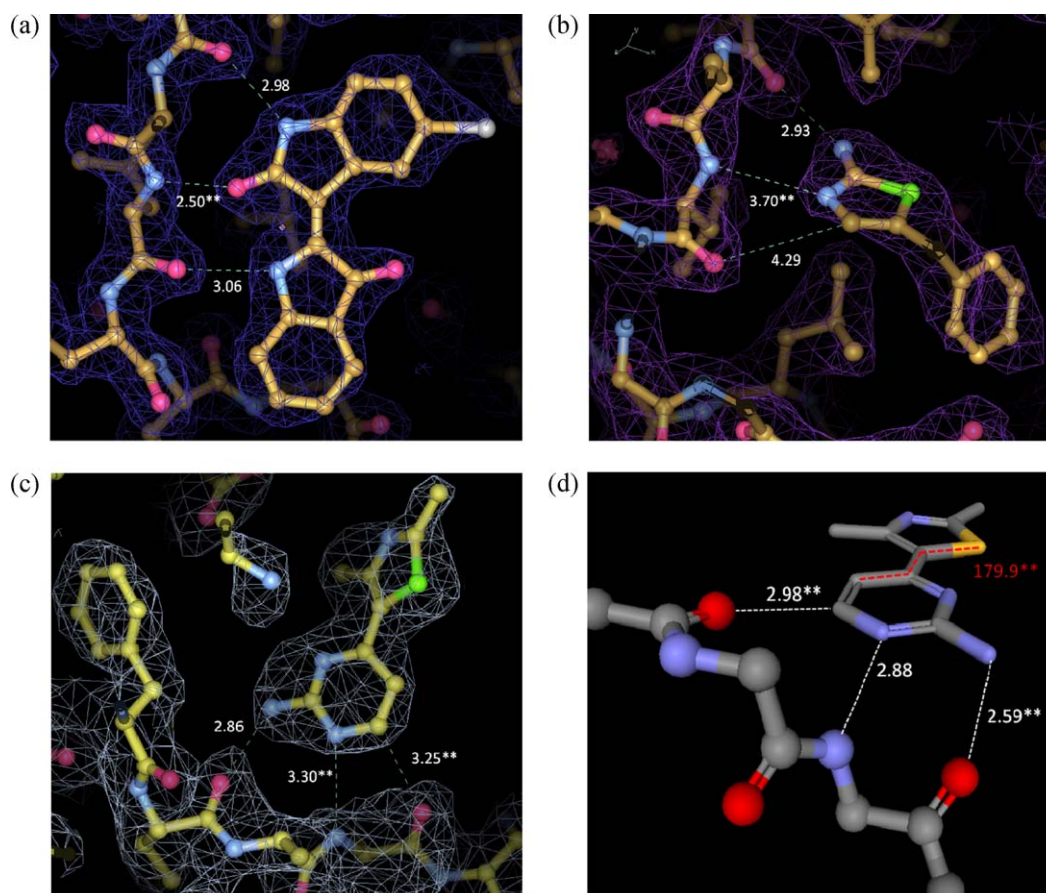


Fig. 5. Experimental atomic coordinates for 4 of the PDBs under consideration here. Clockwise from top left are 2BHE, 3DND, 1PXJ and 2C5O. Uncharacteristically short or long H-bond distances and an unnatural dihedral angle are denoted by **. Density contoured using Coot [52] at $\sim 1.5\sigma$ level. No density has been reported for 2C5O.

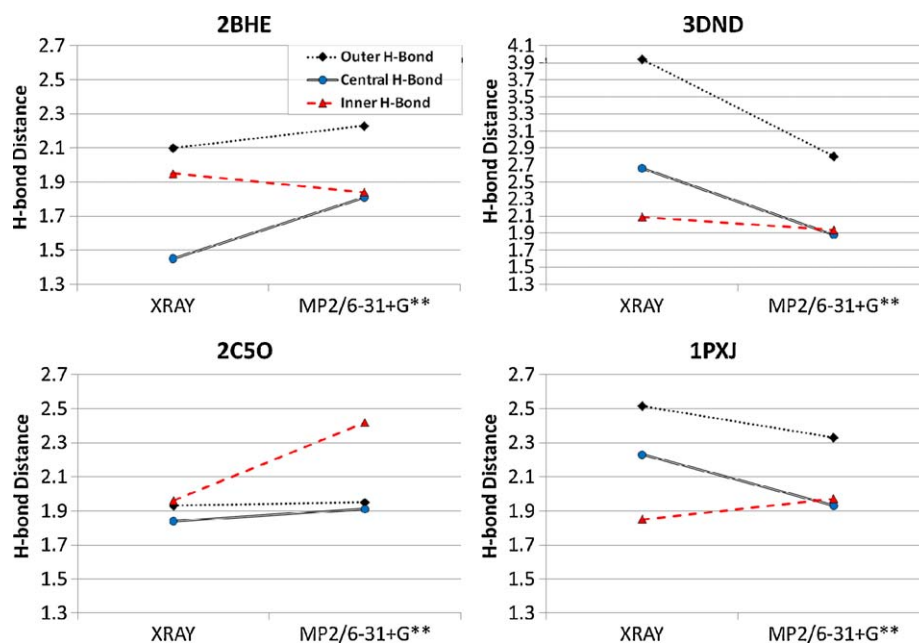


Fig. 6. A comparison of the X-ray and QM model H-bond distances for cases where poor agreement is observed Clockwise from top left are 2BHE, 3DND, 1PXJ and 2C50. The outer H-bond for 1PXJ and 3DND and the inner H-bond distance for 2C50 correspond to CH mediated interactions. Distances are given in Å. Legend the same as that in Fig. 4.

resolution kinase PDB set (3.1 Å) suggesting it is an error arising from the protein refinement step. Analysis of the electron density in Fig. 5a certainly does not rule this out. The density around the Br atom towards the rear of the pocket is relatively poor compared to other coordinates, and it may be that attempts to better incorporate this less tightly, more flexible (relatively), part of the molecule into the observed density led to the unrealistically short H-bond. Note, while this short distance does not affect the interpretation from a visual, qualitative perspective, it will have profound implications for theoretically derived energetics if computed on this geometry.

The MP2/6-31+G** optimized model displays a central H-bond distance within expected parameters, while the H-bonds made to the outer and inner hinge are comparable to those observed in the original structure. Moreover, the RMSD associated with the QM optimized ligand is just 0.2 Å which suggests that the initial fitting of the ligand to the electron density may have introduced this subtle error.

3DND: The 2.3 Å 2BHE structure is the least resolved of all structures studied here, albeit only marginally so. Thus the difference in resolution is unlikely to explain the unnaturally long H-bond distance made by the inhibitor to the central hinge donor. The N–N distance in the X-ray structure is 3.7 Å which is almost 0.6 Å longer than the mean value displayed by the high resolution kinase PDB structures that all make an equivalent interaction. Given this is the key interaction made by almost all type-1 inhibitors it is surprising to find this distance over 3 SDs away from the mean. Analysis of the electron density in Fig. 5b suggests the ligand position is sub-optimal as it is clear the density surrounding the ligand towards the solvent exposed surface is less well defined. It is also clear that density surrounding the nitrogen donor of the inhibitor clearly supports a much shorter N–N distance. This result highlights the problems of relying on empirical solutions to what is often poor quality ligand density. In such cases ligands can be forced into unrealistic conformation that fit the density better (as in this case) but which are clearly chemically rather suspect.

The 3DND structure is relatively difficult to simulate using QM methods as the presence of a number of amino acid sidechains and water molecules towards the mouth of the pocket help to constrain the conformation of the benzyl ring. In the gasphase calculations

this constraint is lost leading to the rotation and translation of the molecule such that it can form an additional interaction between the outer hinge donor and its 2-position phenyl hydrogen atom (2.3 Å). While this deviation is clearly an artifact of the simplistic QM model, the results are interesting in that they highlight that the interaction between the hinge and the ligand in the protein are sub-optimal (or wrong) given that the H-bond distance decreases by 0.8 Å on optimization. While this structure is incongruent with the electron density, it should also be noted that so too is the original ligand solution reported in the PDB (Fig. 5). In a case such as this a QM/MM model would be more appropriate to assess the validity of the experimental coordinates as the effects of the amino acid sidechains and water molecules towards the mouth of the pocket could be directly taken into account. Nonetheless, the QM calculation still provides useful information from a SBD perspective as it highlights that the interaction of the inhibitor is sub-optimal with the hinge.

1PXJ: The 2.3 Å 1PXJ structure differs from 2C50 in that the CDK2 protein is in an active conformation, leading to a new binding mode to the hinge. The N–N distance associated with the central H-bond is 3.3 Å, corresponding to a relatively weak H-bond of 2.2 Å length on addition of hydrogen atoms. The N–N distance is close to 2 standard deviations from the average value of 2.9 Å derived from 57 high resolution kinase structures suggesting the structure may also have been placed in a somewhat suboptimal position. Furthermore, the outer hinge C=O...HC seems somewhat short, displaying a comparable heavy atom distance to that of the central hinge NH...N based interaction (3.3 Å). Analysis of the electron density is equivocal as the interaction towards the outer hinge is less well defined (Fig. 5c). This suggests this part of the molecule is more flexible meaning the atomic solution of the ligand derived from the electron density will be less reliable.

The relatively long central H-bond distance (2.23 Å) in the 1PXJ PDB structure decreases to 1.93 Å in the QM model, closer to the experimental average expected from the high resolution kinase structures (1.97 Å). The C=O...HC distance also decreases by 0.2 Å in the QM model which appears relatively short at 2.33 Å. The overall RMSD is just 0.2 Å which is within experimental error for a PDB of this overall resolution. The results from the QM analysis, database

searches of comparable high resolution kinase complexes, and a re-investigation of the original electron density suggests the relatively long interaction distances may be an artifact of the protein refinement process.

2C5O: No electron density has been reported for 2.1 Å 2C5O structure so a structural evaluation can only be performed on the atomic solution of the density only (Fig. 5c). Analysis of these coordinates reveals apparent flaws both in terms of the interaction between ligand and hinge and the ligand conformation. Firstly, it is expected that the bi-cyclic ligand molecule will not lie in the same plane to minimise unfavourable VDW repulsion (as is observed in 1PXJ). However, in this case the two rings lie in the same horizontal plane suggesting the initial ligand parameters in the refinement process were suboptimal (~5 kcal/mol energy cost at B3LYP/6-31G* level). This error will have a knock on effect on other aspects of the ligand placement and may explain the non-ideal O–C intermolecular distance observed at the inner hinge (2.98 Å). The resulting H-bond distance is 1.96 Å, almost 0.6 Å shorter than that expected from the 57 high resolution kinase structures (i.e. 2SDs). Furthermore, the outer hinge O–N distance is comparatively short (2.6 Å), 0.3 Å shorter than the ideal distance for a typical kinase C=O...HN bond.

2C5O has been explored by us from both a QM and QM/MM perspective, however both models show rather large ligand RMSDs (0.9 vs 1.1 Å, respectively). It is apparent from the X-ray coordinates that the ligand positioning is rather poor given that the addition of hydrogen atoms to the structure leads to a very short inner C=O...HC interaction distance of 2.0 Å. Furthermore, the two aromatic rings of the ligand lie in the same plane which is in contrast to the same molecule in the active CDK2 structure of 1PXJ. In this case it is difficult to comment on the reliability of the QM or QM/MM approach due to the relatively poor quality of the experimental data. Nonetheless, it is interesting to note that the QM and QM/MM derived distances are more akin to those expected from an analysis of high resolution kinase PDBs (Table 2 and Fig. 6).

In summary, we find that for 4 out of the 8 inhibitors studied here, the ligand binding conformation in the gasphase QM model is essentially identical to that observed in the experimental X-ray crystal structure. In these cases it appears that the binding mode is optimal for the given polar interaction. This is perhaps unsurprising given that the source of the structures has been fragment based screening efforts which focus on ligand efficient hits [53,54]. In the remaining 4 cases the inhibitor binding conformation differs significantly from the X-ray structure to form increased interaction with the hinge region. This suggests steric factors associated with the protein environment force these 4 inhibitors into a less optimal binding conformation compared to gasphase/solution.

4. Conclusions

In this investigation, we have assessed the utility of QM and QM/MM calculations to probe the nature of the ligand interactions found in kinase protein–ligand. QM methods are more accurate than MM based alternatives due to the fact they consider the underlying physics of a given molecular system more precisely [15,30–35]. However, they are rarely used for this purpose due to their perceived high computational demand and the relatively few cases studies reported in the literature. This type of information could certainly prove useful in the optimization of ligand affinity, and efficiency [53,54].

Information from model QM calculations could be used to specifically target molecular changes for a particular ligand to either improve the interactions made with the key polar moieties in an active site, or should they be shown to be essentially optimal, focus on alterations elsewhere to probe for additional polar

interactions, or hydrophobic pockets to allow a targeted increase in lipophilicity. In addition, QM models could also be used to assess the core templates associated with large kinase inhibitors. Calculations could be employed to study the interaction of the N-phenylquinazolin-4-amine core of the EGFR inhibitor Tykerb with the 3 hinge residues. Furthermore, alterations to the structure could be assessed in-silico to see whether simple structural changes could be made to either improve the interaction with the hinge directly, or alter the angle associated with the anilino substituent to improve its trajectory into the relatively tight opening of the backpocket.

Finally, X-ray crystal structures play an absolutely critical role in SBD and the impact of computational chemistry methods would be severely restricted without the information they provide. While this work is not representative due to the focus on small numbers of structures from a single target class, from a more general consideration of the literature [21,20,24–27,51] it does appear that the atomic coordinates derived from X-ray analyses, while capable of demonstrating the overall ligand binding modes, should not necessarily be taken as a “perfect” reference given the intrinsic errors that arise from protein mobility and choices made during structural refinement [21]. This is particularly important for the validation of ligand docking methods since the X-ray coordinates are often used in quantitative way to assess the performance of software [51]. Comparison of the observed interactions to results from comparable interactions from the CSD or RCSB can also prove useful, as can results from QM and QM/MM optimizations to determine the relative strength of interactions found in a particular protein–ligand complex.

Acknowledgments

Special thanks are due to Ben Tehan for his help during the study. This work was supported by the Thailand Research Fund grants RTA5080005 (S.H.) and RMU5180032 (D.G.). Support for research facilities at KU provided by the National Center of Excellence in Petroleum, Petrochemical Technology and Advanced Materials is gratefully acknowledged. Finally, M.P.G. would like to thank the Faculty of Science, Kasetsart University (K.U.) for support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.09.012.

References

- [1] M. Von itzstein, W.Y. Wu, G.B. Kok, M.S. Pegg, J.C. Dyason, B. Jin, T.V. Phan, M.L. Smythe, H.F. White, S.W. Oliver, P.M. Colman, J.N. Varghese, D.M. Ryan, J.M. Woods, R.C. Bethell, V.J. Hotham, J.M. Cameron, C.R. Penn, Rational design of potent sialidase-based inhibitors of influenza-virus replication, *Nature* 363 (1993) 418–423.
- [2] R.M. Angell, T.D. Angella, P. Bamborough, M.J. Bamford, C. Chung, S.G. Cockerill, S.S. Flack, K.L. Jones, D.I. Laine, T. Longstaff, S. Ludbrook, R. Pearson, K.J. Smith, P.A. Smee, D.O. Somers, A.L. Walker, Biphenyl amide p38 kinase inhibitors 4: DFG-in and DFG-out binding modes, *Bioorg. Med. Chem. Lett.* 18 (2008) 4428–4432.
- [3] C. Bissantz, B. Kuhn, M. Stahl, A medicinal chemist's guide to molecular interactions, *J. Med. Chem.* 53 (2010) 5061–5084.
- [4] K.A. Brameld, B. Kuhn, D.C. Reuter, M. Stahl, Small molecule conformational preferences derived from crystal structure data. A medicinal chemistry focused analysis, *J. Chem. Inf. Model.* 48 (2008) 1–24.
- [5] B. Grimme, Do special noncovalent p–p stacking interactions really exist? *Angew. Chem. Int. Ed.* 47 (2008) 3430–3434.
- [6] Y. Zhao, H.T. Ng, E. Hanson, Benchmark data for noncovalent interactions in HCOOH...benzene complexes and their use for validation of density functionals, *J. Chem. Theory Comput.* 5 (2009) 2726–2733.
- [7] L. Estevez, N. Otero, R.A. Mosquera, Computational study on the stacking interaction in catechol complexes, *J. Phys. Chem. A* 113 (2009) 11051–11058.
- [8] R.S. Paton, J.M. Goodman, Hydrogen bonding and π -stacking: how reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions, *J. Chem. Inf. Model.* 49 (2009) 944–955.

- [9] P. Zhou, J. Zou, F. Tian, Z. Shang, Fluorine bonding – how does it work in protein–ligand interactions? *J. Chem. Inf. Model.* 49 (2009) 2344–2355.
- [10] Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang, W. Zhu, Halogen bonding – a novel interaction for rational drug design? *J. Med. Chem.* 52 (2009) 2854–2862.
- [11] Y.N. Imai, Y. Inoue, Y. Yamamoto, Propensities of polar and aromatic amino acids in noncanonical interactions: nonbonded contacts analysis of protein–ligand complexes in crystal structures, *J. Med. Chem.* 50 (2007) 1189–1196.
- [12] M.-H. Hao, O. Haq, I. Muegge, Torsion angle preferences of small-molecule ligands bound to proteins, *J. Chem. Inf. Model.* 47 (2007) 2242–2252.
- [13] G.M. Spitzer, B. Wellenzohn, P. Markt, J. Kirchmair, T. Langer, K.R. Liedl, Hydrogen-bonding patterns of minor groove–binder–DNA complexes reveal criteria for discovery of new scaffolds, *J. Chem. Inf. Model.* 49 (2009) 1063–1069.
- [14] R. Wu, T.B. McMahon, Investigation of cation– π interactions in biological systems, *J. Am. Chem. Soc.* 130 (2008) 12554–12555.
- [15] G.J. Chang, J.Z.H. Zhang, NMR scalar coupling constant reveals that intraprotein hydrogen bonds are dynamically stabilized by electronic polarization, *J. Phys. Chem. B* 113 (2009) 13898–13900.
- [16] M. Nocker, S. Handschuh, C. Tautermann, K.R. Liedl, Theoretical prediction of hydrogen bond strength for use in molecular modeling, *J. Chem. Inf. Model.* 49 (2009) 2067–2076.
- [17] J.J.-L. Liao, Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors, *J. Med. Chem.* 50 (2007) 1–16.
- [18] R. Vargas, J. Garza, D.A. Dixon, B.P. Hay, How strong is the $\text{C}\alpha\text{--H}\cdots\text{O}=\text{C}$ hydrogen bond? *J. Am. Chem. Soc.* 122 (2000) 4750–4755.
- [19] A.C. Pierce, K.L. Sandretto, G.W. Bemis, Kinase inhibitors and the case for $\text{CH}_3\cdots\text{O}$ hydrogen bonds in protein–ligand binding, *Proteins* 49 (2002) 567–576.
- [20] G.J. Kleywegt, K. Henrick, E.J. Dodson, D.M.F. van Aalten, Pound-wise but penny-foolish: how well do micromolecules fare in macromolecular refinement? *Structure* 11 (2003) 1051–1059.
- [21] A.M. Davis, S.A. St-Gallay, G.J. Kleywegt, Limitations and lessons in the use of X-ray structural information in drug design, *Drug Discovery Today* 13 (2008) 831–841.
- [22] S. Wlodek, A.G. Skillman, A. Nicholls, Automated ligand placement and refinement with a combined force field and shape potential, *Acta Cryst. D* 62 (2006) 741–749.
- [23] N.W. Moriarty, R.W. Grosse-Kunstleve, P.D. Adams, Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Cryst. D* 65 (2009) 1074–1080.
- [24] N. Yu, H.P. Yennawar, K.M. Merz Jr., Refinement of protein crystal structures using energy restraints derived from linear-scaling quantum mechanics, *Acta Crystallogr. D Biol. Crystallogr.* (61) (2005) 322–332.
- [25] X. Li, S.A. Hayik, K.M. Merz Jr., QM/MM X-ray refinement of zinc metalloenzymes, *J. Inorg. Biochem.* (104) (2010) 512–522.
- [26] M.P. Gleeson, D. Gleeson, QM/MM as a tool in fragment based drug discovery, *J. Chem. Inf. Model.* 49 (2009) 1437–1448.
- [27] M.P. Gleeson, D. Gleeson, Assessing the utility of hybrid quantum mechanical/molecular mechanical methods in structure-based drug design, *J. Chem. Inf. Model.* 49 (2009) 670–677.
- [28] G.L. Warren, C.A. Andrews, A. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, A critical assessment of docking programs and scoring functions, *J. Med. Chem.* 49 (2006) 5912–5931.
- [29] I.J. Enyedy, W.J. Egan, Can we use docking and scoring for hit-to-lead optimization? *J. Comput. -Aided Mol. Des.* 22 (2008) 161–168.
- [30] J.A. Platts, Theoretical prediction of hydrogen bond basicity, *Phys. Chem. Chem. Phys.* 2 (2000) 3115–3120.
- [31] J. Reynisson, E. McDonald, Tuning of hydrogen bond strength using substituents on phenol and aniline: a possible ligand design strategy, *J. Comput. -Aided Mol. Des.* 18 (2004) 421–431.
- [32] M.-H. Hao, Theoretical calculation of hydrogen-bonding strength for drug molecules, *J. Chem. Theory Comput.* 2 (2006) 863–872.
- [33] S. Raub, C.M. Marian, Quantum chemical investigation of hydrogen-bond strengths and partition into donor and acceptor contributions, *J. Comput. Chem.* (28) (2007) 1503–1515.
- [34] J. Schwobel, R.-U. Ebert, R. Kuhne, G. Schuurmann, Prediction of the intrinsic hydrogen bond acceptor strength of organic compounds by local molecular parameters, *J. Chem. Inf. Model.* 49 (2009) 956–962.
- [35] P.W. Kenny, Hydrogen bonding, electrostatic potential, and molecular design, *J. Chem. Inf. Model.* 49 (2009) 1234–1244.
- [36] P.E.M. Siegbahn, F. Himo, Recent developments of the quantum chemical cluster approach for modeling enzyme reactions, *J. Biol. Inorg. Chem.* 14 (14) (2009) 643–651.
- [37] A. Warshel, M. Levitt, Theoretical studies of enzymatic reactions: dielectric electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *J. Mol. Biol.* 103 (1976) 227–249.
- [38] U.C. Singh, P.A. Kollman, A combined ab initio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular-systems – applications to the $\text{CH}_3\text{Cl}+\text{Cl}^-$ exchange-reaction and gas-phase protonation of polyethers, *J. Comput. Chem.* 7 (1986) 718–730.
- [39] K. Raha, K.M. Merz Jr., Large scale validation of a quantum mechanics scoring function: predicting the binding affinity and the binding mode of a diverse set of protein–ligand complexes, *J. Med. Chem.* 48 (2005), 4558–2475.
- [40] F.H. Allen, The Cambridge structural database: a quarter of a million crystal structures and rising, *Acta Cryst. B* 58 (2002) 380–388.
- [41] RCSB Protein databank. <http://www.rcsb.org/>.
- [42] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
- [43] S. Dapprich, I. Komaromi, K.S. Byun, K. Morokuma, M.J. Frisch, A new ONIOM implementation in Gaussian 98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives, *J. Mol. Struct.* 1 (1999) 461–462.
- [44] T. Vreven, K.S. Byun, I. Komaromi, S. Dapprich, J.A. Montgomery, K. Morokuma, M.J. Frisch, Combining quantum mechanics methods with molecular mechanics methods in ONIOM, *J. Chem. Theory Comput.* 2 (2006) 815–826.
- [45] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard III, W.M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *J. Am. Chem. Soc.* 114 (1992) 10024–10035.
- [46] K. Senthilkumar, J.I. Mujika, K.E. Ranaghan, F.R. Manby, A.J. Mulholland, J.N. Harvey, Analysis of polarization in QM/MM modelling of biologically relevant hydrogen, *J. R. Soc. Interface* 5 (2008) S207–S216.
- [47] I.J. Bruno, J.C. Cole, J.P.M. Lommerse, R.S. Rowland, R. Taylor, M.L. Verdonk, A library of information about non-bonded interactions, *J. Comput. -Aided Mol. Des.* 11 (1997) 525–537.
- [48] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.* 117 (1995) 5179–5197.
- [49] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comp. Chem.* 25 (2004) 1157–1174.
- [50] ROCS 2.3.1: OEChem version 1.5.1 OpenEye Scientific Software Inc., Santa Fe NM, USA. www.eyesopen.com.
- [51] D. Yusuf, A.M. Davis, G.J. Kleywegt, S. Schmitt, An alternative method for the evaluation of docking performance: RSR vs RMSD, *J. Chem. Inf. Model.* 48 (2008) 1411–1422.
- [52] Coot 0.6.1, <http://www.biop.ox.ac.uk/coot/>.
- [53] A.L. Hopkins, C.R. Groom, A. Alex, Ligand efficiency: a useful metric for lead selection, *Drug Discovery Today* 9 (2004) 430–431.
- [54] C. Abad-Zapatero, J.T. Metz, Ligand efficiency indices as guideposts for drug discovery, *Drug Discovery Today* 10 (2005) 464–469.



Effects of the CN and NH₂ substitutions on the geometrical and optical properties of model vinylfluorenes, based on DFT calculations

Wichanee Meeto^a, Songwut Suramitr^a, Vladimír Lukeš^b, Peter Wolschann^c, Supa Hannongbua^{a,*}

^a Department of Chemistry, Faculty of Science, and Center of Nanotechnology, Kasetsart University, Bangkok 10900, Thailand

^b Department of Chemical Physics, Slovak University of Technology, Radlinského 9, SK-81 237 Bratislava, Slovakia

^c Institute for Theoretical Chemistry, University of Vienna, Währingerstrasse 17, A-1090 Wien, Austria

ARTICLE INFO

Article history:

Received 16 July 2009

Received in revised form 21 September 2009

Accepted 21 September 2009

Available online 6 October 2009

Keywords:

Bifluorenevinylene

Density functional theory

Torsional potential

Vertical excitation energy

Conducting polymer

Substitutional effect

ABSTRACT

A systematic study on the structural and photo-physical properties of model bifluorenevinylene compounds based on the density functional theory (DFT) and its time-dependent (TD-DFT) version is presented. The main aim of this work is to investigate the influence of substitution on bifluorenevinylene using strong electron acceptor CN or electron donor NH₂ groups on: (a) the optimal geometry, (b) torsional potentials and (c) photo-physical properties. Our results indicate that the substitution on the vinylene bridge, leads to the twisting of molecular fragment on the side of added group and are in good overall agreement with experiment. In the case of the amino mono-substituted bifluorenevinylene, the amino group leads to non-planarity at the non-substituted portion of the molecule. The chemical modification also have a pronounced impact on the electronic properties. The shape of the potential energy curves evaluated for the lowest vertically excited states is heavily dependent on the molecular conformation. Finally, we discuss how the structural and electronic information presented here can be useful in designing of novel optical materials as well as understanding of excitation–relaxation phenomena which may occur in various time-dependent optical experiments.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Polyfluorenes, with their excellent photoluminescence quantum characteristics, excellent solubility in common organic solvents, can be used as blue-shift emitters [1,2]. Additionally, their thermal and chemical stability can be modified and improved by adding of different alkyl substituents at the 9-th position of the fluorene ring [3]. However, the polyfluorenes systems containing simple connected fluorene units tend to aggregate in the condensed phase and provide less effective performance. Fortunately, these problems can be limited with the modification of their molecular structure, such as the polyfluorene-2,7-vinylenes (PFVs) [4]. Moreover, the various electron-donating or electron-accepting groups can be added to the vinylene bridge located within the aromatic molecular chain which subtly alters the intermolecular steric interactions which can be used to effectively tune the optical spectra from blue to the infrared region without the necessity to perturb the planarity of aromatic units [5,6].

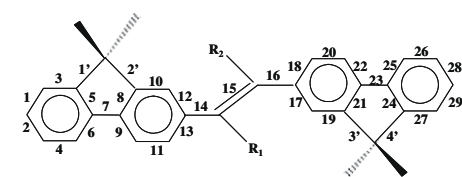
Theoretical quantum studies of π -conjugated oligomers or polymers can provide a fundamental understanding of the physical process occurring, and make a considerable contribution to the design of novel optoelectronic materials. These include theoretical

studies done using semiempirical, *ab initio* and density functional theory (DFT) methods, focusing on the electronic properties based on the ground-state equilibrium geometries [7–10]. Among them, the DFT approach and its time-dependent (TD) extension [11] for the excited states have been successfully demonstrated in investigations of optical and electronic properties of moderate to large organic conjugated oligomers [12]. These results are often used in cooperation with the experimental data to characterize the structural, electronic and optical properties of conjugated polymers and represent primary information to aid in our understanding of phenomena connected with conformation relaxation processes occurring during the electronic excitation and/or de-excitation [13]. Bifluorenevinylene derivatives are the focus of this research investigation, being shortest computationally efficient representation of oligo- and poly-vinylfluorenes which are of significant scientific interest.

The main aim of this work is the investigation of the direct influence of substitution using strong electron acceptor CN or electron donor NH₂ groups on the optimal geometry, torsional potentials and photo-physical properties. We do this by modifying the electronic structure of these model systems by the symmetric or asymmetric addition of the representative electron-accepting cyano and electron-donating amino groups on vinyl positions (see Fig. 1). We focus on the *all-trans* conformations and the electronic ground-state torsional potential curves in this investigation. The

* Corresponding author. Tel.: +66 2 5625555x2140; fax: +66 2 5793955.

E-mail address: fscisph@ku.ac.th (S. Hannongbua).



Molecule	R ₁	R ₂
F ₂ V	H	H
F ₂ V-(CN) ₂	CN	CN
F ₂ V-(NH ₂) ₂	NH ₂	NH ₂
F ₂ V-CN	CN	H
F ₂ V-NH ₂	NH ₂	H
F ₂ V-CN-NH ₂	CN	NH ₂

Fig. 1. Schematic structure, bond and dihedral angle numbering of studied systems in *all-trans* conformation.

corresponding vertical excitation characteristics of these molecules are calculated using (TD-)DFT method. Finally, the physical origin of the lowest electronic transitions will be explained using molecular orbital analysis.

2. Methodology

The geometries and torsion potential of studied dimers F₂V, F₂V-(CN)₂, F₂V-(NH₂)₂, F₂V-CN, F₂V-NH₂ and F₂V-CN-NH₂ (formula and abbreviations are given in Fig. 1) are optimized by the DFT method using the Becke three parameter hybrid (B3LYP) [14] functional in conjunction with the 6-31G(d) basis set [15]. The torsion potentials were calculated for the fixed angles from the

Table 1
The B3LYP/6-31G(d) optimal dihedral angles (in deg) and BLA parameters (in Å) for *all-trans* conformations.

Molecule	Θ ₁	Θ ₂	BLA
F ₂ V	0	0	0.224
F ₂ V-(CN) ₂	38	−38	0.218
F ₂ V-(NH ₂) ₂	45	−45	0.236
F ₂ V-CN	28	−7	0.212
F ₂ V-NH ₂	36	−31	0.227
F ₂ V-CN-NH ₂	52	−44	0.223

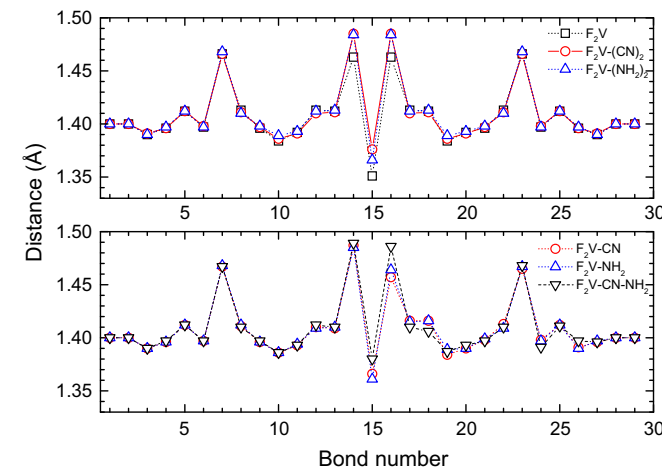


Fig. 2. Computed B3LYP/6-31G(d) bond lengths of the molecules under study in *all-trans* conformation. For notation see Fig. 1.

Table 2
The lowest excitation energies in eV and oscillator strengths (values in parentheses) for the optimal *all-trans* geometries. The values written in italics stand for the excitation contributions in percentage involved in each calculated transition (H denotes HOMO and L is LUMO).

Molecule	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇
F ₂ V	3.23 (1.870) 99%: H → L	3.98 (0.000) 60%: H-1 → L 38%: H → L+1	4.18 (0.000) 50%: H → L+3	4.19 (0.006) 68%: H → L+2	4.41 (0.00) 43%: H → L+2	4.44 (0.019) 33%: H-2 → L	4.44 (0.000) 60%: H-3 → L
F ₂ V-(CN) ₂	2.82 (0.907) 99%: H → L	3.22 (0.000) 97%: H-1 → L	3.64 (0.009) 70%: H-2 → L	3.67 (0.000) 75%: H-3 → L	3.67 (0.000) 68%: H-4 → L	3.88 (0.000) 73%: H-5 → L	4.29 (0.336) 72%: H-6 → L
F ₂ V-(NH ₂) ₂	2.90 (0.586) 99%: H → L	3.30 (0.000) 97%: H → L+1	3.53 (0.009) 94%: H → L+2	3.61 (0.000) 92%: H → L+3	4.21 (0.016) 91%: H → L+4	4.24 (0.000) 90%: H → L+5	4.41 (0.713) 93%: H-2 → L+1
F ₂ V-CN	3.08 (1.424) 99%: H → L	3.72 (0.096) 91%: H-1 → L	4.08 (0.002) 98%: H-2 → L	4.08 (0.016) 62%: H-3 → L	4.22 (0.007) 42%: H-5 → L	4.26 (0.006) 44%: H-4 → L	4.39 (0.063) 75%: H → L+1
F ₂ V-NH ₂	3.29 (1.099) 99%: H → L	3.86 (0.251) 87%: H → L+1	4.03 (0.013) 67%: H → L+2	4.08 (0.021) 69%: H → L+3	4.40 (0.079) 83%: H-1 → L	4.54 (0.023) 39%: H → L+6	4.58 (0.027) 41%: H → L+5
F ₂ V-CN-NH ₂	3.53 (0.824) 98%: H → L	4.01 (0.165) 73%: H → L+1	4.25 (0.006) 70%: H-1 → L	4.29 (0.085) 67%: H → L+1	4.38 (0.010) 61%: H → L+3	4.41 (0.131) 70%: H-2 → L	4.63 (0.015) 70%: H-4 → L

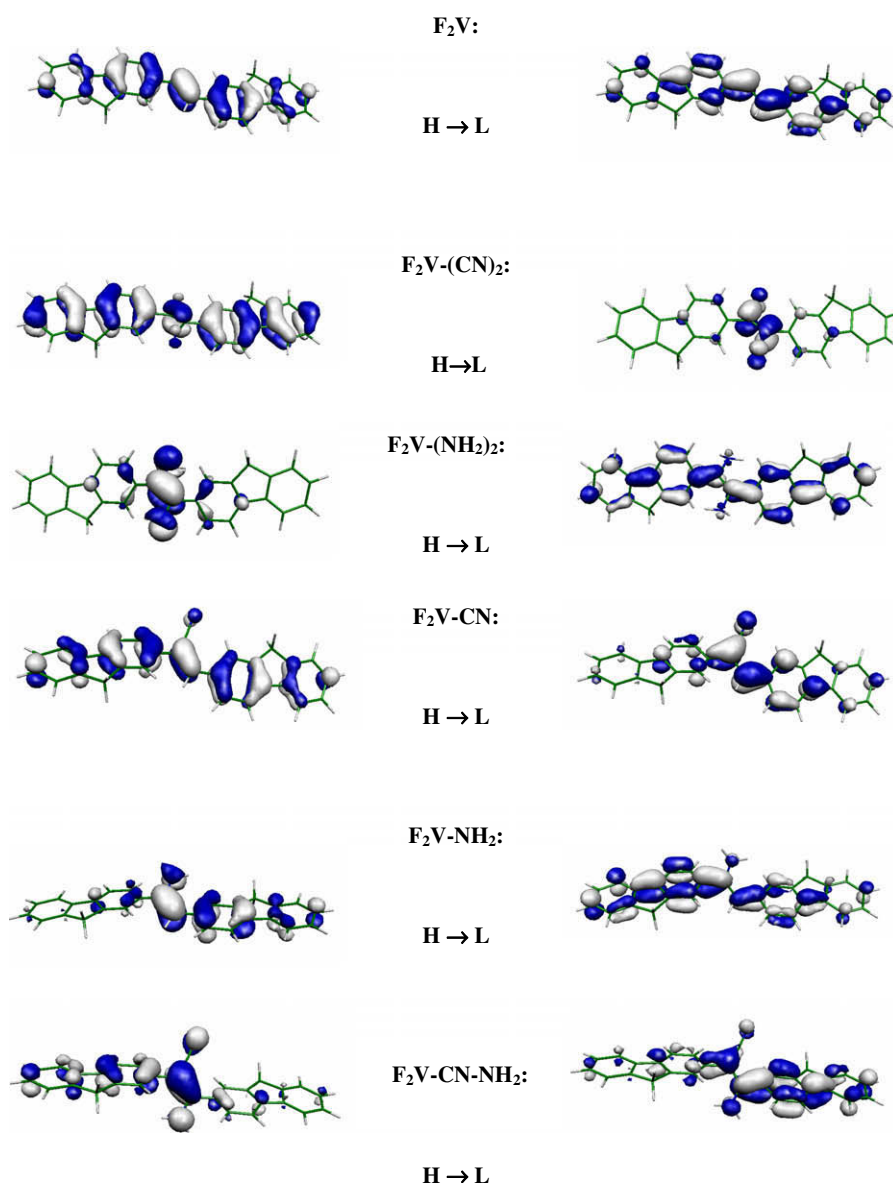


Fig. 3. Plots of the B3LYP/6-31G(d) molecular orbitals contributing significantly to the lowest energy transitions of studied molecules in *all-trans* conformation. H denotes HOMO and L is LUMO.

interval 0° to 180° using a 10° step size. In the case of asymmetric molecules, dihedral angles for both sides were investigated due to the different steric effects. Singlet vertical excitation energies are calculated from the optimized geometries using the TD-B3LYP method. All calculations are performed using the Gaussian 03 program package [16]. Due to the computational cost reduction, the alkyl groups at 9-th position on fluorene ring were replaced by hydrogen atoms. This is because reports suggest that substituents at the 9-th position play an important role in the thermal stability and solubility but do not affect the electronic structure and optical property of fluorene-based polymer [17,18]. All minima were confirmed as such through normal mode analysis, displaying no imaginary vibration frequencies.

3. Results and discussion

Comparison of the optimized ground-state 1¹A geometries in terms of bond lengths and torsional angles can help us to understand the structural and energetic differences observed between

the different systems. Schematic visualization of the molecular structure of the bifluorenevinylene skeleton studied here are given in Fig. 1. The data presented in Table 1 show the energy profile for the dihedral angles Θ_1 (between the bond nos. 12, 14 and 15) and Θ_2 (between the bond nos. 15, 16 and 17) between fluorene unit and the vinylene bridge on the substitution. The B3LYP/6-31G(d) calculation indicates the structure of F₂V molecule is completely planar. Substitution at the vinylene bridge leads to the perturbation of planarity due to the steric hindrance and the effect of electronegativity from nitrogen atoms in the vicinity of the newly added substituents. In the case of mono-substituted molecules, the presence of CN substituents increases the angle Θ_1 to the value of 30° while the torsion for NH₂ group is higher, 35°. Interestingly, the second angle Θ_2 is evidently non-planar only for molecule F₂V-NH₂. The symmetric bi-substitution is responsible for a torsion angle increase of ~10° with respect to the mono-substituted systems. The most distorted structure is obtained for the asymmetric F₂V-CN-NH₂ molecule, where the torsion on the side of cyano group results in a dihedral angle of 52°.

The computed bond lengths for the molecules studied here are presented in Fig. 2. In all cases the smallest C–C bond distances are located in the central part of the fluorene rings (see bond nos. 3, 10, 19 and 27) and double bond on vinylene bridge (see bond no. 15). The largest bond lengths are found for the inter-ring distances (bond nos. 7 and 23) which have a larger single bond character than the next bonds in the fluorene unit. The single bond nos. 14 and 16 on vinylene bridge exhibit also the largest distances. The substitution of vinylene bridge affects only the bonds in the vicinity of added group, especially the bonds of vinylene bridge. The double bond no. 15 as well as the neighbouring single bond nos. 14 and 16 are elongated while the bonds on rigid fluorene rings are very slightly shortened. Our calculations indicate that the differences in bond length changes for CN and NH₂ substitutions are significant only for bond no. 15. The CN substitution elongates it at 0.03 Å while the amino group changes it only 0.02 Å with respect to the F₂V molecule. Similar changes in bond lengths are observed for the mono-substituted NH₂ and CN compounds. However, in this case the bonds on the side of the substituted part are affected. The asymmetric substitution of F₂V–CN–NH₂ molecule leads to the formation of electronic system where NH₂ acts as electron donor to the aromatic system and the fluorene atoms as electron acceptor. In comparison with previously discussed molecules the bonds on the side of CN group and located on fluorene rings (bond nos. 17, 18, 24 and 26) are more influenced than in the F₂V–CN and F₂V–(CN)₂ molecules.

Bond length changes in aromatic systems can be efficiently described using the bond length alternation (BLA) definition [19]. The BLA values for a given molecular fragment is defined as the difference in length between single and double bonds between non-hydrogen atoms. The positive sign of BLA for example indicates that the molecular unit has an aromatic character. The BLA value for central vinylene bridge may be evaluated using equation 1,

$$\text{BLA} = (d_{14} + d_{16}) - 2d_{15}$$

where d symbols denote the bond lengths determined in Fig. 1.

Brédas et al. [20] studied the relationship between band gap and bond length alternation of conjugated polymers and their results indicate that in aromatic-based conjugated polymers, the energy gap decreases as a function of increasing quinoid character of polymer backbone. As can be seen from data in Table 1, the mono-substitution with electron-withdrawing CN group leads to decrease of the BLA value by 0.006 Å with respect to the F₂V molecule. The next presence of CN group has an additive influence on the BLA decrease. The presence of CN groups leads to increasing quinoid character. Consequently, the lowest excitation energies of F₂V–CN and F₂V–(CN)₂ are decreased when compared to the original compound. On the other hand, a small increase of BLA value is obtained for amino derivatives F₂V–NH₂ and F₂V–(NH₂)₂. For the push–pull system of F₂V–CN–NH₂, the resulting influence of both groups has a compensatory effect since the BLA value is similar to the non-substituted F₂V molecule.

The structural changes in the central part of the molecule, induced by the strong electron-withdrawing cyano or donating amino groups, are also reflected in the vertically excited electronic states, reflecting the balance between the perturbation of conjugation due to the distortion of the planar scaffold and the electronic effects of the substituents. The first seven vertical excitation energies with non-negligible oscillator strengths are summarized in Table 2. The lowest excitation energy for non-substituted molecule F₂V is 3.23 eV, compared to 3.08 eV [21] for tetrahydrofuran, suggesting the model systems employed and the level of theory employed is sufficiently accurate to gain useful insights.

The mono- or symmetric bi-substitutions lead to a bathochromic energy shift. Only in the case of F₂V–NH₂ molecule, the lowest excitation is blue-shifted by 0.06 eV. The global hypsochromic shift of excitation energies is indicated for the F₂V–CN–NH₂ system. The

lowest excitation energy is higher about 0.20 eV with respect to the F₂V. It seems that the torsion of molecular chains has more dominant influence of the photo-physical properties than the global push–pull effects of used substituents. Finally, in all cases, the substitution of vinylene bridge leads to a decrease in the oscillator strength for the lowest singlet excitation transition (S₁) and to the increase of the oscillator strengths for the next optical transitions.

In order to understand the physical origin of optical transitions for the selected excitation energies, it is useful to examine the (highest) occupied (HOMO) and lowest unoccupied molecular orbitals (LUMO). As reported in Table 2, the lowest energy electronic excitation is from the HOMO to the LUMO for all molecules and displays $\pi\pi^*$ character. However, the electronic distribution observed for the next transitions, as different as given by the molecular orbitals, are very different for the different substitutions. As presented in Fig. 3, the HOMO orbital of F₂V molecule is located mostly on the double bond of vinylene bridge (bond no. 15) and neighbouring bonds on fluorene units perpendicular oriented with respect to the chain (see bond nos. 12, 13, and 17, 18). The LUMO orbital is spread over the single bonds of vinylene bridge (bond nos. 14 and 16) and bonds on the fluorene units which are parallel with the molecular chain (see bond nos. 10, 11 and 19, 20). The presence of electron-withdrawing cyano group in F₂V–CN and F₂V–(CN)₂ molecules increases the electron delocalization in HOMO over the fluorene units. On the other hand, the electron-donating amino groups in F₂V–NH₂

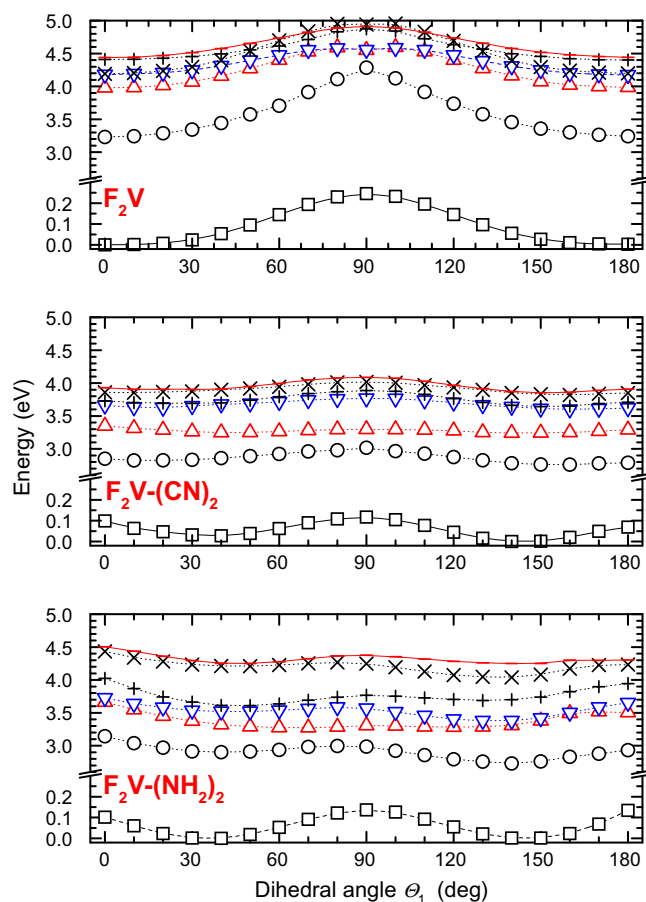


Fig. 4. One-dimensional dependence of the electronic ground-state and vertically excited energies of studied symmetric systems on the torsion calculated at the B3LYP/6-31G(d) theoretical level. The ground-state energy minimum is taken as energy reference (see also Table 3). The open square denote electronic ground-state and the next symbols indicate the first six excited states (S₁, ○; S₂, △; S₃, ▽; S₄, +; S₅, ×; S₆, *).

and $F_2V-(NH_2)_2$ molecules is responsible for the electron delocalization over the vinyne bond no. 15.

The opposite situation is observed in the LUMOs for both types of derivatives. With respect to this fact, we can deduce that the optical transition for cyano derivatives is oriented from the fluorene units to the CN chromophore while for the amino derivatives it is spread from the central part to the fluorene units. This character is also reflected for the higher vertical excitations. The transi-

tions from lower occupied orbitals to the LUMO orbital play important role for the CN derivatives. The transitions from HOMO to the next unoccupied orbitals occur for the amino derivatives. In the case of the $F_2V-CN-NH_2$ molecule, the presence of the strong push-pull systems leads to the combined effect with strongly determined direction of optical transition. This electron transition starts from the fluorene unit and neighbouring NH_2 group and goes to the CN group and opposite fluorene unit.

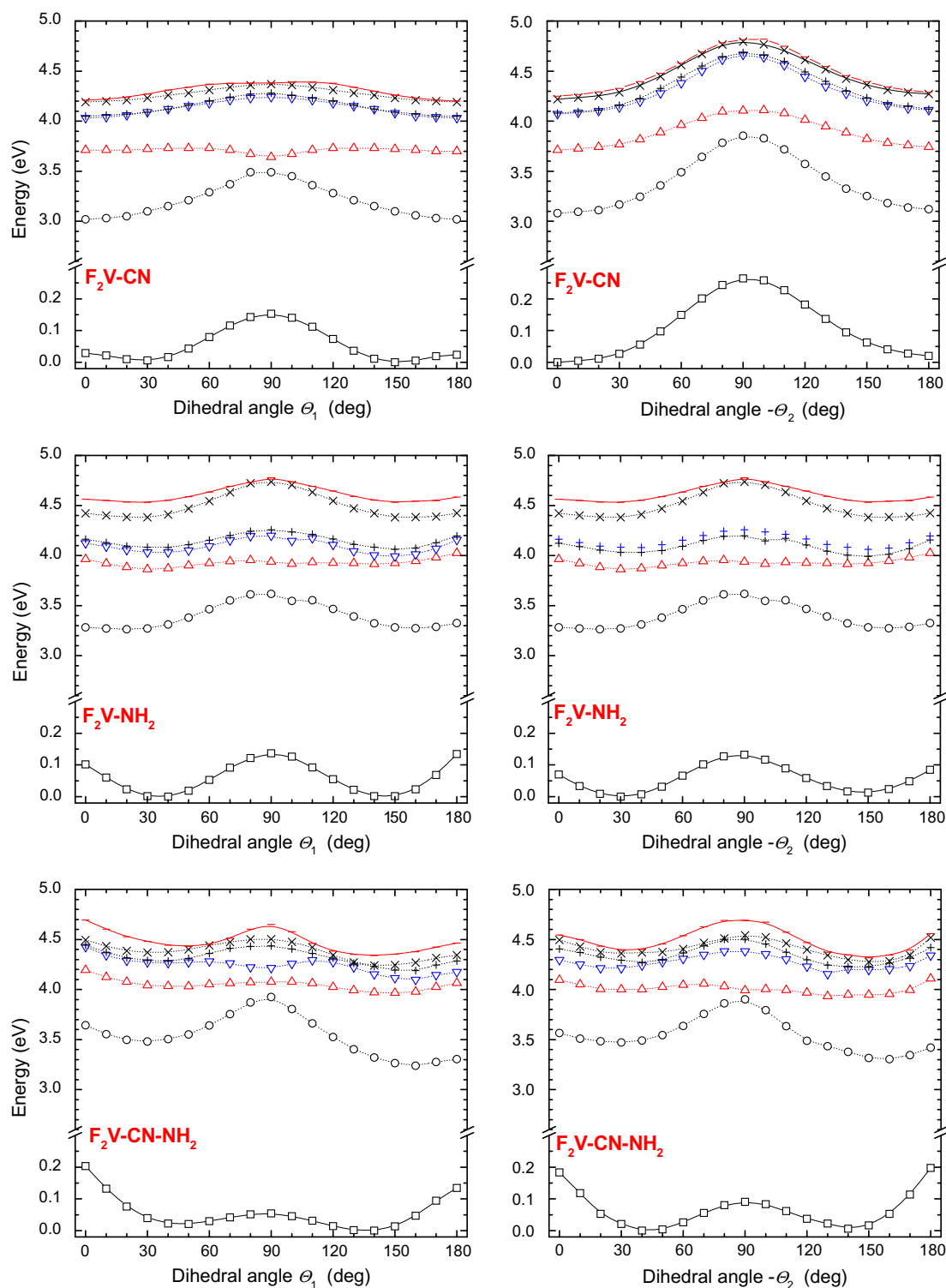


Fig. 5. One-dimensional dependence of the electronic ground-state and vertically excited energies of studied asymmetric systems on the torsion calculated at the B3LYP/6-31G(d) theoretical level. The ground-state energy minimum is taken as energy reference (see also Table 3). The open square denote electronic ground-state and the next symbols indicate the first six excited states (S_1 , $\circ$; S_2 , Δ ; S_3 , ∇ ; S_4 , $+$; S_5 , $\times$; S_6 , $-$).

The electronic ground-state torsional potentials of the fluorene unit around the common single bond were also investigated using the B3LYP/6-31G(d) method with the results being depicted in Figs. 4 and 5 and the geometrical and energetic values in Table 3. As can be seen in Fig. 4, the non-substituted molecule F_2V has two planar minima and the barrier for perpendicular arrangement ($\Delta E = 0.245$ eV or 5.65 kcal mol⁻¹). We note that the similar shape of potential and barrier location with $\Delta E = 0.22$ eV (5.0 kcal mol⁻¹) were obtained for the torsion of phenylene ring around the vinylene single bond at the B3LYP/cc-pVDZ theoretical level [22]. The presence of cyano or amino groups on the vinylene bridge leads to structures that are more stable to a non-planar conformation. The obtained potential curves exhibit two non-planar minima corresponding to the most stable *trans* or *cis* conformations and two first-order saddle points. The most stable conformation of $F_2V-(NH_2)_2$ molecule is at 45° while the $F_2V-(CN)_2$ molecule prefers *trans* conformation (144°). The mutual comparison of the energy barriers with respect to the F_2V molecule shows that the vinylene bridge substitution decreases the barrier at perpendicular arrangement. The cyano substitution decreases the barrier relative to F_2V by 0.092 eV (for F_2V-CN) and by 0.128 eV (for $F_2V-(CN)_2$). Although the amino substitution leads to the higher decrease of the perpendicular barrier relative to F_2V (0.108 eV for F_2V-NH_2 and 0.169 eV for $F_2V-(NH_2)_2$), the torsion at the planar arrangements is more restricted. The electronic ground-state potential curves for mono-substituted molecules around the second dihedral

angle Θ_2 have different shape. For the F_2V-CN molecule, the shape is similar to the potential curve of non-substituted F_2V molecule. A different situation occurs for the F_2V-NH_2 molecule in that the second evaluated curve copies the shape of the dependence around the first dihedral angle. It seems that the strong electron donor amino group is able directly to affect through the double bond no. 15 the torsional motion on the opposite molecular side. The modification of the torsional potential with the amino or cyano groups is also observed for mixed $F_2V-CN-NH_2$ molecule. As can be seen in Fig. 5, the torsional potential profile for the Θ_1 and Θ_2 dihedral angles are two-times lower in energy for the perpendicular arrangement compared to the planar arrangement. In addition, the potential curve for the torsion around the angle Θ_1 (at the side of CN substitution) exhibits more stable conformation at 132° . The energy difference with respect to the second minimum at 52° is 0.021 eV. This is a different situation to that presented for F_2V-CN and F_2V-NH_2 . It appears that these features have a different impact on the excitation–relaxation phenomena which occur in various time-dependent optical experiments.

The excited-state potential energy curves calculated at the TD-B3LYP/6-31G(d) level based on ground-state optimized geometries are also shown in Figs. 4 and 5. The curve for the lowest excited state of S_1 with dominant oscillator strengths for F_2V exhibits the energy minimum for a planar geometry and a maximum for the perpendicular one. The energy difference between these points is ca 1.06 eV. In the case of F_2V-CN and F_2V-NH_2 , the S_1 state practically reflects the curve for the F_2V molecule, but the energy differences between the perpendicular and planar arrangements are approximately two-times lower. The symmetric bi-substitution decreases very effectively the sensitivity of the evaluated potential curves on the torsion. Moreover, in all investigated molecules, the lowest S_1 state does not cross the next higher states. The potential energy curves for the next higher excited states of F_2V molecule are quite closely spaced to each other and show multiple intersections (around 70 – 120°). On the other hand, the substitution can cause the small separation of crossing region of potential curves between the next lowest excited states.

4. Conclusions

The (TD)-B3LYP method has been used for the systematic theoretical investigation of the photo-physical properties of substituted model bifluorenevinylene compounds. The substitutions of the vinylene unit by strong electron-accepting CN and/or electron-donating NH_2 groups were considered. Our calculations indicate that the non-substituted F_2V compound is planar and that substitution leads to the twisting of the molecular fragment on the side of substitution. In the case of the F_2V-NH_2 molecule, the amino group is also responsible for the non-planarity on the non-substitute side. Additionally, the chemical modification of vinylene bridge affects the electronic ground-state torsional potentials of the fluorene unit around the single bond. The non-substituted molecule F_2V has two planar minima and the barrier for perpendicular arrangement. The presence of cyano or amino groups on vinylene bridge is responsible for the restriction of the torsional motion at the barriers located for planar and perpendicular arrangements. The energy barrier highs are also dependent on the substitution.

The TD-DFT torsional potential energy curves in the vertically excited states were also investigated in this work. From our calculations, we do not find any indication of state crossings of the S_1 state with higher ones for all investigated systems. The symmetric bi-substitution markedly decreases the potential curves on the torsion. The potential energy curves for the next higher excited states of F_2V molecule are quite closely spaced to each other and show multiple intersections (around 70 – 120°). On the other hand, the substitution can modulate the separation of crossing region of

Table 3

Relative energies ΔE and dihedral angles Θ_1 and Θ_2 of external points for torsional dependencies with respect to the most stable structure. The angles are in degrees and energies in eV or kcal mol⁻¹ (values in parentheses). See also Figs. 4 and 5.

Molecule	Θ_1	Θ_2	ΔE
F_2V	0	0	0.000 (0.00)
	90	0	0.245 (5.65)
	180	0	0.003 (0.07)
$F_2V-(CN)_2$	0	-46	0.099 (2.28)
	38	-38	0.000 (0.00)
	90	-29	0.117 (2.70)
	144	-37	0.001 (0.02)
	180	-45	0.069 (1.59)
$F_2V-(NH_2)_2$	0	-57	0.224 (5.17)
	45	-45	0.000 (0.00)
	90	-39	0.076 (1.75)
	128	-37	0.038 (0.88)
	180	-45	0.298 (6.88)
F_2V-CN	0	0	0.029 (0.67)
	28	-7	0.001 (0.02)
	90	0	0.153 (3.53)
	148	-7	0.000 (0.00)
	180	-1	0.024 (0.55)
	25	-90	0.293 (6.76)
	29	-180	0.049 (1.13)
F_2V-NH_2	0	-31	0.102 (2.35)
	36	-31	0.000 (0.00)
	90	-28	0.137 (3.16)
	146	-24	0.002 (0.05)
	180	-38	0.134 (3.09)
	36	0	0.070 (1.61)
	35	-90	0.132 (3.04)
	35	-148	0.013 (0.30)
	39	-180	0.085 (1.96)
$F_2V-CN-NH_2$	0	-31	0.205 (4.73)
	52	-44	0.021 (0.50)
	90	-28	0.056 (1.29)
	132	-43	0.000 (0.00)
	180	-38	0.124 (2.65)
	57	0	0.184 (4.24)
	42	-90	0.091 (2.10)
	43	-133	0.024 (0.57)
	58	-180	0.198 (4.57)

potential curves between the next lowest excited states. The absence of the intersections of excited states around the stable structures and relative well separation of the lowest excited state around the minima enable us to perform the molecular dynamics studies of investigated molecules based on an adiabatic approach. These results show that theoretical studies can help us to understand the relationship between the torsional broadening of absorption spectra, chemical structure and time-dependent optical phenomena, having implications for the design and synthesis of novel optical materials.

Acknowledgements

The authors thank Prof. Harald Kauffmann for valuable comments and discussion. This work was supported by the Thailand Research Fund (RTA5080005 to S.H. and MRG5180287 to S.S.). W.M. is grateful to Thai Graduate Institute of Science and Technology (TGIST), the Austrian Science Fund within the framework of the Special Research Program F16, Advanced Light Sources (ADLIS), and Bilateral Research Cooperation (BRC) from Faculty of Science, Kasetsart University for scholarships and supporting. The National Center of Excellence in Petroleum, Petrochemical Technology and Advanced Materials is gratefully acknowledged for research supporting and facilities. The calculations were performed in part on the Schrödinger III cluster of the University of Vienna. Thanks are also due to Dr. Matthew Paul Gleeson for helpful comments and reading of the manuscript.

References

- [1] M. Ranger, D. Rondeau, M. Leclerc, *Macromolecules* 30 (1997) 7686.
- [2] V.P. Barberis, J.A. Microyannidis, *Synth. Met.* 156 (2006) 1408.
- [3] C. Chi, C. Im, V. Enkelmann, A. Ziegler, G. Lieser, G. Wegner, *Chem. Eur. J.* 11 (2005) 6833.
- [4] S.H. Jin, H.J. Park, J.Y. Kim, K. Lee, S.P. Lee, D.K. Moon, H.J. Lee, Y.S. Gal, *Macromolecules* 35 (2002) 7532.
- [5] Y. Jin, J.Y. Kim, S. Song, Y. Xia, J. Kim, H.Y. Woo, K. Lee, H. Suh, *Polymer* 49 (2008) 467.
- [6] L. Zhang, Q. Zhang, H. Ren, H. Yan, J. Zhang, H. Zhang, J. Gu, *Sol. Energy Mater. Sol. Cells* 92 (2008) 581.
- [7] V. Lukes, A. Aquino, H. Lischka, *J. Phys. Chem. A* 109 (2005) 10232.
- [8] W. Meeto, S. Suramitr, S. Vannarat, S. Hannongbua, *Chem. Phys.* 349 (2008) 1.
- [9] S. Suramitr, T. Kerdcharoen, T. Srihirin, S. Hannongbua, *Synth. Met.* 155 (2005) 27.
- [10] Y. Yang, J.K. Feng, Y. Liao, A.M. Ren, *Polymer* 46 (2005) 9955.
- [11] F. Furche, R. Ahlrichs, *J. Chem. Phys.* 117 (2002) 7433.
- [12] J. Gierschner, J. Cornil, H.J. Egelhaaf, *Adv. Mater.* 19 (2007) 173.
- [13] S. Tretiak, A. Saxena, R.L. Martin, A.R. Bishop, *Phys. Rev. Lett.* 89 (2002) 97402.
- [14] A.D. Becke, *J. Chem. Phys.* 98 (1993) 5648.
- [15] P.M.W. Gill, B.G. Johnson, J.A. Pople, M.J. Frisch, *Chem. Phys. Lett.* 197 (1992) 499.
- [16] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, 567 revision B.05. Gaussian, Inc., Pittsburgh, 2003.
- [17] P. Poolmee, M. Ehara, S. Hannongbua, H. Nakatsuji, *Polymer* 46 (2005) 6474.
- [18] K. Sriwichitkamol, S. Suramitr, P. Poolmee, S. Hannongbua, *J. Theor. Comput. Chem.* 5 (2006) 595.
- [19] D. Jacquemin, E.A. Perpète, H. Chermette, I. Ciofini, C. Adamo, *Chem. Phys.* 332 (2007) 79.
- [20] J.L. Brédas, *J. Chem. Phys.* 82 (1985) 3808.
- [21] Q. Liu, W. Liu, B. Yao, H. Tian, Z. Xie, Y. Geng, F. Wang, *Macromolecules* 40 (2007) 1851.
- [22] S.P. Kwasniewski, L. Claes, J.P. Francois, M.S. Deleuze, *J. Chem. Phys.* 118 (2003) 7823.

Excited State Properties, Fluorescence Energies, and Lifetimes of a Poly(fluorene-phenylene), Based on TD-DFT Investigation

RUNGTIWA CHIDTHONG,¹ SUPA HANNONGBUA^{2,3}

¹Chemistry Program, Faculty of Science and Technology, Nakhon Pathom Rajabhat University, Nakhon Pathom 73000, Thailand

²Department of Chemistry, Faculty of Science, Kasetsart University, Jatuchak, Bangkok 10900, Thailand

³Center of Nanotechnology, Kasetsart University, Bangkok 10900, Thailand

Received 28 May 2009; Revised 21 August 2009; Accepted 27 August 2009

DOI 10.1002/jcc.21429

Published online 27 October 2009 in Wiley InterScience (www.interscience.wiley.com).

Abstract: The structural and electronic properties of fluorene-phenylene copolymer (FP)_n, *n* = 1–4 were studied by means of quantum chemical calculations based on density functional theory (DFT) and time dependent density functional theory (TD-DFT) using B3LYP functional. Geometry optimizations of these oligomers were performed for the ground state and the lowest singlet excited state. It was found that (FP)_n is nonplanar in its ground state while the electronic excitations lead to planarity in its S₁ state. Absorption and fluorescence energies were calculated using TD-B3LYP/SVP and TD-B3LYP/SVP+ methods. Vertical excitation energies and fluorescence energies were obtained by extrapolating these values to infinite chain length, resulting in extrapolated values for vertical excitation energy of 2.89 and 2.87 eV, respectively. The S₁ ← S₀ electronic excitation is characterized as a highest occupied molecular orbital to lowest unoccupied molecular orbital transition and is distinguishing in terms of oscillator strength. Fluorescence energies of (FP)_n calculated from TD-B3LYP/SVP and TD-B3LYP/SVP+ methods are 2.27 and 2.26 eV, respectively. Radiative lifetimes are predicted to be 0.55 and 0.51 ns for TD-B3LYP/SVP and TD-B3LYP/SVP+ calculations, respectively. These fundamental information are valuable data in designing and making of promising materials for LED materials.

© 2009 Wiley Periodicals, Inc. J Comput Chem 31: 1450–1457, 2010

Key words: fluorene-phenylene; TD-DFT; excited state; vertical excitation energy; fluorescence energy; lifetime

Introduction

Conducting polymers have been increasing interest in novel π -conjugated materials exhibiting light emitting properties. Conjugated polymers are widely investigated for used in light emitting diodes (LED),^{1–7} solar cells,^{7–9} field-effect transistors,^{10–13} etc. Among these applications, the development of tunable and flexible polymeric light emitting diodes has been extensively studied for both basic structural properties and industrial applications.^{1,14–16} Olydialkylfluorenes are found to be highly fluorescent compounds, high efficient blue emission, reasonable mobility, and good thermal and chemical stabilities, which make them suitable for applications in LEDs.^{17–24} Oly(9,9-dialkylfluorene) reveals a good and nondispersive hole transport.^{18,25} However, it was observed that the excimers formation and poor electron mobility limit their application in blue LEDs.^{26–29} To overcome these problems, the inclusion of various monomer such as pyridine,^{30–35} phenylene,^{30,36–40} and thiophene^{7,36,41,42} in fluorene oligomer or polymer backbone is recommended.

Recently, novel fluorene-phenylene-based copolymers displaying interesting blue electroluminescence have been synthesized and characterized.^{30,36–40} Moreover, theoretical investigation on both ground and excited states of fluorene-phenylene oligomers has been carried out. Belletête et al. reported the theoretical methods and spectroscopic and photophysical properties of fluorene-based dimers.⁴⁰ Fluorene-phenylene-based oligomers, which can be used as polyester were also investigated theoretically using HF/6-31G* and ZINDO/S.⁴³ While poly(9,9-dihexylfluorene-1,4-phenylene) unit cells were investigated on the influence of electron acceptors on the geometric and electronic properties by quantum chemical calculations.⁴⁴ In addition the

Correspondence to: R. Chidthong; e-mail: rungtiwa@npru.ac.th

Contract/grant sponsor: The Thailand Research Fund; contract/grant numbers: MRG5080324 to RC, RTA5080005 to SH.

Contract/grant sponsor: The Higher Education Commission, The Research and Development Institute, Nakhon Pathom Rajabhat University.

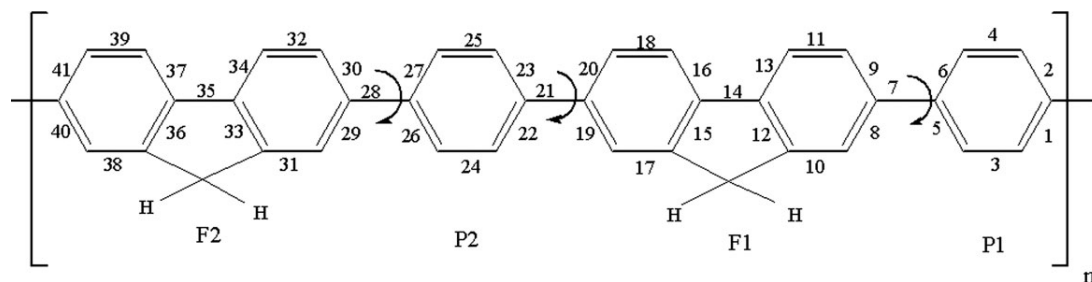


Figure 1. Molecular structure and numbering scheme of poly[2,7-(9,9-difluorene)-*co-alt-p*-phenylene], (FP)_n used in this study.

ground state and first excited state properties of oligofluorenes using RHF/6-31G* and RCIS/6-31G* were studied, respectively.⁴⁵ Gong et al. calculated the HOMO-LUMO energy differences, singlet-singlet electronic transitions and excited state optimization of fluorene-phenylene monomer.⁴⁶ However, the investigation of ground and excited state properties, fluorescence energy, and radiative lifetime of poly(fluorene-phenylene) are still a very challenging task. Therefore, this study is the first time to calculate the excited state geometry of poly(fluorene-phenylene) and predict fluorescence energy and radiative lifetime of this oligomer.

The main objectives of this work are to investigate the ground and excited state structures, electronic and optical properties of poly[2,7-(9,9-difluorene)-*co-alt-p*-phenylene], (FP)_n. Up to now, one of the best methods for obtaining excited state properties of large molecules is configuration interaction singles (CIS). However, due to the computational cost of CIS applied to oligomers, this method is able to use for only monomer in many applications.^{45–50} Thus, in this study, the lowest singlet excited state optimization of (FP)_n, $n = 1–4$ were investigated using TD-B3LYP, which has been acceptable and successfully applied for polymer molecules.^{51–56} Finally, fluorescence energies and radiative lifetimes are also predicted. These information are needed and useful for the development of LEDs materials.

Methods

In this study, fluorene (F) and phenylene (P) oligomer was theoretical studied. The ground and excited conformational curves for the interring torsion angle between the fluorene (F) and phenylene (P) units were investigated in the steps of 30°. Geometry optimizations in the ground and excited states were calculated using the split valance polarization (SVP) basis set.⁵⁶ The ground and excited state conformation analysis were performed using density functional theory (DFT) and time dependent density functional theory (TD-DFT) calculations, respectively. In addition, the ground and excited state geometry optimizations of (FP)_n, $n = 1–4$ were fully optimized by DFT and TD-DFT calculations, respectively. The B3LYP^{57–59} functional was used in the DFT and TD-DFT calculations, respectively. To reduce computational time, alkyl groups at the 9 position were replaced by hydrogen-atom on the five membered ring of the fluorene unit. Sriwichitkamol et al.⁶⁰ studied the effect of the alkyl group on

fluorene oligomers and they proved that the length of alkyl group does not affect the structural and electronic properties of fluorene oligomers. In fact, the alkyl side chain is helpful in increasing the solubility of the polymer. On the basis of optimized geometries, the excitation energies and fluorescence energies were calculated at the TD-B3LYP/SVP and TD-B3LYP/SVP+ levels of theory. The SVP+ basis set was constructed by augmenting the SVP basis with a set of s and p functions on carbon. The exponents of these additional basis functions were obtained by dividing the smallest respective exponent of the SVP basis set by the factor of three.^{53,61} Vertical excitation energies were performed on the basis of ground state optimized geometries. The fluorescence energies were computed as the vertical de-excitation based on the lowest excited state optimized structures. The first five singlet-singlet electronic transition ($S_1 \leftarrow S_0$) were calculated at the TD-B3LYP/SVP and TD-B3LYP/SVP+ levels of calculation. An extrapolating technique was used in the electronic calculations to estimate the energy gaps of the polymer. Moreover, the results were compared with the experimental data. All calculations were done using Turbomole version 5.7⁶² program packages, running on Linux 3.4 GHz PC. In this work, the properties (excitation energy, fluorescence energy, and radiative lifetime) of a series of oligomers with increasing chain length were calculated following by extrapolation to infinite chain length.^{35,47,49,51–53,55,60}

Results and Discussion

Geometries of Poly(fluorene-phenylene) Copolymer

The molecular structure of FP copolymer, containing the bond numbering is illustrated in Figure 1. The ground state (S_0) and first lowest excited state (S_1) conformational curves around the interring bond number 7 of the FP monomer were calculated by the B3LYP/SVP and TD-B3LYP/SVP methods and depicted in Figure 2. According to the calculations, the S_0 and S_1 states show similar two minima. The energy barrier against planarity for ground state (2.20 kcal/mol) is lower than that of excited state (13.14 kcal/mol). Whereas torsion energy at planar structures for S_0 (1.42 kcal/mol) is higher than S_1 state (0.01 kcal/mol). The ground state optimized bond distance of F1-P1 (bond number 7) computed by B3LYP/SVP is 1.488 Å and optimized torsion angle is 35.7°. The results implied that monomer struc-

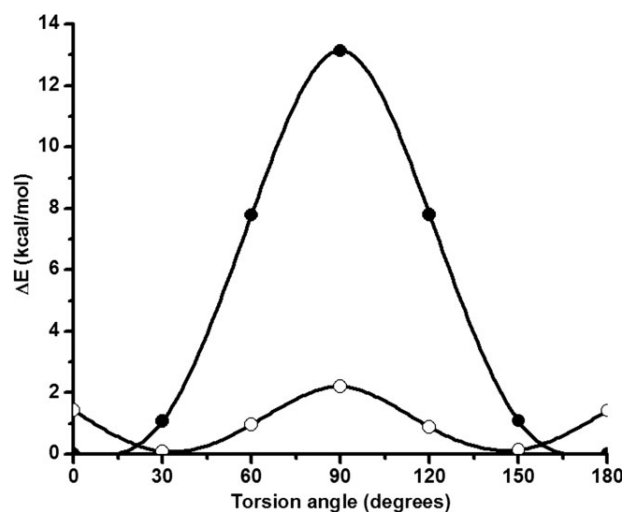


Figure 2. Torsional potential energy curves around bond number 7, ϕ_{P1-F1} (see Fig. 1) for the ground state of FP monomer as obtained from the B3LYP/SVP (open circles) calculations and the lowest excited state as calculated using the TD-B3LYP/SVP method (solid circles).

ture is nonplanar at ground state. Comparing these results to the optimized interring distance (bond number 7) and torsion angle of FP monomer as calculated by HF/6-31G* level, we observed that the interring distance and torsion angle calculated from B3LYP/SVP are close to that computed by HF/3-21G* ($r = 1.485$ Å and torsion angle $= 38.1^\circ$).⁴⁶ The excited state interring distance (bond number 7) and torsion angle calculated by TD-B3LYP/SVP are 1.443 Å and 7.7° , respectively. The obtained results indicate that excited state monomer prefers planarity than that of the ground state.

The ground state and first lowest excited state potential energy curves of FP dimer calculated by B3LYP/SVP method is further studied as shown in Figure 3. The ground state energy barriers as obtained from F1-P1 (bond number 7), P2-F1 (bond number 21), and P2-F2 (bond number 28) are 2.21, 2.40, and 2.38 kcal/mol, respectively, whereas excited state energy barriers are 4.08, 15.54, and 8.81 kcal/mol, respectively. Interestingly, the torsion angles between F1-P1 (bond number 7), P2-F1 (bond number 21), and P2-F2 (bond number 28) are decreased from -37.2° to -27.9° , 37.1° to 9.6° , and -34.6° to -10.6° , respectively. These indicated that the torsion angles between the two adjacent units are reduced in excited state when compared with the ground state. Consequently, the excited state structure of the dimer comes closer to planarity.

To understand the structural properties of fluorene-phenylene oligomer $(FP)_n$, the comparison between ground state and first lowest excited state geometries in terms of bond length and torsion angle were extended studied using B3LYP/SVP and TD-B3LYP/SVP calculations, respectively. Optimized bond lengths and torsion angles of $(FP)_n$, $n = 1-4$, calculated by B3LYP/SVP and TD-B3LYP/SVP methods are shown in Table 1. The relative optimized interring distances and torsion angles of $(FP)_n$, $n = 1-4$, in the ground and excited states are illustrated in Figure

4. It was found that bond distances do not suffer appreciable variation with the oligomers size in the series of $(FP)_n$. And it implied that we can depict the basic bond distances of the polymers as their oligomers. For torsion angle, the ground state interring torsion angles of fluorene and phenylene oligomers are in the range of $34.6-37.5^\circ$. Similarly as found in the monomer and dimer, the torsion angles in excited state are smaller than that of the ground state. Therefore, the conjugation is well found in the excited state and it is also obviously that excited structure has a strong coplanar tendency. These behaviors have been also found in the case of poly(fluorene-vinylene) copolymer,⁵⁵ poly(fluorene-pyridine) copolymer,^{47,51,63} and polycarbazole.^{48,64,65}

Vertical Excitation Energy

From the previous study, the electronic excitation of fluorene-phenylene (FP) monomer was theoretically studied and it was found that the excitation energy, investigated by CIS/6-31G* method is 4.04 eV.⁴⁵ In addition, the singlet-singlet excitation energy of FP monomer performed by TD-B3LYP/6-31G* is 4.12 eV.⁴⁶ In our study, the vertical excitation energies (E_{excit}) of FP copolymer were focused. The TD-B3LYP/SVP and TD-B3LYP/SVP+ were used to obtain the vertical excitation energies of the first five singlet-singlet transitions of $(FP)_n$. In each case, the ground state optimized structures were used to investigate the vertical excitation energy. The first five vertical excitation energies and oscillator strengths (f) of $(FP)_n$, $n = 1-4$, were listed in Table 2. For all oligomers, excitation energies were decreased as the increase of conjugation length indicating delocalizing of the electron along the chain. Therefore, the chain lengths of polymer have effect on electronic transition. The $S_1 \leftarrow S_0$ electronic transition appears to have the highest intensity

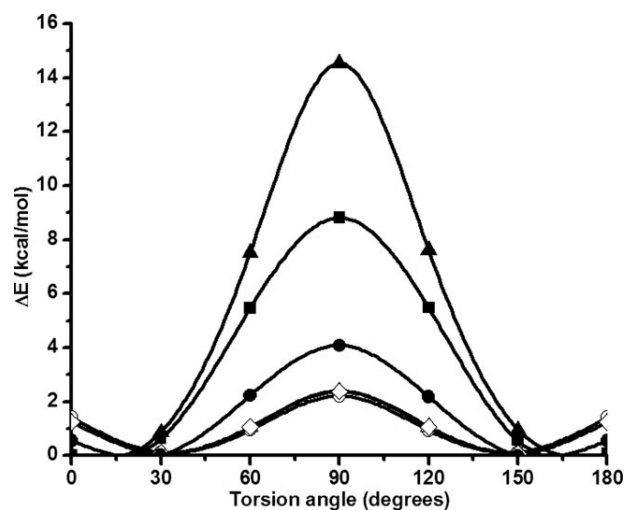


Figure 3. Torsional potential curves around bond number 7, ϕ_{P1-F1} (open circles) 21, ϕ_{F1-P2} (open triangles), and 28, ϕ_{P2-F2} (open squares) (see Figure 1) for the ground state of FP dimer as computed from B3LYP/SVP calculations and for the lowest excited state using the TD-B3LYP/SVP approach (solid circles, solid triangles, and solid squares, respectively).

Table 1. Optimized Interring Distances, d (Å) and Torsion Angle, φ (deg) of $(FP)_n$ ($n = 1-4$) in the Ground State (S_0) and the Lowest Singlet Excited State (S_1), Obtained from B3LYP/SVP and TD-B3LYP/SVP Calculations.

	Monomer		Dimer		Trimer		Tetramer	
	S_0	S_1	S_0	S_1	S_0	S_1	S_0	S_1
	B3LYP/SVP	TD-B3LYP/SVP	B3LYP/SVP	TD-B3LYP/SVP	B3LYP/SVP	TD-B3LYP/SVP	B3LYP/SVP	TD-B3LYP/SVP
d_7	1.488	1.443	1.488	1.475	1.488	1.482	1.488	1.486
d_{14}	1.470	1.425	1.468	1.436	1.468	1.451	1.468	1.462
d_{21}			1.486	1.446	1.486	1.458	1.486	1.474
d_{28}			1.486	1.452	1.486	1.451	1.486	1.466
d_{35}			1.470	1.451	1.468	1.433	1.468	1.438
d_{42}					1.486	1.461	1.486	1.453
d_{49}					1.486	1.471	1.486	1.456
d_{56}					1.470	1.463	1.468	1.445
d_{63}							1.486	1.473
d_{70}							1.486	1.480
d_{77}							1.470	1.467
φ_{P1-F1}	-35.7	-7.7	-37.2	-27.9	37.0	33.0	37.5	36.0
φ_{F1-P2}			37.1	9.6	-35.5	-16.6	35.6	27.3
φ_{P2-F2}			-34.6	-10.6	35.5	13.3	35.0	22.2
φ_{F2-P3}					-36.25	-18.4	-35.4	-13.7
φ_{P3-F3}					-36.1	-25.4	35.5	16.1
φ_{F3-P4}							-35.4	-26.5
φ_{P4-F4}							36.1	31.42

P denotes the phenylene ring and F the fluorene ring.

as determined by its large oscillator strength contribution. On the other hand, the S_2 , S_3 , S_4 , and S_5 electronic transition showed small oscillator strengths. From the data in Table 2, TD-B3LYP/SVP excitation energies are found to be similar to that obtained from TD-B3LYP/SVP+ calculations as the difference are less than 0.1 eV. These results are in the same trend as found in our previous study.⁵¹

It is also interesting to see the relationship between excitation energies and the inverse chain length as plotted in Figure 5. The linear relation was observed between excitation energies by both methods and the inverse chain length. Excitation energies of fluorene-phenylene copolymer predicted from TD-B3LYP/SVP and TD-B3LYP/SVP+ calculations are 2.89 and 2.87 eV, respectively, which are underestimated than those of experimental results^{47,49,66} (3.34 eV in film and 3.36 eV in chloroform solution).³⁰ Several investigations displayed that TD-DFT calculations give underestimate excitation energies when compared with experimental data. For example, calculated excitation energies of carbazole derivatives from the group of Suramitr⁶⁴ showed lower E_{excit} than experimental results of about 0.5 eV. Additionally, Brière et al. also reported the underestimate TD-DFT values of the poly(2,7-fluorene), poly(2,7-carbazole), and poly(p-phenylene).⁶⁷ However, there are some successful investigations on excitation energy calculated by TD-DFT level of calculation which agreed well with experimental data.^{46,47,68,69}

From the analysis of the wave function in Table 3, it is clearly to explain that the $S_1 \leftarrow S_0$ electronic transition corresponds mainly to the promotion of an electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied

molecular orbital (LUMO) as indicated by the highest dominant contribution. It is reasonable that since $\text{LUMO} \leftarrow \text{HOMO}$ excitation is strong in the $S_1 \leftarrow S_0$ electronic transition and as analysis above that with the elongating of molecular chain, E_{excit} decrease. Additionally, the $S_2 \leftarrow S_0$ revealed smaller character as compared with $S_1 \leftarrow S_0$. Dominant contributions for the monomer, dimer, trimer, and tetramer are characterized predominantly by the transition of an electron from $\text{LUMO} + 1 \leftarrow$

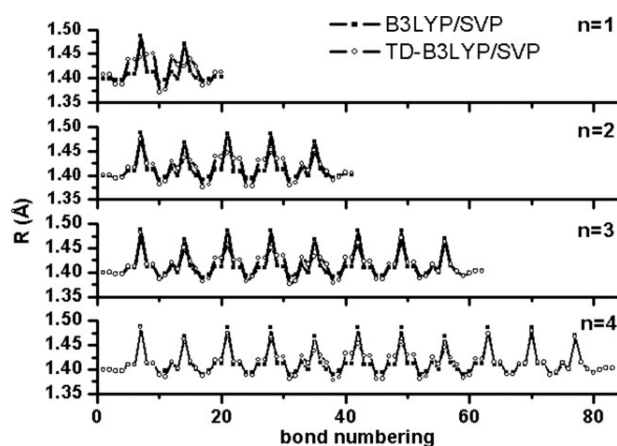
**Figure 4.** Optimized bond lengths (Å) of $(FP)_n$ in the ground state (S_0) and the lowest excited state (S_1), obtained from B3LYP/SVP (solid squares) and TD-B3LYP/SVP (open circles) calculations.

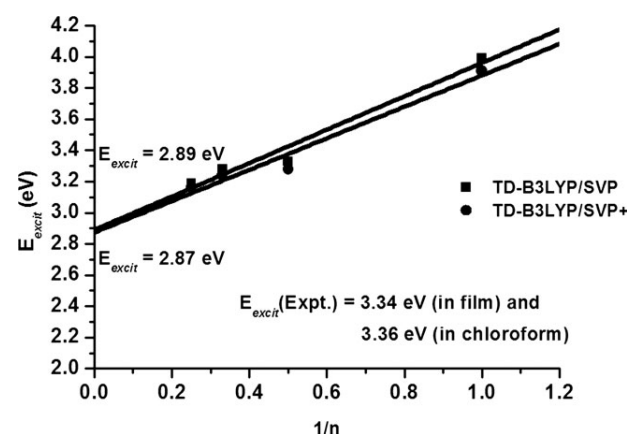
Table 2. Calculated and Experimental Excitation Energies and Oscillator Strengths of Fluorene-Phenylene Oligomers.

	State	TD-B3LYP/SVP		TD-B3LYP/SVP+	
		Excitation energy (eV)	Oscillator strength	Excitation energy (eV)	Oscillator strength
$n = 1$	S ₁	3.99	0.904	3.91	0.874
	S ₂	4.33	0.009	4.22	0.025
	S ₃	4.66	0.004	4.60	0.003
	S ₄	4.91	0.004	4.83	0.004
	S ₅	4.95	0.005	4.87	0.007
$n = 2$	S ₁	3.33	2.040	3.28	1.990
	S ₂	3.81	0.000	3.77	0.000
	S ₃	3.95	0.091	3.87	0.082
	S ₄	4.11	0.005	4.01	0.013
	S ₅	4.15	0.002	4.03	0.000
$n = 3$	S ₁	3.28	3.015	3.24	2.969
	S ₂	3.66	0.163	3.62	0.172
	S ₃	3.71	0.032	3.67	0.014
	S ₄	3.98	0.114	3.94	0.413
	S ₅	3.98	0.590	3.94	0.263
$n = 4$	S ₁	3.19	3.909	3.16	3.843
	S ₂	3.45	0.442	3.42	0.440
	S ₃	3.58	0.002	3.54	0.003
	S ₄	3.71	0.642	3.67	0.612
	S ₅	3.77	0.035	3.74	0.011
$n = \infty$	S ₁	2.89		2.87	
Expt. ³⁰	S ₁	3.34 (in film) 3.36 (in chloroform)			

Geometries were optimized at B3LYP/SVP level.

HOMO, LUMO $\leftarrow$ HOMO-1, LUMO $\leftarrow$ HOMO-1, and LUMO $\leftarrow$ HOMO - 1, respectively.

It is important to analyze the HOMO-LUMO describing the lowest singlet excitation. These orbitals are displayed in Figure 6 for the monomer to dimer. The HOMO and LUMO are delo-

**Figure 5.** Lowest excitation energies (E_{excit}) computed by the TD-B3LYP/SVP (solid squares) and TD-B3LYP/SVP+ (solid circles) methods as a function of reciprocal chain length n of the oligomer (FP) _{n} .

calized practically along the entire π -conjugated chain with the HOMO more concentrated on the side towards the other end. Therefore, a charge transfer toward the phenylene end will occur on the excitation to the LUMO. From HOMO in Figure 6, there is interring antibonding between the bridged carbon atoms. In contrast, LUMO is characterized by the interring bonding. These may explain the singlet excitation involving mainly the promotion of an electron from the HOMO to LUMO and the planarization in the excited state.

Fluorescence Energy and Lifetime

On the basis of S₁-optimized structures based on TD-B3LYP/SVP calculation, the fluorescence energy of fluorene-phenylene oligomers were computed, using TD-B3LYP/SVP and TD-B3LYP/SVP+ calculations. Fluorescence energies and oscillator strengths (in parentheses) are collected in Table 4. All fluorescence energies were decreased with the elongation of fluorene-phenylene chain. The opposite trend was found for oscillator strengths. The fluorescence energies are ranged between 3.53 to 2.63 eV and 3.45 to 2.60 eV by TD-B3LYP/SVP and TD-B3LYP/SVP+ calculations, respectively. TD-B3LYP/SVP results are found to be slightly higher than those of TD-B3LYP/SVP+ results. Additionally, fluorescence energies between TD-B3LYP/SVP and TD-B3LYP/SVP+ calculations are differently less than 0.1 eV. The addition of fluorene-phenylene units leads to decreasing in fluorescence energies when compared between TD-B3LYP/SVP and TD-B3LYP/SVP+ calculations. According to the data in Table 4, the relationship between fluorescence energies and the reciprocal chain lengths as demonstrated in Fig-

Table 3. Dominant Orbital Contributions for the First Five Excitation Energies of FP Oligomers Using TD-B3LYP/SVP Method.

Oligomers	State	MO/character (dominant contributions)
		TD-B3LYP/SVP
$n = 1$	S ₀ →S ₁	HOMO→LUMO (97%)
	S ₀ →S ₂	HOMO→LUMO+1 (79%)
	S ₀ →S ₃	HOMO-1→LUMO (50%)
	S ₀ →S ₄	HOMO-3→LUMO (30%)
	S ₀ →S ₅	HOMO-2→LUMO (52%)
$n = 2$	S ₀ →S ₁	HOMO→LUMO (98%)
	S ₀ →S ₂	HOMO-1→LUMO (59%)
	S ₀ →S ₃	HOMO→LUMO+1 (56%)
	S ₀ →S ₄	HOMO→LUMO+2 (74%)
	S ₀ →S ₅	HOMO→LUMO+3 (72%)
$n = 3$	S ₀ →S ₁	HOMO→LUMO (94%)
	S ₀ →S ₂	HOMO-1→LUMO (80%)
	S ₀ →S ₃	HOMO→LUMO+1 (80%)
	S ₀ →S ₄	HOMO-2→LUMO (62%)
	S ₀ →S ₅	HOMO-1→LUMO+1 (80%)
$n = 4$	S ₀ →S ₁	HOMO→LUMO (88%)
	S ₀ →S ₂	HOMO-1→LUMO (56%)
	S ₀ →S ₃	HOMO→LUMO+1 (55%)
	S ₀ →S ₄	HOMO-1→LUMO+1 (77%)
	S ₀ →S ₅	HOMO-2→LUMO (74%)

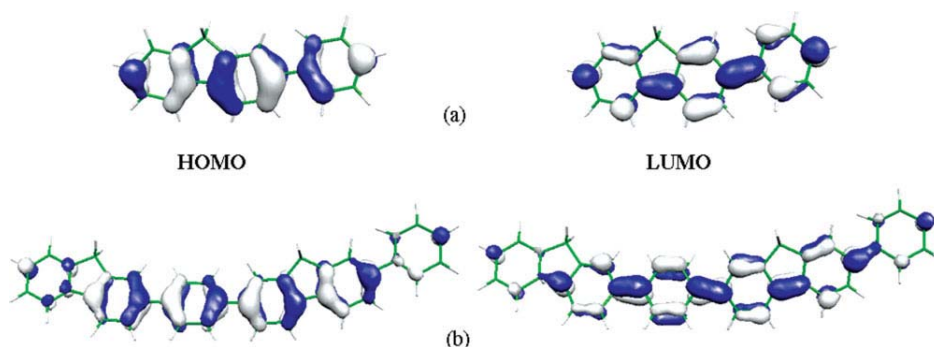


Figure 6. Plots of the HOMO to LUMO of (a) (FP) and (b) (FP)₂ performed by the B3LYP/SVP and TD-B3LYP/SVP approaches.

ure 7. The extrapolated fluorescence energies calculated by TD-B3LYP/SVP and TD-B3LYP/SVP+ methods are 2.27 and 2.26 eV, respectively. Linear dependence on the inverse number of repeating units are clearly found. The diffuse function leads to very small decreased in fluorescence energy. The basis set dependence of fluorescence energies is somewhat equal to the one found for vertical excitation energy (see earlier). In this work, a comparison of theoretical prediction and experimental results were also studied. It can be seen that fluorescence energies presented in Table 4 are lower than that of experimental results. However, the Stokes shift of 0.62 and 0.61 eV (from 2.27 and 2.26 eV, respectively), computed as the difference between absorption and fluorescence energies were observed. These values are higher than that of experimental values, 0.40 and 0.32 eV, respectively, (E_{excit} equals to 3.24 eV in film and 3.36 eV in CHCl₃ solution and E_{Flu} equals to 2.94 eV in film and 3.04 eV in CHCl₃ solution).³⁰

Finally, the corresponding fluorescence energies and oscillator strengths were used in the calculations of the radiative lifetimes according to the formula (in au).^{70,71}

$$\tau = \frac{c^3}{2(E_{\text{Flu}})^2 f} \quad (1)$$

where c is the velocity of light, E_{Flu} is the fluorescence transition energy, and f is oscillator strength. This formula is

achieved for investigating radiative lifetime of conjugated polymers.^{51–53,55} From the results, radiative lifetimes are summarized in Table 4. Extension of oligomer chain leads to a decrease of radiative lifetime. Lifetime extrapolated to $n \rightarrow \infty$ from TD-B3LYP/SVP and TD-B3LYP/SVP+ results are 0.55 and 0.51 ns, respectively. The difference between the two basis sets is only 0.04 eV. For the sake comparison, a chemical similar system, fluorescence lifetime in THF solution of poly(9,9-dihexylfluorene) is 0.46 ns.⁷² Radiative lifetime of poly(fluorene-pyridine) calculated by TD-B3LYP/SVP and TD-B3LYP/SVP+ are 0.25 and 0.38 ns, respectively.⁵¹ Radiative lifetime values can be used for more understand of emission and fluorescence energies. Moreover, fluorescence lifetime gives useful information in discrimination of particles.

Conclusions

A systematic theoretical study was performed on fluorene-phenylene oligomers for ground and excited states properties starting from monomer to tetramer. Full B3LYP and TD-B3LYP geometry optimizations on both ground and lowest excited states were performed. The behavior of torsion angles of monomer and dimer are similar that non-planar structures were obtained in ground state. In the first excited state, electronic excitation leads to a quinoid-like character, shortening of interring distances and

Table 4. Calculated Fluorescence Energies (Oscillator Strengths in Parentheses) and Radiative Lifetimes of FP Oligomers as Obtained from TD-B3LYP/SVP and TD-B3LYP/SVP+ Calculations.

	TD-B3LYP/SVP		TD-B3LYP/SVP+	
	Fluorescence energy (eV)	Lifetime (ns)	Fluorescence energy (eV)	Lifetime (ns)
$n = 1$	3.53 (1.049)	1.77	3.45 (1.036)	1.87
$n = 2$	2.84 (2.416)	1.18	2.79 (2.354)	1.26
$n = 3$	2.67 (3.351)	0.96	2.63 (3.280)	1.02
$n = 4$	2.63 (4.008)	0.83	2.60 (3.918)	0.87
$n = \infty$	2.27	0.55	2.26	0.51
Expt. ³⁰	2.94 (in film)			
	3.04 (in chloroform)			

Geometries were optimized at TD-B3LYP/SVP level.

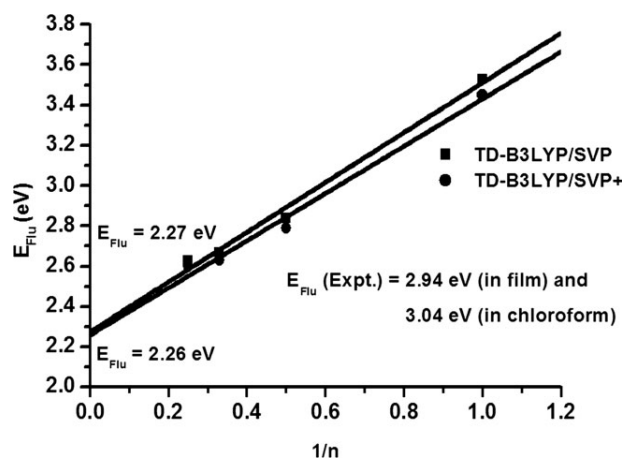


Figure 7. Dependence of fluorescence energies on the reciprocal chain length of $(\text{FP})_n$ as computed by TD-B3LYP/SVP (solid squares) and TD-B3LYP/SVP+ (solid circles) calculations.

lowering torsion angles. Vertical excitation from S_0 to S_1 corresponds to the HOMO-LUMO transition and possesses much higher oscillator strength. Vertical excitation energies and fluorescence energies indicate very good linear relation with the reciprocal chain length. The extrapolated vertical excitation energies are lower than the experimental results. The fluorescence energies are quite underestimated to those of the available experimental data. The computed radiative lifetimes are in the range of 0.51–0.55 ns. Our investigations have shown that detailed and reliable calculations on the properties of electronically ground and excited states of oligomer sizes relevant for comparison with experiment are possible nowadays. According to our study, we suggested that these studies should also be regarded as a first step toward the fluorescence spectra and radiative lifetime.

Acknowledgments

The authors acknowledge A. W. Sheppard and N. Sheppard for the reading of the manuscript.

References

- Burroughes, J. H.; Bradley, D. D. C.; Brown, A. R.; Marks, R. N.; MacKay, K.; Friend, R. H.; Burns, P. L.; Holmes, A. B. *Nature* 1990, 347, 539.
- Kraft, A.; Grimsdale, A. C.; Holmes, A. B. *Angew Chem Int Ed* 1998, 37, 402.
- Onoda, M. *J. Appl Phys* 1995, 78, 1327.
- Mitschke, U.; Bäuerle, P. *Mater Chem* 2000, 10, 1471.
- Kim, D. Y.; Cho, H. N.; Kim, C. Y. *Prog Polym Sci* 2000, 25, 1089.
- Leclerc, M. *J Polym Sci Part A: Polym Chem* 2001, 39, 2867.
- Beaupré, S.; Leclerc, M. *Adv Funct Mater* 2002, 12, 192.
- Sariciftci, N. S.; Braun, D.; Zhang, C.; Srdanov, V. I.; Heeger, A. J.; Stucky, G.; Wudl, F. *Appl Phys Lett* 1993, 62, 585.
- Brabec, C. J.; Sariciftci, N. S.; Hummelen, J. C. *Adv Funct Mater* 2001, 11, 15.
- Garnier, F.; Hajlaoui, R.; Yassar, A.; Srivastava, P. *Science* 1994, 265, 1684.
- Dimitrakopoulos, C. D.; Malenfant, P. R. L. *Adv Mater* 2002, 14, 99.
- Lovinger, A. J.; Rothberg, L. *J Mater Res* 1996, 11, 1581.
- Bao, Z.; Lovinger, A. J.; Brown, J. *J Am Chem Soc* 1998, 120, 207.
- Skotheim, T. A.; Elsenbaumer, R. L.; Reynolds, J. R. *Handbook of Conducting Polymers*, 2nd ed.; Marcel Dekker: New York, 1998.
- Gustafsson, G.; Cao, Y.; Treacy, G. M.; Klavetter, F.; Colaneri, N.; Heeger, A. J. *Nature* 1992, 357, 477.
- Burn, P. L.; Holmes, A. B.; Kraft, A.; Bradley, D. D. C.; Brown, A. R.; Friend, R. H.; Gymer, R. W. *Nature* 1992, 356, 47.
- Grell, M.; Bradley, D. D. C.; Inbasekaran, M.; Woo, E. P. *Adv Mater* 1997, 9, 798.
- Redecker, M.; Bradley, D. D. C.; Inbasekaran, M.; Woo, E. P. *Appl Phys Lett* 1998, 73, 1565.
- Janietz, S.; Bradley, D. D. C.; Grell, M.; Giebeler, C.; Inbasekaran, M.; Woo, E. P. *Appl Phys Lett* 1998, 73, 2453.
- Yu, W.; Cao, Y.; Pei, J.; Huang, W.; Heeger, A. J. *Appl Phys Lett* 1999, 75, 3270.
- Bernius, M. T.; Inbasekaran, M.; O'Brien, J.; Wu, W. S. *Adv Mater* 2000, 12, 1737.
- Weinfurter, K.; Fujikawa, H.; Tokito, S.; Taga, Y. *Appl Phys Lett* 2000, 76, 2502.
- Virgili, T.; Lidzey, D. G.; Bradley, D. D. C. *Adv Mater* 2000, 12, 58.
- Scherf, U.; List, E. J. W. *Adv Mater* 2002, 14, 477.
- Campbell, A. J.; Bradley, D. D. C.; Antoniadis, H. *J Appl Phys* 2001, 89, 3343.
- Pei, Q.; Yang, Y. *J Am Chem Soc* 1996, 118, 7416.
- Grice, A. W.; Bradley, D. D. C.; Bernius, M. T.; Inbasekaran, M.; Wu, W. W.; Woo, E. P. *Appl Phys Lett* 1998, 73, 629.
- Lee, J. L.; Lee, V. Y.; Miller, R. D. *ETRI J* 2002, 24, 409.
- Klämer, G.; Davey, M. H.; Lee, J. I.; Miller, R. D. *Adv Mater* 1999, 11, 115.
- Liu, B.; Yu, W. L.; Lai, Y. H.; Huang, W. *Chem Mater* 2001, 13, 1984.
- Liu, B.; Yu, W. L.; Pei, J.; Liu, S. Y.; Lai, Y. H.; Huang, W. *Macromolecules* 2001, 34, 7932.
- Liu, S. P.; Chan, H. S. O.; Ng, S. C. *J Polym Sci Part A: Polym Chem* 2004, 42, 4792.
- Aubert, P. H.; Knipper, M.; Grommelaar, L.; Lutsen, L.; Manca, J.; Vanderzande, D. *Macromolecules* 2004, 37, 4087.
- Liu, S. P.; Ng, S. C.; Chan, H. S. O. *Synth Met* 2005, 149, 1.
- Yang, L.; Feng, J. F.; Ren, A. M.; Sun, J. Z. *Polymer* 2005, 47, 1397.
- Zeng, G.; Yu, W. L.; Chua, S. J.; Huang, W. *Macromolecules* 2002, 35, 6907.
- Donat-Bouillud, A.; Lévesque, I.; Tao, Y.; D'Iorio, M. *Chem Mater* 2000, 12, 1931.
- Kreyenschmidt, M.; Klaerner, G.; Fuhrer, T.; Ashenurst, J.; Karg, S.; Chen, W. D.; Lee, V. Y.; Scott, J. C.; Miller, R. D. *Macromolecules* 1998, 31, 1099.
- Ranger, M.; Rondeau, D.; Leclerc, M. *Macromolecules* 1997, 30, 7686.
- Belletête, M.; Beaupré, S.; Bouchard, J.; Blondin, P.; Leclerc, M.; Durocher, G. *J Phys Chem B* 2000, 104, 9118.
- Liu, B.; Yu, W. L.; Lai, Y. H.; Huang, W. *Macromolecules* 2000, 32, 8945.
- Alguino, A. C.; Chung, W. C.; Bantaculo, R. V.; Vequizo, R. M.; Miyata, H.; Ignacio, E. W.; Bacala, A. M. *AM NECTEC Tech J* 2000, II(a), 218.

43. Belletête, M.; Morin, J. F.; Beaupré, S.; Leclerc, M.; Durocher, G. *Synth Met* 2002, 126, 43.
44. Zeng, G.; Chua, S. J.; Huang, W. *Thin Solid Films* 2002, 417, 194.
45. Tirapattur, S.; Belletête, M.; Leclerc, M.; Durocher, G. *J Mol Struct (THEOCHEM)* 2003, 625, 141.
46. Gong, Z.; Lagowski, J. B. *J Mol Struct (THEOCHEM)* 2005, 729, 211.
47. Yang, L.; Feng, J. K.; Liao, Y.; Ren, A. M. *J Phys Chem A* 2005, 109, 7764.
48. Belletête, M.; Bédard, M.; Leclerc, M.; Durocher, G. *J Mol Struct (THEOCHEM)* 2004, 679, 9.
49. Yang, L.; Feng, J. K.; Ren, A. M. *Polymer* 2005, 46, 10970.
50. Yang, L.; Liao, Y.; Feng, J. K.; Ren, A. M. *J Mol Struct (THEOCHEM)* 2006, 758, 29.
51. Chidthong, R.; Hannongbua, S.; Aquino, A.; Wolschann, P.; Lischka, H. *J Comput Chem* 2007, 28, 1735.
52. Lukeš, V.; Aquino, A.; Lischka, H. *J Phys Chem A* 2005, 109, 10232.
53. Lukeš, V.; Aquino, A.; Lischka, H.; Kauffmann, H. F. *J Phys Chem B* 2007, 111, 7954.
54. Lukeš, V.; Šolc, R.; Sperling, J.; Kauffmann, H. F. *Chem Phys* 2008, 349, 226.
55. Meeto, W.; Suramitr, S.; Vannarat, S.; Hannongbua, S. *Chem Phys* 2008, 349, 1.
56. Schäfer, A.; Huber, Ch.; Ahlrichs, R. *J Chem Phys* 1994, 100, 5829.
57. Becke, A. D. *Phys Rev A* 1988, 38, 3098.
58. Lee, C.; Yang, W.; Parr, R. G. *Phys Rev B* 1988, 37, 785.
59. Miehlisch, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem Phys Lett* 1989, 157, 200.
60. Sriwichitkamol, K.; Suramitr, S.; Poolmee, P.; Hannongbua, S. *J Theor Comput Chem* 2006, 5, 1.
61. Schäfer, A.; Horn, H.; Ahlrichs, R. *J Chem Phys* 1992, 97, 2571.
62. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem Phys Lett* 1989, 162, 165.
63. Yang, L.; Feng, J. K.; Liao, Y.; Ren, A. M. *Opt Mat* 2007, 29, 642.
64. Suramitr, S.; Hannongbua, S.; Wolschann, P. *J Mol Struct (THEOCHEM)* 2007, 807, 109.
65. Gong, Z.; Lagowski, J. B. *J Mol Struct (THEOCHEM)* 2008, 866, 27.
66. Puschning, P.; Ambrosch-Draxl, C.; Heime, G.; Zofer, E.; Resel, R.; Leising, G.; Kriechbaum, M.; Graupner, W. *Synth Met* 2001, 116, 327.
67. Brière, J. F.; Côté, M. *J Phys Chem B* 2004, 108, 3123.
68. Ma, J.; Li, S.; Jiang, Y. *Macromolecules* 2002, 35, 1109.
69. Poolmee, P.; Hannongbua, S. *J Theor Comput Chem* 2006, 3, 481.
70. Brandsen, B. H.; Joachain, C. J. *Physics of Atoms and Molecules*; Longman: London, 1983.
71. Litani-Barzilai, I.; Bulatov, V.; Schechter, I. *Anal Chem Acta* 2004, 501, 151.
72. Zhou, X. H.; Zhang, Y.; Xie, Y. Q.; Cao, Y.; Pei, J. *Macromolecules* 2002, 39, 3830.



Original article

Investigating the structural basis of arylamides to improve potency against *M. tuberculosis* strain through molecular dynamics simulationsAuradee Punkvang^a, Patchreenart Saparpakorn^b, Supa Hannongbua^b, Peter Wolschann^c, Anton Beyer^c, Pornpan Pungpo^{a,*}^a Department of Chemistry, Ubon Ratchathani University, 85 Sthollmark Road, Warinchamrap, Ubonratchathani 34190, Thailand^b Department of Chemistry, Kasetsart University, Chatuchak, Bangkok 10900, Thailand^c Institute for Theoretical Chemistry, University of Vienna, A-1090 Vienna, Austria

ARTICLE INFO

Article history:

Received 11 May 2010

Received in revised form

13 August 2010

Accepted 4 September 2010

Available online 17 September 2010

Keywords:

Molecular dynamics simulations

Arylamide

InhA

M. tuberculosis

ABSTRACT

Arylamides have been identified as direct InhA inhibitors which overcome the drug-resistance problem of isoniazid, the first-line drug for tuberculosis treatment. However, arylamide properties are not yet optimal against *Mycobacterium tuberculosis*. Arylamides show high potency in InhA enzyme assay, but they fail in antimycobacterial assay. To achieve the structural basis to improve antimycobacterial activity, the dynamic behavior of arylamide inhibitors and a substrate, *trans*-2-hexadecenoyl-(*N*-acetylcysteine)-thioester, were carried out by molecular dynamics (MD) simulations. Arylamide inhibitors and a substrate are positioned at the same site which indicates the competitive inhibitor function of arylamides. Based on our findings, the amide carbonyl oxygen causes the selectivity of arylamide inhibitors for InhA inhibition. Moreover, this moiety is crucial for the affinity of the arylamide–InhA interactions with Tyr158 and NADH to form hydrogen bonds. It is possible to enhance the selectivity of arylamide inhibitors to reach the InhA target by introducing a hydrophilic substituent into the aryl ring A. In order to increase the membrane permeability of arylamide inhibitors, more lipophilic properties should be incorporated into the substituent B. Therefore, based on the obtained results, the correct balance between the selectivity and the membrane permeability of arylamide inhibitors should improve their inhibitory activity against *M. tuberculosis* strain.

© 2010 Elsevier Masson SAS. All rights reserved.

1. Introduction

Multidrug resistant-tuberculosis (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB) do not respond to the standard treatment with first-line anti-TB drugs, which makes tuberculosis treatment complicated and expensive [1]. Accordingly, to address these problems the research on novel and more potent drug candidates is very important. The enoyl-acyl-ACP reductase (InhA) catalyzing the NADH-specific reduction of 2-*trans*-enoyl-ACP [2] is an attractive target for designing novel antibacterial agents [3–8]. InhA has been identified as the primary target of isoniazid (INH), one of the most effective first-line anti-TB drugs [9–14]. As a prodrug, INH must first be activated by catalase-peroxidase (KatG) to generate the reactive acyl radical [15–21]. Thereupon, the reactive species binds covalently to nicotinamide adenine dinucleotide (NAD⁺) to form the active adduct (INH-NAD) that functions as a highly potent inhibitor of InhA [22,23]. However,

the high potency of INH for tuberculosis treatment is diminished by drug resistance. The INH resistance is related to the mutation in several genes of *Mycobacterium tuberculosis* involving *inhA*, *ahpC*, *kasA*, *katG* and *ndh* [24–30]. Commonly found in *M. tuberculosis* clinical isolates, high levels of INH resistance arise from mutations in *katG* [31,32]. Thus, to overcome the INH resistance associated with mutations in the KatG enzyme, compounds which directly inhibit the InhA enzyme without requiring activation by KatG are to be seen as very promising new agents against tuberculosis. Many compounds functioning as direct InhA inhibitors have been discovered and identified [33–37]. A series of arylamides is one of the novel classes of potent InhA inhibitors that circumvent the resistance mechanism to INH prodrug [38]. Arylamides show high potency for inhibiting the InhA enzyme. The highest InhA inhibitory activity with IC₅₀ of 0.09 μM could be observed. On the other hand it has to be taken into account that the majority of arylamides exhibits a lower *M. tuberculosis* growth inhibition with the minimum inhibitory concentrations (MIC) against *M. tuberculosis* strain above 125 μM. However, it can be reasonably assumed that these compounds are extruded from the bacterial cell by efflux pumps. The above given data, especially the direct InhA inhibitor

* Corresponding author. Tel.: +66453534014124; fax: +6645288379.
E-mail address: pornpan_ubu@yahoo.com (P. Pungpo).

property of arylamides justify a more detailed examination of the structural basis to improve antimycobacterial activity.

MD simulations have been widely used for inhibitor–enzyme complexes in order to investigate the structural features and the dynamic behavior providing detailed information about flexibility, conformation and inhibitor–enzyme interactions [39–47]. In the present study, MD simulations have been performed for arylamide–InhA complexes to gain an insight into their structural and dynamic features. Moreover, to compare the dynamic behavior of the interaction between arylamide inhibitors and substrate, MD simulations of a substrate, *trans*-2-hexadecenoyl-(*N*-acetylcysteamine)-thioester, were also carried out. The binding free energies of arylamide–InhA complexes were calculated to gain quantitative insights into the binding affinity of arylamide inhibitors in the InhA binding site. In addition, the binding free energy of the substrate was also calculated to be compared with that of arylamide inhibitors. The interaction energies between arylamide inhibitors and particular amino acids were also evaluated to explore the key residues that are crucial for the binding affinity of arylamide–InhA complexes.

2. Materials and methods

2.1. Starting structures for MD simulations

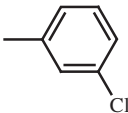
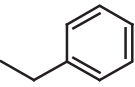
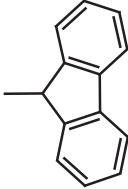
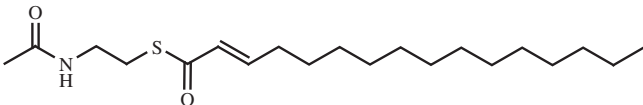
Two X-ray crystal structures of arylamide B3–NADH–InhA and the substrate–NADH–InhA were taken from the Protein Data Bank

with pdb codes of 2NSD [38] and 1BVR [48], respectively. Because of the homotetramer characteristic of InhA [48], only chain A of both X-ray crystal structures served as the initial coordinates for MD simulations. In the case of arylamide A10 and P2 bound complexes, their starting structures were taken from molecular docking calculations using Autodock 3.05 [49]. In this work, the chemical structures and InhA inhibitory activities (the half maximal inhibitory concentration, IC_{50}) of arylamides taken from literature [38] are shown in Table 1.

2.2. Molecular docking calculations

The X-ray crystal structure of arylamide B3 complexed with InhA (pdb code 2NSD) was employed for molecular docking calculations. Docking calculations of arylamide inhibitors were carried out by the Autodock 3.05 program using Lamarckian Genetic Algorithm (LGA) [49]. All hydrogen atoms and Kollman charges were added to the protein using AutoDockTools. Solvation parameters were generated for the protein using the Addsol utility of Autodock. The grid maps representing the protein in the actual docking process were calculated with Autogrid. The dimensions of the grids were $60 \times 60 \times 60$ points with a spacing of $0.375 \text{ \AA}$ and the center close to the ligand. Docking parameters were used as default values, except for the number of docking runs which was set to 50. Arylamide B3 was docked back into the InhA binding pocket to validate the docking method, and subsequently the other

Table 1
The chemical structures and InhA inhibitory activities of arylamides and the chemical structure of substrate.

Compound	X	R ₁	B	IC ₅₀ (μM) ^a
A10	N	4-t-Bu		>100
B3	C	4-CH ₃		5.16
P2	N	H		0.09
Substrate				

^a IC_{50} of all inhibitors were taken from literature [36]. The nomenclature of the inhibitors was taken from the original literature and is also given in Table 1.

compounds were docked. The ligand pose with the lowest final docked energy and the greatest number of members in the cluster was selected as the best binding mode of arylamide inhibitors in the InhA binding pocket.

2.3. Molecular dynamics simulations

Four MD simulations for the substrate-bound complex (substrate-NADH-InhA) and the arylamide inhibitor-bound complexes (A10-NADH-InhA, B3-NADH-InhA and P2-NADH-InhA) were carried out using GROMACS 4.0.4 [50–52] with GROMOS96 43a2 force field [53]. AnteChamber PYthon Parser interface (ACPYPE) tool [54] was employed to generate parameters for the MD force field of NADH, the substrate and arylamide inhibitors. Each of the starting complexes was inserted and centered into the cubic periodic box of SPC216 water molecules [55]. A cubic box extended at 3 nm from the starting complexes was set for all simulations. Each simulation system was neutralized by four Na^+ counterions. To remove bad contacts before MD simulations, an energy minimization with all bonds constrained using the steepest descent algorithm (2000 steps) was performed for each simulation system. Afterwards, a 500 ps simulation of each system was performed in the NVT ensemble at 300 K with a time step of 0.002 ps under the conditions of position restraints and LINCS constraints [56]. Finally, 6 ns MD simulations without the position restraints were performed under the same conditions. Long-range electrostatic interactions were evaluated by the particle-mesh Ewald method [57]. Coulomb and van der Waals interactions were cut off at 1.0 nm. The last 1 ns simulation of each system was selected for detailed analysis. The crude average structure was refined using the steepest descent and conjugate gradient minimization.

2.4. Calculations of binding free energies

The linear interaction energy (LIE) method [58] was chosen to calculate the binding free energies for the substrate and arylamide inhibitors in the InhA enzyme. To calculate the binding free energy with LIE method, MD simulations of the substrate and arylamide

inhibitors in water were also performed under the same conditions as described above. The binding free energy based on LIE method can be expressed by the following equation;

$$\Delta G_{\text{bind}} = \alpha [(V_{\text{LJ}})_{\text{bound}} - (V_{\text{LJ}})_{\text{free}}] + \beta [(V_{\text{CL}})_{\text{bound}} - (V_{\text{CL}})_{\text{free}}] \quad (1)$$

Where $(V_{\text{LJ}})_{\text{bound}}$ = average Lennard–Jones energy for ligand/solvent interaction; $(V_{\text{LJ}})_{\text{free}}$ = average Lennard–Jones energy for ligand/water interaction; $(V_{\text{CL}})_{\text{bound}}$ = average electrostatic energy for ligand/solvent interaction; $(V_{\text{CL}})_{\text{free}}$ = average electrostatic energy for ligand/water interaction; α , β = scaling factors with $\alpha = 0.18$ and $\beta = 0.50$.

It is noted that ligand/solvent interaction denotes interaction of inhibitor with receptor, cofactor and waters.

3. Results and discussion

3.1. Structural stability and flexibility during MD simulations

A molecular dynamics simulation of arylamide B3 in InhA was performed. In order to compare the binding behavior of arylamides that show the highest and lowest InhA inhibitory activities, arylamides P2 and A10 in a InhA binding pocket were selected for MD simulations. Moreover, to compare the binding affinities of the substrate, *trans*-2-hexadecenoyl-(*N*-acetylcysteamine)-thioester, and of the arylamide inhibitors which are competitive inhibitors, a MD simulation of the substrate was carried out. The root mean square deviations (RMSD) as a function of the simulation time of each complex with respect to the starting structure were analyzed as shown in Fig. 1. RMSDs of all atoms of InhA in four complexes, InhA/NADH/substrate, InhA/NADH/B3, InhA/NADH/P2 and InhA/NADH/A10, reach the plateau characteristic at 0.5 ns, 1.5 ns, 1.0 ns and 2.0 ns, respectively. These results indicate that 6 ns unrestrained simulation is enough for stabilizing the fully relaxed systems.

The root mean square fluctuation (RMSF) of residues around the ligand-binding site of InhA (residues 96–104, 149–165, 192–223) was calculated to reveal the mobile flexibility of these residues.

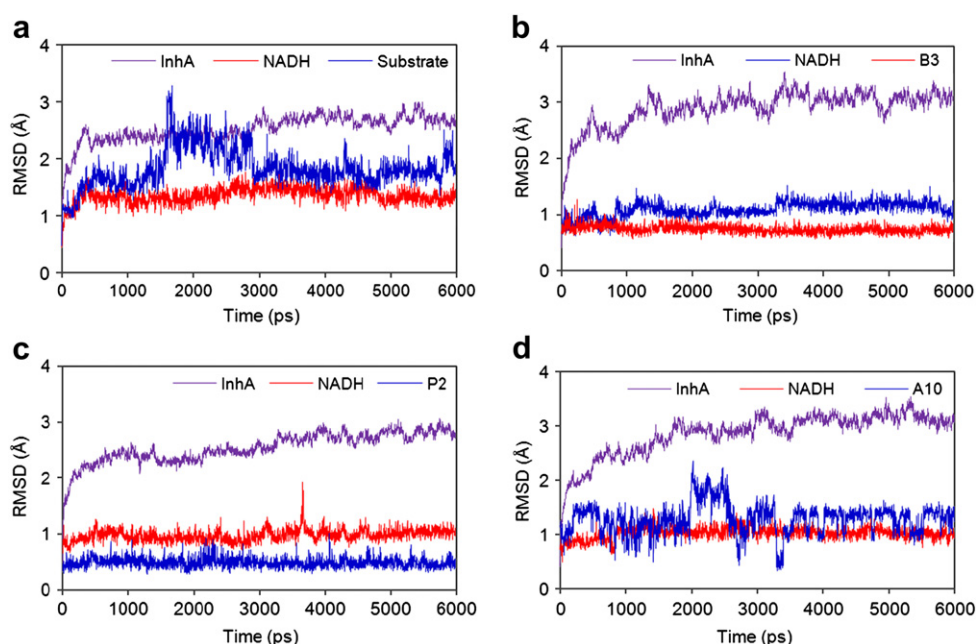


Fig. 1. RMSD of all atoms of each molecule in complexes of (a) InhA/NADH/substrate, (b) InhA/NADH/B3, (c) InhA/NADH/P2, and (d) InhA/NADH/A10.

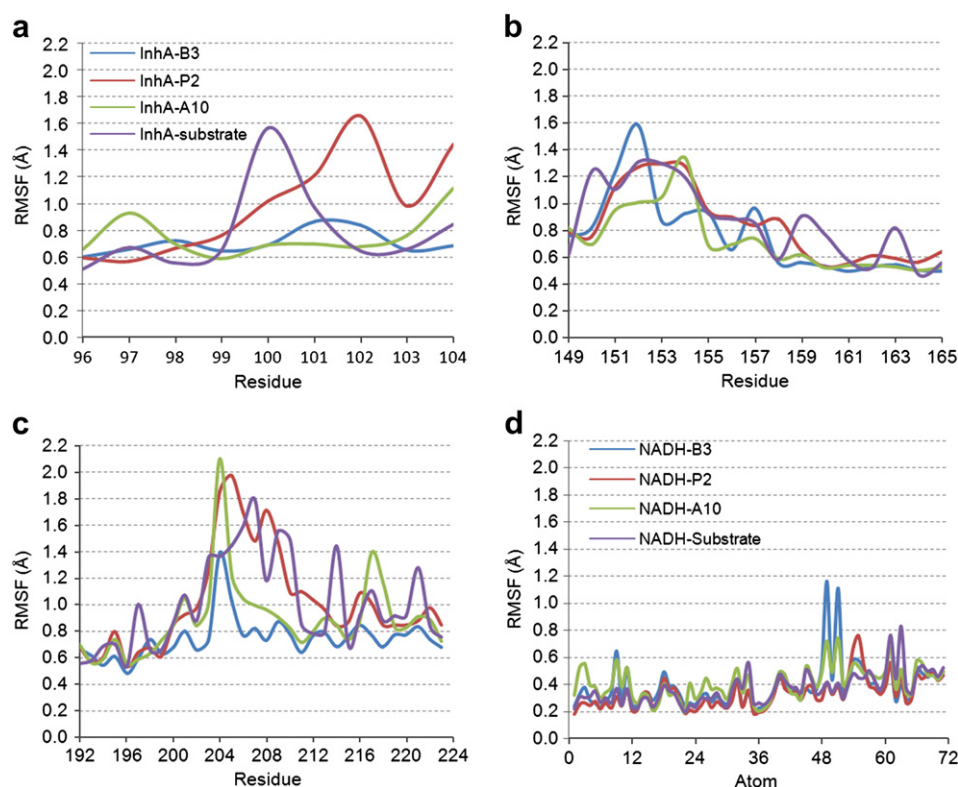


Fig. 2. RMSDF of residues (a) 96–104, (b) 149–165, (c) 192–223 and (d) NADH in four complexes.

RMSF of these residues are shown in Fig. 2(a)–(c). RMSF of all atoms of NADH that directly interacts with inhibitors and the substrate were also calculated as shown in Fig. 2(d). Fig. 2(a)–(c) clearly depict different flexibilities in the binding site of InhA when the substrate and arylamides B3, P2, A10 are bound to their binding sites. All residues of the InhA binding site that bind with the substrate and arylamide inhibitors show a small degree of flexibility with the RMSF less than 2.2 Å. Particularly, residues 96–99, 155–165 and 192–200 fluctuate less than 1 Å indicating that these residues seem to be rigid. These results indicate that the residues are not flexible enough to bind the substrate or arylamide inhibitors. Most of these rigid residues are located around the binding pocket of the aryl ring A of the arylamide inhibitors as shown in Fig. 3. In addition, NADH cofactor is also located adjacent to the aryl ring A of arylamide inhibitors as shown in Fig. 3. The dynamics flexibility of NADH co-binding with a substrate and arylamides B3, P2, A10 is shown in Fig. 2(d). All atoms of NADH seem to be rigid with a RMSF value less than 0.8 Å, except atoms 49 and 51 of NADH-B3 (hydrogen atoms at hydroxyl group of adenine ribose). Thus, the pocket size for the binding of the aryl ring A is limited by NADH cofactor and rigid residues surrounding this pocket. With regard to the substituent size on the aryl ring A of arylamides B3, P2, A10, arylamide P2, bearing the small substituent on the aryl ring A, shows an inhibitory activity better than arylamides B3 and A10 bearing the bulky substituents on the aryl ring A. These results suggest that the aryl ring A of arylamides should not contain too large substituents.

As illustrated in Fig. 2(a)–(c), RMSF of residues 100–104, 149–154 and 201–223 fluctuate in each complex in a wide range from 0.6 Å to 2.1 Å. These results indicate that the residues could be flexible enough for the binding of the substrate and arylamides B3, P2, A10. Most of these flexible residues colored by orange are located around the binding pocket of the substituent B of arylamide

inhibitors as shown in Fig. 3. Focusing on the substituent B of arylamides, arylamide P2 bearing the large substituent of the fluorene ring shows the inhibitory activity better than arylamides B3 and A10 bearing the smaller substituent of the phenyl ring. These results imply that the bulky substituent could be incorporated into the substituent B of the arylamide inhibitors and seems to be favorable for inhibitory activity. However, the substituent B should not be too large because some rigid residues (residues 155–158) are also located near this substituent as shown in Fig. 3.

3.2. Structural change of InhA bound to the inhibitor and substrate

To study the structural change of InhA when the inhibitor and the substrate are bound, the superimposition of two X-ray crystal structures of arylamide B3 bound InhA (pdb code 2NSD) [38] and the substrate bound InhA (pdb code 1BVR) [48] was performed. The conformational differences of residues 198–223 including two α -helices and one loop are shown in Fig. 4. The differences may be affected by new positional rearrangements of these residues to accommodate the binding of the substrate and arylamide inhibitors. Therefore, the results derived from the superimposition of the experimental data [38,48] and the MD simulations as described above are consistent. The results obtained from both approaches indicate the flexibility of these residues for the binding of each ligand. NADH in two complexes is held at the same position implying that ligand binding has no effect on the binding of NADH cofactor. Arylamide B3 is bound in the same binding site with the substrate above the binding site of NADH. The obtained results clearly indicate that arylamides function as competitive inhibitors of InhA. This is supported by the kinetic study of arylamide proving that arylamide is competitive with the enoyl-CoA substrate [59].

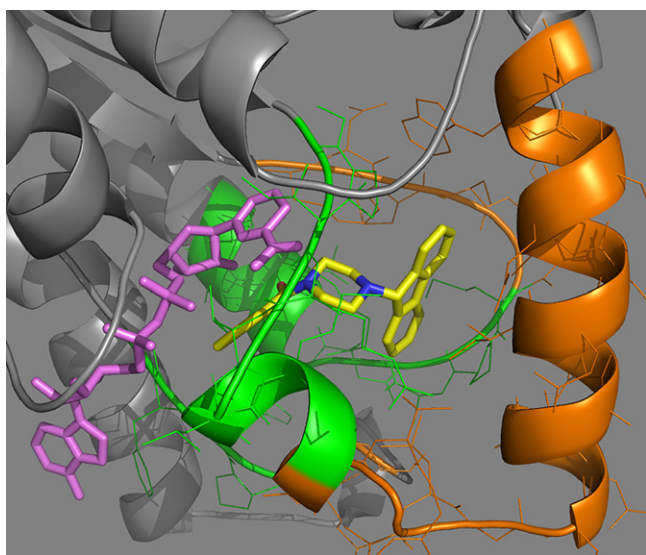


Fig. 3. The residues around the ligand-binding site of InhA (residues 96–104, 149–165, 192–223). The rigid residues colored by green, the flexible residues colored by orange, NADH colored by purple and compound P2 colored by yellow.

3.3. Ligand–InhA interactions

The competitive inhibitor function of arylamides should mimic the behavior of the substrate. When comparing structures of arylamide inhibitors and the substrate, the carbonyl moiety was found to be the only similar part. To compare the binding modes of the arylamide inhibitors and the substrate in an InhA binding pocket, the X-ray crystal structure of arylamide B3 was superimposed on the X-ray crystal structure of the substrate as shown in Fig. 5. The amide carbonyl oxygen of arylamide B3 and the thioester carbonyl

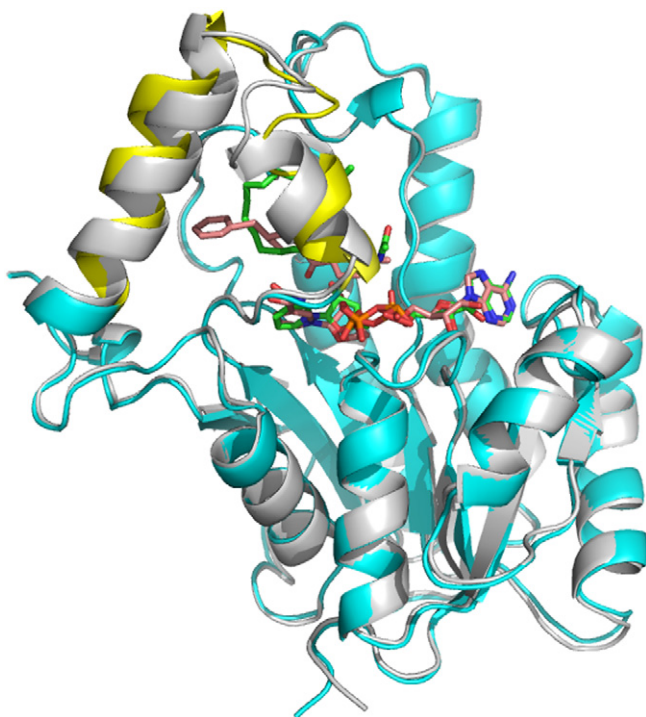


Fig. 4. Superimposition of InhA/B3/NADH of pdb code 2NSD (cyan and yellow), B3 and NADH labeled by orange. InhA/substrate/NADH of pdb code 1BVR (grey), substrate and NADH labeled by green.

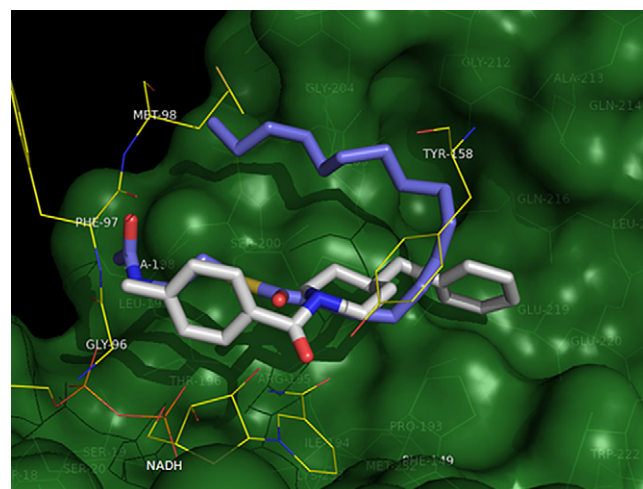


Fig. 5. Superimposition of X-ray crystal structures of arylamide B3 (grey) and substrate (purple) in the InhA binding pocket.

oxygen of the substrate are positioned at the same place between two hydroxyl groups of Tyr158 and nicotinamide ribose of NADH. Therefore, the amide carbonyl oxygen is the selective part of arylamide inhibitors to reach the InhA binding pocket and to act as the competitive inhibitor. MD results are consistent with the experimental data which show the same binding site of the arylamide inhibitor and the substrate. The obtained results from MD simulations reveal that the amide carbonyl oxygen of arylamide B3 could form two hydrogen bonds with Tyr158 and NADH with distances of 1.68 Å and 1.76 Å, respectively. In the case of the substrate, the thioester carbonyl oxygen could not form hydrogen bonds with Tyr158 and NADH. However, this moiety is held in the same position with that of arylamide B3 by the hydrogen bond with a NH₃ sidechain of Lys165.

The aryl ring A of arylamide B3 lies in the same site as the acetylcysteine moiety of the substrate. This binding site is surrounded by hydrophilic groups of pyrophosphate, nicotinamide ribose and amino acid backbones of Gly96, Phe97, Met98 and Ala198 as shown in Fig. 5. Thus, this binding site seems to be favored for the hydrophilic substituent. From MD simulations, NH and the acetyl oxygen of the acetylcysteine moiety of the substrate form hydrogen bonds with the pyrophosphate oxygen of NADH and the NH backbone of Ala198 with the distance of 2.26 Å and 1.96 Å, respectively. On the other hand, the hydrogen bond interaction could not be found in the aryl ring A of arylamide B3. Besides the carbonyl moiety, it is possible to increase the selectivity of competitive arylamide inhibitors by introducing NH moiety and acetyl oxygen into the aryl ring A of arylamide inhibitors.

To support the influence of the NH moiety at the aryl ring A on the selectivity of competitive arylamide inhibitors, arylamides P4 and P6 were taken into account [38]. InhA inhibition activities of arylamides P4 and P6 are comparable with IC₅₀ of 1.04 μM and 2.04 μM, respectively. Interestingly, arylamide P4 having aryl ring A substituted by indolyl group represents the best activity against the *M. tuberculosis* strain with the minimum inhibitory concentrations (MIC) of 62.5 μM. On the other hand, arylamide P6 having aryl ring A substituted by a phenyl group shows higher MIC of 125 μM. The results indicate that a NH moiety at the aryl ring A could increase the selectivity of competitive arylamide inhibitors to reach the InhA target leading to increase the MIC value. To clarify this fact in detail, arylamide P3 where an indolyl group is located at the aryl ring A [38] was also taken into account. Because of the availability of the X-ray structure of arylamide P3 [59], a superimposition of the X-ray crystal structure of P3 bound InhA (pdb code 1P44) and the X-ray

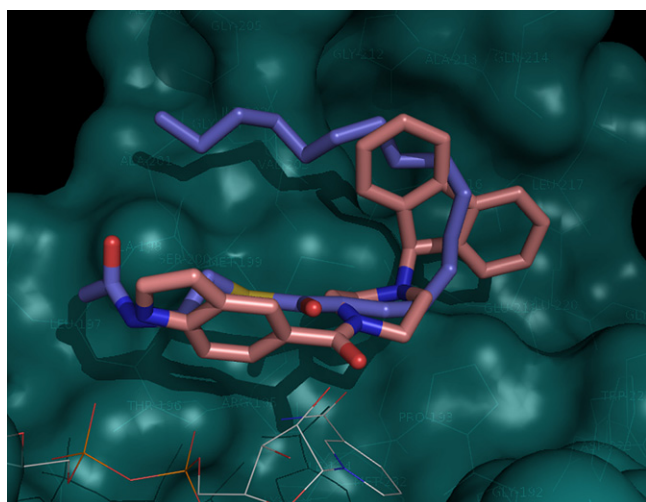


Fig. 6. Superimposition of X-ray crystal structures of arylamide P3 (pink) and substrate (purple) in the InhA binding pocket.

crystal structure of substrate-bound InhA (pdb code 1BVR) was performed (Fig. 6). It was found that NH moiety of arylamide P3 is located at the same position as the NH moiety of the substrate confirming the role of the NH moiety in the aryl ring A of arylamide inhibitors. However, arylamide P3 possessing the lower MIC maybe affected by its poor membrane permeability [38].

The substituent B of arylamide B3 lies in the same site as the fatty acyl chain of the substrate surrounded by the hydrophobic residues 100–104, 149–154 and 201–223 as shown in Fig. 5. This result reveals that the lipophilic substituent is favored for this binding site. Therefore, a greater lipophilicity of the substituent B as well as of the fatty acyl chain of the substrate should enhance the binding of arylamide inhibitors. Moreover, it should increase the membrane permeability of arylamide inhibitors.

For a better understanding of the binding behaviors of arylamide inhibitors in InhA binding pocket, MD simulations of arylamides P2 and A10 were performed. The binding mode of arylamide P2, with the highest InhA inhibitory activity, is shown in Fig. 7. This arylamide is bound to the InhA binding pocket in the same fashion as in the X-ray crystal structure of arylamide B3. The amide carbonyl oxygen of arylamide P2 is held by two hydrogen bonds of Tyr158 and NADH with distances of 1.97 and 1.69 Å, respectively.

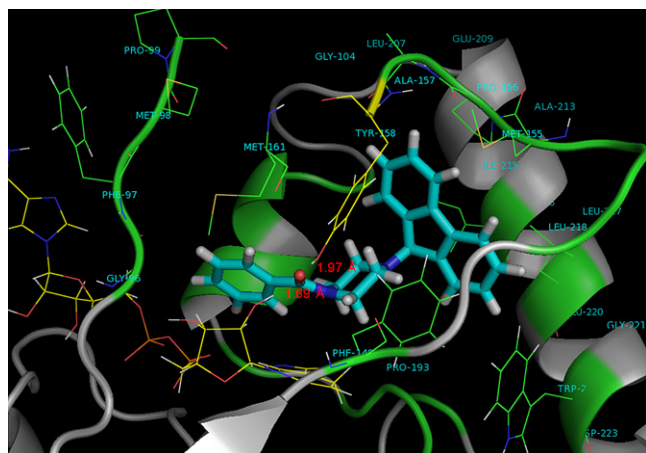


Fig. 7. The binding mode of arylamide P2 in the InhA binding pocket obtained from MD simulation.

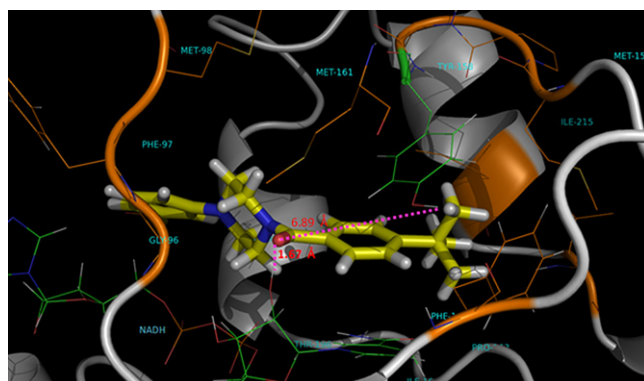


Fig. 8. The binding mode of arylamide A10 in the InhA binding pocket obtained from MD simulation.

The aryl ring A is placed in the hydrophilic pocket, whereas the substituent B, the fluorene ring, is located in the hydrophobic pocket. In the case of the compound with the lowest InhA inhibitory activity, arylamide A10, its binding mode is very different from those of arylamides B3 and P2 as presented in Fig. 8. The aryl ring A of arylamide A10 is switched to bind to the hydrophobic pocket of substituent B. This result clearly shows that the aryl ring A with the bulky substituent as well as the t-butyl group cannot be occupied in the hydrophilic pocket because of the rigidity of this binding site as previously explained. Resulting from the conformational change, the amide carbonyl oxygen of arylamide A10 loses a crucial hydrogen bond contact with Tyr158. Only the hydrogen bond with the hydroxyl group of nicotinamide ribose (1.67 Å) could be observed during the MD simulation as shown in Fig. 8. To confirm the influence of the bulky substituent of the aryl ring A on the InhA inhibitory activity, arylamide A9 [38], one of the lowest active compounds was selected for further studies. Based on the obtained MD simulation results, arylamide A9 is bound to the InhA binding pocket in the same way as arylamides B3 and P2 (Fig. 9). In contrast to the MD simulation structure of arylamide A10 (Fig. 8), the aryl ring A of arylamide A9 could be located in the hydrophilic pocket. However, the iso-propyl substituent of the aryl ring A causes the loss of a crucial hydrogen bond with the hydroxyl group of nicotinamide ribose as shown in Fig. 9. It is obvious, that the bulky substituent of the aryl ring A is not favorable for InhA binding of arylamide inhibitors. These obtained results could successfully

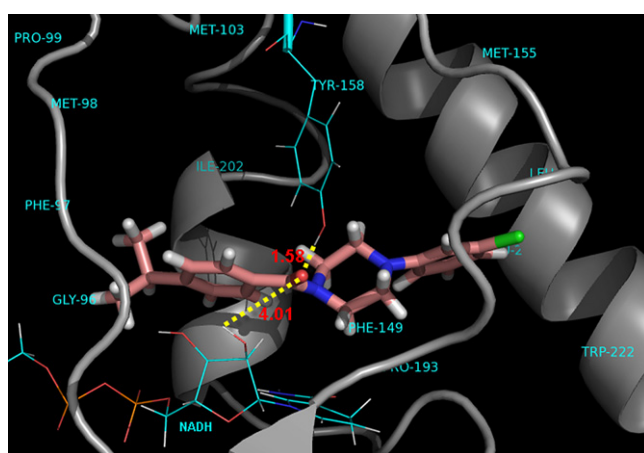


Figure 9. The binding mode of arylamide A9 in the InhA binding pocket obtained from MD simulation.

confirm the structural requirements of the substituent of the aryl ring A of arylamide inhibitors.

3.4. Interaction energy

The interaction energies between arylamide inhibitors and particular amino acids within 6 Å of each arylamide inhibitor were calculated to explore the contribution of each residue. Arylamides P2 and B3 show the highest attractive interaction energies with NADH and Tyr158 as shown in Table 2. These high attractive energies are consistent with the two strong hydrogen bonds observed between the arylamide inhibitors and NADH and Tyr158. In the case of arylamide A10, the interaction energies with Tyr158 and NADH are related to results obtained from MD simulations. A lower attractive interaction energy with Tyr158 can be found because of the loss of the hydrogen bond, whereas the interaction energy with NADH is still high. These results reveal that two hydrogen bonds are essential for the affinity of the arylamide inhibitors in the InhA binding pocket. The substituent B of arylamide P2 interacts with hydrophobic residues more than those of arylamides B3 and A10 as shown in Table 2. Therefore, more interactions with hydrophobic residues of the substituent B could enhance the binding of arylamides in InhA which is clearly shown by the InhA inhibition activity of arylamides P2, B3 and A10.

3.5. The binding free energy

To gain quantitative insights into the affinity for binding of arylamide inhibitors in the InhA binding site, the binding free energies of arylamide–InhA complexes were calculated by the LIE method. In addition, the binding free energy of the substrate is also calculated to compare the binding affinities of the substrate and the competitive arylamide inhibitors. Only the binding free energy of

Table 2
Average interaction energies between arylamide inhibitors and the particular residue in InhA binding pocket.

The inhibitor portion	Residue	Interaction energy (kJ/mol)		
		A10	B3	P2
The aryl ring A	Gly96	−12.56	−6.49	−7.43
	Phe97	−3.72	−5.17	−1.67
	Met98	−1.75	−2.64	−0.62
	Met103	−12.69	−15.15	−14.18
	Met161	−14.54	−10.20	−6.02
	Lys165	−5.64	—	−3.69
	Leu197	—	—	−5.76
	Ala198	−17.40	−8.07	−21.05
	Ala201	−2.82	—	−1.00
	Ile202	−10.30	−13.69	−1.86
	NADH	−149.46	−98.69	−104.95
	Tyr158	−27.21	−68.36	−57.51
The amide carbonyl	Phe149	−9.89	−20.94	−32.79
	Ser152	—	−0.04	−6.31
The substituent B	Arg153	—	−2.23	−0.32
	Ala154	—	−8.14	−2.94
	Met155	−7.09	−11.80	−7.27
	Pro156	—	−11.75	−21.27
	Ala157	—	−2.76	−11.32
	Gly192	−3.52	—	−0.66
	Pro193	−10.37	−8.92	−10.67
	Ile194	−10.88	—	−2.09
	Thr196	−3.62	−2.53	−2.87
	Met199	−21.86	−20.46	−0.48
	Val203	—	−7.97	—
	Ile215	−11.16	−5.35	−17.05
	Leu218	—	−4.08	−2.28
	Trp222	—	—	−9.06
	Met232	—	—	−4.92

Table 3

The estimate binding free energies (kJ/mol) of arylamide inhibitors and substrate calculated by LIE method.

Compound	IC ₅₀ (μM) ^a	(V _{LJ}) _{bound}	(V _{LJ}) _{free}	(V _{CL}) _{bound}	(V _{CL}) _{free}	ΔG _{bind}
A10	>100	−294.39	−184.34	−63.84	−90.50	−6.59
B3	5.16	−288.96	−184.93	−81.18	−100.48	−9.18
P2	0.09	−327.50	−160.36	−77.13	−87.65	−24.99
Substrate	—	−343.36	−188.50	−97.06	−129.78	−11.67

^a IC₅₀ of all inhibitors were taken from literature [36]. The nomenclature of the inhibitors was taken from the original literature and is also given in Table 1.

arylamide P2 (−24.99 kJ/mol) shows a higher value than that of substrate (−11.67 kJ/mol) (Table 3). Consequently, the InhA inhibition activity of arylamide P2 is better than those of arylamides B3 and A10 (binding free energies −9.18 kJ/mol and −6.59 kJ/mol, respectively). It is notable that the calculated free binding energies of inhibitors are in the correct order as compared with the IC₅₀ values. The obtained results could be successfully used to validate the MD procedure in this study.

3.6. The structural basis for inhibitor design

Many arylamides show high potency for inhibiting the InhA enzyme [38]. The highest InhA inhibitory activity with IC₅₀ of 0.09 μM, could be observed for the arylamide P2. However, most of the arylamides with the best InhA inhibitory activities show lower *M. tuberculosis* growth inhibition with MIC against *M. tuberculosis* strain above 125 μM. These results suggest that arylamide inhibitors have poor membrane permeability and maybe lower selectivity to reach the enzyme target. Based on our finding, the amide carbonyl oxygen is selective for arylamide inhibitors for InhA inhibition. When simulating the behavior of the acetylcysteamine moiety of the substrate, it is possible to increase the selectivity of arylamide inhibitors by introducing NH or an acetyl oxygen into the aryl ring A. However, the hydrophilic pocket related to the aryl ring A seems to be rigid and unfavorable for a bulky substituent. Thus, the substituent of the aryl ring A should not be too large. Substituent B is located close to hydrophobic residues. Therefore, substituents B, with a higher lipophilicity as well as the fatty acyl chain of the substrate, should enhance the binding affinity of arylamide inhibitors. Moreover, it could help to increase the membrane permeability of arylamide inhibitors leading to improve *M. tuberculosis* growth inhibition. As seen from MD simulations, a bulky lipophilic substituent could be incorporated into the substituent B, because its hydrophobic pocket seems to be flexible enough for ligand binding.

4. Conclusion

The dynamic behavior in terms of flexibility, conformation and the inhibitor–enzyme interaction of arylamide inhibitors and the substrate in the InhA binding pocket was successfully explained by MD simulations using the program package Gromacs. Based on MD simulations, only the hydrophobic binding pocket of InhA is flexible enough to bind arylamide inhibitors and the substrate. These results allow us to gain insight into some structural requirements of arylamide inhibitors to bind to InhA. When comparing the dynamic behavior of competitive arylamide inhibitors and the substrate in the InhA binding pocket, some substituents of arylamide inhibitors should be imitated from the substrate structure. The amide carbonyl oxygen is highly essential for the affinity of the arylamide inhibitor InhA interactions. The presence of small hydrophilic substituents at the aryl ring A of the arylamide inhibitors, as well as NH and the acetyl oxygen in the acetylcysteamine moiety of the

substrate should improve arylamide selectivity. A bulky substituent with higher lipophilicity as well as the fatty acyl chain of the substrate, should enhance the binding affinity of the arylamide inhibitors. Moreover, it could increase their membrane permeability. Based on our findings, the correct balance between the hydrophilic property of the aryl ring A and the lipophilic property of the substituent B should improve suitable therapeutic activity against *M. tuberculosis* strain. Therefore, this study should facilitate the design of new and more potentially effective antitubercular agents.

Acknowledgements

This research was supported by the Thailand Research Fund (DBG5380006, RTA5080005 and MRG5080267). A. Punkvang is grateful to the grant from under the program Strategic Scholarships for Frontier Research Network for the Ph.D. ASEA-UNINET and Vienna Scientific Cluster are gratefully acknowledged for research support.

References

- [1] World Health Organization, Multidrug and Extensively Drug-resistant TB (M/XDR-TB) Global report on surveillance and response, WHO, 2010.
- [2] A. Quémard, J.C. Sacchettini, A. Dessen, C. Vilcheze, R. Bittman, W.R. Jacobs Jr., J.S. Blanchard, Enzymatic characterization of the target for isoniazid in *Mycobacterium tuberculosis*, *Biochemistry* 34 (1995) 8235–8241.
- [3] J.W. Campbell, J.E. Cronan Jr., Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery, *Annu. Rev. Microbiol.* 55 (2001) 305–332.
- [4] R.J. Heath, C.O. Rock, Fatty acid biosynthesis as a target for novel antibacterials, *Curr. Opin. Invest. Drugs* 5 (2004) 146–153.
- [5] S.W. White, J. Zheng, Y.M. Zhang, C.O. Rock, The structural biology of type II fatty acid biosynthesis, *Annu. Rev. Biochem.* 74 (2005) 791–831.
- [6] Y.M. Zhang, Y.J. Lu, C.O. Rock, The reductase steps of the type II fatty acid synthase as antimicrobial targets, *Lipids* 39 (2004) 1055–1060.
- [7] L. Wen, J.N. Chmielowski, K.C. Bohn, J.K. Huang, Y.N. Timsina, P. Kodali, A.K. Pathak, Functional expression of *Francisella tularensis* FabH and FabI, potential antibacterial targets, *Protein Exp. Purif.* 65 (2009) 83–91.
- [8] H.T. Wright, K.A. Reynolds, Antibacterial targets in fatty acid biosynthesis, *Curr. Opin. Microbiol.* 10 (2007) 447–453.
- [9] D.A. Rozwarski, G.A. Grant, D.H. Barton, W.R. Jacobs Jr., J.C. Sacchettini, Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*, *Science* 279 (1998) 98–102.
- [10] C. Vilcheze, F. Wang, M. Arai, M.H. Hazbón, R. Colangeli, L. Kremer, T.R. Weisbrod, D. Alland, J.C. Sacchettini, W.R. Jacobs Jr., Transfer of a point mutation in *Mycobacterium tuberculosis* inhA resolves the target of isoniazid, *Nat. Med.* 12 (2006) 1027–1029.
- [11] A. Dessen, A. Quémard, J.S. Blanchard, W.R. Jacobs Jr., J.C. Sacchettini, Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*, *Science* 267 (1995) 1638–1641.
- [12] B. Lei, C.J. Wei, S.C. Tu, Action mechanism of antitubercular isoniazid. Activation by *Mycobacterium tuberculosis* KatG, isolation, and characterization of inhA inhibitor, *J. Biol. Chem.* 275 (2000) 2520–2526.
- [13] K. Johnsson, D.S. King, P.G. Schultz, Studies on the mechanism of action of isoniazid and ethionamide in the chemotherapy of tuberculosis, *J. Am. Chem. Soc.* 117 (1995) 5009–5010.
- [14] A. Quémard, A. Dessen, M. Sugantino, W.R. Jacobs Jr., J.C. Sacchettini, J.S. Blanchard, Binding of catalase-peroxidase-activated isoniazid to wild-type and mutant *Mycobacterium tuberculosis* enoyl-ACP reductases, *J. Am. Chem. Soc.* 118 (1996) 1561–1562.
- [15] B. Saint-Joanis, H. Souchon, M. Wilming, K. Johnsson, P.M. Alzari, S.T. Cole, Use of site-directed mutagenesis to probe the structure, function and isoniazid activation of the catalase/peroxidase, KatG, from *Mycobacterium tuberculosis*, *Biochem. J.* 338 (1999) 753–760.
- [16] X. Zhao, H. Yu, S. Yu, F. Wang, J.C. Sacchettini, R.S. Magliozzo, Hydrogen peroxide-mediated isoniazid activation catalyzed by *Mycobacterium tuberculosis* catalase-peroxidase (KatG) and its S315T mutant, *Biochemistry* 45 (2006) 4131–4140.
- [17] C. Metcalfe, I.K. Macdonald, E.J. Murphy, K.A. Brown, E.L. Raven, P.C. Moody, The tuberculosis prodrug isoniazid bound to activating peroxidases, *J. Biol. Chem.* 283 (2008) 6193–6200.
- [18] B.K. Sinha, enzymatic activation of hydrazine derivatives, *J. Biol. Chem.* 258 (1983) 796–801.
- [19] M. Nguyen, C. Claparois, J. Bernadou, B. Meunier, A fast and efficient metal-mediated oxidation of isoniazid and identification of isoniazid-NAD(H) adducts, *Chem. Bio. Chem.* 2 (2001) 877–883.
- [20] B. Heym, Y. Zhang, S. Poulet, D. Young, S.T. Cole, Characterization of the katG gene encoding a catalase-peroxidase required for the isoniazid susceptibility of *Mycobacterium tuberculosis*, *J. Bacteriol.* 175 (1993) 4255–4259.
- [21] K. Johnsson, P.G. Schultz, Mechanistic studies of the oxidation of isoniazid by the catalase peroxidase from *Mycobacterium tuberculosis*, *J. Am. Chem. Soc.* 116 (1994) 7425–7426.
- [22] G.S. Timmins, V. Deretic, Mechanisms of action of isoniazid, *Mol. Microbiol.* 62 (2006) 1220–1227.
- [23] K. Johnsson, W.A. Froland, P.G. Schultz, Overexpression, purification, and characterization of the catalase-peroxidase KatG from *Mycobacterium tuberculosis*, *J. Biol. Chem.* 272 (1997) 2834–2840.
- [24] C. Vilcheze, H.R. Morbidoni, T.R. Weisbrod, H. Iwamoto, M. Kuo, J.C. Sacchettini, W.R. Jacobs Jr., Inactivation of the inhA-encoded fatty acid synthase II (FASII) enoyl-acyl carrier protein reductase induces accumulation of the FASII end products and cell lysis of *Mycobacterium smegmatis*, *J. Bacteriol.* 182 (2000) 4059–4067.
- [25] A.S. Piatek, A. Telenti, M.R. Murray, H. El-Hajj, W.R. Jacobs Jr., F.R. Kramer, D. Alland, Genotypic analysis of *Mycobacterium tuberculosis* in two distinct populations using molecular beacons: implications for rapid susceptibility testing, *Antimicrob. Agents Chemother.* 44 (2000) 103–110.
- [26] R.A. Slayden, C.E. Barry 3rd, The genetics and biochemistry of isoniazid resistance in *Mycobacterium tuberculosis*, *Microbes Infect.* 2 (2000) 659–669.
- [27] M.J. Torres, A. Criado, J.C. Palomares, J. Aznar, Use of real-time PCR and fluorimetry for rapid detection of rifampin and isoniazid resistance-associated mutations in *Mycobacterium tuberculosis*, *J. Clin. Microbiol.* 38 (2000) 3194–3199.
- [28] Y. Zhang, B. Heym, B. Allen, D. Young, S. Cole, The catalaseperoxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*, *Nature* 358 (1992) 591–593.
- [29] J.S. Oliveira, J.H. Pereira, F. Canduri, N.C. Rodrigues, O.N. de Souza, W.F. de Azevedo Jr., L.A. Basso, D.S. Santos, Crystallographic and pre-steady-state kinetics studies on binding of NADH to wild-type and isoniazid-resistant enoyl-ACP(CoA) reductase enzymes from *Mycobacterium tuberculosis*, *J. Mol. Biol.* 359 (2006) 646–666.
- [30] M.R. Marques, J.H. Pereira, J.S. Oliveira, L.A. Basso, D.S. Santos, W.F. de Azevedo Jr., M.S. Palma, The inhibition of 5-enolpyruvylshikimate-3-phosphate synthase as a model for development of novel antimicrobials, *Curr. Drug Targets* 8 (2007) 445–457.
- [31] A.I. De La Iglesia, H.R. Morbidoni, Mechanisms of action of and resistance to rifampicin and isoniazid in *Mycobacterium tuberculosis*: new information on old friends, *Rev. Argent. Microbiol.* 38 (2006) 97–109.
- [32] A. Banerjee, E. Dubnau, A. Quemard, V. Balasubramanian, K.S. Um, T. Wilson, D. Collins, G. de Lisle, W.R. Jacobs Jr., InhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*, *Science* 263 (1994) 227–230.
- [33] J.S. Freundlich, F. Wang, C. Vilcheze, G. Gulten, R. Langley, G.A. Schiehser, D.P. Jacobus, W.R. Jacobs Jr., J.C. Sacchettini, Triclosan derivatives: towards potent inhibitors of drug-sensitive and drug-resistant *Mycobacterium tuberculosis*, *Chem. Med. Chem.* 4 (2009) 241–248.
- [34] C.W. Am Ende, S.E. Knudson, N. Liu, J. Childs, T.J. Sullivan, M. Boyne, H. Xu, Y. Gegina, D.L. Knudson, F. Johnson, C.A. Peloquin, R.A. Slayden, P.J. Tonge, Synthesis and in vitro antimycobacterial activity of B-ring modified diaryl ether InhA inhibitors, *Bioorg. Med. Chem. Lett.* 18 (2008) 3029–3033.
- [35] M.E. Boyne, T.J. Sullivan, C.W. Am Ende, H. Lu, V. Gruppo, D. Heaslip, A.G. Amin, D. Chatterjee, A. Lenaerts, P.J. Tonge, R.A. Slayden, Targeting fatty acid biosynthesis for the development of novel chemotherapeutics against *Mycobacterium tuberculosis*: evaluation of A-ring-modified diphenyl ethers as high-affinity InhA inhibitors, *Antimicrob. Agents Chemother.* 51 (2007) 3562–3567.
- [36] P.J. Tonge, C. Kisker, R.A. Slayden, Development of modern InhA inhibitors to combat drug resistant strains of *Mycobacterium tuberculosis*, *Curr. Top. Med. Chem.* 7 (2007) 489–498.
- [37] X. He, A. Alian, R. Stroud, P.R. Ortiz de Montellano, Pyrrolidine carboxamides as a novel class of inhibitors of enoyl acyl carrier protein reductase from *Mycobacterium tuberculosis*, *J. Med. Chem.* 49 (2006) 6308–6323.
- [38] X. He, A. Alian, P.R. Ortiz de Montellano, Inhibition of the *Mycobacterium tuberculosis* enoyl acyl carrier protein reductase InhA by arylamides, *Bioorg. Med. Chem.* 15 (2007) 6649–6658.
- [39] S. Alcaro, A. Artese, F. Ceccherini-Silberstein, F. Ortuso, C.F. Perno, T. Sing, V. Svicher, Molecular dynamics and free energy studies on the wild-type and mutated HIV-1 protease complexed with four approved drugs: mechanism of binding and drug resistance, *J. Chem. Inf. Model* 49 (2009) 1751–1761.
- [40] H.A. Wahab, Y.S. Choong, P. Ibrahim, A. Sadikun, T. Scior, Elucidating isoniazid resistance using molecular modeling, *J. Chem. Inf. Model* 49 (2009) 97–107.
- [41] A. Ivetac, J.A. McCammon, Elucidating the inhibition mechanism of HIV-1 non-nucleoside reverse transcriptase inhibitors through multicopy molecular dynamics simulations, *J. Mol. Biol.* 388 (2009) 644–658.
- [42] J.J. Tan, R. Kong, C.X. Wang, W.Z. Chen, Molecular dynamics simulation on the complexes of N-terminal region of HIV-1 gp41 and its C-peptide inhibitors, *Theochem* 682 (2004) 9–15.
- [43] S. Weerasinghe, R. Samantha Dassanayake, Simulation of structural and functional properties of mevalonate diphosphate decarboxylase (MVD), *J. Mol. Model.* 16 (2010) 489–498.
- [44] V. Nukoolkarn, V.S. Lee, M. Malaisree, O. Aruksakulwong, S. Hannongbua, Molecular dynamic simulations analysis of ritonavir and lopinavir as SARS-CoV 3CL(pro) inhibitors, *J. Theor. Biol.* 254 (2008) 861–886.

- [45] M. Liu, X.J. Cong, P. Li, J.J. Tan, W.Z. Chen, C.X. Wang, Study on the inhibitory mechanism and binding mode of the hydroxycoumarin compound NSC158393 to HIV-1 integrase by molecular modeling, *Biopolymers* 91 (2009) 700–709.
- [46] Y. Zhao, W. Li, J. Zeng, G. Liu, Y. Tang, Insights into the interactions between HIV-1 integrase and human LEDGF/p75 by molecular dynamics simulation and free energy calculation, *Proteins* 72 (2008) 635–645.
- [47] S. Han, Molecular dynamics simulations of thioredoxin with S-glutathiolated cysteine-73, *Biochem. Biophys. Res. Commun.* 362 (2007) 532–537.
- [48] D.A. Rozwarski, C. Vilchèze, M. Sugantino, R. Bittman, J.C. Sacchettini, Crystal structure of the *Mycobacterium tuberculosis* enoyl-ACP reductase, InhA, in complex with NAD⁺ and a C16 fatty acyl substrate, *J. Biol. Chem.* 274 (1999) 15582–15589.
- [49] G.M. Morris, D.S. Goodshell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function, *J. Comput. Chem.* 19 (1998) 1639–1662.
- [50] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible and free, *J. Comput. Chem.* 26 (2005) 1701–1719.
- [51] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun.* 91 (1995) 43–56.
- [52] E. Lindahl, B. Hess, D. van der Spoel, Package for molecular simulation and trajectory analysis, *J. Mol. Model.* 7 (2001) 306–317.
- [53] W.R.P. Scott, P.H. Hunenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, A.E. Torda, T. Huber, P. Kruger, W.F. van Gunsteren, The GROMOS biomolecular simulation program package, *J. Phys. Chem. A* 103 (1999) 3596–3607.
- [54] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general AMBER force field, *J. Comput. Chem.* 25 (2004) 1157–1174.
- [55] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, The missing term in effective pair potentials, *J. Phys. Chem.* 91 (1987) 6269–6271.
- [56] B. Hess, H. Bekker, H.J.C. Berendsen, J. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [57] T. Darden, D. York, L. Pedersen, Particle mesh Ewald – an N·Log(N) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [58] J. Aqvist, C. Medina, J.E. Samuelsson, A new method for predicting binding affinity in computer-aided drug design, *Protein Eng.* 7 (1994) 385–391.
- [59] M.R. Kuo, H.R. Morbidoni, D. Alland, S.F. Sneddon, B.B. Gourlie, M.M. Staveski, M. Leonard, J.S. Gregory, A.D. Janjigian, C. Yee, J.M. Musser, B. Kreiswirth, H. Iwamoto, R. Perozzo, W.R. Jacobs Jr., J.C. Sacchettini, D.A. Fidock, Targeting tuberculosis and malaria through inhibition of enoyl reductase: compound activity and structural data, *J. Biol. Chem.* 278 (2003) 20851–20859.

List of abbreviation

ACPYPE: AnteChamber PYthon Parser interface
IC₅₀: the half maximal inhibitory concentration
INH: isoniazid
LGA: Lamarckian genetic algorithm
LIE: linear interaction energy
MD: molecular dynamics
MDR-TB: multidrug resistant-tuberculosis
MIC: minimum inhibitory concentrations
NADH: nicotinamide adenine dinucleotide
RMSD: root mean square deviations
RMSF: root mean square fluctuation
TB: tuberculosis
XDR-TB: extensively drug-resistant tuberculosis



Short communication

¹H NMR chemical shifts of some DMSO-solvated amines using MD-ONIOM2Veeramol Vailikhit^{a,*}, Wilhelm J. Holzschuh^a, Supa Hannongbua^b^a Faculty of Liberal Arts and Science, Kasetsart University, Kamphaeng Saen Campus, Kamphaeng Saen, Nakhon Pathom 73140, Thailand^b Department of Chemistry, Faculty of Science, and Center of Nanotechnology, Kasetsart University, Jatuchak, Bangkok 10900, Thailand

ARTICLE INFO

Article history:

Received 21 December 2009

Received in revised form 30 December 2009

Accepted 30 December 2009

Available online 14 January 2010

Keywords:

MD-ONIOM2

¹H NMR

Solvent effect

ABSTRACT

MD-ONIOM2 has been previously used to model nevirapine in the polar solvent DMSO, accurately predicting the ¹H NMR chemical shifts of all the protons, including the acidic amine proton that suffers significant deshielding due to hydrogen bonding. In this work, MD-ONIOM2 is shown to be a more generally applicable model by predicting the NMR shifts of four amines in DMSO solution with acceptable accuracy. The solutes *N*-methylaniline and valerolactam with a single acidic proton, and aniline and acetamide with two acidic protons were used, and the NMR predictions compared to experimental values. Gas-phase models and IEFPCM are shown to not predict the acidic proton shifts well in comparison to MD-ONIOM2.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The use of the Molecular Dynamics-Our own *N*-layered Integrated molecular Orbital and Molecular mechanics, with *N* = 2, (MD-ONIOM2) method for the prediction of ¹H NMR chemical shifts has been previously reported for nevirapine in DMSO [1], with good results when compared to an experimentally measured spectrum. This included the acidic proton that forms a hydrogen bond with a polar DMSO solvent molecule and suffers significant deshielding as a result. In comparison, using a gas-phase model, while cheap to calculate in terms of computer resources, gives good results for all hydrogen NMR shifts except those of the acidic protons. This is expected as there are no solvent interactions included in such a simple model, and hydrogen bonding is a significant interaction.

To show that the MD-ONIOM2 method has general application for modelling acidic protons in polar solvents, the predicted ¹H NMR shifts of four simple amine compounds in DMSO are reported here and compared to experimentally measured shifts. The molecules were chosen to represent a limited range of different electronic environments for the acidic protons. *N*-methylaniline is the simplest with only a single acidic proton with an adjacent phenyl group. Aniline has a second equivalent acidic proton attached to the amine nitrogen. δ -Valerolactam has a carbonyl group adjacent to the amine and its one acidic proton, and acetamide has two acidic protons on the amine nitrogen adjacent to the carbonyl group. Acetamide has the extra complication of keto-enol

tautomerism preventing rotation around the C–N bond, and thus the *cis*- and *trans*-amine protons are not equivalent.

The results were also compared to the Integral Equation Formalism Polarizable Continuum Model (IEFPCM) which has also been proposed to include solvent effects on the solute molecule. This model treats the solvent as a continuum of polarisable dielectric described by the solvent's dielectric constant, and only considers relatively long range electrostatic interactions between the solute molecule and the continuum. This may be appropriate for non-polar solvents, but fails to include the localised strong effect of hydrogen bonding between a solute and discrete solvent molecules. The MD-ONIOM2 method, in comparison, attempts to model the solute molecule with discrete solvent molecules bonded to it, and so it incorporates short-range, localised interactions such as hydrogen bonding.

2. Method

Modelling of the solute molecules was performed in two stages: first MD of the molecule in a box of solvent molecules, followed by ONIOM2 of sample snapshots from the MD stage to calculate ¹H NMR shifts. This was described in detail previously [1] and only a summary and differences given below.

The AMBER9 software package was used for the MD with this work. Each solute molecule, like DMSO previously, was generated with the SYBYL7.0 program [2] and optimised with the Gaussian03 program [3] at B3LYP/6-31G** level. The molecule ESP was generated using single point calculation at HF/6-31G* level with the Merz-Kollman-Shigh charge scheme (MK). The Antechamber module was used to generate the “prep” input file and atomic charge, using the AMBER force-field parameters [4]. The solute

* Corresponding author. Tel.: +66 34 281 105.

E-mail address: veeramol.v@ku.ac.th (V. Vailikhit).

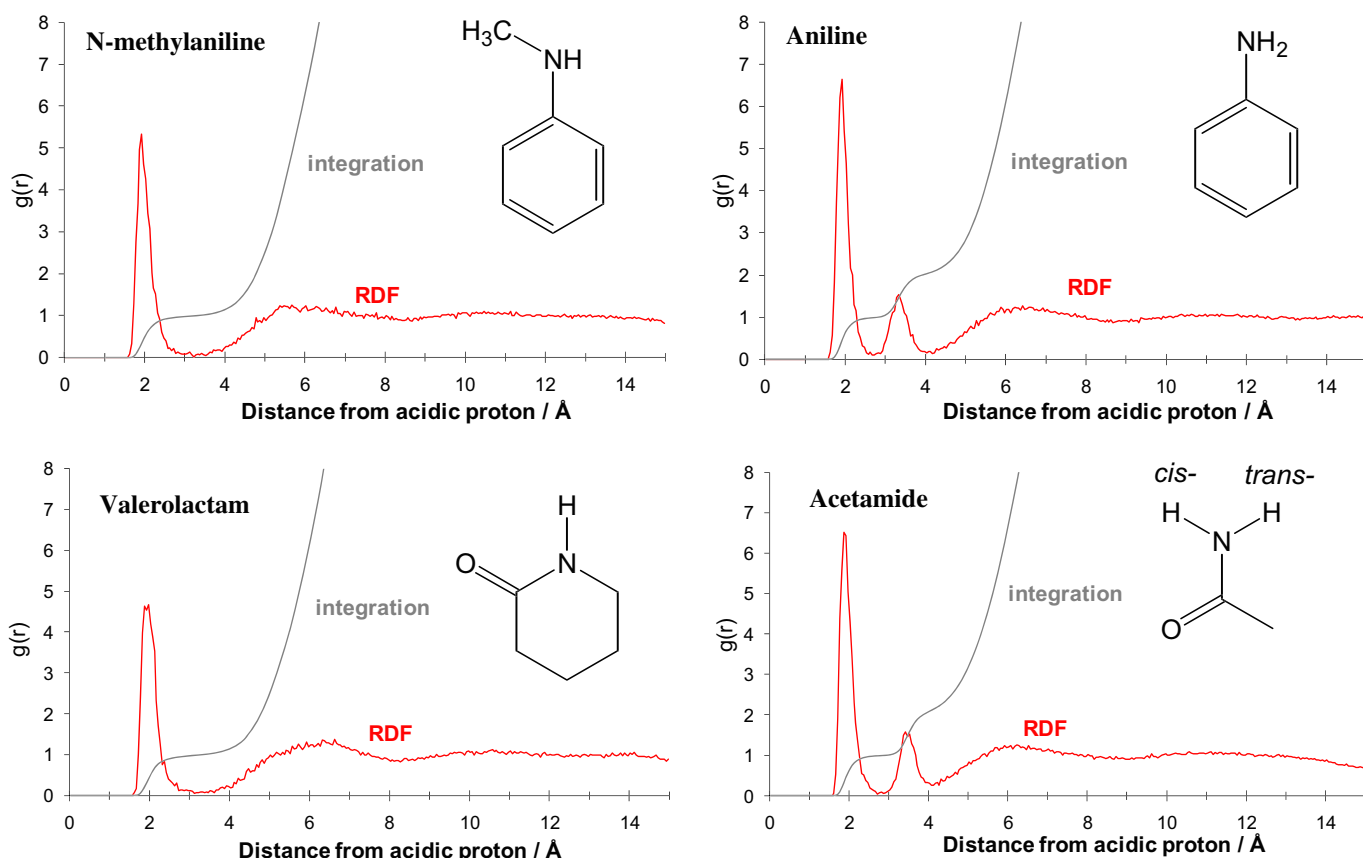


Fig. 1. From the molecular dynamics simulations of each solute molecule in DMSO, the radial distribution of DMSO-oxygen atoms can be determined. The peaks show the extent of the solvation shell, used for the cut-off in ONIOM2, and integration determined that one DMSO molecule was bonded to each acidic proton.

molecule model was placed in a box of DMSO molecules [1] for the MD run and a radial distribution function (RDF) was used to determine the average distribution of DMSO-oxygen atoms around the acidic protons. The RDF was used to determine the discrete models

(Fig. 2), from 10 snapshots taken every 100 ps during the final 1 ns of the production period, for the ONIOM2 modelling with the Gaussian03 software package at B3LYP/6-311++G**/B3LYP/6-31G**::HF/STO-3G//PM3, using TMS as the standard NMR reference. The predicted ^1H NMR shifts from each of the ten snapshots were averaged at the end.

For comparison, the solutes were modelled as isolated molecules in the gas phase, and in with DMSO solvent using SCRF-IEFPCM in Gaussian03. The accuracies of the predicted chemical shifts were evaluated against experimentally measured NMR data reported in the literature [5].

3. Results

During the MD simulations, the root mean square deviation (RMSD) of heavy atoms in the solute molecules from the starting geometry and the total energy of the whole system were observed to ensure the MD simulations were stable.

For each solute MD run, the RDF of the acidic proton to the DMSO-oxygen atoms was used to determine the radius of the solvation shell; that is, the distance within which the DMSO molecules are bonded to the acidic protons. This was obvious from peaks in the RDF shown in Fig. 1, and integration of the area under the peaks showed that on average only one DMSO molecule was bonded to each acidic proton. The solvation shell radius was used as the cut-off for the ONIOM2 model with only DMSO molecules that were at least partially inside the radius included. All other DMSO molecules were discarded. The cut-off radius was approximately 3 Å for the molecules with only one acidic proton, and approximately 4 Å for those with two acidic protons.

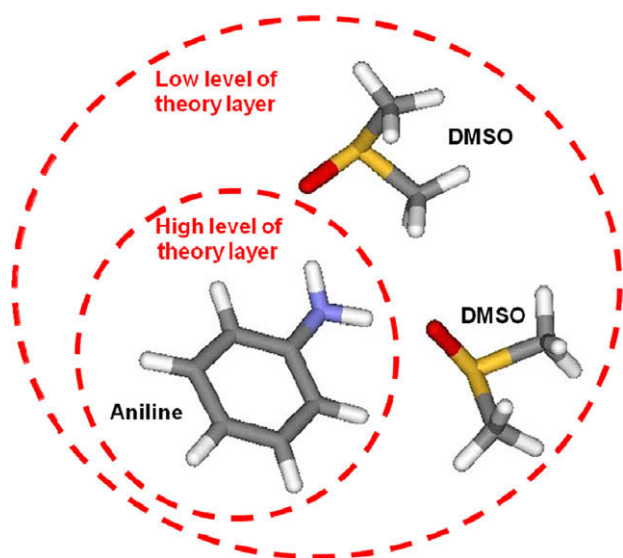


Fig. 2. Schematic of the two layers in the ONIOM2 model showing, in this case, an aniline molecule which is treated at a high level of quantum mechanics theory, and the combined aniline and H-bonded DMSO molecules treated at a low level of quantum mechanics theory.

Table 1

Shown here are the experimentally measured NMR shifts of the acidic amine protons of the four solutes, and the predicted values from each of the three models. The difference, experimental value–predicted value, is given to make comparison easier. The shifts for all protons are given as online [Supplementary data](#).

	Expt.	ONIOM2		Gas phase		IEFPCM	
	ppm	ppm	Difference	ppm	Difference	ppm	Difference
N-methylaniline	5.52	5.00	0.52	3.35	2.17	4.52	1.00
Aniline	4.99	5.16	−0.17	3.20	1.79	3.96	1.03
	4.99	5.39	−0.40	3.20	1.79	3.96	1.03
Valerolactam	7.34	6.96	0.38	4.62	2.72	5.81	1.53
Acetamide	6.70	6.28	0.42	4.38	2.32	5.28	1.42
	7.30	6.85	0.45	4.74	2.56	5.95	1.35

The ONIOM2 calculated NMR shifts for the acidic amine protons can be seen in [Table 1](#). Also seen are the experimental NMR shift values, and the corresponding gas-phase model and IEFPCM re-

sults. The differences from the experimental value are shown for clarity. The shifts from all protons are shown in [Supplementary Data](#) available online.

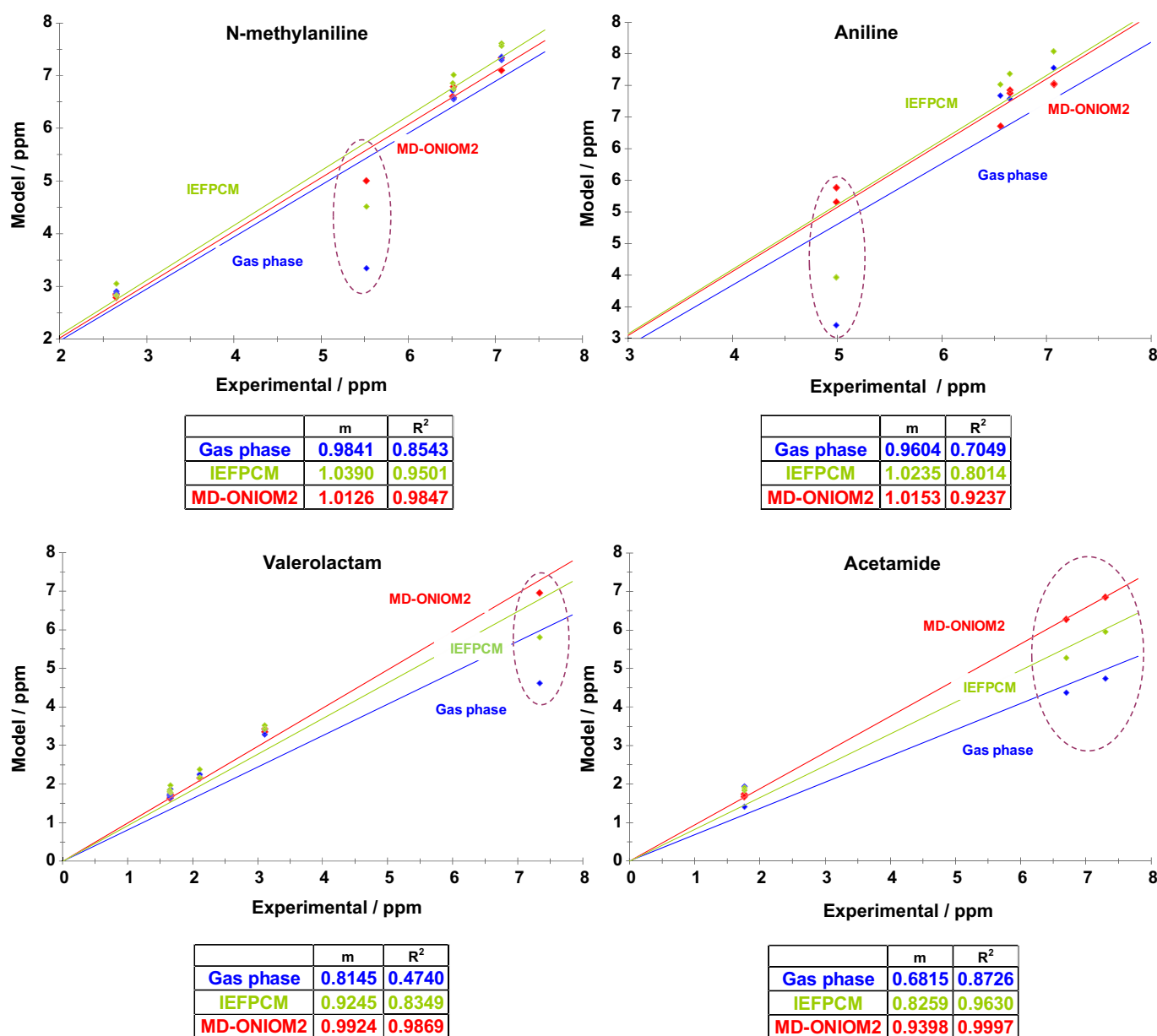


Fig. 3. Modelled NMR shifts plotted against experimental values are shown. The slope and correlation are also given for least squares regression lines which are constrained to go through the origin, since all shifts are relative to TMS. It can be seen that the MD-ONIOM2 results are very close to the ideal line which would have $m = 1$ and $R^2 = 1$. The encircled points are the shifts of the acidic protons and show the large variation in their predicted values.

The three models give similar predictions of NMR shifts for all the non-acidic protons. However, for the acidic amine protons, the gas-phase model gives the poorest predictions and ONIOM2 the best. The IEFPCM predictions are approximately halfway between the other two. This can be seen in the graphs of Fig. 3 which plot the predicted shifts against the experimentally measured shifts of all the protons. Using least squares regression analysis, the slope and correlation for a line fitted to each set of points is also given. The regression line is constrained to go through the origin as both modelled and experimental shifts are relative to TMS with a shift of 0 ppm. For a perfect model, the resultant points would lie on a line with a slope, $m = 1$, and correlation, $R^2 = 1$. MD-ONIOM2 gives results closest to the ideal. The gas-phase model especially but also IEFPCM consistently give shifts for the acid protons that are too low. The IEFPCM also consistently overestimates the shifts of all the non-acid protons.

It should also be noted that the *cis*- and *trans*-protons in acetamide are also well modelled and different values for each predicted. Additionally, either acid proton can be chosen as the centre for the RDF as all the bonded DMSO molecules will still be observed and the same snapshots used in ONIOM2.

4. Conclusion

The MD-ONIOM2 method has been shown to be capable of modelling the interaction between a solute with acidic protons and a polar solvent, including the hydrogen bonding. MD shows that in the solvation shell, a DMSO molecule is generally attached to each acidic proton, and there is exchange of solvent molecules over time. Averaging the ONIOM2 results from a sample set of

MD snapshots gives the final NMR shifts, with the desheilding from H-bonding properly included, and the other protons well described. The other two models are adequate if it is acceptable that the acidic protons and their complex local electronic environments do not need to be accurately described. This may be the case if modelling a solute in an apolar solvent such as chloroform.

Acknowledgement

The Thailand Research Fund (MRG5180282) and Kasetsart University Research and Development Institute are thanked for financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.theochem.2009.12.038](https://doi.org/10.1016/j.theochem.2009.12.038).

References

- [1] V. Vailikhit, W. Tressuwan, S. Hannongbua, J. Mol. Struct. (THEOCHEM) 806 (2007) 99.
- [2] SYBYL 7.0, Tripos Associates Inc., 1699 South Hanley Road, Suite 303, St. Louis, Missouri 63144, USA.
- [3] M.J. Frisch et al., Gaussian03, Inc., Wallingford CT, 2004.
- [4] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, P.A. Kollman, AMBER 9, University of California, San Francisco, 2006.
- [5] R.J. Abraham, J.J. Byrne, L. Griffiths, M. Perez, Magn. Reson. Chem. 44 (2006) 491.