TASK QUARTERLY 5 No 3 (2001), 301–310

PREDICTION OF BINDING AFFINITIES FOR DNA INTERCALATORS BY MOLECULAR DYNAMICS SIMULATIONS

JAN MAZERSKI

Department of Pharmaceutical Technology and Biochemistry,
Technical University of Gdansk,
Narutowicza 11/12, 80-952 Gdansk, Poland
janmaz@chem.pg.gda.pl

(Received 29 November 2000; revised manuscript received 31 January 2001)

Abstract: The computer modelling and simulation methods are widely used in rational drug design to obtain information necessary for understanding interactions between a ligand (drug molecule) and its cellular macromolecular target on molecular level. The determination of free energies of binding for ligand-target complexes is one of the crucial points in those studies. In recent years several methods have been proposed to solve this problem. The majority of them use molecular dynamics (MD) simulations. Two, most popular methods: (i) a free energy perturbation method (FEP), and (ii) a linear response (LR) method, are shortly presented in this paper together with their limitations and advantages. In this work I presented the first attempt to use LR approach to 10 anti-tumour agents able to intercalate into DNA. The LR relationship obtained in the present study indicated that in the system studied the electrostatic term has no influence on the free energy of binding. The relationship is now successfully used in our research group in further molecular modelling studies concerning DNA intercalators with similar structure.

Keywords: molecular modelling, free energy of binding, intercalation into DNA, linear responce method, free energy perturbation method

1. Introduction

The application of computer modelling and simulation methods to rational drug design is a rapidly growing field. The general aim of the research is the development of ways of determining knowledge necessary for understanding interactions between a ligand (drug molecule) and its cellular macromolecular target on molecular level. That knowledge should contain information about:

- 3D structure of a ligand-target complex,
- types of interactions responsible for the complex formation and structure,
- thermodynamics properties of the complex which determine complex stability.

Typical way in which the ligand-target complex is constructed during molecular modelling calculations bases on 3D structures of isolated complex components: the ligand and the target molecule. One may think that this step is relatively easy because the ligand

should fit the target binding site according to the "key and lock" rule. However, in practice this problem is not so simple. The macromolecular target may possess more than one site with which the ligand could interact. Moreover, the ligand and/or target molecule might change the 3D structure during complex formation. Those problems can be easier solved when molecular dynamics (MD) simulations are used instead of molecular mechanics (MM) energy minimization. In a typical approach dynamic simulations of isolated molecule of the ligand and the target are performed and several probable conformations of the molecules are selected. Next, several possible complex configurations are constructed by combination of selected conformers of the complex constituents. Each of the complex configurations constructed in this way is relaxed by MD simulation and the most stable one is chosen for further studies. The determination of a thermodynamic stability of each configuration is the crucial point at this stage.

When the 3D structure of the complex between a lead compound and its macromolecular target was determined that information could be used for selecting which of a series of possible derivatives might be the most interesting to synthesize. Also on this step the determination of the thermodynamic stability of complexes with different ligands is crucial for correct prediction.

The change of free energy between bound and free states of the ligand (free energy of binding) is a thermodynamic function, which directly decides about stability of the complex. The same function can also be used to calculate a ligand binding constant.

The development of efficient computational methods for an estimation of free energies of binding for ligand-target complexes has been an area of active investigation. MM approaches to the current problem offer the advantage of being relatively rapid and in some cases have yielded valuable information [1–3]. These methods, however, generally rely upon single structures for the unbounded ligand and the complex and thus suffer from the difficulty of locating an appropriate local minimum for these species. Of necessity, a single structure is incapable of representing the range of ligand and target flexibility found in practice.

In recent years two methods relied upon free energies calculated from averages based on thermally equilibrated collections of configurations have been proposed. These collections of configurations are typically generated through MD simulations.

The free energy perturbation method (FEP) is well suited to calculate the free energy difference (the relative free energy) between two systems differing by a moderately small change, Figure 1. (For reviews see [4–8]). The relative binding energy, $\Delta\Delta G_{\text{binding}}$, is calculated as free energy change associated with an unphysical transmutation of a ligand A to a ligand B inside a complex:

$$\Delta\Delta G_{\text{binding}} = \Delta G_{\text{bound,B}} - \Delta G_{\text{bound,A}} \cong \Delta G_{\text{A}\rightarrow\text{B,complex}}. \tag{1}$$

This unphysical step could be simulated by quasi-reversible, multi-step transformation of potential V_A for the ligand A into potential V_B for the ligand B according to the perturbation equation [9]:

$$V_i = \lambda_i \times V_A + (1-\lambda_i) \times V_B, \qquad i \in <0 \ldots n>, \tag{2}$$

where $\lambda_0 = 0$ for the complex with the ligand A and $\lambda_n = 1$ for the complex with the ligand B, Figure 2. The approximation (1) is valid as long as difference between potential energy of two adjacent λ-points is small enough. This condition makes very small λ-steps necessary.

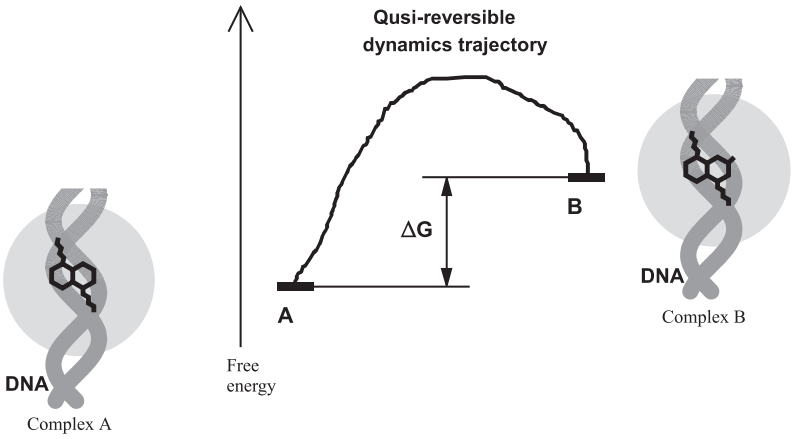


Figure 1. Two systems used by free energy perturbation (FEP) approach to estimate relative free energy of binding

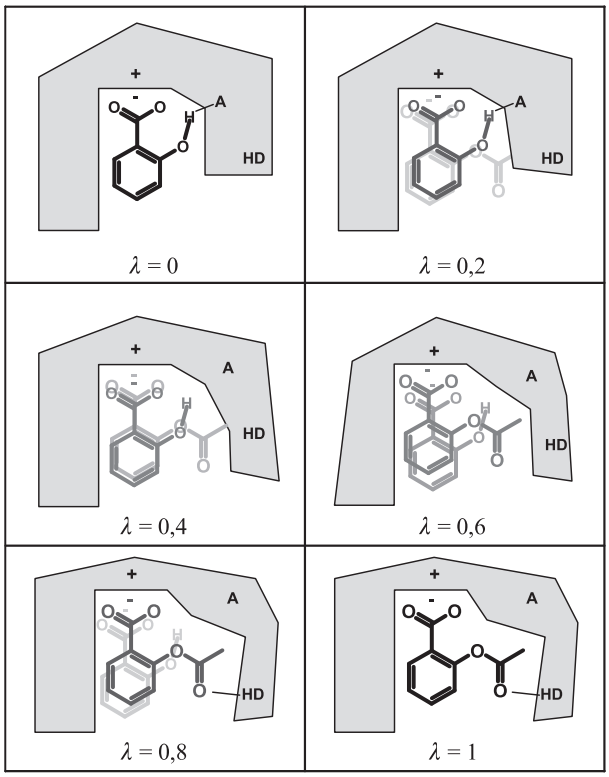


Figure 2. Transmutation from ligand *A* (λ = 0) to ligand *B* (λ = 1). Note that the macromolecular target changes also its 3D configuration during the transmutation of the ligand

In each λ-step one hundred or more MD iterations are performed to relax the system after the potential change.

Determinations of relative free energies for systems differing by a small change are nowadays almost routine MD calculations using FEP approach. However, in rational drug

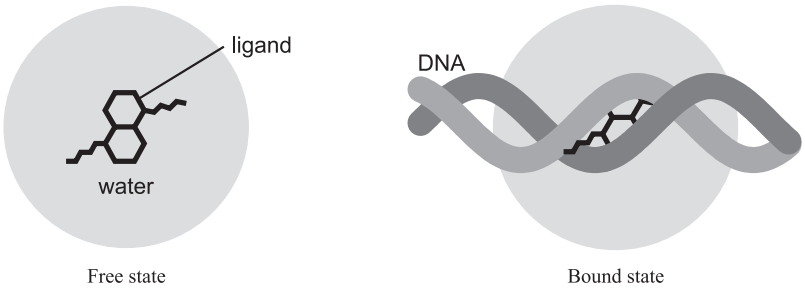


Figure 3. Two systems used by the linear response (LR) method to estimate absolute free energy of binding

design practice series of compounds with significantly different structures are met very often.

Åqvist and co-workers have recently presented another method relied upon MD simulations more suitable to solve that problems. Reasonable accurate absolute binding energies have been estimated by this method from a linear response (LR) assumption for a combination of electrostatic and van der Waals non-bonded interaction terms [9–17]. The LR method uses time averages from two simulations, one for the ligand in water and the other for the ligand bounded inside the macromolecule, Figure 3.

Thus, typical LR equation takes the form:

$$\Delta G_{\text{bind}} = \alpha \times (< E^{\text{vdw}} >_{\text{bound}} - < E^{\text{vdw}} >_{\text{free}}) + \beta \times (< E^{\text{elec}} >_{\text{bound}} - < E^{\text{elec}} >_{\text{free}}). \qquad (3)$$

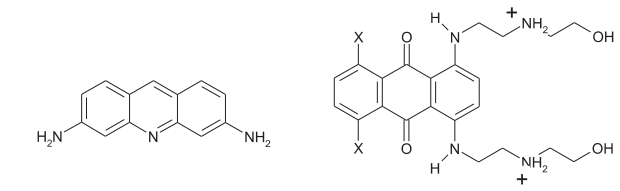
The terms $< E^{\text{vdw}} >$ and $< E^{\text{elec}} >$ refer to the Lenard-Jones and electrostatic average interaction energies for the bound and free states of the ligand. The coefficients α and β are either assigned or empirically determined to obtain the best fit to the experimental observable, the free energy of binding. Jorgensen and co-workers have proposed [18–21] to develop Equation (3) by taking into account changes in solvent accessible surface area, $\Delta SASA$, of the ligand during complex formation:

$$\begin{aligned} \Delta G_{\text{bind}} = {} & \alpha \times (< E^{\text{vdw}} >_{\text{bound}} - < E^{\text{vdw}} >_{\text{free}}) \\ & + \beta \times (< E^{\text{elec}} >_{\text{bound}} - < E^{\text{elec}} >_{\text{free}}) \\ & + \gamma \times (< SASA >_{\text{bound}} - < SASA >_{\text{free}}), \end{aligned} \qquad (4)$$

where γ is a third coefficient to be fit empirically to the experimental data.

Up till now the LR approach has been successfully used for different type of ligands forming complexes with proteins. In this work I present the first attempt to use this approach for compounds able to intercalate into DNA, Figure 4. One could imagine that DNA is a very simple molecular target and due to its well known 3D structure it is suitable for molecular modelling. It is true that 3D structure for native DNA is very well known, Figure 5. However, this structure is highly flexible [22, 23] and could be easily changed in different manner during interactions with ligands. In addition, DNA might interact with ligands in several different ways due to a large number of polar functional groups. Site and type of interactions depend on chemical constitution of a ligand. Polycyclic aromatic compounds interact with DNA by intercalation between base-pairs. Relatively long, polar compounds able to form hydrogen bonds prefer fitting to the minor groove. More spherical ligands interact with DNA from the major groove side. Moreover, cationic compounds like aliphatic amines may

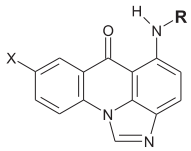
interact with negatively charged phosphate groups by strong ionic forces. A prediction of a structure of a DNA complex is particularly difficult when a ligand consists of parts with different or opposite interaction preferences. Thus, a site and type of interaction with DNA should be determined separately for nearly each family of compounds (chemotype).



Proflavine

Anthracenodions

MIT: X = OH
AMT: X = H



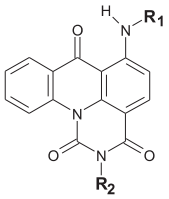
Imidazoacridinons

C1311: X = OH
R = $(CH_2)_2N^+H(CH_2CH_3)_2$

C1330: X = OCH_3
R = $(CH_2)_2N^+H(CH_2CH_3)_2$

C1492: X = OH
R = $(CH_2)_2N^+H(CH_2CH_3)_2$

C1558: X = OH
R = $(CH_2)_2N^+H(CH_2CH_3)_2$



Pyrimidoacridinetrions

PA36: R_1 = Cl
R_2 = $(CH_2)_2N^+H(CH_3)_2$

PA36: R_1 = $NH(CH_2)_2N^+H(CH_3)_2$
R_2 = $(CH_2)_2N^+H(CH_3)_2$

PA36: R_1 = $NH(CH_2)_2N^+H(CH_3)_2$
R_2 = H

Figure 4. Chemical structures of DNA ligands used to determine coefficients of LR model

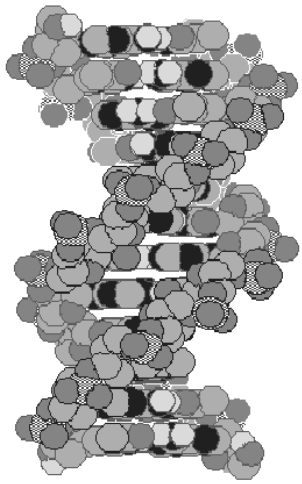


Figure 5. 3D structure of DNA fragment in native form

The mode of action of the majority of anti-tumour compounds bases on their interactions with DNA. In the present paper we studied synthetic anti-tumour agents belonging to anthracenedione and acridine chemotypes, Figure 4. It has been determined in

our earlier modelling studies [24–26] that compounds used in this study interact with DNA in the same way: a flat aromatic core intercalates between DNA base-pairs in such manner that side chain(s) interact with DNA surface from the minor groove side. However, orientation of a long axis of the aromatic core inside the intercalation cavity is determined by:

- chemical constitution of the aromatic core,
- length and chemical structure of the side chain(s),
- site in the core to which the chain is attached,
- presence and type of additional functional group in the aromatic core.

In these works the complexes were formed by dodecamer duplex d(GCGCGC|GCGCGC)$_2$ with ligands intercalated between central base-pairs.

2. Computational methods

The dynamic simulations reported in this work were performed with GROMOS 96, a set of MD programs from the University of Groningen, Netherlands [27], implemented on INDIGO 2 (SG) workstation.

The anthracenedione, pyrimidoacridinetrione, and imidazoacridinone ligands were simulated as mono- (C1311, C1330, C1492, C1558, PA36 and PA51) or dication (MIT, AMT and PA39) with protonated side chain amino group(s), whereas proflavine was treated as a neutral molecule. The periodic boundary condition was applied, all bond lengths were constrained to their equivalence values using SHAKE algorithm [28], and distances related to Watson-Crick hydrogen bonds were restricted to 0.21 nm [29] during all simulations. A dielectric constant of 1 and a cut-off of 0.8 nm were used for non-bounded interactions, whereas a cut-off of 1.2 nm was used for long term electrostatic interactions. An appropriate number of counterions, 0, 1 or 2 Cl^- for ligand and 22 Na^+ for the dodecamer of DNA, and of water molecules, about 1700 for free ligand simulations and about 1300 for the complex, were added to the system. In order to save computation time the intercalation complexes were constructed with DNA dodecamer in which the intercalation cavity was preformed [24] between central base-pairs. The initial co-ordinates of ligand atoms were obtained from molecular modelling of the molecules by MolBuilder module of INSIGHT II (MSI) program with cvff force field.

The following protocol of calculation was applied for each intercalator studied:

(ia) For simulations in water an intercalator molecule was placed in the middle of a rectangular 3×3×5 nm box. About 1700 water molecules and appropriate number of Cl^- counterions were inserted into the box.

(ib) For simulations of the intercalation complex the DNA fragment with an intercalator inserted between central base-pairs was placed in the middle of a rectangular 3×3×5 nm box. About 1400 water molecules, 22 Na^+, and Cl^- counterions were inserted into the box.

(ii) 100 steps of the steepest descent minimisation were applied to the system. The list of non-bonded interactions was updated after each step.

(iii) The system was thermalised at 300 K by 10 ps MD simulation (5000 steps of 2 fs). The atomic velocities were reassigned according to a Maxwell-Boltzmann distribution at the beginning of the thermalisation. The co-ordinates of an intercalator (and DNA) atoms were constrained to their initial positions. The list of non bonded interactions

was updated every 10 steps. The system was coupled to an external thermal and pressure bath of 300 K and 1 bar with time constants equal to 10 fs (for temperature) and 50 fs (for pressure).

(iv) The system obtained in step (iii) was relaxed by next 10 ps MD simulation. The atomic velocities were also reassigned at the beginning of this step and the system was coupled to thermal and pressure bath. An additional potential which restrained distances related to Watson-Crick hydrogen bonds was used in this and all further steps. The list of non bonded interactions was updated every 20 steps during relaxation and all further simulations.

(v) The system prepared in such a way was next simulated by 100 ps with initial reassignment of atomic velocities. The conditions of simulation were the same as for relaxation. The co-ordinates of the system were saved every 0.4 ps and formed trajectory of 250 geometries.

The energies of Lenard-Jones and electrostatic interactions between ligands and their environment in free as well as bound states used in the LR method were obtained as time averaged values over all 100 ps of the main simulations.

All statistical calculations were done by package STATISTICA PL v.5.1G (StatSoft, Inc.).

3. Results and discussion

The apparent binding constants to calf thymus DNA, K_{app}, were published earlier for all ligands studied [30, 31]. These values have been recalculated to free energy of binding according to the formula:

$$(\Delta G_{bind})^{exp} = -RT \ln K_{app} \tag{5}$$

and are presented in Table 1.

Table 1. Data for and results of LR calculations

Compound	ΔG^{exp}	$<E^{vdw}>_{bound}$	$<E^{vdw}>_{free}$	ΔE^{vdw}	$<E^{elec}>_{bound}$	$<E^{elec}>_{free}$	ΔE^{elec}	ΔG^{calc}
Proflavine	–16.41	–175	–78	–97	–70	–95	25	–18.03
Anthracenedions								
MIT	–44.07	–301	–97	–204	–770	–867	97	–45.38
AMT	–40.00	–286	–113	–173	–773	–737	–36	–37.46
Imidazoacridinones								
C1311	–29.17	–244	–116	–128	–737	–746	9	–25.96
C1330	–28.00	–262	–131	–131	–729	–701	–28	–26.72
C1492	–28.03	–263	–120	–143	–684	–723	39	–29.79
C1558	–24.15	–292	–143	–149	–671	–713	43	–31.32
Pyrimidoacridinetrions								
PA36	–28.37	–292	–147	–145	–715	–667	–48	–30.30
PA39	–38.10	–301	–145	–156	–706	–670	–36	–33.11
PA51	–28.00	–290	–129	–161	–386	–377	–9	–26.21

The time averaged energies of Lenard-Jones and electrostatic interactions between ligands and their environment in free as well as bound states used in this study have

been determined during 100 ps simulations reported earlier [24–26]. The obtained values are presented in Table 1 together with the differences used by LR approach: $\Delta E^{vdw} = <E^{vdw}>_{bound} - <E^{vdw}>_{free}$ and $\Delta E^{elec} = <E^{elec}>_{bound} - <E^{elec}>_{free}$. Åqvist and co-workers in their early works [9–13] determined values of α and β coefficients, Equation (3), as equal to 0.161 and 0.5, respectively. However, next works of these as well as other researchers indicated that the coefficient in LR approach depends upon structural/polar nature of a ligand and a target as well as the particular force field employed [17–21]. Thus, the values of the LR coefficients should be fit empirically by least square regression for each new ligand/target system.

Application of the regression analysis to the data presented in Table 1 results in Equation (6):

$$\Delta G_{bind} = 0.212(\pm 0.008) \times \Delta E^{vdw} + 0.032(\pm 0.027) \times \Delta E^{elec} \quad (6)$$

$n = 10,\ R^2 = 0.989,\ F = 369.89,\ s = 3.63$ kJ/mol.

This equation is characterized by high statistical significance, high value of F statistic and the determination coefficient $R^2 \approx 1$, and predicts free energy of binding with acceptable level of error, standard error of the equation, $s = 3.6$ kJ/mol. However, the β coefficient in Equation (6) is insignificant ($p > 0.26$) and from the statistical point of view should be removed. After deletion of the electrostatic term ΔE_{elec} from Equation (6) the new equation has been obtained:

$$\Delta G_{bind} = 0.211(\pm 0.008) \times \Delta E^{vdw} \quad (7)$$

$n = 10,\ R^2 = 0.987,\ F = 702.41,\ s = 3.72$ kJ/mol.

This equation is characterized by practically the same high statistical significance as Equation (6), however, it contains the significant term, only. This means that Equation (7) should possess better predictability than Equation (6).

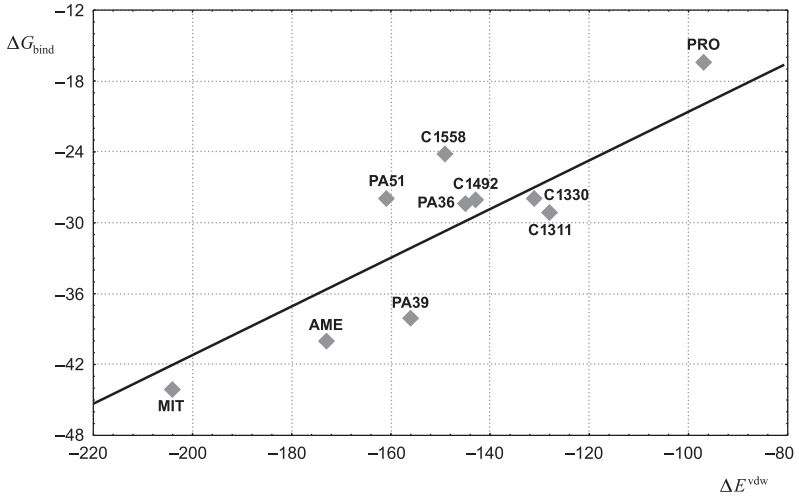


Figure 6. Plot of relationship between experimental ΔG_{bind} values and calculated ΔE^{vdw} terms (Equation (7))

The LR relationship (7) obtained in the present study, Figure 6, is now successfully used in our research group in further molecular modelling studies concerning DNA ligands

with similar structure. However, the fact that in the system studied the electrostatic term seems to have no influence on the free energy of binding is so exceptional that it needs some comments. A more detailed statistical analysis of the data indicated that for the compounds studied the electrostatic term is not correlated with experimental free energies of binding (results not shown). In my opinion, it does not mean that electrostatic interactions are negligible during formation of the complex between cationic ligand and anionic target. This result indicates rather that the cationic and highly polar parts of the ligands arrange complex configuration in such manner that they are surrounded by very similar environment in free as well as bounded states. High and very similar values of $< E^{\mathrm{elec}} >_{\mathrm{bound}}$ and $< E^{\mathrm{elec}} >_{\mathrm{free}}$ terms presented in Table 1 strongly support this opinion. To check the validity of Equation (7) experimental evaluation of free energies of binding as well as molecular dynamics simulations are performed for a new set of intercalators with similar mode of action.

Acknowledgements

This work was supported by the State Committee for Scientific Research (KBN, Poland) grant No. 3 T09A 124 14 and in part by the Faculty of Chemistry, Technical University of Gdansk (Poland).

References

[1] Smith M B K, Rouzer C A, Taneyhill L A, Smith N A, Hughes S H, Boyer P L, Janssen P A J, Moereels H, Koymans L, Arnold A, Ding J, Das K, Zhang W, Michejda C J and Smith R H 1995 *Protein Sci.* **4** 2203
[2] Smith M B K, Michejda C J, Hughes S H, Boyer P L, Janssen P A J, Andries K, Buckheit R W and Smith R H 1997 *Protein Eng.* **10** 1379
[3] Smith R H, Michejda C J, Hughes S H, Arnold E, Janssen P A J and Smith M B K 1998 *J. Mol. Struct.* **423** 67
[4] Beveridge D L and DiCapua F M 1989 *Annu. Rev. Biophys. Biophys. Chem.* **18** 431
[5] Jorgensen W L 1989 *Acc. Chem. Res.* **22** 184
[6] van Gunsteren W F and Berendsen H J C 1990 *Angew. Chem., Int. Ed. Engl.* **29** 992
[7] Stratsma T P and McCammon J A 1992 *Annu. Rev. Phys. Chem.* **43** 407
[8] Kollmann P 1993 *Chem. Rev.* **93** 2395
[9] Åqvist J, Fothergill M and Warshel A 1993 *J. Amer. Chem. Soc.* **115** 631
[10] Åqvist J, Medina C and Samuelsson J E 1994 *Protein Eng.* **7** 385
[11] Åqvist J 1994 *J. Phys. Chem.* **98** 8253
[12] Åqvist J and Mowbray S L 1995 *J. Biol. Chem.* **270** 9978
[13] Hansson T and Åqvist J 1995 *Protein Eng.* **8** 1137
[14] Åqvist J and Hansson T 1996 *J. Phys. Chem.* **100** 9512
[15] Åqvist J 1996 *J. Comput. Chem.* **17** 1587
[16] Hulten J, Bonham N M, Nillroth U, Hansson T, Zuccarello G, Bouzide A, Åqvist J, Classon B, Danielson U H, Karlen A, Kvarnstrom I, Samuelsson B and Hallberg A 1997 *J. Med. Chem.* **40** 885
[17] Hansson T, Marelius J and Åqvist J 1998 *J. Comput.-Aided Mol. Design* **12** 27
[18] Carlson H A and Jorgensen W L 1995 *J. Phys. Chem.* **99** 10667
[19] JonesHertzog D K and Jorgensen W L 1997 *J. Med. Chem.* **40** 1539
[20] McDonald N A, Carlson H A and Jorgensen W L 1997 *J. Phys. Org. Chem.* **10** 563
[21] Smith R H, Jorgensen W L, Tirado-Rives J, Lamb M L, Janssen P A J, Michejda C J and Smith M B K 1998 *J. Med. Chem.* **41** 5272
[22] Prabhakaran M and Harvey S C 1985 *J. Phys. Chem.* **89** 5767
[23] Goodfellow J M, Norberto de Souza O, Parker K and Cruzeiro-Hansson L 1993 *Simulation of Biomolecular Systems* in Computer Simulation of Biomolecular Systems, Ed. van Gunsteren W F, Weiner P K and Wilkinson A J ESCOM, Leiden

[24] Mazerski J, Martelli S and Borowski E 1998 *Acta Biochim. Polon.* **45** 1
[25] Mazerski J, Antonini I and Martelli S 1999 *Acta Biochim. Polon.* **47** 47
[26] Mazerski J and Muchewicz K 1999 *Acta Biochim. Polon.* **47** 65
[27] van Gunsteren W F, Billeter S R, Eising A A, Hunenberger P H, Kruger P, Mark E A, Scott W R P and Tiron I G 1996 *Biomolecular simulation: GROMOS 96 manual and user guide* Vdf Hochschulverlag AG Zurich
[28] Ryckeart J P, Ciccotti G and Berendsen J H C 1977 *J. Comput. Phys.* **23** 327
[29] Subraminian P S, Ravishanker G and Beveridge D L 1991 *J. Amer. Chem. Soc.* **113** 5027
[30] Antonini I, Polucci P, Jenkins T C, Kelland L R, Menta E, Pescalli N, Stefańska B, Mazerski J and Martelli S 1997 *J. Med. Chem.* **40** 3749
[31] Dzięgielewski J 1998 *Mode of antitumor activity of imidazoacridinones*, PhD thesis, Faculty of Chemistry, Technical University of Gdansk, Gdansk