COMPUTING OF CATHEPSIN B – DRUG INTERACTION

Karel Huml, Vladimír Šubr a Karel Ulbrich

Institute of Macromolecular Chemistry, Academy of Sciences of the Czech Republic, 162 06 Praha 6, Czech Republic

Abstract

Empirical energy calculations are carried out to study the binding of two substrates, Gly-Phe-Gly-L-Lys-NAp and Gly-Phe-Gly-D-Lys-NAp, to cathepsin B. The computer-aided molecular modelling methods were chosen to predict the influence of L and D configuration of the lysine in the substrate on the selectivity of formation covalent and noncovalent complexes between the substrates and the active site of the enzyme.

1 Introduction

This work is a part of the study of polymeric drugs pursued in the Institute [1]. Attachment of drugs to macromolecular carriers alters their rate of excretion from the body and provides the possibility of targeting and sustained release over a prolonged period. In most cases, drugs bound via a covalent bond directly to the polymer chain exhibit either a reduced or zero biological activity. For this reason, drugs should be separated from the polymeric backbone by means of a biodegradable spacer. Once the drug conjugate, tailored as a substrate for enzymes present in the site of required action, reaches the target compartment, the drug can be split off more readily in its active form [2]. Prediction of energy of intermolecular and intramolecular interactions is essential for understanding biochemistry at the molecular level or for designing new molecules having predefined properties. Particularly, the L and D stereo selectivity of the cathepsin B-catalysed hydrolysis of peptides was of interest, when L or D residues are in the P_1 position [3] of the substrate (Fig.1). We have studied interactions of two closely related substrates, Gly-L-Phe-Gly-L-Lys-NAp (briefly GFGK-NAp) and Gly-L-Phe-Gly-D-Lys-NAp (briefly GFGk-NAp) with the active site of the enzyme cathepsin B. These two situations will be labelled as L-type and D-type in this article. 4-Nitroaniline, NAp, was used as a drug model which is to be released from the tetrapeptide spacer GFGK (or GFGk) in the cell breaking the peptide bond between lysine and NAp via enzymatic reaction with the intracellular enzyme cathepsin B. X-ray atomic coordinates of considered molecules were chosen from the Cambridge Structural Database, CSD [4] and from the Protein Data Bank, PDB [5]. Energy minimization (EM), quenched molecular dynamics (MD) based on the force field method, and absolute free energy calculation were applied utilizing the BIOSYM suite of computer programs [6]. The influence of the polymer carrier was neglected in all calculations.

2 Theory

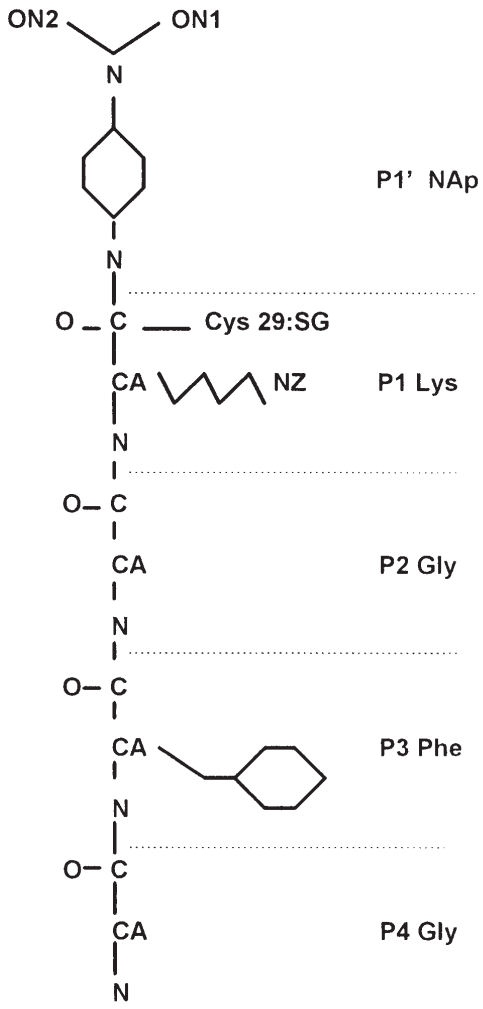

Figure 1: Scheme of the P_4 - P_1' substrate positions: Gly-Phe-Gly-Lys-NAp (numbering according to [3]).

2.1 Reaction path

The mechanism of the cathepsin B catalysis can be divided into two processes : acylation and deacylation. The first process was found to be decisive and, therefore, we have concentrated our attention on its mechanism. During acylation there are, at least, four states which can be modelled :

$$E^f + S^f \rightarrow ES^{nc} \rightarrow ES^{cc} \rightarrow EP^2 + P^1 \quad (1)$$

The first state, $\mathbf{E^f + S^f}$, represents free enzyme, $\mathbf{E^f}$, and free substrate, $\mathbf{S^f}$, coming into the reaction. Then $\mathbf{ES^{nc}}$ is a noncovalent (Michaelis or van der Waals) complex, and $\mathbf{ES^{cc}}$ is a covalent complex represented by a tetrahedral intermediate close to the transition state. Finally, $\mathbf{EP^2}$ is the acyl enzyme, and $\mathbf{P^1}$ is the first product (released drug) of the reaction [7] .

2.2 Covalent complex

Molecular mechanics cannot treat the formation/breaking of covalent bonds, i.e., the electronic structure, correctly. Hence, potential energy of covalently bound substrates, E^{cc}, will be poorly represented [8]. Despite this fact, it is helpful to start the modelling with this step as X-ray data show information concerning the scissile bond position and general features of contacts of the individual subsites of the enzyme active cleft for different inhibitors. The first step in determining the covalent complex was to bring the two molecules together with the knowledge of the valence bond between S_1 Cys29:SG and the C atom of the P_1 Lys. A combination of molecular graphics and quenched MD was used to search for the relative geometry studies of substrate and enzyme.

2.3 Noncovalent complex

For the reasons mentioned above, our main focus was on noncovalent interaction between the enzyme and substrate. Calculation of the binding energy of the noncovalent complex requires minimizing the structures of the complex and the separate species. The resulting binding energy, DE_B, is formally given by

$$\Delta E_B = E_{pl} + (E_1^b - E_1^f) + (E_1^b - E_1^f) \quad (2)$$

where E_{pl} is interaction energy of the minimized protein-ligand (enzyme-substrate) noncovalent complex, E_1^b and E_p^b are internal energies of the bound minimized ligand (substrate) and protein (enzyme) molecule, respectively (Fig.2). Similarly, E_1^f and are internal energies of the free optimized ligand and protein molecules, respectively [9]. The quantity E_1^f is generally smaller than E_1^b, i.e., the positive internal energy of the bound ligand is larger than that of the free ligand. The same situation is true for the protein, so that the binding energy, ΔE_B is smaller in magnitude than E_{pl} .

Let us assume that the binding energy given by Eq.(2) is known for both L- and D-type noncovalent complexes, $\Delta E_B(L)$ and $\Delta E_B(D)$, respectively. Then the difference in binding of the two substrates is given by

$$\Delta\Delta E_B = \Delta E_B(L) - \Delta E_B(D) \quad (3)$$

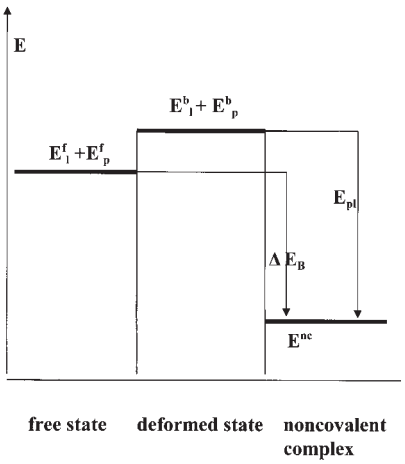

Figure 2: Scheme of the potential energy course along the pathway of the cathepsin B - substrate noncovalent complex formation. The acylation step was formally split into two substeps : conformation (intramolecular deformation), and interaction (intermolecular noncovalent binding). Symbols of energy levels : free protein, E^f_p, free ligand, E^f_l, bound protein, E^b_p, bound ligand, E^b_l, noncovalent complex, E^{nc}.Binding energy, ΔE_B, interaction energy, E_{pl}.

2.4 Free energy

Although the potential energy calculations are of interest, they have number of limitations. The first is inherent in the inaccuracies of the empirical potential energy functions used. Further, there are corrections of thermodynamic properties of the system due to the fact that it is not fixed at the potential energy minimum but is undergoing thermal motion. This means that the average energy is not that associated with the minimum energy structure or even with the average structure, which is the one obtained by X-ray analysis. The average energy corresponds to that obtained by calculating the energy for each of a series of structures and averaging them with the Boltzmann weights appropriate for the system temperature. This average can be determined by molecular dynamics and can be used to estimate the related partition function, Q. Consequently, the free energy A related to the partition function is

$$A = -kT \ln Q \quad (4)$$

where k is the Boltzmann constant and T is the absolute temperature. Integrating the partition function analytically may be possible for simple Hamiltonians. However, for more realistic systems including many non-bond and bond interactions between atoms, an analytic solution is impossible. For this reason, we used thermodynamic integration to determine the change in free energy [6]. Appropriately averaging the results of a molecular dynamics trajectory enables us to calculate the expression

$$A_0 = A_1 + \int < V_0 - V_H > d\lambda \quad (5)$$

where A_1 is the free energy of the Einstein solid identified with the minimum potential energy conformation of the system, V_0 is the normal potential energy function (includ-

ing bonds, angles, torsions, etc), and V_H is a harmonic oscillator site potential. By performing several calculations for many values of λ between 0 and 1, the function can eventually be numerically integrated.

Similarly to Eq.(3), we can compare the free energy of both L and D complexes by the subtraction

$$\Delta A = A_0(L) - A_0(D) \qquad (6)$$

3 Experimental

All computations were performed under the conditions of dielectric constant 4, pH 6, and temperature 300 K. The cvff system of potentials was utilized. In the case of two closely related substrates the water environment is supposed to be the same. Therefore, influence of water molecules was neglected to simplify our calculation.

3.1 Modelling of the covalent complex

(1) Modelling of the L-type substrate was based on the CSD [4] and BIOSYM fragments library [6]. The P_1 Lys:C was set to sp^3 hybridization.

(2) The minimum potential energy model of the L-type substrate was found with the MD method (10 000 steps, 2 fs step, each 20th frame stored).

(3) Cathepsin B model based on the PDB X-ray data [5] was relaxed with the conjugated-gradients energy minimization method.

(4) The minimum potential energy substrate was introduced into the cathepsin B cleft with the manual docking method. The P_1 Lys:C atom of the L-type substrate was covalently bonded at the distance of 181 pm to the Cys 29:SG atom of cathepsin B.

(5) The starting minimum-energy covalent complex was chosen using the MD method when all cathepsin B atoms beyond the 500 pm limit from the substrate were fixed.

(6) Under the same 500 pm constraint, the steepest descent and the conjugated- gradients energy minimization lead to the final model of the covalent complex (Fig.3) defining the starting model for the most probable noncovalent complex calculation.

3.2 Modelling of the noncovalent complex

(7) The P_1 Lys:C - Cys 29:SG bond was broken and the P_1 Lys:C atom was changed into the sp^2 hybridization. Simultaneously, corresponding charges and potentials were modified. Two relatively weak distance restraints were introduced to protect the complex from possible strong repulsion at the beginning of the EM procedure (P_1' NAp:ON2 - His111:NE2, and P_4 Gly:O - Phe174:N). During the EM of the L-type substrate in the covalent complex, the P_1' and

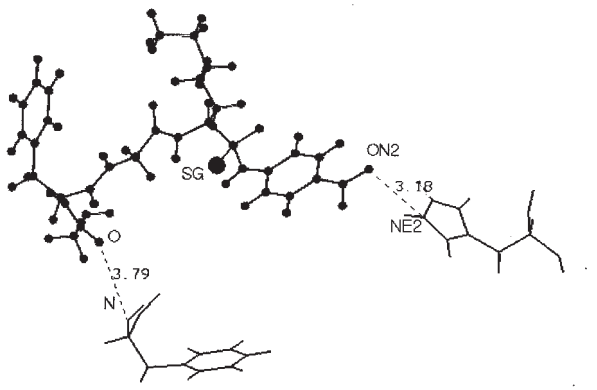

Figure 3: Conformation of the substrate at the active site of the enzyme in the covalent complex cathepsin B-GFGK-NAp. Short distances of P_1' NAp:ON2 - His111:NE2 of 318 pm and P_4 Gly:O - Phe174:N of 379 pm show the orientation of the molecules. The sulfur atom, Cys29:SG, forms the covalent bond with P_1 L-Lys:C of 183 pm characteristic of the tetrahedral intermediate.

the P_4 fragments moved only a little. On the other hand, for the D-type, the P_4 Gly was situated in a different position and, therefore, only the first restraint was applied.

(8) The conjugated-gradients method was used to calculate the minimum-energy noncovalent complex and the corresponding potential energy level E^{nc}.

3.3 Modelling of the free substrate and enzyme

(9) The substrate and the enzyme taken from the noncovalent complex were isolated.

(10)(Both molecules were individually optimized using the conjugated-gradients method. The input data represent isolated but deformed molecules and gives the values of E^b_p and E^b_l. However, the final data represent the relaxed molecules and give the values of E^f_p and E^f_l.

The covalent complex of the L-type substrate with cathepsin B was chosen as the starting model for the D-type substrate. After the change of P_1 L-Lys into the P_1 D-Lys, a procedure similar to the previous (5)-(10) steps was applied.

3.4 Absolute free energy calculation

(11) (For setting up the absolute free energy calculation, the following model was used : (i) all enzyme atoms were fixed, (ii) all substrate atoms were tethered to the minimum potential energy structure with the spring constants of 10 kcal/mol.Ĺ,

(iii) the Gaussian-Legendre quadrature algorithm was applied to integrate Eq.(5) for six different values of λ, and 1000 steps, 1 fs each, of molecular dynamics after 100 steps of equilibration.

4 Results and discussion

4.1 Covalent complex

For the L-type covalent complex, some hydrogen bonds were detected in the P_1 position: Lys:N - Cys29:SG of 288 pm, Lys:N - Gly198:O of 289 pm, Lys:O - Cys29:N of 296 pm, where the Cys:29:N belongs to the "oxyanion hole" [10]. The valence bond P_1 Lys:C - Cys29:SG converged to the value of 183 pm. For the D-type covalent complex, the following hydrogen bonds were detected in the P_1 position: Lys:N - Cys29:SG of 291 pm, Lys:O - Cys29:N of 312 pm, where the Cys29:N belongs to the oxyanion hole. In addition, there is a possible salt bridge Lys:NZ - Asn72:OD1 of 315 pm. In the P_1' position, the NAp:N - Gly198:O hydrogen bond of 299 pm was formed. The values (energy in kcal/mol) calculated for the E^{cc}(L) and E^{cc}(D) are 243.84 and 195.10, respectively. Therefore, the D- covalent complex has lower potential energy by 48.74 with respect to the L one.

4.2 Noncovalent complex

For the L-type noncovalent complex, no hydrogen bond between the enzyme and the substrate was detected. The P_1 Lys:C – Cys 29:SG distance converged to the value of 333 pm. For the D-type substrate, two hydrogen bonds appeared : P_1'NAp:ON2 – His 111:NE2 of 308 pm, and P_1 Lys:N – Gly 198:O of 301 pm. In addition, a salt bridge can be formed, such as P_1 Lys:NZ – Asn72:OD1 of 305 pm. The P_1 Lys:C – Cys 29:SG distance was 355 pm. From Table 1 (energy in kcal/mol) it follows that the deformation potential energy produced in the conformational change is 12.46 for the L-type and 17.05 for the D-type substrates, respectively. Similarly, the potential part of the interaction energy is -72.14 and -118.14 for the L-type and D-type substrates, respectively. The potential part of the binding energy is -59.68 and -101.09 for the L-type and D-type, respectively. Therefore, the $\Delta\Delta E_B$, according to Eq.(3) is 41.41. In accord with these results, we can expect the D-type substrate to be better than the L-type for the noncovalent complex.

Table 1. MD potential energy levels (kcal/mol) for free protein, E^f_p, free ligand, E^f_l, bound protein, E^b_p, bound ligand, E^b_l, and noncovalent complex, E^{nc}, for both the L-type, and D-type models.

Model	E_p^f	E_1^f	E_p^b	E_1^f	E^{nc}
L-type	160.96	123.45	167.87	129.00	224.73
D-type	156.25	128.27	160.49	141.08	183.34

4.3 Absolute free energy

The final value of A_0 for the L-type noncovalent complex was 211 kcal/mol, and that for the D-type, 177 kcal/mol, the difference of 34 kcal/mol supporting the D-type noncovalent complex as the low-free-energy system.

5 Conclusion

Our results show that both L- and D-type ligands are accepted as substrates in the cathepsin B active cleft and suggest that the D-type ligand is a better substrate in comparison with the L-type for the noncovalent complex. Our computing was based on rather rough approximations mentioned above. Therefore, to support the results obtained, it is necessary to use more extensive calculations, e.g., 1-fs step, more frequent frame storage, longer simulation time (at least 10 ns) at higher temperatures followed by the annealing procedure, and to consider the influence of water surrounding the system. Particularly, the model would be more reliable if all the low potential energy conformations obtained by molecular dynamics were considered after application of the cluster analysis. In any case, the final check of the model will be a biochemical experiment, which, unfortunately, is expensive and time consuming.

The authors are indebted to Professor E. Chiellini and Professor A.M. Bianucci (University of Parma) for their help in their computer modelling. This work was supported by the Grant Agency of the Czech Republic (grant 307/96/K226 and grant 203/96/0111) and by COPERNICUS (grant ERB 3512 PL 941009).

References

1. K. Ulbrich, *J.Bioact.Compat.Polym.*, **6** (1991) 348.
2. K. Huml, *Mater. Struct.*, **3** (1996) 283.
3. I. Schechter & A. Berger, Biochem. *Biophys. Res. Commun.*, **27** (1967) 157.
4. E.H. Allen, S. Bellard, M.D. Brice, B.A. Cartwright, A. Doubleday, H. Higgs, Hummelink, B.H.Hummelink- Peters, O. Kennard, W.D.S. Motherwell, J.R. Rodgers & D.G. Watson, *Acta Crystallogr.*, **B35** (1979) 2331.
5. F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F.Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi & M. Tasumi, *J.Mol.Biol.*, **112** (1977) 535.
6. BIOSYM/MSI, release 95 (1995),San Diego, CA 92121-3752, USA.
7. G. Wipff, A. Dearing, P.K. Weiner, J.M. Blaney & P.A. Kollman, *JACS*, **105** (1983) 997.
8. Ajay & M.A. Murcko, *J.Med.Chem.*, **38** (1995) 4954.
9. Brooks,C.L., Karplus,M. & Pettitt,B.M., *Adv.Chem.Phys.*, **71** (1988) 1.
10. D. Musil, D. Zucic, D. Turk, R.A. Engh, I.Mayr, R. Huber, T. Popovic, V.Turk, T. Towatari, N. Katunuma & W. Bode, *EMBO Journal,* **10** (1991) 2321.