Prediction of activities in enzyme-catalyzed reactions from *ab initio* calculations

M. Fermeglia[1], S. Pricl[1], L. Gardossi[2], P. Braiuca2, C. Eberth[2], P. Halling[3]

[1]*Computer-aided System Laboratory, Department of Chemical, Environmental and Raw Materials Engineering, University of Trieste – Piazzale Europa 1, 34127 Trieste, Italy*
2. Department of Pharmaceutical Sciences, University of Trieste - Piazzale Europa 1, 34127 Trieste, Italy
3. sòekjldfboàe

URL: www.caslab.units.it *e-mail: maufp@dicamp.units.it*

Biocatalysis in organic media is an important technique for the synthesis of stereoselective compounds. An important feature of biocatalysis in organic solvents is that it overcomes the difficulty of dissolving hydrophobic substances in aqueous media, and it also essentially eliminates hydrolytic reactions. Many of the reactions involve relatively simple species (esters, acids, alcohols, and water) and for such systems classical methods such as UNIFAC might give good predictions of activity data. There are many other systems, however, which can be somewhat tricky - e.g. glycerol esters and amide and peptide hydrolysis - for which UNIFAC has been shown to give poor predictions or where UNIFAC group parameters are not well founded.
In this work we apply a procedure, based on the integration of concepts from quantum chemistry, dielectric continuum models, electrostatic surface interactions and statistical thermodynamics, for the a priori calculation of activities in enzyme-catalyzed reactions, and compared the results with both experimental evidences and UNIFAC predictions

Keywords: free energy calculations, molecular simulations, biological systems, protein stability, binding

1 INTRODUCTION

Free energy is arguably the most important general concept in physical chemistry. The free energies of molecular systems describe their tendency to associate and react. Thus, being able to predict this quantity using molecular theory in general would be an enormously important advance and is a seductive goal [1,2] The statistical mechanical definition of free energy is in term of the partition function, a sum of Boltzmann weights of all the energy levels of the systems. However, only for the simplest model systems can this free energy be represented by an analytical function. One can write a classical analog of the quantum mechanical partition function where the energy is viewed as a continuous function, rather that discrete. This is likely to be a good approximation in most systems involving non-covalent interactions near room temperature. Unfortunately, the free energy thus represented requires an integration over all $3N$ degrees of freedom, where N = number of atoms in the system. However, if one focuses on free energy differences between related systems A and B ($\Delta G = G_B - G_A$) represented by the respective Hamiltonians H_B and H_A, this free energy can be represented as:

$$G_B - G_A = \Delta G = -RT \ln \left\langle e^{-\Delta H / RT} \right\rangle_A \qquad (1)$$

where $\Delta H = H_B - H_A$ and $< >_A$ refers to an ensemble average over a system represented by Hamiltonian H_A. Equation (1) is the fundamental equation of free energy perturbation (FEP) calculations.
If systems A and B differ in more that a trivial way, then Eq. (1) will not lead to a sensible free energy. One can, however, generalize the problem and describe the Hamiltonian $H(\lambda)$ as in Eq. (2):

$$H(\lambda) = \lambda H_B + (1-\lambda) H_A \qquad (2)$$

where λ can vary from 0 ($H = H_A$) to 1 ($H=H_B$). One can then generalize Eq. (1) as follows:

$$\Delta G = G_B - G_A = \sum_{\lambda=0}^{1} - RT \ln \left\langle e^{-\Delta H'/RT} \right\rangle_\lambda \quad (3)$$

where $\Delta H' = H_{\lambda+d\lambda} - H_\lambda$. One breaks up the free energy calculations into windows, each one involving a small enough interval in λ to allow the free energy to be calculated accurately.

An alternative to free energy perturbation calculations is thermodynamic integration (TI), where the free energy difference between two systems (one characterized by $H = H_A$ or λ = 0, and the other by $H = H_B$ or λ = 1 in Eq. (2) is given by:

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (4)$$

The application of Eq. (4) requires one to evaluate the ensemble average of the derivative of the Hamiltonian with respect to λ, $\langle \partial H/\partial \lambda \rangle_\lambda$ at various values of λ.

The third commonly used method for free energy calculations is called slow growth, in which the Hamiltonian is changed an infinitesimal amount over each step of the simulation:

$$\Delta G = \sum_{\text{n. steps}} \sum_{\lambda=0}^{\lambda=1} \left(H_{n+1} - H_n \right) \quad (5)$$

where H_n is the Hamiltonian for a given λ and H_{n+1} is the Hamiltonian for the next larger λ. Slow growth equation can be derived both from FEP or TI equations, using the assumption in Eq. (1) that ΔG is small, and in Eq. (4) that $\partial H/\partial \lambda = \Delta H/\Delta \lambda$.

If evaluated accurately enough, ΔG should be independent of path or simulation protocol; thus, there are just a number of practical reasons for using one of these three approaches.

All methods above, although highly precise, are extremely CPU-intensive, and hence are to be used only when a few data are to be obtained. In the field of drug discovery, for instance, many compounds are to be screened for activity towards a given receptor. Accordingly, alternative methods for free energy calculations must be considered to confine simulations into reasonable times.

The ability to calculate structure and free energy of binding of a complex system for a large screening is mainly based on the molecular mechanics/Poisson Boltzmann surface area approach (MM/PBSA) [2]. In this method, one carries out a molecular dynamics simulation, typically in a periodic box with water and counterions, and correct representation of long-range electrostatics such as PME, saving a set of representative structures. Then, one postprocesses these structures, removing any solvent and counteion molecules, and calculates the free energy, $\underline{G}$, according to the following equation:

$$\underline{G} = \underline{E}_{MM} + \underline{G}_{PBSA} - TS_{MM} \quad (6)$$

in which $\underline{G}$ is the calculated average free energy, and $\underline{E}_{MM}$ is the average molecular mechanical energy:

$$\underline{E}_{MM} = \underline{E}_B + \underline{E}_A + \underline{E}_T + \underline{E}_{vdW} + \underline{E}_{ele} \quad (7)$$

where these correspond to the bond, angle, torsion, van der Waals and electrostatic terms in the molecular mechanical force filed, evaluated with no nonbonded cutoff. $\underline{G}_{PBSA}$ is the solvation free energy calculated with a numerical solution of the Poisson-Boltzmann equation, and an estimate of the nonpolar free energy with a simple surface area term. $-TS_{MM}$ is the solute entropy, which can be estimated by quasi harmonic analysis of the trajectory or, in selected cases, by using normal mode analysis.

In what follows, I will present a discussion of some applications of the above outlined free energy calculations to molecules of biological interest. Here I highlight a number of themes which will be presented: (1) ligand binding to proteins [3-7], (2) sequence-dependent stabilities of proteins [8-11], and (3) enantioselective enzymatic catalysis [9-12].

2 APPLICATIONS

2.1 The case of RNR inhibitors [4]

The ribonucleotide reductase (RNR) is a crucial enzyme in the *de novo* synthesis of DNA; it converts all ribonucleoside diphosphates (NDP) into the corresponding 2'-deoxyribonucleotides (dNDP) in prokaryotic and eukaryotic cells. Viral RNRs are also known: Herpes simplex virus (HSV)- and Varicella zoster virus (VZV)-infected cells express a viral ribonucleotide reductase, distinct from that present in uninfected cells, which is endowed with a proper enzymatic activity that ensure sufficient 2'-deoxynucleotide supply and DNA synthesis. Over the last decade, this enzyme has gained increasing interest as an important target in the control of the replication of neoplastic cells as well as of pathogenic viral agents.

Accordingly, we have applied the MM/PBSA method to verify both substrate specificity and contribution to biological activity of a series of 5'-

phosphono acetic acid, amide and ester analogs of adenosine, uridine and cytidine. From this standpoint, the enzyme was studied with different isosteric nucleotide analogues (heterocycles, phosphonic acid, esters and amides) of the four natural substrates (ADP, GDP, CDP and UDP). These were docked into the binding site of the R1 subunit by modifying the 5'-position of the GDP contained in the crystallographic structure, by substitution of the diphosphate group with the appropriate bioisosteric residue. Figure 1 reports, as a graphical example, the comparison between the structures of the **5a**/R1 and the crystallographic GDP/R1 complexes, respectively.

The resulting structures of the enzyme, complexed with the considered nucleotide analogues, presented interesting data on the point of view of the binding energy, as shown in Table 1. In fact, the relevant calculations indicated an high stability of the substrate-enzyme complexes in the case of the amide and ester derivatives (e.g., compounds **8b**, **5a** and **14**); in particular, a detailed analysis of these structures revealed an increased number of hydrogen bonds as compared to the natural substrate suggesting a potential inhibition activity. The relevant experiments revealed an excellent agreement with the data predicted by modeling (compare columns 2 and 3 in Table 1), thus yielding an extremely useful tool for the development of further, more potent drugs starting from these leads and with ad hoc modifications, as suggested by further modeling studies.

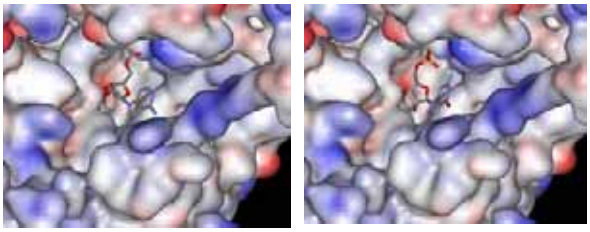


Fig. 1. The active-binding site cleft of the RNR R1 subunit with the docked nucleotide analogue **5a** (left) and with the natural substrate molecule GDP (right). The docked molecules are shown in stick representation.

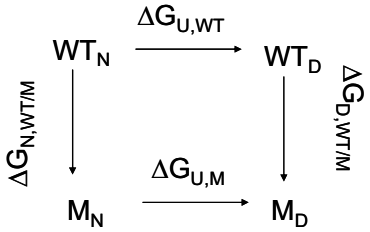
Table 1. Inhibitory activity and binding energies of the selected compounds. NI = not inhibitory

Drug	IC_{50} (μM).	$\Delta G_{calc.}$ (kcal/mol)
8a	655	- 5.35
8b	300	- 7.74
5a	500	- 6.29
5b	NI	- 5.15
6a	760	- 5.08
6b	NI	- 4.94
7a	880	- 5.20
7b	NI	- 4.70
14	600	- 6.14
14b	NI	- 5.04
15	773	- 5.06
15b	NI	- 4.92

2.2 The guardian of the genome: p53 [9-12]

Inherent in models for progression of mammalian cells towards malignancy is the concept that an initial mutation in an important regulatory gene (protein) may be pivotal in this process. Once the initial mutation is introduced, loss of normal gene function or the acquisition of deleterious functions may lead to additional mutations furthering the malignant transformation of the cell. A candidate for the involvement in this process is the tumor suppressor gene, p53. The P53 protein provides one of the key regulatory elements monitoring genomic integrity in mammalian cells and is involved in a multiplicity of cellular functions. Tumor-specific p53 mutations were first identified in 1989. Loss of p53 function is the most common event in human cancer, with more than half of all invasive tumors involving the decrease or total loss of p53 functions. In contrast to many other tumor suppressors, which are often inactivated by deletion or frameshift mutations, most of mutations in P53 are point mutations (missense mutations: 75%; nonsense mutations: 8%). These mutations are exceptionally diverse in their nature and position. Thus, it is possible to draw tumor-specific mutation spectra that show significant differences from one type of cancer to the other. This observation has two very important implications: first, the spectrum of mutations reveals information on the mutagenic process that cause human cancer, and second, the whole set of mutations observed in cancer can be analyzed as an immense, in vivo, random mutagenesis experiment aimed at identifying residues which are important in the maintenance of the tumor suppressive function of the protein.

By applying the following thermodynamic cycle:



we were able to calculate, in outstanding agreement with the experimental evidences, the relative

thermodynamic stability of a large series of P53 mutants, as reported in Table 2.

Table 1. Comparison of experimental and calculated free energy of denaturation for different P53 mutants.

Mutation	ΔΔG exp. (kcal/mol)	ΔΔG calc. (kcal/mol)
R248Q	1.87 ± 0.09	2.04 ± 0.14
R273H	0.45 ± 0.04	0.22 ± 0.15
F134L	4.78 ± 0.08	4.90 ± 0.11
G245S	1.21 ± 0.03	1.50 ± 0.19
R249S	1.92 ± 0.04	2.20 ± 0.15
R282W	3.30 ± 0.10	3.42 ± 0.16
R175H	3.52 ± 0.06	3.69 ± 0.11
M237I	3.18 ± 0.06	3.30 ± 0.14
C242S	3.07 ± 0.05	3.27 ± 0.15
C238Y	-	3.20 ± 0.17

Mutation C238Y was discovered and analyzed for the first time by our jointed group of research Cas-Lab/Istituto Tumori di Milano, and is involved in malignant peripheral nerve sheath tumors. As P53 carrying the C238Y mutation was found to be in complex with overexpressed MDM2 *in vivo*, we verified the hypothesis that this mutation is able to preserve the general structure of the protein, in agreement with the cytogenetical evidence.

2.3 A genial tool in enantioselective catalysis: α-chymotrypsin [13-17]

In a series of work, the interaction of different enantiomeric couples β and γ lactamic esters, with the α-chymotrypsin active site has been simulated using a fine-tuned automated docking procedure, subsequently refined by quenched molecular dynamics.

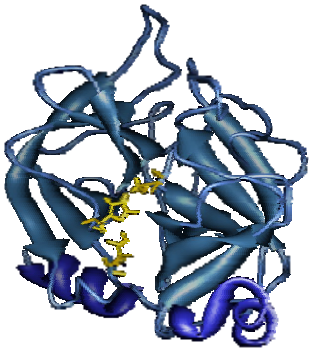


Fig. 2. Energy-relaxed model of three-dimensional structure of α-CT (ribbon model). The catalytic amino acid triad is highlighted using a stick style.

By applying a combination of molecular mechanics energy derived from MD simulations in explicit solvent, and solvation free energy derived from a continuum solvation model, we have calculated reasonable absolute free energies of binding for all α-CT/enantiomer complexes formation. Further, we applied a quantum mechanical/free energy perturbation method (QM/FE), coupled with free energy component analysis, aimed at investigating the roles played by different parts of the catalytic center in the experimentally observed enantioselective power of α-chymotrypsin towards the lactamic esters. According to the resulting evidence, we speculated that the enantioselectivity in the hydrolysis of α-CT towards these compounds may arise also from the fact that, in the case of the recovery of esters, the ester bond, once cleaved, is much more likely to re-ligate to form the intact lactam ester than to form the acyl-enzyme and have the newly formed methyl alcohol molecule diffuse away. In contrast to this, for hydrolyzing compounds the corresponding cleaved lactamic ester substrate moves away from the vicinity of the acyl C=O group, favoring hydrolysis of the acyl-enzyme intermediate.

ACKNOWLEDGMENTS

I am very much indebted to several collaborators, students and postdocs, who have been the source of the research in which I had the pleasure to participate. I will just mention only those whose work is featured in this paper: Drs. Silvana Pilotti, Elena Tamborini, Domenico Delia and Marco A. Pierotti of the Istituto Nazionale per lo Studio e la Cura dei Tumori di Milano are long-time collaborators in several, stimulating common project on mutated proteins in tumors. Dr. Fulvia Felluga and the group of Prof. Ennio Valentin of the Chemistry Department of Trieste University, offer superb study opportunities on enantioselective catalysis. Prof. Stefano Manfredini, Paolo La Colla and all people from Idenix Pharmaceuticals (USA) are wonderful persons to set up interdisciplinary research, leading new drugs to the clinical trials. Finally, the staff of the Computer-aided Systems Laboratory, Prof. Maurizio Fermeglia and the PhD students Alessandro Coslanich, Marco Ferrone and Maria Silvia Paneni, without whom this growing reality of molecular simulations in chemical engineering would have never existed. To all of the above, and many others, I give my thanks.

I also gratefully acknowledge the generous, financial research support from MIUR (FIRB), AIRC, ISS and EU.

This work is dedicated to the memory of Prof. Peter A. Kollman, unparalleled master of free energy calculation-based molecular simulations in biological systems, prematurely deceased of cancer on May, 2002.

BIBLIOGRAFIA

1. P.A. Kollman, *Chem. Rev*. 93 (1993) 2395.

2. P.A. Kollman et al., *Acc. Chem. Res*. 33 (2000) 889.
3. S. Pricl et al., *Science* (2004) submitted.
4. S. Pricl et al., *Antiviral Res*. 30 (2003) 30.
5. S. Pricl et al., *Arkivoc* 5 (2004) 231.
6. S. Pricl et al., *Antiviral Chem. Chemother*. 14 (2003) 181.
7. S. Pricl et al., *J. Mol. Graph. & Mod*. 22 (2004) 229.
8. S. Pricl et al., *Proceedings of the 225th ACS National Meeting*, New Orleans, LA, USA, paper # MEDI 74 (2003).
9. S. Pricl et al., *Oncogene* (2004) in press.
10. S. Pricl et al., *Proceedings of the 225th ACS National Meeting*, New Orleans, LA, USA, paper # MEDI 140 (2003).
11. S. Pricl et al., *Proceedings of the AIChE Annual Meeting*, San Francisco, CA, USA, paper # 107g (2003).
12. S. Pricl et al., *Proceedings of the AIChE Annual Meeting*, Indianapolis, IN, USA, paper # 299c (2002).
13. S. Pricl et al., *Tetrahedron: Asymmetry* 14 (2003) 3385.
14. S. Pricl et al., *Proceedings of the 224th ACS National Meeting*, Boston, MT, USA, paper # ORG 0401 (2002).
15. S. Pricl et al., *Helv. Chim. Acta* 85 (2002) 4046.
16. S. Pricl et al., *Tetrahedron: Asymmetry* 13 (2002) 475.
17. S. Pricl et al., Tetrahedron Asymmetry 12 (2001) 3241.