Pergamon

Bioorganic & Medicinal Chemistry Letters 12 (2002) 267–270

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

Estimation of Binding Affinities for Celecoxib Analogues with COX-2 via Monte Carlo-Extended Linear Response

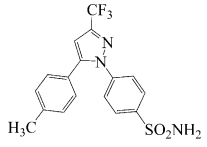
Steven S. Wesolowski and William L. Jorgensen*

Department of Chemistry, Yale University, New Haven, CT 06520-8107, USA

Received 6 November 2001; accepted 6 December 2001

Abstract—Monte Carlo (MC)-extended linear response (ELR) calculations have been used for prediction of binding affinities of celecoxib analogues with the COX-2 enzyme. Three physically motivated descriptors from the MC simulations were used in a regression equation to fit 45 experimental activities with $r^2 = 0.71$ and $q^2 = 0.68$. The ELR approach provides a promising screen for optimization of enzyme inhibitors.

Selective inhibition of the COX-2 enzyme has been the central theme in the recent development of nonsteroidal antiinflammatory drugs that treat osteo- and rheumatoid arthritis while simultaneously providing gastrointestinal tolerability.[1] In addition to its expansive role in arthritis treatment, COX-2 inhibition has been recognized as one of the most promising hypotheses for increased protection against various forms of colorectal cancers including those stemming from familial adenomatous polyposis (FAP).[2] Known COX-2 inhibitors such as celecoxib are currently being used in numerous clinical tests that are exploring not only the connection between COX-2 inhibition and cancer, but also its effect on Alzheimer's and other diseases.[3–5] New and potentially more effective classes of COX-2 inhibitors are under continual investigation and development.



celecoxib

Rigorous computational methods such as free energy perturbation (FEP) approaches have demonstrated that experimental binding affinities for celecoxib analogues can be reproduced well.[6] These simulations have also clarified the preferred conformation of the phenylsulfonamide moiety in the binding site and the origin of COX-2/COX-1 selectivity.[6,7]

*Corresponding author. Tel.: +1-203-432-6278; fax: +1-203-432-6299; e-mail: bill@adrik.chem.yale.edu

While FEP calculations can provide accurate relative free energies of binding, the number of inhibitors and range of structural changes investigated are often limited due to the relatively high computational demands of the method. In an effort to explore larger sets of COX inhibitors while retaining the three-dimensional structural and energetic information of Monte Carlo methods, a more efficient extended linear response (ELR) approach was pursued here for a series of 50 celecoxib analogues. A corresponding scoring function based on several physically motivated properties from the MC simulations was developed.

Theoretical Methods

Binding affinities were computed using a modification of the linear response approach originally proposed by Åqvist.[8] This protocol is considerably more efficient than the FEP or thermodynamic integration alternatives since a series of intermediate transformation processes is not necessary to compute the binding affinities. In the original linear response approach, the free energy of binding (ΔG_b) was cast simply as a linear combination of two physically motivated descriptors:

$$\Delta G_b = \alpha\langle\Delta E_{vdW}\rangle + \beta\langle\Delta E_{Coul}\rangle \quad (1)$$

In eq (1), α and β are constants and $\langle\ \rangle$ represents an ensemble average of the difference in van der Waals (Lennard–Jones) and Coulombic interaction energies (ΔE) in the bound and unbound states as shown in Figure 1.

0960-894X/02/$ - see front matter
PII: S0960-894X(01)00825-3

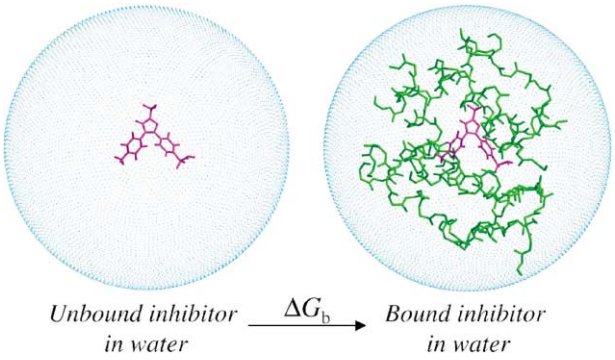


Figure 1. Schematic representation of the bound and unbound environments for COX-2 inhibitors. Simulations were performed in both the unbound state (inhibitor + water) and in the bound state (inhibitor + enzyme + water). Although not illustrated, explicit water molecules are using during the simulations.

This approach has been generalized to include other potentially important descriptors beyond Lennard–Jones and Coulombic contributions.[9,10] So-called 'extended' linear response (ELR) may be expressed as

$$\Delta G_b = \sum_n c_n \xi_n + \text{constant} \qquad (2)$$

where c_n are optimized coefficients and ξ_n are values of physical descriptors from MC or molecular dynamics simulations. The increased flexibility of the descriptors provides a means for including additional relevant terms such as changes in solvent-exposed surface areas and numbers of hydrogen bonds.[9,10]

A set of experimental data[11] was used to generate a multivariate fit of ΔG_b as a linear function of descriptors from the MC simulations. Although the ELR approach relies on experimental data to train the scoring function, no additional experimental data are required for prediction of binding free energies of novel compounds once a reliable fit is obtained. Only simulations of the bound and unbound inhibitors are needed to obtain values for the descriptors of new inhibitors which are used in the regression equation to predict ΔG_b.

The ELR approach was examined using a 148-residue COX-2 enzyme model of the murine 1cx2 crystal structure[12] from the Brookhaven Protein Data Bank, which is described in detail elsewhere.[6] Ensemble averages of descriptors in the bound state were obtained via MC simulations consisting of 1 million configurations of solvent-only equilibration, 10 million configurations of full equilibration, and 10 million configurations of averaging. Each unbound simulation consisted of an initial 1 million configurations of solvent-only equilibration at the experimental temperature of 25 °C. Subsequently, the unbound systems were subjected to an annealing protocol including heating, equilibration and averaging stages in order to facilitate convergence of average energetics.[9] The first stage consisted of 5 million configurations of equilibration with an elevated inhibitor temperature of 727 °C (1000 K) to promote greater conformational sampling. During this stage, the bond

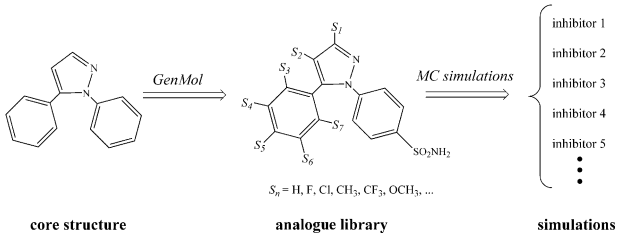


Figure 2. Schematic of the combinatorial capacity of the *GenMol*-ELR approach. Both the free and enzyme-bound inhibitors are generated using *GenMol*.

lengths and bond angles of the inhibitor were held fixed (i.e., only the dihedral angles, translations and total rotations of the inhibitors were sampled). An additional 5 million configurations of full equilibration at 25 °C was performed followed by 10 million configurations of averaging. The heating, equilibration and averaging stages of the annealing cycle were repeated five times.

The core structure of the 1,5-diarylpyrazole was initially positioned in the binding pocket to match the 1cx2 crystal structure. Celecoxib analogues were generated from the core structure using the *GenMol* program[13] which automatically docks and selects complexes with energetically favorable ligand conformations (see Fig. 2).

A brief conjugate-gradient minimization using a distance-dependent dielectric of $4r$ was performed on each complex prior to MC sampling. Each complex and unbound inhibitor was encapsulated in a 22-Å sphere of 725 and 1483 TIP4P water molecules, respectively. The OPLS-AA force field[14] was used throughout, and the partial charges on atoms at points attaching the inhibitor core to the substituent were adjusted to maintain the overall neutrality of the inhibitors. All MC simulations were performed using the *MCPRO* program.[15]

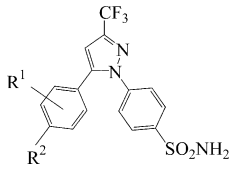
Results and Discussion

The model was trained on 45 inhibitors which were partitioned into two classes. Class 1 consists of 5-aryl-substituted celecoxib analogues and preserves a trifluoromethyl group at the 3-position of the heterocycle (Table 1).

Class 2 consists of 3,4-substituted compounds with variations at the para position in the 5-aryl ring (Table 2). The focus here was on testing the ELR method for optimization of small substituents on aromatic rings. Previously, the utility of the ELR method for larger structural changes has been demonstrated.[9,10]

Functional groups that may hydrolyze (methyl esters) or change protonation state (carboxylic acids) were specifically avoided. We note that five additional compounds (**C03**, **C11**, **C17**, **C40**, and **C43**) were considered, but they were excluded from the fitting set. These are all predicted to bind more strongly than the experimentally derived values. There are unusual patterns in the experimental data for these compounds. For example, the activities for the **C02**, **C03** pair versus **C07**, **C08** or

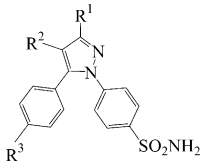
Table 1. Class 1: 5-Aryl-substituted celecoxib analogues



No.	R^1	R^2	IC_{50}[a]	ca. ΔG_{exptl}[b]	ΔG_{calcd}[c]
C01	H	H	0.032	−10.22	−9.75
C02	2-F	H	0.058	−9.87	−10.39
C03	3-F	H	7.73	−6.97	—
C04	H	F	0.041	−10.08	−9.59
C05	2-Cl	H	0.056	−9.89	−10.37
C06	H	Cl	0.01	−10.91	−11.39
C07	2-Me	H	0.069	−9.77	−9.90
C08	3-Me	H	0.11	−9.49	−10.58
C09	H	NO_2	2.63	−7.61	−8.45
C10	H	Me	0.040	−10.09	−11.17
C11	H	Et	0.86	−8.28	—
C12	2-OMe	H	0.29	−8.92	−9.06
C13	H	OMe	0.008	−11.05	−9.78
C14	H	SMe	0.009	−10.98	−10.08
C15	H	NH_2	0.34	−8.82	−8.28
C16	2-NMe_2	H	14.3	−6.61	−7.06
C17	H	CF_3	8.23	−6.94	—
C18	H	NHMe	0.016	−10.64	−8.96
C19	H	NMe_2	0.0047	−11.36	−11.04
C20	H	CH_2OH	93.3	−5.50	−7.04
C21	3-Me	OMe	0.0093	−10.96	−9.95
C22	3-OMe	OMe	0.60	−8.49	−9.17
C23	3-Et	OMe	0.43	−8.67	−10.42
C24	3-Me	SMe	0.0037	−11.50	−10.80
C25	3-F	NMe_2	0.0057	−11.25	−9.87
C26	3-Cl	NHMe	0.027	−10.33	−9.68
C27	3-Cl	Cl	0.015	−10.67	−10.83
C28	2-Cl	Cl	0.056	−9.89	−10.00
C29	2-Me	Me	0.12	−9.44	−10.35
C30	3-Cl, 5-Me	OMe	0.066	−9.80	−10.23

[a]Enzyme assay IC_{50} values at 25 °C (see ref 11).
[b]Estimated experimental binding free energies $\Delta G_{exptl} \approx RT \ln (IC_{50})$ in kcal/mol.
[c]Computed binding free energies (kcal/mol) from extended linear response fit.

Table 2. Class 2: 3,4,(5-aryl)-substituted celecoxib analogues



No.	R^1	R^2	R^3	activity[a]	ca. ΔG_{exptl}[b]	ΔG_{calcd}[c]
C31	CH_2OH	H	Cl	0.83	−8.30	−7.13
C32	CN	H	F	0.34	−8.82	−9.38
C33	CF_3	Cl	Cl	0.0053	−11.29	−10.95
C34	CF_3	F	H	0.0017	−11.96	−10.23
C35	CF_3	Me	Cl	0.022	−10.45	−11.05
C36	CF_3	Et	Cl	0.028	−10.30	−10.75
C37	CF_3	OMe	H	0.08	−9.68	−9.37
C38	CF_3	OH	H	3.58	−7.43	−8.09
C39	H	Cl	H	0.049	−9.97	−10.06
C40	Me	H	H	62.8	−5.73	—
C41	H	Br	Cl	0.031	−10.24	−10.16
C42	H	F	H	4.66	−7.27	−8.61
C43	H	Me	H	47.1	−5.90	—
C44	H	CN	Me	0.076	−9.71	−8.74
C45	H	SO_2Me	Cl	19.8	−6.42	−6.70
C46	H	NH_2	H	29.7	−6.18	−6.18
C47	Me	Cl	H	0.028	−10.30	−10.50
C48	CH_2OH	Cl	Cl	0.34	−8.82	−8.49
C49	CN	Cl	Cl	0.010	−10.91	−10.11
C50	$CONH_2$	Cl	Cl	1.09	−8.13	−8.21

[a]Enzyme assay IC_{50} values at 25 °C (see ref 11).
[b]Estimated experimental binding free energies $\Delta G_{exptl} \approx RT \ln (IC_{50})$ in kcal/mol.
[c]Computed binding free energies (kcal/mol) from extended linear response fit.

C28, **C27** seem inconsistent as do the activities for **C39**, **C40** and **C47**, and for **C39**, **C43** versus **C33**, **C35**.

The best regression equation obtained utilizes only three physical descriptors, which have high statistical significance (Prob F < 0.007).

$$\Delta G_{calcd} = 0.16\langle EXX_{LJ}\rangle - 1.65\langle \Delta HB_{tot}\rangle + 0.007 \times \langle \Delta FOSA\rangle - 7.87$$

Here, EXX_{LJ} is the inhibitor–protein Lennard–Jones (LJ) interaction energy, ΔHB_{tot} is the change in the total number of hydrogen bonds for the inhibitor upon binding, and $\Delta FOSA$ is the change in hydrophobic component of the solvent-accessible surface area of the inhibitor upon binding. All three descriptors and their associated signs are physically intuitive. More favorable inhibitor–protein LJ interactions indicate a good steric fit and enhance binding. Loss of hydrogen bonds in the binding pocket relative to the freely hydrated state is energetically unfavorable, and a 1.65 kcal/mol penalty is applied for each H-bond lost. Compounds **C20**, **C45**, and **C31** show the greatest loss of hydrogen bonding with penalties of 8.9–9.6 kcal/mol. Finally, solvent-accessible surface area is always reduced upon binding, and burial of the hydrophobic component enhances binding. This is simply a manifestation of the hydrophobic effect.

The dimethylamino analogue **C19** is observed and predicted to be one of the best inhibitors in part due to the positioning of the methyl groups with respect to Y385 and W387. The LJ energy for **C19** is among the most favorable for the inhibitors studied. In contrast, the amino derivative **C15** has one of the weakest LJ interactions among the *para*-substituted inhibitors, which is reflected in its poor observed and predicted activity. The methylamino analogue **C18** has intermediate LJ interactions and a corresponding activity between these two extremes. Loss of hydrogen bonding is also relevant as the penalty for the amino derivative is 0.6 kcal/mol greater than that of the dimethylamino analogue.

Overall, for this set of 45 inhibitors, a squared correlation coefficient $r^2 = 0.71$ was obtained and provides one measure of the quality of the fit (Fig. 3). The average error is only 0.67 kcal/mol. Furthermore, cross-validation via the leave-one-out procedure gave $q^2 = 0.68$, which illustrates the reasonable predictive utility of the model for new analogues.

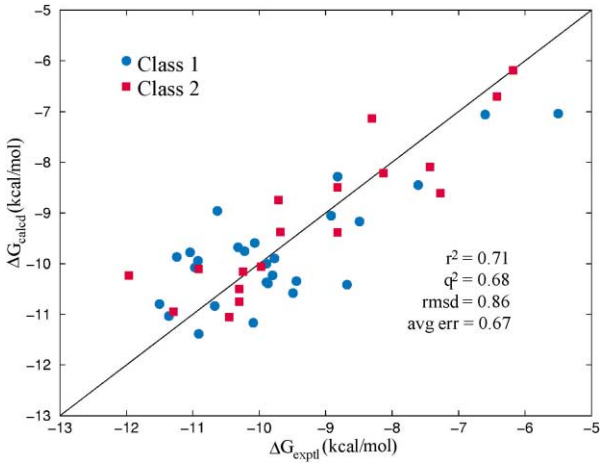


Figure 3. Calculated binding affinities (ΔG_{calcd}) using the ELR model versus experimental activities (ΔG_{exptl}) for 45 celecoxib analogues with COX-2.

Interestingly, the descriptors and quality of the ELR fit of COX-2 inhibitors parallel those of an ELR fit of forty HEPT and nevirapine analogues with HIV-1 reverse transcriptase.[9] The physically intuitive descriptors EXX_{LJ}, ΔHB_{tot} and $\Delta FOSA$ are recurring themes in several of the ELR models to date.[9,10,16] Using only a few descriptors, ELR has potential for screening out poor inhibitors and identifying compounds likely to have high affinity for their target proteins. Coupled with the earlier work, the present study establishes that the ELR method with adequate MC or MD sampling can successfully accommodate both small and large structural variations for inhibitor series.

Conclusion

A viable scoring function for inhibition of COX-2 has been developed using the Monte Carlo-Extended Linear Response approach. The model reproduces well trends in experimental activities of a series of celecoxib analogues. Extensions to multiple chemotypes, COX-2/COX-1 selectivity, and automated design of new inhibitors are subjects of ongoing research.

Acknowledgements

The authors thank Shoshannah Pearlman for programming assistance and the National Institutes of Health (GM32136) for financial support.

References and Notes

1. Vane, J. R.; Botting, R. M. In *Clinical Significance and Potential of COX-2 Inhibitors*; William Harvey: London, 1998.
2. Lynch, P. M. *Oncology* **2001**, *15*, 21.
3. Bresalier, R. S. *Gastroenterology* **2000**, *119*, 1797.
4. Blain, H.; Jouzeau, J. Y.; Blain, A.; Terlain, B.; Trechot, P.; Touchon, J.; Netter, P.; Jeandel, C. *Presse Med.* **2000**, *29*, 267.
5. Grubbs, C. J.; Lubet, R. A.; Koki, A. T.; Leahy, K. M.; Masferrer, J. L.; Steele, V. E.; Kelloff, G. J.; Hill, D. L.; Seibert, K. *Cancer Res.* **2000**, *60*, 5599.
6. Price, M. L. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 9455.
7. Price, M. L. P.; Jorgensen, W. L. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1541.
8. Åqvist, J.; Medina, C.; Samuelsson, J.-E. *Protein Eng.* **1994**, *7*, 385.
9. Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. *J. Med. Chem.* **2001**, *44*, 145.
10. Pierce, A. C.; Jorgensen, W. L. *J. Med. Chem.* **2001**, *44*, 1043.
11. Penning, T. D.; Talley, J. J.; Bertenshaw, S. R.; Carter, J. S.; Collins, P. W.; Docter, S.; Graneto, M. J.; Lee, L. F.; Malecha, J. W.; Miyashiro, J. M.; Rogers, R. S.; Rogier, D. J.; Yu, S. S.; Anderson, G. D.; Burton, E. G.; Cogburn, J. N.; Gregory, S. A.; Koboldt, C. M.; Perkins, W. E.; Seibert, K.; Veenhuizen, A. W.; Zhang, Y. Y.; Isakson, P. C. *J. Med. Chem.* **1997**, *40*, 1347.
12. Kurumbail, R. G.; Stevens, A. M.; Gierse, J. K.; McDonald, J. J.; Stegeman, R. A.; Pak, J. Y.; Gildehaus, D.; Miyashiro, J. M.; Penning, T. D.; Seibert, K.; Isakson, P. C.; Stallings, W. C. *Nature* **1996**, *384*, 644.
13. Jorgensen, W. L. *GenMol Version 1.5*; Yale University: New Haven, 2001.
14. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
15. Jorgensen, W. L. *MCPRO Version 1.65*; Yale University: New Haven, CT, 2000.
16. Rizzo, R. C.; Udier Blagovic, M.; Wang, D. P.; Watkins, E. K.; Smith, M. B. K.; Smith, R. H.; Tirado-Rives, J.; Jorgensen, W. L. Submitted for publication.