Pergamon

Bioorganic & Medicinal Chemistry 8 (2000) 11–17

BIOORGANIC & MEDICINAL CHEMISTRY

Building a Common Feature Hypothesis for Thymidylate Synthase Inhibition

Seung-Gull Kim, [b,*] Chang-Ju Yoon, [a] Sei-Hwan Kim, [b] Yoon-Je Cho [b] and Dong-Il Kang [b]

[a]*Department of Chemistry, The Catholic University of Korea, San 43-1 Yokkok 2-Dong Wonmi-Gu, Puchon City 420-743, South Korea*
[b]*Central Research Laboratory, R&D Division, Choongwae Pharma Co., PO Box 61, Suwon, South Korea*

Received 28 April 1999; accepted 1 August 1999

Abstract—A set of 21 highly flexible competitive inhibitors of thymidylate synthase (TS; EC 2.1.1.45) covering a wide activity range ($IC_{50} = 6$ nM–100 μM) has been investigated by three-dimensional quantitative structure–activity relationship (3D-QSAR). CATALYST™ was used to generate three-dimensional hypotheses to study the common interaction features among a set of thymidylate synthase inhibitor. The verification of the hypothesis was achieved by using the molecules outside the training set.

Introduction

Thymidylate synthase (TS; EC 2.1.1.45) is involved in the catalytic reaction of deoxyuridylate (dUMP) to deoxythymidylate (dTMP); a one-carbon transfer to the 5′-position of uridylate. In this process N5,N10-methylenetetra-hydrofolate plays a critical role as a cofactor, where this folate derivative delivers one-carbon as a methylene unit from serine to uridylate.[1]

Since DNA contains thymine as a base component instead of uracil, rapidly proliferating cells require an abundant supply of deoxythymidylate for the biosynthesis of DNA. For the synthesis of dTMP this reaction is the sole de novo pathway. Therefore, without an exogenous supply of thymidine, blockade of this step by the inhibition of TS would ultimately lead to a 'thymineless cell death'.

CATALYST™,[2] using relationships between chemical structures and their biological activities, can be useful in the discovery of new compounds. In theory, this is an integrated system that speeds the iterative process of drug discovery, in which new compounds bring new information.

During the beginning phase of an automated hypothesis generation, CATALYST™ provides two numbers to help the chemist assess the trustworthiness of a hypothesis. The first of these numbers is the cost of an ideal hypothesis, which is a lower bound on the cost of the simplest possible hypothesis that still fits the data perfectly. CATALYST™ also provides the cost of null hypothesis, which presumes that there is no structure in the data and that the experimental activities are normally distributed about their mean. These models can be considered upper and lower bounds for the training set. The cost values for them are useful guides for estimating the chances for a successful experiment and are available in the hypothesis generation directory. What is important is the difference between these two costs. The greater the difference, the higher the probability for finding useful models.

In terms of hypothesis significance, what really matters is the magnitude of the difference between the cost of any returned hypothesis and the cost of the null hypothesis. In general, if this difference is greater than 60 bits, there is an excellent chance the model represents a true correlation. Since most returned hypotheses will be higher in cost than the fixed cost model, a difference between fixed cost and null cost of 70 or more will be necessary in order to achieve the 60 bit difference. If a returned hypothesis has a cost that differs from the null hypothesis by 40–60 bits, there is a high

*Corresponding author.

0968-0896/00/$ - see front matter
PII: S0968-0896(99)00246-1

probability (a 75–90% chance) of it representing a true correlation in the data. Under these conditions, it may be difficult to find a model that can be shown to be predictive.[3]

Computational Methods

The study was performed using the CATALYST™ software packages on SGI indigo^2 4400. A training set of 21 molecules with binding affinities distributed over a range of four orders of magnitude for thymidylate synthase was selected from the literature and the data of Choongwae Pharma Co. (Table 1).

Results and Discussion

Conformation models were generated for each compound that emphasized conformational diversity under constraint of 10 kcal/mol energy threshold above the estimated global minimum.[4–6] The 'training set' of molecules, along with its associated conformational models, was submitted to CATALYST™ hypothesis generation. The chemical function used in the hypothesis generation step included a hydrogen bond donor, hydrogen bond acceptor and hydrophobic. We constrained the hypothesis generator to report only hypotheses with four features. The measured activities and the number of conformers generated for each molecule are listed in Table 2.

The cost of the ideal hypothesis was 89.4 and the cost of the null hypothesis was 171.0. Since the 10 generated hypotheses have costs that range from 107.1 to 121.1 with an average of 116.7, all of these were statistically significant: these were evaluated visually. The large difference between fixed and null cost indicates that there was a strong signal. The closeness of the hypothesis to the fixed cost meant that the signal could be interpreted (Table 3, Fig. 1).

Table 1. The picking-rule of training sets for CATALYST™

1. At least 16 compounds are necessary to assure statistical power.
2. Activities should span four orders of magnitude.
3. Each order of magnitude should be represented by at least three compounds.
4. No redundant information.
5. No excluded volume problems.
6. Include the most active compounds you have.
7. Each compound must possess something new to teach CATALYST™.
8. If two compounds have similar structures (collections of features), they must differ in activity by an order of magnitude to be included; otherwise, pick only the most active of the two.
9. If two compounds have similar activities (within one order of magnitude), they must be structurally distinct (from a chemical feature point of view) in order that both be included; otherwise, pick only the most active of the two.
10. Fixed and mull cost differential must be more than 20.
11. The entropy of hypothesis space value (config.) must be less than 18; a thorough analysis of all models will be carried out. If higher, CATALYST™ will truncate the list and some models will not be considered.

A regression line of 'measured' versus 'estimated' TSI activity for the training set, expressed as log IC_{50}, based on the lowest cost CATALYST™-generated TSI hypothesis exhibited a correlation coefficient (r^2) = 0.94 with root-mean-square (rms) deviation = 1.09 (Fig. 2). The good score value indicated a reliable ability to predict activities within the training set.

According to the concept of active site directed irreversible inhibitor,[13] the binding site is composed of the active site region, hydrophobic region and polar region. In the active site region of thymidylate synthase, folate analogues form favorable stacking with substrate dUMPs. Additionally, this region is composed of a hydrogen bonding and hydrophobe region. The hydrogen bonding region is composed of dUMP, Ala263, and Asp169 that has hydrogen bonding with hydrogen donor, and acceptor parts of quinazoline ring of TSI.[14] The hydrophobe region is composed of Trp83, Trp80. Additional active site residues Glu58, Ser180, and Asn177 interact with the quinazoline moiety. Also, Ile79, Trp83, Leu172, Ile176, and Val262 residues form a hydrophobic cluster in the shape of a doughnut in the active site of TS. Residues Thr78, His51, Ile258, Leu172 form the polar region of TS, in which hydrogen bonding with glutamate residue of TSIs takes place. In particular Ile258 and Leu172 make contact with TSI through a water molecule. In addition, the carbonyl oxygen of the peptide bond portion of TSI makes hydrogen bonding with Ala260 through water.

Figure 3 shows JMC-34-1594-7c and JMC-34-1594-8a mapped to a hypothesis that is characterized by one hydrophobic and three hydrogen bond acceptors. In the active site region, the oxygen and nitrogen group of the quinazolinone ring is aligned with the hydrogen bond acceptor. The phenyl ring of the quinazolinone ring is aligned within the hydrophobic areas. In the thymidylate synthase case, sulfur of bridged thiophene ring aligned the hydrogen bond acceptor areas in the hydrophobic region of receptor, because they have the residue (Leu172) of hydrogen bond character. The estimated activities of JMC-34-1594-7c and JMC-34-1594-8a, 5.9 nM and 9.5 M, respectively, accurately reflect the experimental values of 7.0 and 8.0 nM.

Figure 4 shows JMC-35-2761-2aa mapped to this hypothesis, characterized by only two hydrogen bond acceptor functions. This hypothesis clearly indicates that the low affinity of the JMC-35-2761-2aa was due to the consequence of having of the bridged phenyl and phenyl ring of quinazolinone in a bad location.

The comparison of JMC-34-1594-7c with JMC-35-2761-2aa mapped to the hypothesis is shown in Figure 5. JMC-34-1594-7c and JMC-35-2761-2aa are shown in a similar alignment to the four-feature hypothesis. JMC-35-2761-2aa belongs to the same chemical class as JMC-34-1594-7c, but it is 10,000 times less active than JMC-34-1594-7c. JMC-34-1594-7c can match most of the regions resulting in a higher estimated activity, the 5.9 nM value for estimated activity is in good agreement with the measured value of 7 nM. However,

Table 2.[7–12] Structures, number of conformers generated, and nM activities for the molecules in the thymidylate synthase inhibitor training set. Compounds were tested as inhibitors of bacterial cell line (L1210; mouse lymphocytic leukemia)

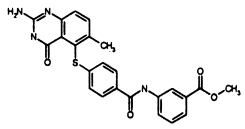
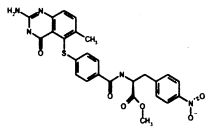
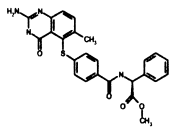
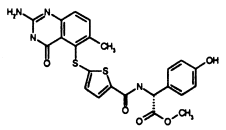
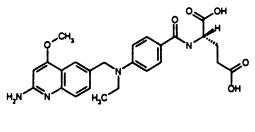
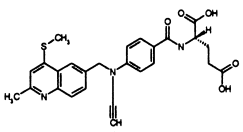
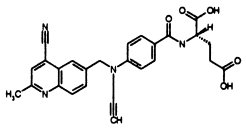
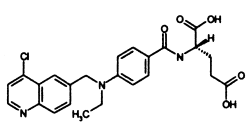
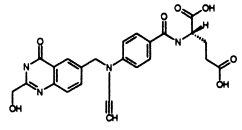
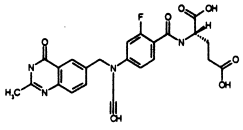
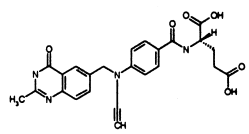
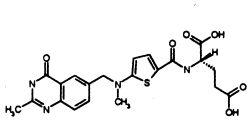
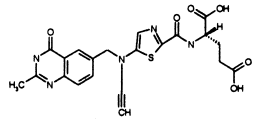
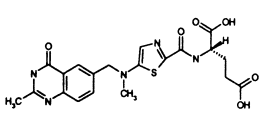
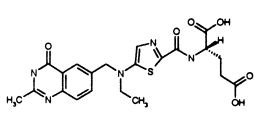
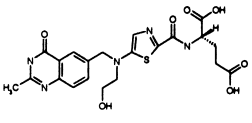
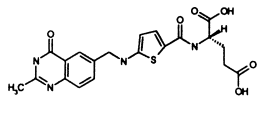
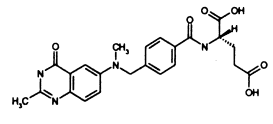
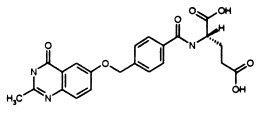
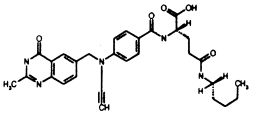
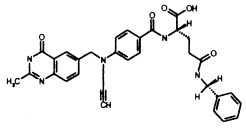
			
CW252001	cw252038	cw252021	Cw252032
Conf 30 Act 450	Conf 92 Act 3400	Conf 34 Act 390	Conf 52 Act 7300
			
JMC-35-2761-2aa[7]	JMC-35-2761-2d[7]	JMC-35-2761-2f[7]	JMC-35-2761-2l[7]
Conf 102 Act 100000	Conf 93 Act 100000	Conf 22 Act 75000	Conf 18 Act 30000
			
JMC-35-3067-7[8]	JMC-33-3072-2a[9]	JMC-34-1594-2a[10]	JMC-34-1594-7C[10]
Conf 109 Act 5000	Conf 120 Act 27	Conf 118 Act 90000	Conf 144 Act 7
			
JMC-34-1594-8a[10]	JMC-34-1594-8c[10]	JMC-34-1594-8d[10]	JMC-34-1594-8h[10]
Conf 143 Act 8	Conf 159 Act 6	Conf 142 Act 8000	Conf 150 Act 20
			
JMC-34-2209-12[11]	JMC-34-2209-15[11]	JMC-34-2209-29[11]	JMC-37-3294-84[12]
Conf 136 Act 3000	Conf 106 Act 700	Conf 107 Act 40	Conf 125 Act 20000
			
JMC-34-3294-85[12]			
Conf 77 Act 22000	Conf Act	Conf Act	Conf Act

Table 3. The ten lowest-cost hypotheses from the TSI training set

No.	1	2	3	4	Cost	RMS	Correlation
1	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	107.112	1.08779	0.940792
2	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	107.659	1.15824	0.930637
3	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	114.651	1.49528	0.876556
4	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	117.416	1.50339	0.878204
5	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	118.442	1.59378	0.858952
6	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	119.109	1.48707	0.884852
7	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	120.243	1.61979	0.855053
8	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	120.555	1.64173	0.850096
9	HB acceptor	HB acceptor	HB Donor	Hydrophobic	120.764	1.58848	0.863571
10	HB acceptor	HB acceptor	HB acceptor	Hydrophobic	121.056	1.69277	0.838059

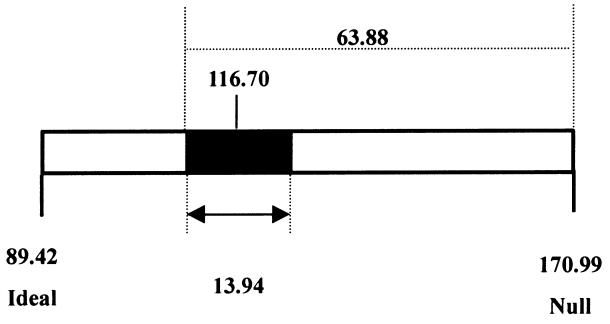


Figure 1. Hypothesis costs of TSI training set.

JMC-35-2761-2aa cannot reach two of the functional regions (hydrogen bond acceptor and hydrophobic) resulting in a lower estimated activity. The 230,000 nM value for estimated activity is in good agreement with the measured value of 100,000 nM.

The validity and the predictive character of this four-feature hypothesis was further assessed using the test molecules outside the training set. The resulting four-feature hypothesis explained the major SAR points and was predictive with respect to molecules outside the training set. Fourteen molecules outside the training set were align to hypothesis 1 for verification of generated-hypothesis. Figure 6 shows cw252051 and cw252041 mapped to the hypothesis, which is characterized by one hydrophobic and three hydrogen bond acceptor functions. In the active site region, the oxygen and nitrogen group of the quinazolinone ring aligned the hydrogen bond acceptor and the phenyl ring of the quinazolinone ring aligned hydrophobic areas. In this case, the carboxyl group of the peptide bridge aligned with the hydrogen bond acceptor area, whereas, because sulfur bridged cw252-series compounds have more short distance of compound than the other compounds. In the set of test molecules, cw252051 is shown in higher

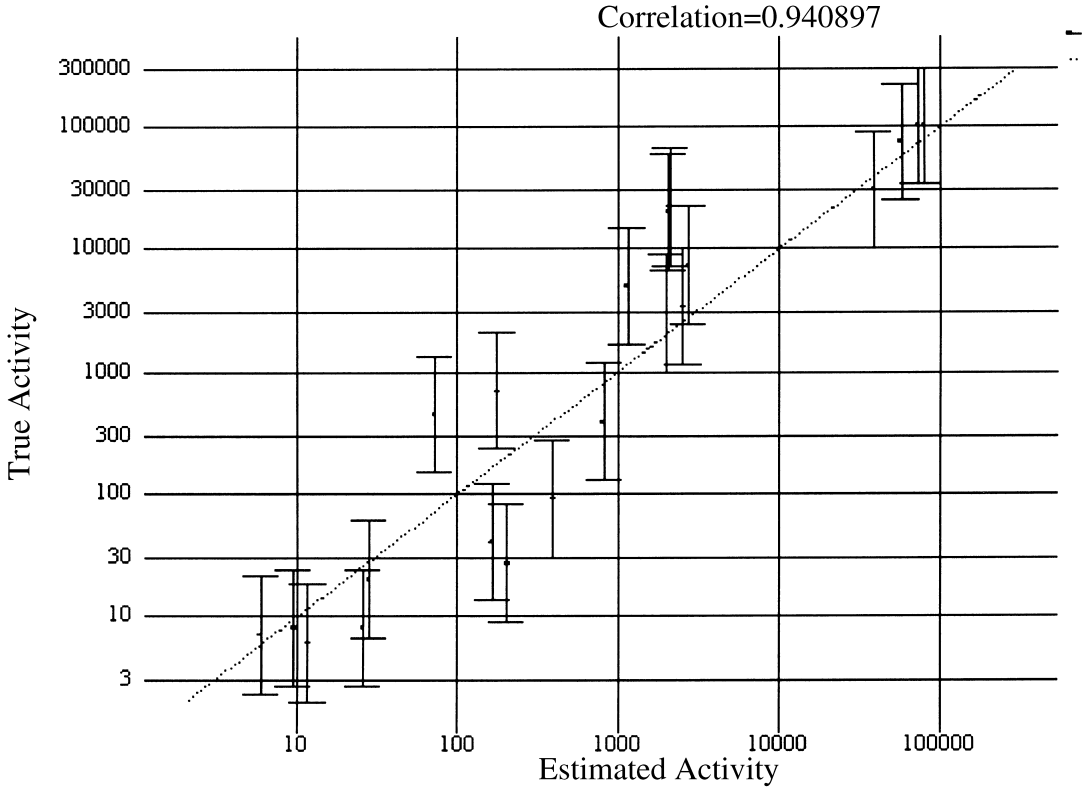


Figure 2. The regression lines for hypothesis 1 comparing actual to estimated activities for each number of the training set.

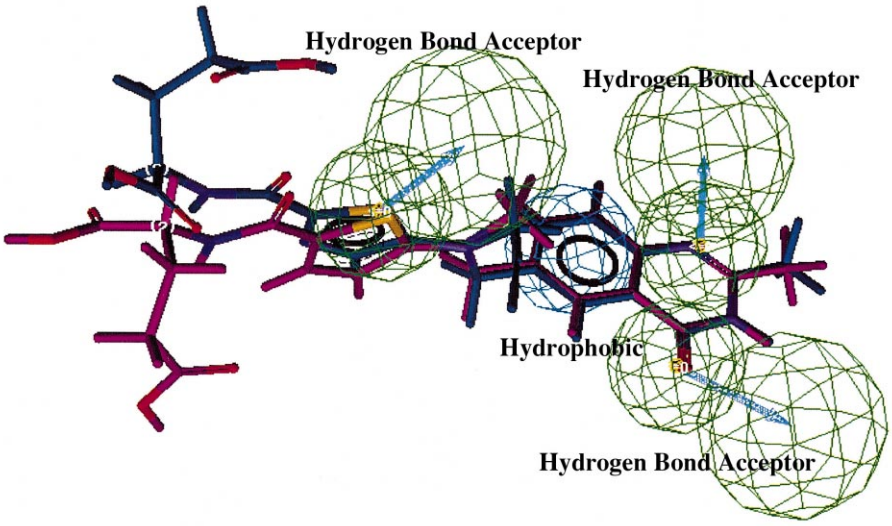


Figure 3. JMC-34-1594-8a and JMC-34-1594-7c mapped to hypothesis 1.

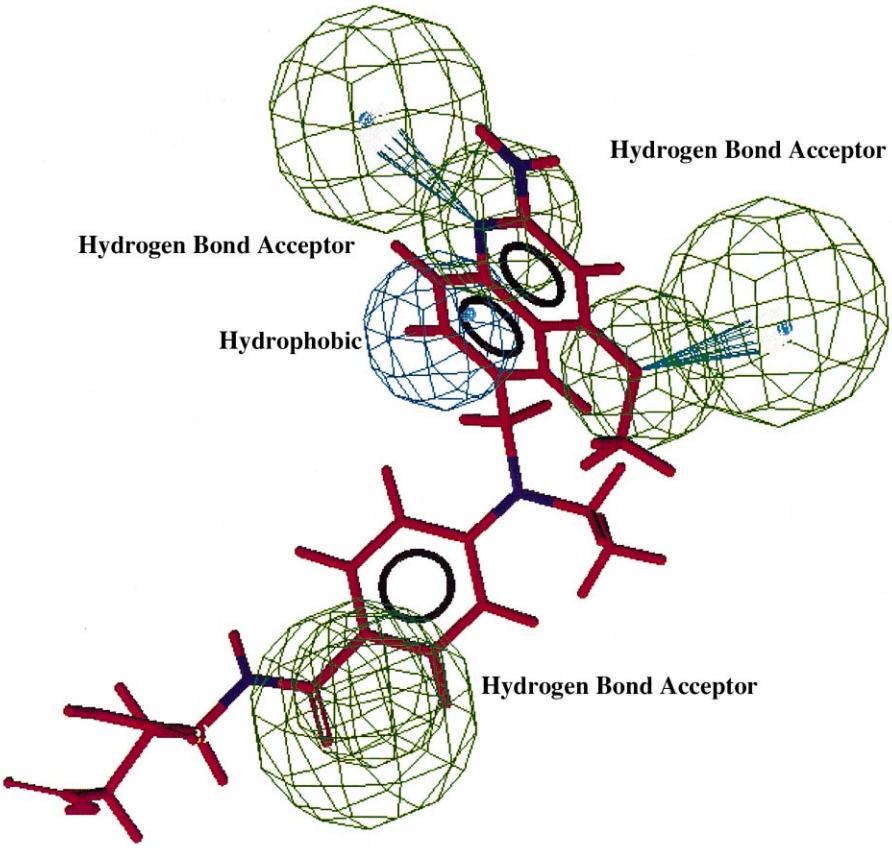


Figure 4. JMC-35-2761-2aa mapped to hypothesis 1.

estimated activity; however, the 540 nM value for estimated activity is in good agreement with the measured value of 710 nM. Similarly, the 5400 nM value for estimated activity is in good agreement with the measured value of 2500 nM for cw252041. This model exhibited satisfactory predictivity within an acceptable range of error, and it can be used as a search query to identify alternative structural templates from 3D databases.

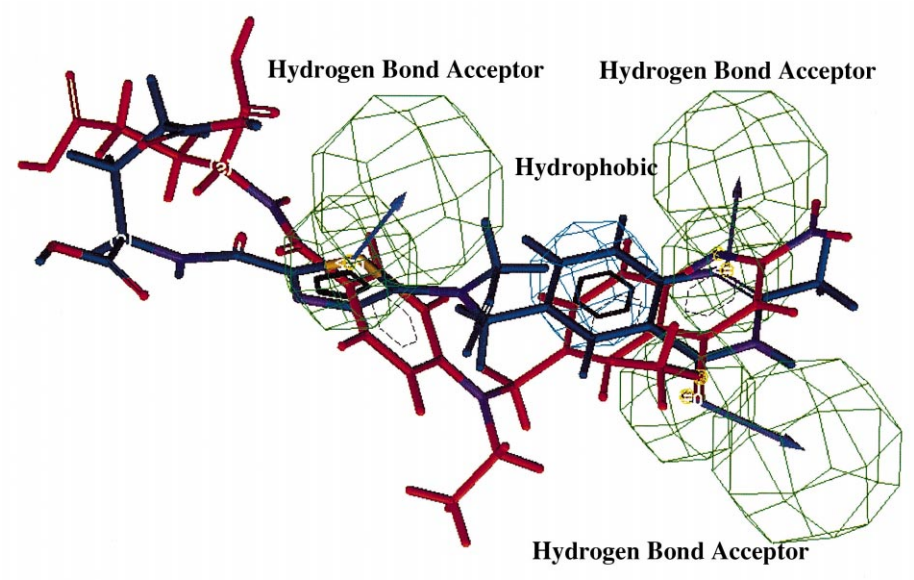


Figure 5. JMC-35-2761-2aa (red) and JMC-34-1594-7c (green) mapped to hypothesis 1.

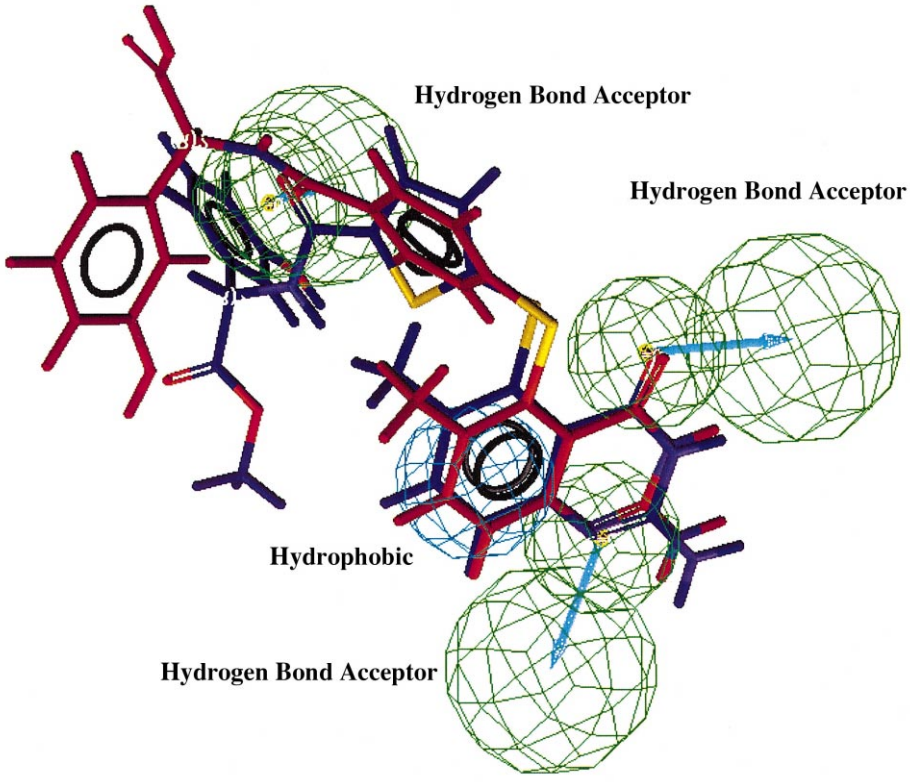


Figure 6. An overlay of cw252041 (blue) and cw 252051 (red) of test-set mapped to the thymidylate synthase hypothesis.

Acknowledgements

We thank Dr. Du-Jong Baek who kindly provided thymidylate synthase inhibitor activity data, and Dr. Young-Tae Han and Byung-Ha Chang of MSI-Korea for technical support.

References

1. Baek, D. J.; Park, Y. K.; Heo, H. I.; Lee, M. H.; Yang, Z. Y.; Choi, M. H. *Bioorg. Med. Chem. Lett*. **1998**, *8*, 3287.
2. CATALYST 4.0; Molecular Simulation Inc.. San Diego, CA, **1998**.

3. Hypothesis generation in catalyst, http://www.msi.com/support/catalyst/hypogen.html
4. > Smellie, A.; Teig, S. L.; Towbin, P. *J. Comp. Chem.* **1995**, *16*, 171.
5. Smellie, A.; Khan, S. D.; Teig, S. L. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285.
6. Smellie, A.; Khan, S. D.; Teig, S. L. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 295.
7. Warner, P.; Baker, A. J.; Jackman, A. L.; Burroes, K. D.; Roberts, N.; Bishop, J. A. M.; O'Connor, B. N.; Hughes, L. R. *J. Med. Chem.* **1992**, *35,* 15, 2761.
8. Jackman, A. L.; Marsham, P. R.; Thornton, T. J.; Bishop, J. A. M.; O'Connor, B. M.; Hughes, L. R.; Calvert, A. H.; Jones, T. R. *J. Med. Chem.* **1990**, *33*, 3067.
9. Marsham, P. R.; Jackman, A. L.; Oldfield, J.; Hughes, L. R.; Thornton, T. J.; Bisset, G. M. F.; O'Connor, B. M.; Bishop, J. A. M.; Calvert, A. H. *J. Med. Chem.* **1990**, *33*, 3072.
10. Marsham, P. R.; Hughes, L. R.; Hughes, A. L.; Hayter, A. J.; Oldfield, J.; Wardleworth, J. M.; Bishop, J. A. M.; O'Connor, B. M.; Calvert, A. H. *J. Med. Chem.* **1991**, *34*, 1594.
11. Marsham, P. R.; Jackman, A. L.; Jackman, A. J.; Daw, M. R.; Snowden, J. L.; Wardleworth, J. M.; Bishop, J. A. M.; O'Connor, B. M.; Calvert, A. H.; Hughes, L. R. *J. Med. Chem.* **1991**, *34*, 2209.
12. Bisset, G. M. F.; Bavetsias, V.; Thornton, T. J.; Pawelczak, K.; Calvert, A. H.; Hughes, L. R.; Jackman, A. L. *J. Med. Chem.* **1994**, *37*, 3294.
13. Baker, B. R. Design of active site directed irreversible enzyme inhibitors, The organic chemistry of the enzyme active site. John-Wiley and Sons: New York, 1967.
14. Kim, S. G. A 3D-QSAR study on some Thymidylate Synthase Inhibitor, Ph.D. Dissertation, Catholic University of Korea, 1999.