Quantitative Structure−Activity Relationship (5D-QSAR) Study of Combretastatin-like Analogues as Inhibitors of Tubulin Assembly

Sylvie Ducki,*[,†,‡] Grant Mackenzie,[†] Nicholas J. Lawrence,[§,||] and James P. Snyder[‡]

Centre for Molecular Drug Design, Cockcroft Building, University of Salford, Salford M5 4WT, U.K., Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, Georgia 30322, and Department of Chemistry, Cardiff University, P.O. Box 912, Cardiff, CF10 3TB, U.K.

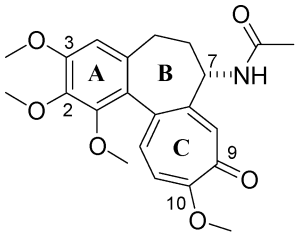
Received July 14, 2004

A molecular modeling study was carried out to develop a predictive model for combretastatin-like analogues populating the colchicine-binding site of β-tubulin. A series of compounds built around a framework including two aromatic groups linked by various moieties such as alkenes (stilbenes), enones (chalcones), or ethers was selected for the study. The 5D-QSAR model was developed stepwise. First a model was generated for the chalcone series (19 compounds, 71 conformations), then for the stilbene series (18 compounds, 59 conformations), and finally for the combined dataset (47 ligands, 160 conformers). Although the models for the chalcone and stilbene series appeared slightly different when represented by QSAR colored surfaces, the combined model seems to reconcile the differences without compromise and represents a highly predictive model for compounds that bind to the colchicine-binding site of tubulin.

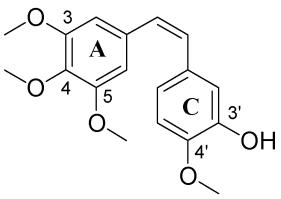
Introduction

Among the cellular structures necessary to maintain the growth and function of normal and malignant cells, the microtubules play a pivotal role exceeded only by that of DNA as the template for transcription and transactivation. Microtubules are of particular importance for the formation of the mitotic spindle, which provides the structural framework for the physical segregation of chromosomes during cell division (mitosis). The formation of microtubules is a dynamic process that involves assembly of the heterodimers formed by α- and β-tubulin subunits and degradation of the linear polymers. Drugs binding to tubulin are known to disrupt this dynamic equilibrium. Some are used as chemotherapeutic agents for a variety of cancers, while others are either investigational drugs or under study as probes of microtubule dynamics in cellular and biochemical processes.[1] Well-known examples include paclitaxel,[2] an antimitotic agent known to stabilize microtubules by binding to β-tubulin,[3,4] and colchicine **1**, an antimitotic agent that binds to a different site of β-tubulin and inhibits its assembly into microtubules.[5] Combretastatin A4 (CA4) **2**, a stilbene isolated by Pettit from the bark of the African bush willow tree *Combretum caffrum*,[6] also binds to the colchicine-binding site and blocks microtubule formation.[7]

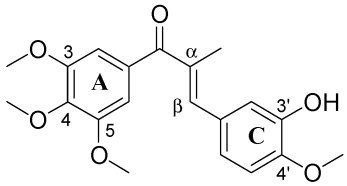
The importance of tubulin as a target has been amplified by the recent discovery that CA4 displays potent and selective toxicity toward tumor vasculature.[8] In this way tumors are starved of oxygen and nutrients; their constituent cells die. It has become clear that agents acting in this manner will have a significant impact on the clinical management of cancer.[9] The antivascular effect of such drug candidates derives from the role tubulin and microtubules play in determining the elongated shape of vascular endothelial cells.



Colchicine **1**



Combretastatin A4 **2**



α-methyl chalcone **3**

The cellular microtubule network—a principal part of the cytoskeleton—plays a major role in maintaining cell shape, particularly in the case of the neovasculature. The drugs cause microtubules to rapidly depolymerize. As a result, endothelial cells round up and very quickly block blood-flow through the vascular network. This effect is most pronounced for agents that bind at the center targeted by colchicine.[10]

The spatial relationship between the two aromatic rings of CA4, colchicine, and similar drugs is an important structural feature that determines their ability to bind to tubulin.[11] We have already reported the synthesis and biological evaluation of a number of colchicine-binding site agents, such as substituted stilbenes,[12] which are CA4 analogues, and chalcones, which

* To whom correspondence should be addressed: Phone: + 44 161 295 4701. Fax: +44 161 295 5111. E-mail: s.ducki@salford.ac.uk.
† University of Salford.
‡ Emory University.
§ Cardiff University.
|| Current address: Drug Discovery Program, H. Lee Moffitt Cancer Center & Research Institute, Department of Interdisciplinary Oncology, University of South Florida, Tampa, Florida 33612.

10.1021/jm049444m CCC: $30.25
Published on Web 12/31/2004

Table 1. Properties of Quasi-Atomistic Descriptors Used in Quasar

descriptor (color)	nb potential type[a]	electronic charge	well depth of nb function (kcal/mol)
hydrophobic, neutral (gray)	6/12		−0.024[b]
hydrophobic, positive (orange)	6/12 + elec	+0.10	−0.09[b]
hydrophobic, negative (brown)	6/12 + elec	−0.10	−0.09[b]
hydrogen-bond donor (green)	10/12		−5.0/−4.1/−2.3[c]
hydrogen-bond acceptor (yellow)	10/12		−5.0/−4.1/−2.3[c]
salt bridge, positive (red)	10/12 + elec	+0.25	−5.0/−4.1/−2.3[c]
salt bridge, negative (blue)	10/12 + elec	−0.25	−5.0/−4.1/−2.3[c]
H-bond flip-flop[d]	10/12		−5.0/−4.1/−2.3[c]
surface solvent	symmetric 10/12[e]		−0.97/−0.80/−0.46[c,f]
void (shallow pocket)			

[a] The values *i, j* refer to the attractive and repulsive coefficients of the nonbonded potential used for the ligand−receptor interaction. The general form of this potential is $E(r) = A/r_i - C/r_j$. [b] This function adopts the form $E(r) = A/r^{12} - C/r^6$. The coefficients A and C are calculated according to $A = -\epsilon(r_i + r_j)^{12}$ and $C = -2\epsilon(r_i + r_j)^6$, respectively, with $\epsilon = (\epsilon_i\epsilon_j)^{1/2}$. The given figure represents ϵ_j; r_i and r_j correspond to the van der Waals radii of the two involved atoms. [c] Values of −O−H···Y, >N−H···Y, and −S−H···Y H-bond interactions, respectively, where "Y" denotes a virtual H-bond acceptor. Identical values are used for the X···O, X···N, and X···S arrangements where "X" denotes a virtual H-bond donor. [d] H-bond flip-flop particles adapt their property (H-bond donor or acceptor) to each ligand molecule within the pharmacophore, depending on its interacting functional group. [e] To avoid repulsive forces between surface solvent and any ligand molecule, a symmetric 10/12 potential (mirrored at $r = r°$) is used. This represents a possible approximation to a mobile solvent. [f] As the virtual particles are different in radius from a water molecule, the associated energy must be corrected for different volumes: $E = (2r_{vp}/2.75)^3E°$; e.g. for $r_{vp} = 0.8$ Å, $E = 0.197E°$. The 2.75 Å corresponds to a mean O−H···O H-bond distance.

can be regarded as keto-CA4 derivatives.[13−15] Chalcone **3**, the most potent member of the series, is unfortunately highly cytotoxic. Most recently, we also demonstrated that chalcones act as antivascular agents.[16]

QSAR (quantitative structure activity relationship) is an area of computational research that constructs models to correlate and predict biological properties from structural parameters of existing molecules. Such models can play an important role in lead structure optimization.[17,18] The idea behind 3-D QSAR is that differences in biological activity are often related to differences in the magnitudes of molecular fields surrounding the ligands. In a recent approach devised by Vedani and Dobler, 5D-QSAR,[19] the field is represented by a color-coded 3D map displaying specific spatial regions where the magnitude of the various atomistic properties (Table 1) are found to be significantly correlated with the designated biological activity.[20] While the distributed properties are identical for all ligand molecules, their exact location on the envelope varies slightly (*rms* fluctuations range from 0.2 to 0.8 Å with maximal individual shifts up to 2.5 Å) depending on the chosen induced fit. Potential H-bond sites are restricted to positions on the receptor surface which are located within reasonable distance and at favorable orientation with respect to any H-bond donor or acceptor moiety of the ligand molecule defining the training set.

The fourth dimension is the possibility to represent each molecule by an ensemble of conformations, orientations, protonation states, and/or enantiomers, thereby reducing the bias associated with the choice of a bioactive conformation.[21] The fifth dimension refers to the option of considering an ensemble of different induced-fit models.[22] Ligand−receptor interactions are estimated on the basis of a directional force field. A family of quasi-atomistic receptor models is then generated using a genetic algorithm combined with weighted cross-validation. In this approach, the map can be considered as a model for the protein binding site where the ligands under study are presumed to exert their initial biological action. If the resulting QSAR is sufficiently robust, it can be employed to predict the biological activities of a test set of ligands.

The aim of the present study was to determine the essential structural properties of selected ligands binding to the colchicine-binding site as they reflect the capacity for tubulin assembly inhibition. In particular, we have employed the 5D-QSAR method to develop a correlation model that will ultimately be applied to the design of novel colchicine-like blockers.

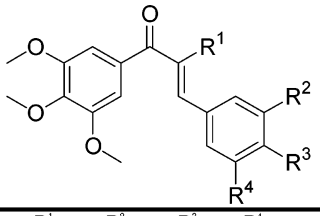
Materials and Methods

Data Set Preparation. The 47 compounds selected for this study represent a series of analogues structurally based on CA4. The series is built around a framework including two aromatic groups (rings A and C) linked by various moieties such as an alkene (stilbene), enone (chalcone), or ether. In most cases, one of the aromatic groups is a 3,4,5-trimethoxyphenyl group, regarded as an important feature in the binding of colchicine to tubulin.[23] The chemical structures of the 47 ligands are shown in Figure 1. The 3D molecular structures were generated and conformational searches were performed using Macromodel 6.5. The energy of the lowest energy structure, the number of conformations found, and the frequency with which the simulation visited the lowest energy structure were monitored to ensure an exhaustive search (Table 2). The chosen conformations were supplemented with MNDO electrostatic potential (ESP) charges scaled to HF/6-31G* ESP values,[24] AMSOL AM1/SM5.4PDA free energies of solvation,[25,26] then aligned using APOLLO. Details are given in the Supporting Information.

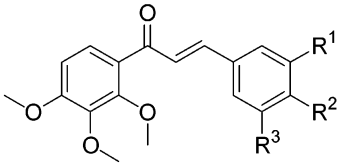
5D-QSAR Analysis. The quasi-atomistic modeling software Quasar 3.5 allows the construction of 5D-QSAR models. This methodology was recently presented in detail[19] and summarized above. Compared with 3D-QSAR, this approach attempts to reduce the bias associated with the choice of the bioactive conformation, the ligand substituent alignment, and the induced-fit model. Quasar allows also for H-bond flip-flop and accounts for solvation phenomena.

Results

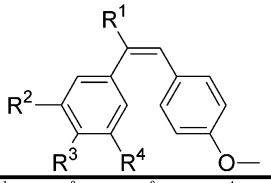
The results for each compound class and the subsequent combined model are depicted in Table 3. The chalcones, stilbenes, and combined dataset consisted of 19, 18, and 47 ligands represented by 71, 59, and 160 conformers, respectively (Table 2). Induced-fit simulated using minimization along the field lines resulted in root-mean-square (rms) deviations ranging from 0.83 to 0.94



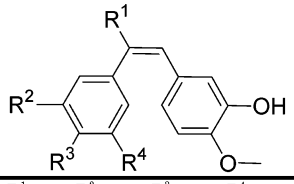
Cpds	R^1	R^2	R^3	R^4	IC_{50} (μM)
M07	CH_3	OH	OCH_3	H	0.46
M08	OCH_3	F	OCH_3	H	0.47
M10	OCH_3	OH	OCH_3	H	0.51
M12	H	OH	OCH_3	H	0.62
M13	CH_3	F	OCH_3	H	0.65
M17	H	F	OCH_3	H	1.20
M20	OCH_3	F	OCH_3	F	1.7
M22	H	NH_2	OCH_3	H	2.0
M24	CH_3	F	OCH_3	F	2.4
M39	CH_3	NO_2	OCH_3	H	>10
M43	H	F	OCH_3	F	12
M67	H	OH	OH	H	>100



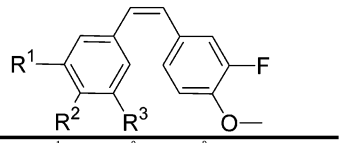
Cpds	R^1	R^2	R^3	IC_{50} (μM)
M06	OH	OCH_3	H	0.45
M35	F	OCH_3	H	7.9
M52	F	OCH_3	F	31
M66	OH	OH	OH	>100



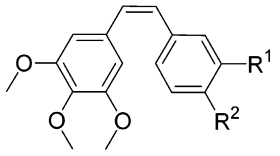
Cpds	R^1	R^2	R^3	R^4	IC_{50} (μM)
M05	H	OCH_3	OCH_3	OCH_3	0.20
M14	H	CH_3	CH_3	CH_3	0.65
M19	CH_3	OCH_3	OCH_3	OCH_3	1.30



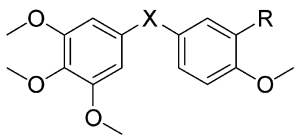
Cpds	R^1	R^2	R^3	R^4	IC_{50} (μM)
M01	H	CH_3	CH_3	CH_3	0.12
M02	CH_3	OCH_3	OCH_3	OCH_3	0.13
M03	Et	OCH_3	OCH_3	OCH_3	0.13
M04	H	OCH_3	OCH_3	OCH_3	0.18
M09	H	OEt	OEt	OEt	0.50
M25	H	H	H	H	2.5
M26	H	CH_3	CH_3	H	3.1
M30	H	F	F	F	4.5



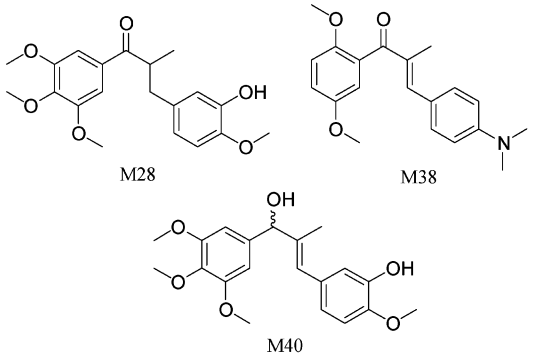
Cpds	R^1	R^2	R^3	IC_{50} (μM)
M15	CH_3	CH_3	CH_3	0.70
M16	OEt	OEt	OEt	1.20
M18	OCH_3	OCH_3	OH	1.25
M32	F	F	F	5.5
M33	CH_3	CH_3	H	6.6



Cpds	R^1	R^2	IC_{50} (μM)
M23	CH_3	CH_3	2.0
M27	H	Et	3.4



Cpds	X	R	IC_{50} (μM)
M11	CH_2-O	CHO	0.56
M34	O-CH_2	OH	7.2
M36	CH=C(CH_3) *E*	OH	9.0
M41	O	OH	11
M48	CH_2-O	OH	21
M54	O-CH_2	O-CH_2-CHO	41
M55	CH=CH *E*	OH	50
M63	NH	OH	>100
M68	O-CH_2	O-CH_2-CH(OH)-CH_2OH	>100
M69	CH_2-O	CH_2OH	>100



M28 M38

M40

Cpds	IC_{50} (μM)
M28	4.0
M38	10
M40	>10

Figure 1. Schematic representation of the 47 selected ligands. IC_{50} represents the concentration of ligand required for 50% inhibition of tubulin polymerization.

Å with associated energies of 0.7−0.9 kcal/mol for the chalcone series, from 0.77 to 1.10 Å with associated energies of 0.7−1.0 kcal/mol for the stilbene series, and from 0.93 to 1.31 Å with associated energies of 0.8−1.2 kcal/mol for the combined series. Using an initial population of 200 receptor models and a transcription-error rate set to 0.02, the different systems were allowed to evolve for 7000 crossover cycles corresponding to 14 generations for the chalcone dataset, 6000 (12 generations) for the stilbene dataset, and 50000 (100 generations) for the combined set.

The varying number of crossover cycles resulted from a desire to avoid overtraining the model during its development.[27] While the combined model yielded an improved predictive p^2 up to 70000 cycles (cf. Tables 3 and 4), the individual chalcone and stilbene models

Table 2. Monte Carlo Conformational Search

compd	no. of conformations found[a]	E_{gmin} (kJ/mol)[b]	visits of E_{gmin}[c]	no. of conformers chosen[d]
M01	20	191.35	263	4
M02	108	290.06	49	3
M03	175	285.59	33	5
M04	110	267.99	33	4
M05	63	287.30	84	4
M06	150	350.36	21	4
M07	217	327.86	31	2
M08	392	463.33	26	4
M09	255	246.13	19	6
M10	285	407.95	18	4
M11	148	338.80	37	2
M12	246	295.21	22	2
M13	527	383.86	18	3
M14	4	211.01	219	2
M15	8	245.18	125	2
M16	256	246.12	29	5
M17	297	350.81	30	5
M18	96	250.58	18	4
M19	71	309.44	87	4
M20	74	469.87	27	5
M22	210	346.30	24	3
M23	65	308.24	154	4
M24	286	389.75	40	6
M25	10	146.38	191	2
M26	20	171.69	95	2
M27	61	292.73	174	3
M28	221	288.09	17	4
M30	20	140.93	81	2
M32	8	195.85	127	2
M33	8	225.56	132	1
M34	159	281.05	27	4
M35	240	405.92	17	5
M36	134	291.05	29	2
M38	56	252.68	63	2
M39	386	449.86	25	5
M40	276	241.92	28	4
M41	124	263.47	36	6
M43	322	356.94	18	5
M48	156	281.73	35	2
M52	320	356.94	24	4
M54	475	382.68	20	3
M55	189	259.52	29	2
M63	118	213.16	33	5
M66	224	275.58	16	2
M67	288	220.41	16	2
M68	447	459.16	20	2
M69	342	288.54	26	2

[a] Number of conformations found in the Monte Carlo conformational searches performed within Macromodel 6.5 using the MMFF force field from 1000 starting structures generating conformations within 20.9 kJ/mol (5 kcal/mol) of the global minimum. During the conformational search, all structures were subjected to the truncated Newton conjugate gradient (TNCG) minimization method to within a derivative convergence criterion of 0.01 kJ/(Å·mol). [b] E_{gmin} is the global minimum energy in kJ/mol. [c] Number of times the global minimum conformation was visited during the Monte Carlo search. [d] Number of conformers chosen as input for multiconformational representation of the ligands.

encountered degradation in p^2 with increased cycles. As a result these two models were examined stepwise, 1000 cycles at a time, to ascertain the optimum p^2 (Table 3).

The simulations reached a cross-validated q^2 of 0.98, 0.96, 0.92, respectively, and a predictive p^2 of 0.71, 0.51, 0.46, respectively, for the chalcone, stilbene, and combined series. These quantities reflect values averaged over the 200 models.

Finally, the validity of the model families was assessed by a series of five scramble tests. The resulting predictive p^2 values of 0.34/0.30/0.27/−0.35/−0.89 (average: −0.06), 0.17/0.15/0.05/−0.01/−0.108 (average: 0.05) and −0.19/−0.14/−0.97/−0.587/−0.386 (average: −0.45), respectively, demonstrate the sensitivity of the respective surrogate families to the experimental IC_{50} data and the inability of a random model to establish a QSAR. By implication, the models described above are not the result of a fortuitous fit of the data.

Table 3. Summary of Receptor Model as Generated by Quasar

parameters	chalcones	stilbenes	combined
	Datasets		
training/test set	14/5	12/6	36/11
conformations	55/16	42/17	128/32
	Induced Fit		
rms (field min.) Å	0.83−0.94	0.77−1.10	0.93−1.31
associated energies in kcal/mol	0.7−0.9	0.7−1.0	0.8−1.2
	Quasar		
size of population	200	200	200
cross-validation groups	4/4/6	4/4/4	12/12/12
crossovers	7000	6000	50 000
cross-validated q^2 [a]	0.98[c]	0.96[c]	0.917[c]
classical r^2	0.982[c]	0.964[c]	0.92[c]
predictive p^2 [b]	0.711[c]	0.51[c]	0.46[c]
scramble p^2 [b]	−0.06[c]	0.05[c]	−0.45[c]
	Models		
rms_{av} training in kcal/mol (IC_{50} factor)	0.1 (1.2)[c]	0.1 (1.2)[c]	0.3 (1.7)[c]
rms_{max} training in kcal/mol (IC_{50} factor)	0.2 (1.4) M10	0.3 (1.6) M15	0.9 (4.9) M43
rms_{av} test in kcal/mol (IC_{50} factor)	0.7 (3.5)[c]	0.6 (2.8)[c]	0.8 (4.0)[c]
rms_{max} test in kcal/mol (IC_{50} factor)	1.0 (5.6) M40	1.0 (5.3) M33	1.6 (15.0) M52
no. of particles	172[c]	210[c]	331[c]
lack of fit[d]	0.327[c]	0.256[c]	0.708[c]
Boltzmann distribution[e]	0.092	0.060	0.150
slope[f]	0.193	0.117	0.252
intercept[f]	−2.979	−4.313	−5.853

[a] q^2 represents the leave-one-out cross-validated correlation coefficient. [b] p^2 represents the predictive correlation coefficient. [c] Data reflects average over 200 models. [d] The lack-of-fit (lof) function (accounting for the number of mapped properties, model uniqueness, and model selectivity) controls the soundness of the receptor family. lof = rms $[\Delta G°_{pred} - \Delta G°_{exp}]/\{1.0 - (p_{part} + p_{diff} + p_{sele})/3.0\}$ where p_{part} represents a penalty for models with relatively many properties mapped on their surface, p_{diff} represents a penalty for relative model similarity when compared with all other models, and p_{sele} represents a penalty for unspecific selection of the conformer/orientomer/protomer ensemble. [e] Contributions of individual conformers are Boltzmann-weighed and averaged over the 200 receptor models. [f] Free energies of ligand binding, $\Delta G°_{pred}$, are predicted by means of a linear regression equation using ligands in the training set: $\Delta G°_{pred} = aE_{bdg} + b$ where $E_{bdg} = E_{lig-rec} - T\Delta S_{bdg} - E_{solv,lig} - \Delta E_{int,lig} - E_{env,lig}$ ($E_{lig-rec}$ represents the force-field energy of the ligand−receptor interaction, $T\Delta S_{bdg}$ the change in ligand entropy upon receptor binding, $E_{solv,lig}$ the ligand desolvation energy, $\Delta E_{int,lig}$ the change in ligand internal energy upon receptor binding, and $E_{env,lig}$ the energy uptake required for modifying the averaged receptor envelope).

Table 4. Cross-Validated q^2 and Predictive p^2 as a Function of Training Using the Number of Cross-Validation Cycles for the Combined Series

crossovers	q^2	p^2
5000	0.661	0.155
10000	0.785	0.239
15000	0.844	0.289
20000	0.859	0.316
50000	0.917	0.464
60000	0.941	0.448
70000	0.948	0.439

Discussion

Data Set. The goal of this study was to construct a 5D-QSAR model for the colchicine-binding site of tubulin using more than one structural class of ligands. To date, only individual QSAR models have been reported for colchicine[28] and combretastatin analogues[29] with respect to inhibition of tubulin polymerization. It is generally believed that compounds that bind at the same site or overlapping sites should at least share a common subset of structural features responsible for activity.[30] The paclitaxel binding center on tubulin might be unique in this regard given that paclitaxel, the epothilones, discodermolide, eleutherobin, and the sarcodictiyns individually stabilize microtubules from this binding site.[31] In spite of the unusual diversity of the ligands, the latter has been characterized by a "common pharmacophore" hypothesis.[32] More recently, however, it has been shown that paclitaxel and epothilone A, at least, do not share a common subset of pharmacophore elements.[33] For the colchicine site we have adopted the common pharmacophore hypothesis and focused on structurally similar classes that inhibit tubulin polymerization through this site. The series is built around a framework including two aromatic groups, linked by simple functionality [ether, alkene (stilbene), or enone (chalcone)].

To establish a QSAR model for colchicine-site-centered tubulin assembly inhibitors, we chose molecules which featured conformational flexibility and functional diversity and spanned 4 orders of magnitude in binding affinity (estimated from IC_{50} 0.1−100 μM). All experimental data were obtained in a single laboratory at the UMIST and the Paterson Institute for Cancer Research (Manchester, England),[34−37] making comparison more reliable than would be the case for literature data scattered across time and geography. Inhibition of tubulin assembly was expressed as IC_{50} (μM) values, which represent the drug concentration that causes 50% inhibition of tubulin assembly. The selected compounds inhibited the binding of [^{3}H] colchicine to tubulin. All the biological data were referenced to an experimental value for combretastatin A4 **2** (cf. M04 Figure 1).

Monte Carlo Conformational Searches. These were performed to permit additional flexibility in the development of a QSAR in the context of the 4D features of the program.[22] In general, a QSAR treatment coupled to pharmacophore development involves a plausible alignment of ligands that yields the best contextual correlation. The same is true in the present study. We attach no particular biological significance to the global minima recorded in Table 2. These low energy structures along with 1−5 other low energy minima were employed to give the Quasar genetic algorithm a maximal opportunity to generate the best possible SAR correlations. It is conceivable that the resulting pharmacophore arising from a subset of structures pictured in Figures 2 and 3 corresponds to the actual bound conformations. In our mind, it is equally likely that they do not. Nonetheless, in the spirit of ligand-based design, we utilize the resulting alignments in anticipation of generating a predictively useful model. In the case of the chalcone series, the choice of conformers proved to be critical. The methyl and methoxy substituents on the bridge between the aromatic rings (α position) in the chalcone series make a sizable difference in the preferred conformation of the ligands. While chalcones usually adopt an s-cis conformation, α-substituted chalcones prefer to adopt the s-trans conformation.[14] The latter is closer to the colchicine and combretastatin A4 preferred conformations. No useful QSAR model could be obtained when the α-substituted chalcones were represented by conformers adopting both the s-cis and s-trans conformations.

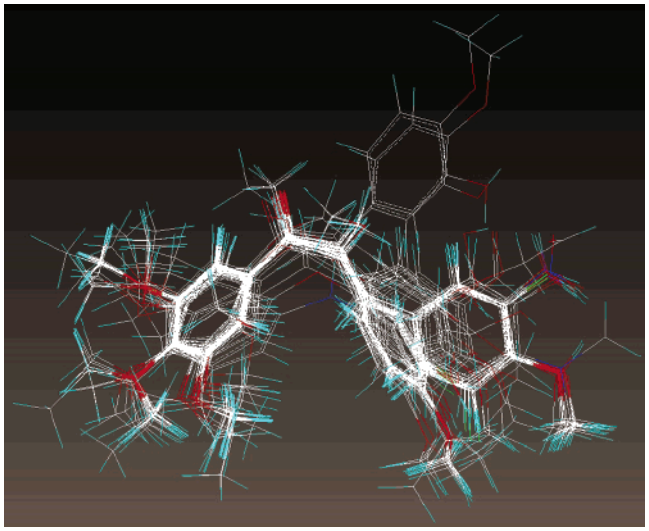


Figure 2. APOLLO alignment of the 47 ligands (160 conformers) in the combined series.

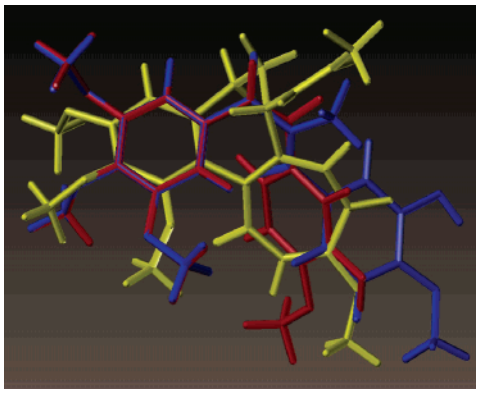


Figure 3. APOLLO alignment of colchicine **1** (yellow), combretastatin A4 **2** (red), and α-methyl chalcone **3** (blue).

Ligands M28 and M40 were tested for their biological activity as racemates. To accommodate this, both enantiomers were represented by two conformations each in the dataset.

APOLLO Alignment. Molecular alignment is one of the most critical aspects in crafting a successful QSAR model. We explored several possible alignments derived from *SEAL*[38] and *FlexS*.[39] These programs attempt to simultaneously fit both volume (or steric) and electrostatic features across the entire molecular frameworks for a pair of structures being superimposed. Sometimes this leads to a tight atomic alignment for substructures such as benzene rings and certain functional groups. More often, however, the fits are a compromise in which the steric and electrostatic requirements are optimally satisfied, but the overall structures do not adopt a neat intuitive or visually pleasing superposition. Such alignments are not accepted with grace by Quasar. Loosely

aligned training sets furnish a reasonable correlation of the biodata with difficulty. The corresponding test sets frequently deliver no useful QSAR, although similar structures are well represented in the training set. In our experience, the key to obtaining a correlation with acceptable statistics within Quasar is to align the training and test molecules such that the bulk of the molecular backbones are tightly superposed, only the substituents being "frayed" around the edges of the pharmacophore. Figure 2 illustrates a suitable alignment. One unfortunate consequence of this strategy appears to be that molecules populating a common pharmacophore must be quite similar in structure in order to deliver a useful QSAR correlation. The paclitaxel mimics mentioned above, for example, would be difficult, if not impossible, to treat in the Quasar framework.

Returning to colchicine, successful models for the chalcone and stilbene series were eventually obtained by using an alignment generated using the APOLLO software (Figure 2). Initially, only three compounds were aligned to find an optimal superposition: colchicine **1**, CA4 **2**, and α-methyl chalcone **3** (Figure 3). The template colchicine conformation is very similar to the global minimum derived from an MMFF conformational search (see Supporting Information for computational details/APOLLO). The conformations resulting in the best fit of the three ligands were kept as a global template for further alignment. The input for the APOLLO superposition searches consisted of a set of conformations for each ligand from the Monte Carlo conformational searches and the **1**−**3** template target. All the selected ligand molecules are diaryl compounds, and most contain the trimethoxyphenyl moiety that is beneficial for tubulin binding activity at the colchicine site. We therefore decided to superpose the three corresponding oxygen atoms of the A-rings tightly while the two oxygen atoms on the C-rings were allowed to superpose with less constraint. The energies of the conformations were used together with the rms atomic deviations to score the best fit. Conformations were selected on the basis of this score as well as diversity in orientation. For the 47 ligand molecules defining the data set, the inclusion of multiple conformers resulted in a total of 160 conformers (Table 2).

5D QSAR Model. The development of a model accounting simultaneously for more than one structural class of compounds also proved to be a challenge. We opted for stepwise model development since the direct approach was unsuccessful. First, individual QSAR models were devised for the chalcone and stilbene series. Then, all the ligands were combined to provide a more general model. One factor that greatly affected the generation of the latter successful models was a balanced selection of member structures for both training and test sets. In some cases, poorly predicted $\Delta G°_{pred}$ resulted from attempts to project this property for analogues with structural features not found, or very poorly sampled, in the training set. The training set molecules were therefore selected to represent the broadest possible spectrum of structural features and biological activities. In addition, the most active compounds were included in the training sets so as to provide critical information concerning pharmacophore

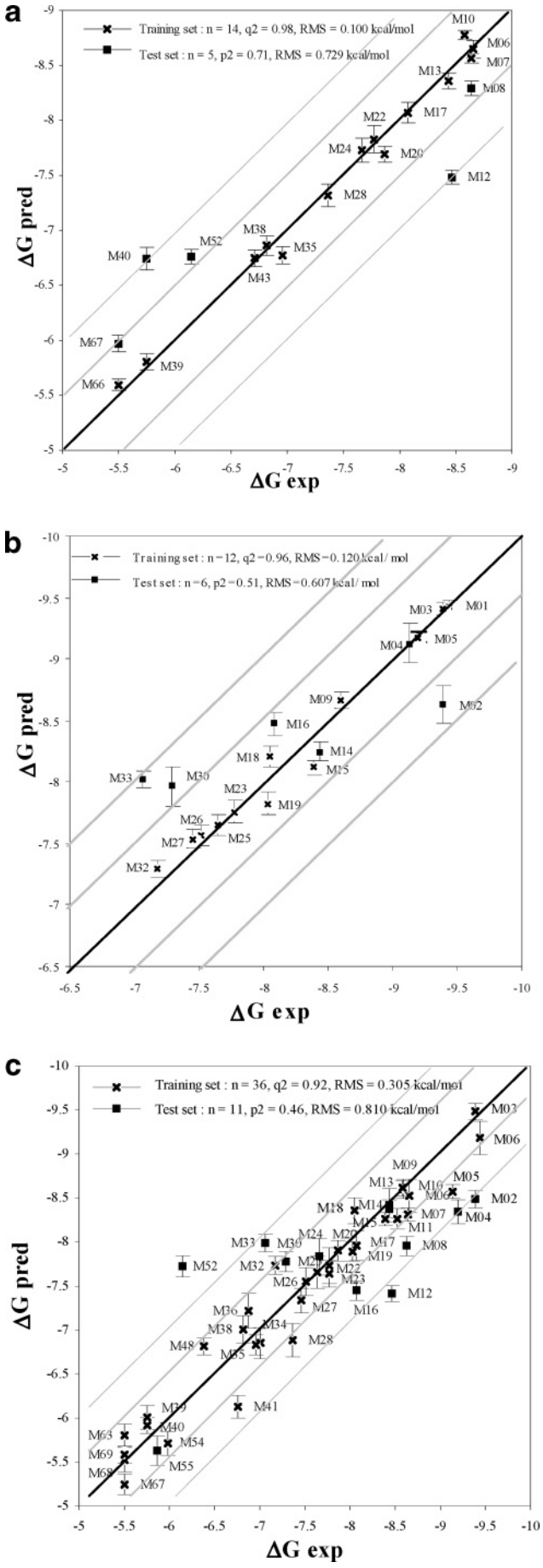


Figure 4. Graphical comparison of $\Delta G°_{exp}$ and $\Delta G°_{pred}$ binding affinities ($\Delta G°$ in kcal/mol).

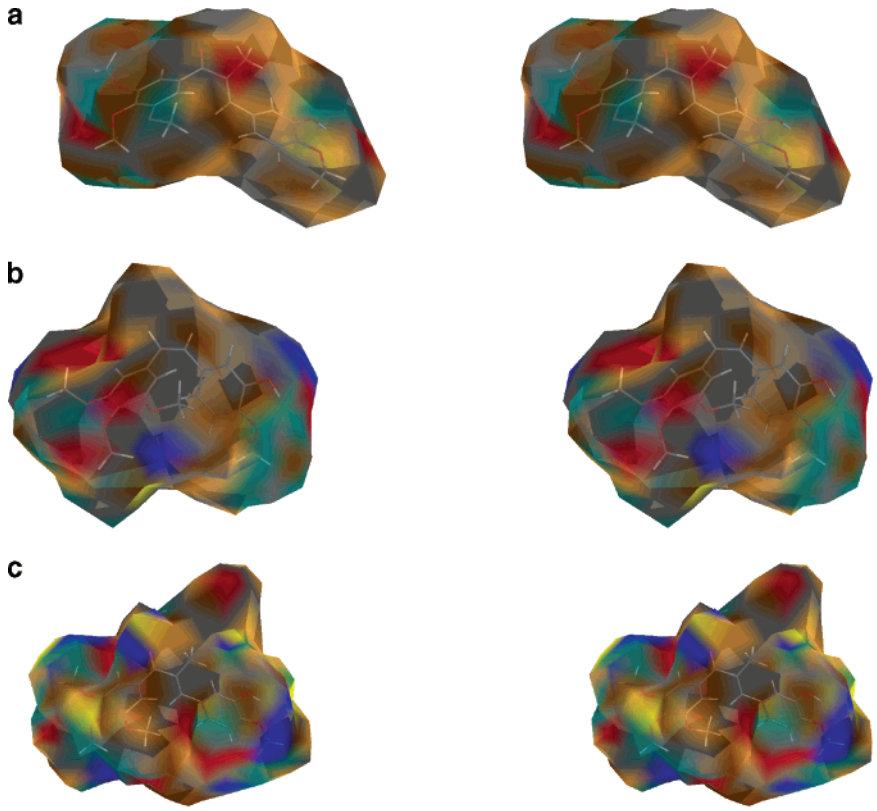


Figure 5. 5D QSAR receptor models. Color-coding: red, positively charged salt bridge; blue, negatively charged salt bridge; yellow, H-bond donor; green, H-bond acceptor; orange, positively charged hydrophobic; brown, negatively charged hydrophobic; gray, neutral hydrophobic. The depicted model corresponds to the most frequently occurring property, averaged over the 200 receptor models. (a) Chalcone series showing chalcone **3** (M07). (b) Stilbene series showing combretastatin **2** (M04). (c) Combined series showing chalcone **3** (M07).

requirements. The test set molecules were selected such that their key structural features were adequately sampled in the training set.[40]

To establish a QSAR model for the inhibition of tubulin assembly by the chalcone series, we selected 19 ligands (14 for the training set and 5 for the test set) which were represented by 71 conformers (55 for the training set and 16 for the test set). On average, the predicted $\Delta G°_{pred}$ for the training set ligands vary slightly by 0.1 kcal/mol from the experiment; the maximal observed deviation is 0.2 kcal/mol for ligand M10. For the test set, the predicted $\Delta G°_{pred}$ value deviates by 0.7 kcal/mol from the experiment; the maximal observed deviation is 1.0 kcal/mol for ligand M40. This indicates that this 5D-QSAR model has good predictive powers. Experimental and predicted values are compared in Figure 4a. Representation of the resulting quasi-atomistic receptor model is depicted in Figure 5a. In general, the best inhibitors of the tubulin assembly in this series were the α-substituted chalcones (M07, M08, M10), which adopt preferentially the s-trans conformation. Also chalcones bearing a hydroxyl group at the 3-position of the C-ring were better inhibitors than those with fluorine at the same position. Moreover, the presence of more than one hydroxyl group on the C-ring proved very detrimental to the biological activity. The only significant outlier was ligand M40 (test set), which was poorly predicted. The ligand was represented not only by different conformers but also by two enantiomers since the compound was tested as a racemate for its ability to inhibit tubulin assembly. This suggests that although M40 as a racemate is a poor inhibitor, one of the enantiomers may be able to inhibit tubulin assembly more effectively than the other.

The QSAR model for the stilbene series was generated from 18 ligands (12 for the training set and 6 for the test set), represented by 59 conformers (42 for the training set and 17 for the test set). On average, the predicted $\Delta G°_{pred}$ values for the training set ligands vary slightly by 0.1 kcal/mol from the experiment; the maximal observed deviation is 0.3 kcal/mol for ligand M15. For the test set, the predicted $\Delta G°_{pred}$ value deviates by 0.6 kcal/mol from the experiment; the maximal observed deviation is 1.0 kcal/mol for ligand M33. This suggests that the 5D QSAR model results in a model with moderate predictive powers. Experimental and predicted values are compared in Figure 4b. The results for this model are visualized in Figure 5b. In general the best inhibitors for the tubulin polymerization were the *cis*-stilbenes. As for the chalcone series, stilbenes bearing a hydroxyl group at the 3-position of the C-ring were better inhibitors than those with fluorine at the same position. Sterically larger substituents on the A-ring were not detrimental to the biological activity, and it was found that oxygenated substituents on the A-ring were not a requirement for high activity. Substituents on the bridge of the stilbene series are also well tolerated.

For the combined ligands, the QSAR model was constructed from all 47 ligands (36 for the training set and 11 for the test set) represented by 160 conformers (128 for the training set and 32 for the test set). On average, the predicted free energies of ligand binding obtained with the 5D QSAR approach vary from experi-

ment somewhat more by comparison with the chalcone and stilbene sets. The training and test set deviations correspond to 0.3 and 0.8 kcal/mol, respectively. Taken at face value, the computed quantities are still within experimental error, implying that 5D QSAR may be reasonably predictive across more than one class of structures, provided that the pharmacophore ligands are very tightly aligned. However, analysis of the experimental and predicted values for the test set in Figure 4c suggests an alternative interpretation. In Figure 4c, apart from the weakly acting M55, all other *test ligands* are predicted to fall within an IC_{50} window spanning a range of 6.5 (ΔG 1.1 kcal/mol) corresponding to an experimental range of 240. It would appear that although combining the two classes of ligands delivers a credible training set correlation (cross validated $q^2 = 0.92$), it also leads to a rather low predictive p^2 (0.46) and a clustering of predicted test ligand IC_{50}'s not observed in the individual chalcone and stilbene models. The result is consistent with the observation made in the APOLLO discussion above; namely, unless the core structures of the ligands are tightly aligned, a good correlation is difficult to achieve. In the combined data set, the core rings obviously do not meet the alignment criterion. Although we have not traced this observation to its source, we speculate that, in spite of the multidimensional aspects of 5D QSAR as implemented in Quasar, the atomistic receptor approach employed by the method is insufficiently flexible to handle ligand diversity that involves multiple ligand classes. We have experienced a similar outcome in applications to a number of other problem areas.[41]

Representations of the resulting quasi-atomistic receptor models are shown in Figure 5a−c. For the present set of tubulin assembly inhibitors, it is obvious that a polar substituent at the C-3 position on the C-ring increases biological activity. The corresponding region of the model is filled with a collection of yellow regions corresponding to areas of H-bond acceptors on the receptor. This is also recognized as a consequence of functional group analysis where the hydroxyl group is colored green, showing the positive contribution of this moiety to the biological activity. It is interesting to note the presence of both yellow and blue regions (H-bond acceptor and negatively charged salt bridge) surrounding the methoxy groups at positions 3 and 4 on the A-rings. When colchicine binds to tubulin, this region of the receptor is thought to be populated by cysteine residues.[42] In general, the model is surrounded by many green and gray regions, corresponding to H-bond acceptor and hydrophobic groups on the receptor, hence confirming the adequate position of the methoxy groups on the ligands and the overall hydrophobic character of the ligands.

The separate models for the chalcone and stilbene series appear slightly different when represented by the symbolic QSAR color-coded surfaces. Although the corresponding surface model derived for the combined model can be interpreted as reconciling the differences, the poor predictivity of this QSAR suggests that the single structure class models should be employed for design of new analogues. Synthesis and testing of genuine chalcone−stilbene hybrids might ultimately lead to a true union of the classes.

Conclusion

In the absence of an experimentally determined receptor structure, we constructed a 5D-QSAR model for a series of tubulin assembly inhibitors interacting at the colchicine-binding site. This study has produced the first predictive model for multiple classes of agents populating the colchicine-binding site of tubulin. It is clear that such a model should be refined to include other classes of ligands that interact at the colchicine-binding site. We are currently developing a version of this model that should prove capable of predicting the biological activities of phenastatins, podophyllotoxins, and, of course, colchicinoids.

Acknowledgment. The authors would like to acknowledge financial support from BBSRC (106/B18017) for a postdoctoral fellowship for G.M. and the Emerson Center for Scientific Computing (Emory University) for a visiting fellowship award for S.D. We are also grateful to Ami Lakdawala, Jim Nettles, Ben Cornett, and Prof Dennis C. Liotta (Emory University) for generous assistance and hospitality during the course of the work and to Prof. Angelo Vedani (University of Basel) for helpful comments and insight.

Supporting Information Available: Computational details. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hadfield, J. A.; Ducki, S.; Hirst, N.; McGown, A. T. Tubulin and Microtubules as target for Anticancer Drugs. *Prog. Cell Cycle Res.* **2003**, 309−325.
(2) Schiff, P. B.; Fant, J.; Horwitz, S. B. Promotion of Microtubule Assembly in vitro by Taxol. *Nature* **1979**, *277*, 665−667.
(3) Nogales, E.; Lowe, J.; Li, H. L.; Downing, K. H. Refined structure of alpha beta-tubulin to 3.5 angstrom. *Biophys. J.* **2002**, *82*, 2488.
(4) Snyder, J. P.; Nettles, J. H.; Cornett, B.; Downing, K. H.; Nogales, E. The binding conformation of Taxol in beta-tubulin: A model based on electron crystallographic density. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5312−5316.
(5) Uppuluri, S.; Knipling, L.; Sackett, D. L.; Wolff, J. Localization of the Colchicine-Binding Site of Tubulin. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11598−11602.
(6) Pettit, G. R.; Singh, S. B.; Hamel, E.; Lin, C. M.; Alberts, D. S.; Garciakendall, D. Antineoplastic Agents. 145. Isolation and Structure of the Strong Cell-Growth and Tubulin Inhibitor Combretastatin-a-4. *Experientia* **1989**, *45*, 209−211.
(7) Woods, J. A.; Hadfield, J. A.; Pettit, G. R.; Fox, B. W.; McGown, A. T. The Interaction with Tubulin of a Series of Stilbenes Based on Combretastatin a-4. *Br. J. Cancer* **1995**, *71*, 705−711.
(8) Tozer, G. M.; Kanthou, C.; Parkins, C. S.; Hill, S. A. The biology of the combretastatins as tumour vascular targeting agents. *Int. J. Exp. Pathol.* **2002**, *83*, 21−38.
(9) Chaplin, D. J.; Hill, S. A. The development of combretastatin A4 phosphate as a vascular targeting agent. *Int. J. Radiat. Oncol. Biol. Phys.* **2002**, *54*, 1491−1496.
(10) Marx, M. A. Small-Molecule, tubulin-binding compounds as vascular targeting agents. *Expert Opin. Ther. Pat.* **2002**, *12*, 769−776.
(11) Li, Q.; Sham, H. L. Discovery and Development of Antimitotic Agents that Inhibit Tubulin Polymerisation for the Treatment of Cancer. *Expert Opin. Ther. Pat.* **2002**, *12*, 1663−1702.
(12) Lawrence, N. J.; Rennison, D.; Woo, M.; McGown, A. T.; Hadfield, J. A. Antimitotic and Cell Growth Inhibitory Properties of Combretastatin A4-like Ethers. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 51−54.
(13) Lawrence, N. J.; McGown, A. T.; Ducki, S.; Hadfield, J. A. The interaction of chalcones with tubulin. *Anti-Cancer Drug Des.* **2000**, *15*, 135−141.
(14) Ducki, S.; Forrest, R.; Hadfield, J. A.; Kendall, A.; Lawrence, N. J.; McGown, A. T.; Rennison, D. Potent antimitotic and cell growth inhibitory properties of substituted chalcones. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1051−1056.
(15) Lawrence, N. J.; Rennison, D.; McGown, A. T.; Ducki, S.; Gul, L. A.; Hadfield, J. A.; Khan, N. Linked parallel synthesis and MTT bioassay screening of substituted chalcones. *J. Comb. Chem.* **2001**, *3*, 421−426.

(16) Lawrence, N. J.; Ducki, S.; Rennison, D.; Mackenzie, G.; McGown, A. Combretastatin-Like Chalcones as Active Microtubule Inhibitors Possessing Potent Tumour Vasculature Targeting Anticancer Activity. Manuscript in preparation.
(17) Kubinyi, H. QSAR and 3D QSAR in drug design Part 2: applications and problems. *Drug Discovery Today* **1997**, *2*, 538−546.
(18) Kubinyi, H. QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discovery Today* **1997**, *2*, 457−467.
(19) Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.* **2002**, *45*, 2139−2149.
(20) Vedani, A.; Dobler, M.; Zbinden, P. Quasi-atomistic receptor surface models: A bridge between 3-D QSAR and receptor modeling. *J. Am. Chem. Soc.* **1998**, *120*, 4471−4477.
(21) Vedani, A.; Dobler, M. Multidimensional QSAR: Moving from three- to five-dimensional concepts. *Quant. Struct.-Act. Relat.* **2002**, *21*, 382−390.
(22) Vedani, A. 4D-QSAR and beyond. *Quant. Struct.-Act. Relat.* **2002**, *21*, 347−347.
(23) Andreu, J. M.; Perez-Ramirez, B.; Gorbunoff, M. J.; Ayala, D.; Timasheff, S. N. Role of the colchicine ring a and its methoxy groups in the binding to tubulin and microtubule inhibition. *Biochemistry* **1998**, *37*, 8356−8368.
(24) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(25) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(26) AMSOL; http://comp.chem.umn.edu/amsol/.
(27) Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959−3969.
(28) Zhang, S. X.; Feng, J.; Kuo, S. C.; Brossi, A.; Hamel, E.; Tropsha, A.; Li, K. H. Antitumor agents. 199. Three-dimensional quantitative structure−activity relationship study of the colchicine binding site ligands using comparative molecular field analysis. *J. Med. Chem.* **2000**, *43*, 167−176.
(29) Brown, M. L.; Rieger, J. M.; Macdonald, T. L. Comparative molecular field analysis of colchicine inhibition and tubulin polymerization for combretastatins binding to the colchicine binding site on [beta]-tubulin. *Bioorg. Med. Chem.* **2000**, *8*, 1433−1441.
(30) Liljefors, T.; Pettersson, I. Computer-Aided Development of Three-Dimensional Pharmacophore Models. In *A Textbook of Drug Design and Development*; Harwood Academic Publishers: Reading, U.K., 1996; pp 63−70.
(31) Jimenez-Barbero, J.; Amat-Guerri, F.; Snyder, J. P. The Solid State, Solution and Tubulin-Bound Conformations of Agents that Promote Microtubule Stabilization. *Curr. Med. Chem.: Anti-Cancer Agents* **2002**, *2*, 91−122.
(32) Ojima, I.; Chakravarty, S.; Inoue, T.; Lin, S. N.; He, L. F.; Horwitz, S. B.; Kuduk, S. D.; Danishefsky, S. J. A common pharmacophore for cytotoxic natural products that stabilize microtubules. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4256−4261.
(33) Nettles, J. H.; Li, H. L.; Cornett, B.; Krahn, J. M.; Snyder, J. P.; Downing, K. H. The binding mode of epothilone A on alpha,beta-tubulin by electron crystallography. *Science* **2004**, *305*, 866−869.
(34) Ducki, S. Anticancer Drugs from Traditional Chinese Herbs. In *Department of Chemistry*; UMIST: Manchester, 1997; p 237.
(35) Woo, M. The Design and Synthesis of Tubulin-Binding Anticancer Agents. In *Department of Chemistry*; UMIST: Manchester, 2001; p 142.
(36) Rennison, D. Design and Synthesis of New Anticancer Drugs. In *Department of Chemistry*; UMIST: Manchester, 2001; p 275.
(37) Gaukroger, K. Structure-Activity Study of Stilbenes Targeted at Tumour Vasculature. In *Department of Chemistry*; UMIST: Manchester, 2001; p 290.
(38) Klebe, G.; Mietzner, T.; Weber, F. Different Approaches toward an Automatic Structural Alignment of Drug Molecules−Applications to Sterol Mimics, Thrombin and Thermolysin Inhibitors. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 751−778.
(39) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: A method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502−4520.
(40) Golbraikh, A.; Shen, M.; Xiao, Z. Y.; Xiao, Y. D.; Lee, K. H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241−253.
(41) Lakdawala, A.; Geballe, M.; Kurtkaya, S.; Snyder, J. P. Unpublished data.
(42) Bai, R. L.; Covell, D. G.; Pei, X. F.; Ewell, J. B.; Nguyen, N. Y.; Brossi, A.; Hamel, E. Mapping the binding site of colchicinoids on beta-tubulin−2-chloroacetyl-2-demethylthiocolchicine covalently reacts predominantly with cysteine 239 and secondarily with cysteine 354. *J. Biol. Chem.* **2000**, *275*, 40443−40452.

JM049444M