

# A Binding Affinity Based Computational Pathway for Active-Site Directed Lead Molecule Design: Some Promises and Perspectives

N. Latha and B. Jayaram\*

FINAL

Department of Chemistry and Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India

Received: November 22, 2004; Accepted: ??????? ??, 2004

**Abstract:** Drug discovery in the 21<sup>st</sup> century is expected to be different in at least two distinct ways: development of individualized medicine utilizing genomic information and emergence of an integrated in silico protocol for facilitating target identification, structure prediction and lead discovery. The expectations from computational methods for developing suggestions on potential leads reliably and expeditiously, are continuously on the increase. Several conceptual and methodological concerns remain before an automation of lead design in silico could be contemplated. The novelty of the candidates generated, their geometries, the partial atomic charges and other force field parameters for enabling energy evaluations is one concern. A proper account of the flexibility of the candidate molecule and the target, a consideration of solvent and salt effects in binding and a reliable methodology for developing quantitative estimates of binding affinities is another. Finally the drug-likeness of the candidates generated is yet another concern. Each of these issues warrants a careful consideration. In this review, we sketch a system independent, binding free energy based, comprehensive computational pathway from chemical templates to lead-like molecules, given the three dimensional structure of the target protein and a definition of its active site, focusing on some emerging in silico trends and techniques. We survey current methods for generation of candidate molecules and some popular protocols for docking candidates in the protein active site. We discuss the theory of protein-ligand binding in the rigorous framework of statistical mechanics and assess the current strategies for affinity based filtering of candidates. We address concerns related to flexibility of the target and the candidate, solvent and salt effects in lead design. We present a realization of the pathway proposed in a high performance computing environment for cyclooxygenase-2 target wherein the computational protocols could sort drugs from non-drugs, assuring the viability of the overall strategy. We highlight a few case studies indicating the current level of agreement between theory and experiment in eliciting binding affinities. Finally, we present a critical assessment of the computational steps involved in binding affinity based active site directed lead molecule design and further improvements envisioned for potential automation.

**Keywords:** Computer-aided drug design, template library, drug-like filters, *ab initio* charges, docking, protein-ligand interactions, binding free energy, molecular dynamics simulations.

## I. INTRODUCTION

The dawn of the 21<sup>st</sup> century heralds a new era in the drug discovery process. Access to the complete human genome sequence as well as to the complete sequences of pathogenic organisms provides us with information that can result in an avalanche of therapeutic targets. As structures of more protein targets become available through crystallography, nuclear magnetic resonance and bioinformatics, the need grows for a set of computational tools that can identify and analyze active sites and suggest compounds that can bind to these sites specifically (Fig. 1).

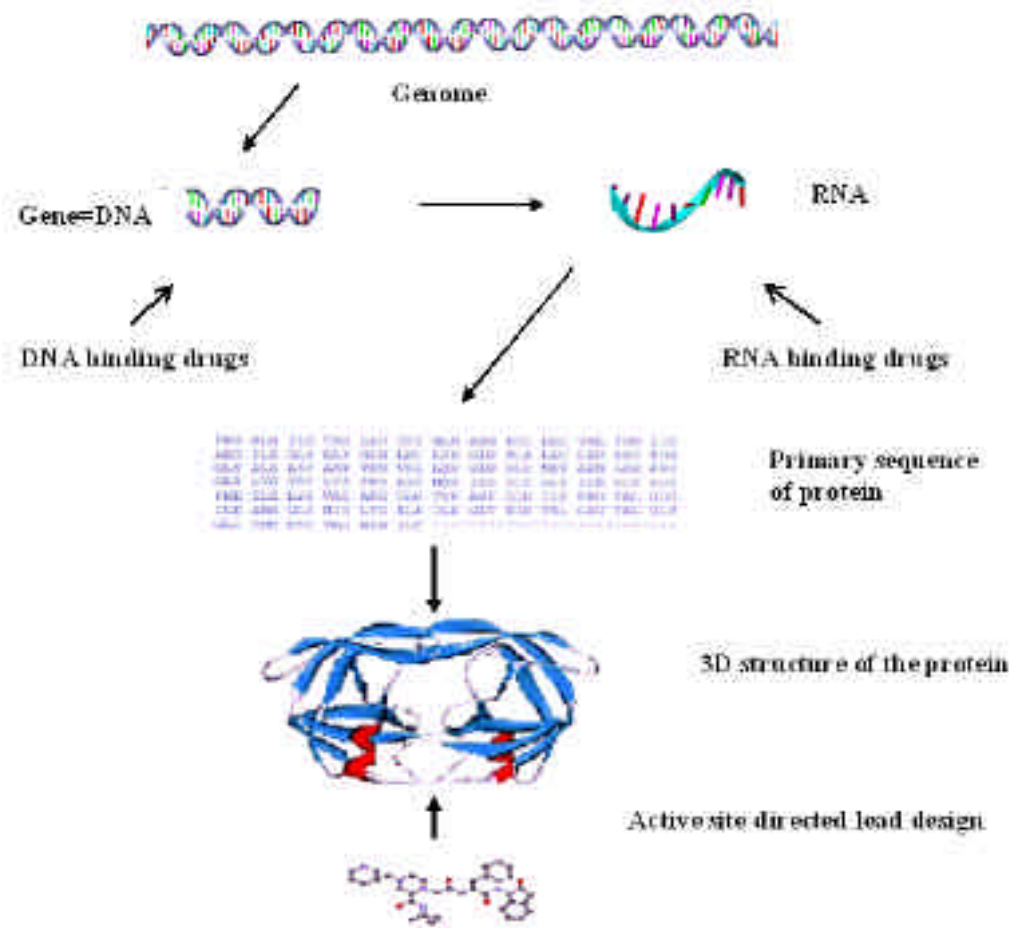
Given the vast size of organic chemical space (>10<sup>18</sup> compounds) [1], drug discovery cannot be reduced to a simple “synthesize and test” drudgery. There is an urgent need particularly for life threatening diseases to identify and / or design lead-like molecules from the vast expanse of what could be synthesized. *In silico* intervention at various stages

in the drug discovery cycle is now well acknowledged for its potential to reduce both time and cost involved.

Over the years, computational approaches have been explored intensely to extract all the relevant information from the available target structures and to use it in an efficient and intelligent manner to design improved ligands. There are approximately 6000 drugs in the market (CMC Database 94.1) [2] directed to about 500 targets [3]. Bioinformatics and structure based drug design are expected to multiply these numbers significantly within the next few years. The reader is referred to some excellent reviews published during the last decade on computer aided drug design [4-47] addressing fragment based libraries for *de novo* design, QSAR methods, docking and ranking candidates etc. Research in the pharmaceutical industry today is highly interdisciplinary in nature and several contributions [48-61] are available covering a rich spectrum of various methods used in drug discovery programmes.

In this review, we sketch a binding affinity based comprehensive computational pathway for lead design starting from an atomic level definition of chemical templates and target structure (Fig. 2). We attempt to collate

\*Address correspondence to this author at the Department of Chemistry and Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India; Tel: 91-11-2659 1505; Fax: 91-11-2658 2037; E-mail: bjayaram@chemistry.iitd.ac.in



**Fig. (1).** The extended arrow of the central dogma of the 21<sup>st</sup> century for drug discovery. The availability of genomic information makes it possible to conceive of a genome to drug discovery pathway. The arrow envisages automated gene finding, protein structure prediction and active site directed lead design as an integrated protocol.

the state of the art collective endeavors and present a general scheme applicable to any biological target with the express goal of paving the way for eventual automation in a high performance-computing environment. We then proceed to examine the merits and limitations of the diverse protocols at each step of the proposed pathway from the standpoint of theoretical rigor and computational expedience. Finally, we demonstrate and discuss the overall viability of the pathway with case studies on a few drug targets.

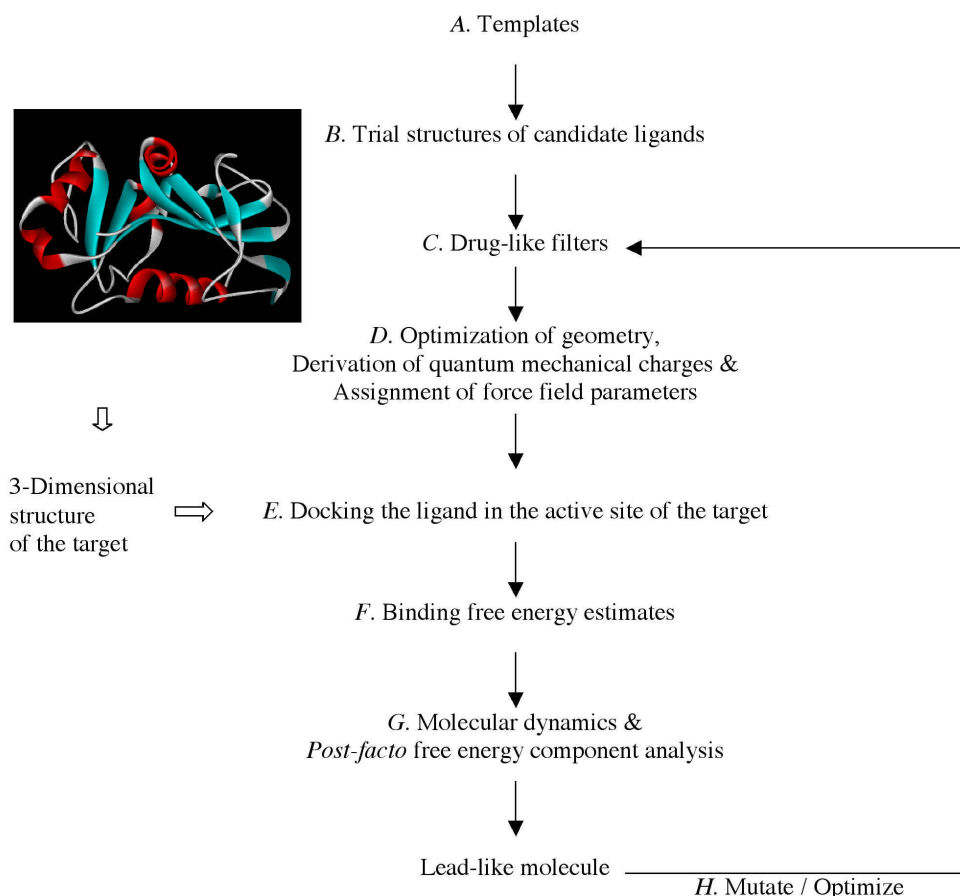
## II. A COMPUTATIONAL PATHWAY FROM CHEMICAL TEMPLATES TO LEAD-LIKE MOLECULES

The proposed computational pathway towards lead design (Fig. 2) envisages making any number of known or new candidate molecules out of a small but versatile set of building blocks called templates. These candidates are then screened for drug-likeness. Their geometry, partial atomic charges and other force field parameters are determined. The candidates are then docked in the active site of a given biological target and a qualitative picture of the binding affinity is developed. In a select few promising cases, molecular dynamics simulations with explicit solvent and salt on the biomolecular target, the candidate and the

complex are performed. This is followed by a rigorous analysis of the binding free energy for further optimization of the candidate. We consider here each of the steps enumerated above.

### Templates for Candidate Molecule Generation

Chemical templates are conceived as building blocks / structural frameworks for assembly and generation of new molecules. A necessary but not sufficient condition for creating a library of templates is that the structural and functional space of all known drugs be sampled. Additional criteria include novelty of the candidates generated upon combination of the templates in the defined set. Several alternative proposals exist in the literature for creating fragment / substructure based libraries [62-82] which typically involve some correlation with drug-likeness or biological activity to a target. Some excellent expositions addressing recent advances in ligand design methods are available [83-85]. Fattori in a recent review [86] opines that the fragment approach is the emerging philosophy in lead discovery. Although not apparent from the nomenclature of template *vis-à-vis* fragment, these are two distinct approaches. Templates deal with chemical space and



**Fig. (2).** A pathway for active site directed lead compound design *in silico*

fragments with biological activity. The goal of the former is to ensure completeness of the coverage of chemical space initiating from building blocks irrespective of the activity. The latter is database driven with built in activity of the fragment, which has its advantages and limitations. It is conceivable that these two approaches could be integrated.

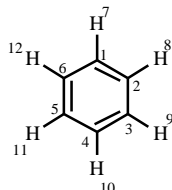
Earlier attempts by Bemis and Murcko on these lines led to an analysis of building blocks in drug design [87-88]. About 32 building blocks were compiled. These templates were identified on the basis of the preferred moieties (rings and side chains) commonly occurring in the drugs found in the Comprehensive Medicinal Chemistry drug database. We considered a chemical template based library approach with the possibility of including active site information. The templates designed, 160 in all as of now, are classified into three groups: rings, side chains and linkers. The appendix illustrates the molecular structural formulas of rings (labeled with prefix 'r') and side chains as well as linkers (labeled with prefix 's'). Most of the known drugs are recovered upon combination of the templates in the library.

Definition of templates to create molecules known or new, is only the first step. To map the templates as well as the candidate molecules generated on to energy space, certain minimum information (Table 1) besides the atoms involved and the connectivity of these atoms has to be generated and stored for each template.

## B. Generation of Candidate Molecules

As a step towards *de novo* lead design, candidates could be generated from chemical templates or fragments introduced in the previous step (section A). Recovery of the known drugs as well as novelty of the new candidates generated besides their activity are some prime considerations at this stage. Bamborough and Cohen [94] recently examined several methods for building candidate molecules. Two broad approaches are conceivable, *viz.* the molecules could be generated either outside or inside the active site. Making molecules outside is tantamount to *in silico* combinatorial chemistry. This is a computationally demanding brute force method involving random combination of templates but capable of generating novel molecules which may mostly be infructuous. This approach needs to be supplemented by methods to tackle the combinatorial explosion. One possibility is to introduce as a pre-processor, active-site information in terms of functional groups required and desired distances between substituents on potential candidates, volume and shape of the candidates etc. The alternative is to grow the candidates inside the active site – in an incremental construction, satisfying the above active-site complementarity criteria [95, 96]. In a way the 3D pharmacophore models [97] adopt this approach. Although this method could be restrictive of the chemical search space in view of the first committed choice of either

Table 1. Molecular Structure File\* for Chemical Template r1 (Benzene)

1	2	3	4	5	6	7	8	9	10	11	12	13
ATOM	1	C1	r1	1	-1.985	-0.471	0.000	0	CA	1.9080	0.0860	-0.1367
ATOM	2	C2	r1	1	-1.985	0.924	0.000	0	CA	1.9080	0.0860	-0.1367
ATOM	3	C3	r1	1	-0.777	1.622	0.000	0	CA	1.9080	0.0860	-0.1367
ATOM	4	C4	r1	1	0.432	0.924	0.000	0	CA	1.9080	0.0860	-0.1367
ATOM	5	C5	r1	1	0.432	-0.471	0.000	0	CA	1.9080	0.0860	-0.1367
ATOM	6	C6	r1	1	-0.777	-1.169	0.000	0	CA	1.9080	0.0860	-0.1367
ATOM	7	H7	r1	1	-2.937	-1.021	0.000	1	HA	1.4590	0.0150	0.1367
ATOM	8	H8	r1	1	-2.937	1.474	0.000	0	HA	1.4590	0.0150	0.1367
ATOM	9	H9	r1	1	-0.777	2.721	0.000	0	HA	1.4590	0.0150	0.1367
ATOM	10	H10	r1	1	1.384	1.474	0.000	0	HA	1.4590	0.0150	0.1367
ATOM	11	H11	r1	1	1.384	-1.021	0.000	0	HA	1.4590	0.0150	0.1367
ATOM	12	H12	r1	1	-0.777	-2.268	0.000	0	HA	1.4590	0.0150	0.1367
CONNECT	1	2	6	7								
CONNECT	2	1	3	8								
CONNECT	3	2	4	9								
CONNECT	4	3	5	10								
CONNECT	5	4	6	11								
CONNECT	6	5	1	12								
CONNECT	7	1										
CONNECT	8	2										
CONNECT	9	3										
CONNECT	10	4										
CONNECT	11	5										
CONNECT	12	6										
END												

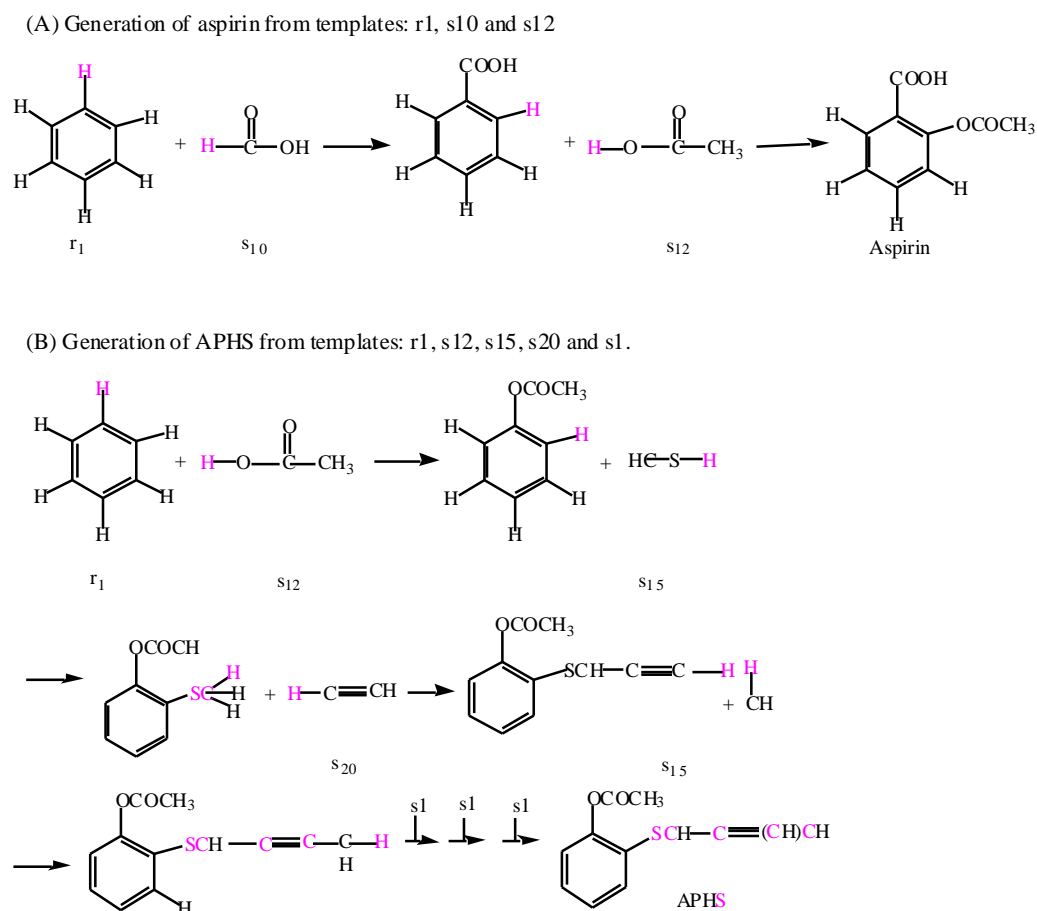
\*The molecular structure files organized in a modified protein data bank (RCSB) format [89], carry information on coordinates, connectivity etc. These are read from the first line-containing ATOM, till the END record is encountered. Column 1 denotes the record ID referred to as ATOM here. This is followed by the atom serial number, atom name, template identification code (assigned in the example as r1) and template sequence number in columns 2, 3, 4 and 5 respectively. The Cartesian coordinates (x, y and z in units of  $\text{\AA}$  obtained via quantum mechanical (AM1) geometry optimization [90]) are given in columns 6, 7 and 8. Atoms used for joining the template with another are indexed as 1 while values for other atoms are set to zero in column 9. This information is helpful in automating linkage of templates. The next column denotes the atom type. The van der Waals parameters ( $R^*$  in  $\text{\AA}$  and  $\epsilon$  in kcal) for each atom to be used in conjunction with a (12,6) potential function for energy calculations are assigned consistent with the AMBER force field [91,92] and tabulated in columns 11 and 12. The *ab initio* partial atomic charges derived with 6-31G\* basis set using GAMESS followed by RESP fitting in a force field compatible manner [93] are listed in column 13. The ATOM records are followed by bond connectivity information for each atom.

the functional groups or fragments decided to fill the active site with, it is attractive from computational expedience.

We illustrate here generation of candidates following the *in silico* combinatorial pathway (Fig. 3 and Table 2). For the purposes of energy based processing, information that each new molecule generated must carry is shown in (Table 2). Note that aspirin is obtained in the second generation (Fig. 3a, Table 2) and *o*-acetoxyphenyl hept-2-ynyl-sulfide (APHS) in the seventh generation (Fig. 3b) with the template library described in the appendix. The scheme presented for

generating new molecules *in silico* is not designed to mimic the *in vitro* conditions and may not represent a plausible synthetic route. This is a computationally convenient method for generating new molecules from building blocks.

The molecules so generated can be used for any biological target after screening them through drug-like filters discussed in the next section. Generating molecules randomly and passing them through drug-like filters to create a virtual molecule library [2] with the hope that they could be synthesized is one option. Constructing a pre-synthesized



**Fig. (3).** *In silico* scheme to combine templates sequentially to make candidates. The templates are connected (points of linkage shown in pink color) satisfying chemical valence and standard bond lengths and angles at each junction. The dihedrals around the junctions are kept at 180°.

molecule / drug library [98-101] is another option. The predominant consideration in most lead-design protocols is activity or ingrained active-site information such that the molecules generated bind well in the active site. The ease of their synthesis, however is important, and the intuition of organic chemist has to be converted into a computational filter particularly in the *in silico* combinatorial approach. Synthesizability, chiral centres, etc. are better handled as filters in the next step.

### C. Drug-Like Filters

A major impetus for a successful lead discovery strategy is provided by investing in techniques that would ensure bioavailability [102] from the very start in generating leads while eliminating wrong candidates from consideration. Some empirical computational filters such as Lipinsky's rules are thus introduced based on drug-like properties of known drugs [102-117]. This comprises estimating molecular weight, number of hydrogen bond donors, hydrogen bond acceptors and logP (the octanol-water partition coefficient) values before embarking on the more compute-intensive follow up steps leading to binding affinity estimates in further processing of the candidates after generation. Additionally, introduction of filters at this stage facilitates

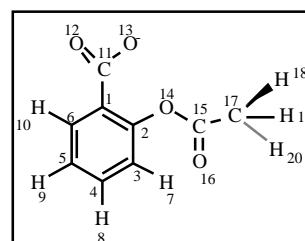
computational tractability by restricting the chemical space for potential candidates. Computer algorithms using just the list of atoms and their connectivity in the candidate molecule can execute the filtering. Some oft-used filters are reported in (Table 3). Ring strain [118] has also been reported recently to act as a new filter. It is desirable that ADME (Absorption, Distribution, Metabolism and Excretion) profiles [119-129] be converted to some empirical rules and introduced at this stage. Reliable predictions of ADME properties are emerging as early tools in lead-design protocols [130-131]. Search for toxicophores at the candidate generation stage eliminates undesirable leads. The development of computer-aided toxicity prediction techniques can be based on QSAR models or expert systems [132-136]. Also, synthesizability [137-138] and shelf life could be converted into useful computational filters for improving the efficiency of lead prediction protocols.

### D. Geometry, Charges, Force Field Parameters for the "Drug-Like" Candidates

To explore the conformational space of the candidate and to arrive at its equilibrium geometry in its free and bound states, a link between structure and energy needs to be

Table 2. Molecular Structure file\* for Aspirin Generated from the Three Templates: r1, s10 and s12

ATOM	1	C1	DRG	1	-1.978	-0.306	0.000	0	CA	1.9080	0.0860	-0.0176
ATOM	2	C2	DRG	1	-1.768	1.092	0.000	0	CA	1.9080	0.0860	0.1468
ATOM	3	C3	DRG	1	-0.480	1.637	0.000	0	CA	1.9080	0.0860	-0.1830
ATOM	4	C4	DRG	1	0.624	0.790	0.000	0	CA	1.9080	0.0860	-0.2443
ATOM	5	C5	DRG	1	0.445	-0.593	0.000	0	CA	1.9080	0.0860	-0.1477
ATOM	6	C6	DRG	1	-0.846	-1.124	0.000	0	CA	1.9080	0.0860	-0.1758
ATOM	7	H7	DRG	1	-0.315	2.724	0.000	0	HA	1.4590	0.0150	0.1775
ATOM	8	H8	DRG	1	1.631	1.224	0.000	0	HA	1.4590	0.0150	0.1425
ATOM	9	H9	DRG	1	1.318	-1.254	0.000	0	HA	1.4590	0.0150	0.1170
ATOM	10	H10	DRG	1	-1.006	-2.217	0.000	0	HA	1.4590	0.0150	0.1449
ATOM	11	C11	DRG	1	-3.361	-0.949	0.000	0	C	1.9080	0.0860	0.7685
ATOM	12	O12	DRG	1	-3.407	-2.215	0.000	0	O	1.6612	0.2100	-0.7852
ATOM	13	O13	DRG	1	-4.382	-0.210	0.000	0	OH	1.7210	0.2104	-0.7905
ATOM	14	O14	DRG	1	-2.949	1.831	0.000	0	OH	1.7210	0.2104	-0.2543
ATOM	15	C15	DRG	1	-3.008	3.188	0.000	0	C	1.9080	0.0860	0.7545
ATOM	16	C16	DRG	1	-4.435	3.644	0.000	0	CT	1.9080	0.1094	-0.3881
ATOM	17	O17	DRG	1	-2.014	3.920	0.000	0	O	1.6612	0.2100	-0.6146
ATOM	18	H18	DRG	1	-4.461	4.757	0.000	0	HC	1.4870	0.0157	0.1164
ATOM	19	H19	DRG	1	-4.946	3.243	-0.908	0	HC	1.4870	0.0157	0.1164
ATOM	20	H20	DRG	1	-4.946	3.243	0.908	0	HC	1.4870	0.0157	0.1164
CONNECT	1			2	6	11						
CONNECT	2			1	3	14						
CONNECT	3			2	4	7						
CONNECT	4			3	5	8						
CONNECT	5			4	6	9						
CONNECT	6			5	1	10						
CONNECT	7			3								
CONNECT	8			4								
CONNECT	9			5								
CONNECT	10			6								
CONNECT	11			12	13	1						
CONNECT	12			11								
CONNECT	13			11								
CONNECT	14			15	2							
CONNECT	15			14	16	17						
CONNECT	16			15	18	19	20					
CONNECT	17			15								
CONNECT	18			16								
CONNECT	19			16								
CONNECT	20			16								
END												



\*The format of the table is as described under Table 1.

**Table 3. Some Commonly Employed Drug Like Filters**

Lipinski's Rule of Five <sup>102</sup>	
Molecular weight	≤ 500
Number of hydrogen bond acceptors	≤ 10
Number of hydrogen bond donors	≤ 5
LogP	≤ 5
Additional Filters	
Molar refractivity <sup>109</sup>	≤ 140
Number of rotatable bonds <sup>110</sup>	≤ 10

established. This necessitates assignment / derivation of partial atomic charges as well as other force field parameters to the candidate molecule. The partial atomic charges may be assigned either *via* some empirical rules of transferability from a library of molecules with known charges or derived from quantum mechanical calculations, semi-empirical or *ab initio*. The commonly employed concept of charge transferability in the case of biomolecules is justifiable in view of the limited templates and bond types at the junctions. Organic molecules / novel candidates pose a challenge in this regard.

Geometry optimization may be handled *via* molecular mechanics (MM) or quantum mechanical (QM) calculations. The computational expense *inter alia*, depends on the number of rotatable bonds. Even with three preferred values (*gauche+*, *gauche-* and *trans*) for each dihedral for a molecule consisting of N rotatable bonds, the search space consists of 3<sup>N</sup> points. For binding free energy estimates, the geometry of the ligand in both unbound and bound forms is necessary. Structure and energetics of small molecules is a well-studied area [139]. QM for small molecules and MM methods for biomolecular target and MM or hybrid approaches for the biomolecule-ligand complex are some popular choices. The accuracies currently attainable with state of the art quantum and molecular mechanics calculations on small molecules in gas phase are on the order of 10 cm<sup>-1</sup> (1 cm<sup>-1</sup> ~ 0.01 kcal/mol). The geometries of the ground state are equally well reproduced by current day theory. The latest version of AMBER force field [140] achieves accuracies on the order of 0.1 Å (RMSD) for geometries in relation to crystal structures. Thus preferred rotamer states and their energetics in gas phase are fairly predictable. Quantum treatments have been extended to account for effects of solvation in some approximate way. Density functional theory is another emerging area for undertaking this task [139]. Thus with the current day QM or MM methods, multiple minimum problem may not be a serious issue for small molecules particularly if they carry aromatic rings / rigid parts [140]. In the context of molecular mechanics calculations, generation of force constants for bond lengths, angles, dihedrals etc. is an area for further research. Attempts have been made to evolve some combination rules [141] from the existing data set of force field parameters [142, 143].

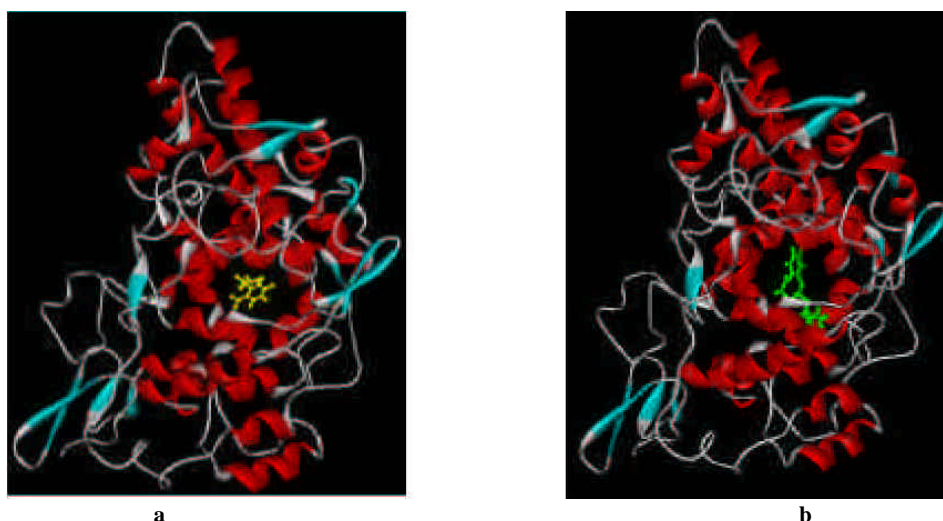
In proceeding along the computational pathway (Fig. 2), semi-empirical (AM1) [144] or molecular mechanical methods for geometry optimization constitute a computationally expedient option. However, derivation of *ab initio* partial atomic charges with 6-31G\* basis set [90] and RESP fitting procedure [92, 145] for the free ligand is necessary for compatibility with the force fields used for protein and nucleic acid targets. For energy optimization and further molecular dynamics simulations, a molecular mechanical treatment of the ligand together with its biomolecular target forms the natural choice. Table 2 lists the Cartesian coordinates (corresponding to AM1 optimized geometry) of aspirin built from chemical templates and partial atomic charges derived from *ab initio* quantum mechanical calculations (as described under Table 1). The complexity as well as the ease of determining the geometry of candidate molecules with current day theory is worth emphasizing.

## E. Docking

It is widely accepted that drug activity is obtained through the binding of one molecule (the ligand) to the pocket of another, usually the larger molecule (the receptor), which in a majority of drug discovery programmes is a protein. In their binding conformations, the molecules exhibit geometric and chemical complementarity, both of which are essential for successful drug activity. The computational process of searching for a ligand that is able to fit in the binding site of a protein both geometrically and energetically is called molecular docking. A docking procedure consists of three inter-related components: identification of the binding site, a search algorithm to position the candidate in the binding site and a scoring function [146]. As can be appreciated from (Fig. 4), the binding pocket of cyclooxygenase looks like an internal cavity (tunnel). The active site of dihydrofolate reductase (not shown), a target for anticancer drugs (eg. methotrexate) [147] looks like a groove on the surface. Experience shows that the exterior sites exposed to solvent in general are slightly more complex to handle in docking experiments.

Most of the docking protocols rely on a predefined binding site so that the search space is limited or confined to a comparatively small region of the target. A scoring function is used to give ranking to the set of final solutions generated by the search [148]. Ideally, the combination of the search algorithm and scoring function should result in a single solution close to the actual ligand position, which is further validated by comparison with the crystal or NMR structure of the protein-ligand complex, if available. The objective of molecular docking is to obtain the structure for the protein-ligand complex corresponding to the lowest interaction or total energy in most practical applications but free energy in a theoretically rigorous sense. Computational strategies for docking to study the formation of stable intermolecular complexes have been the subject of intense research since the days the three dimensional (3D) structures of the targets have become available. The issues encountered in designing docking algorithms have been thoroughly reviewed [149-157]. Finding the low-energy states of ligand-receptor complexes presents a fundamental problem in that the receptor sites have complicated and adjustable shapes





**Fig. (4).** A visual representation of docked and energy minimized structures of (a) COX-2-aspirin complex and (b) COX-2-APHS complex.

and there are many ways of fitting a flexible ligand in them. Various docking programs utilize methods based on the geometric search namely descriptors, grids and fragments and a wide array of protocols for scoring solutions. Taylor *et al.* [158] recently reviewed the merits and limitations of some docking algorithms together with their characterization. Best RMSD obtained with current docking algorithms is less than 1Å with the lowest energy structures. Some of these are listed in (Table 4) [159-172]. The different methods described are not strict divisions, as many docking programs combine aspects from two or more search algorithms. Methods based on artificial intelligence techniques have also been pressed into service. It is worth recalling that algorithms based on Brownian dynamics can mimic diffusion of ligand to the active site obeying laws of physics [173-175]. In view of the computational times involved, these are currently limited to diffusion-controlled reactions and for rigid ligand and active sites. Further exploration of molecular simulations for docking is desirable.

In short, the concerns in the docking are efficiency in sampling the binding pocket on the target, the 6D (three translational and three rotational degrees) search versus incorporating internal (3N-6; N is the number of atoms) degrees of freedom of the candidate and accounting for the flexibility and structural perturbation of the binding site caused by the ligand [177-180]. Other issues include effect of solvation and salt on binding and therefore on docking and a good scoring function which correlates well with theoretical free energies of binding or experimental binding constant data. The LPDB (Ligand-Protein Database) hosted by Brooks and coworkers [181] is oriented towards an improvement of empirical scoring functions by providing structures of a large number of protein-ligand complexes together with experimental binding affinity information. Some of these are addressed in the protocols listed in (Table 4).

We illustrate here the implementation of a Monte Carlo search for docking candidates in the binding sites of targets. The search is confined to 6D space with an empirical potential energy function, which considers electrostatic and

van der Waals interactions at the atomic level and includes solvent implicitly in the electrostatics and in a separate hydrophobic term, to initiate docking [182]. After the preliminary search locates the candidate in the binding pocket with interaction energy as a criterion, several orientations of the candidate around the location identified in the previous step are generated and each energy optimized to arrive at the docked structure with the minimum most energy. Structures of cyclooxygenase-2 (COX-2) complexed with aspirin and APHS corresponding to minimum most energy are shown in (Fig. 4a and b). The structure of the complex thus generated is used for a qualitative estimate of the binding free energy using the methodology (described in section II F below). Nanosecond long, explicit solvent molecular dynamics simulations were conducted on the complex, target and the ligand. The binding free energies were then computed as described in section II G below. This protocol, although computationally intensive, alleviates most of the concerns in docking as the case studies (section III below) reported indicate.

#### **F. Assessing the Suitability of Small Molecules as Lead-Like for Biomolecular Targets: Theory of Binding Affinities**

To judge whether a small molecule is worth pursuing as a candidate for lead generation, one could use energy based methods or develop a set of rules using artificial intelligence techniques and databases or derive effective scoring functions calibrated on a series of known ligands. The former has the advantage of theoretical rigor and the latter two prove powerful in terms of computational expedience and in dealing with millions of molecules. Even in the energy based methods, one can opt for the speedier route of interaction energy based methods. This involves considering only the final state instead of the differential energetics between the final and initial states and further neglecting entropies and Boltzmann averaging. The candidate molecules are ranked based on the interaction energy between the target and the ligand. Although popular this approach involving scoring functions for quantifying interaction energies, has to be tested on a large number of



Table 4. Some Popular Docking Programs

Program Name	Author(s)	Ref.	Algorithm/Descriptor	Scoring Function
AUTODOCK	Goodsell, Olson	[159]	Kinetic / Grid, Monte Carlo, Genetic Algorithm based	Force field
CAVEAT	Bartlett	[160]	Descriptor based /bond vector matching	Geometry
CLIX	Lawrence, Davis	[161]	Descriptor based / Steric effects matching	Force field
COMBIBUILD	Kick, Roe <i>et al.</i>	[162]	Fragment based	Force Field
DOCK	Kuntz <i>et al.</i>	[163]	Distance compatibility graph, Genetic algorithm based	Force field
DIVALI	Clark <i>et al.</i>	[164]	Genetic Algorithm based	Force field
FLEXX	Rarey, Kramer <i>et al.</i>	[165]	Fragment Incrementation / pose clustering and triplet matching	Force field
FLOG	Miller <i>et al.</i>	[166]	Distance Geometry based	Force field
FTDOCK	Gabb <i>et al.</i>	[167]	Fourier Transformation	Empirical
GOLD	Jones <i>et al.</i>	[168]	Genetic Algorithm based	Empirical
GROW	Moon <i>et al.</i>	169	Fragment based	Force field
GLIDE	Friesner <i>et al.</i>	[170, 171]	Geometry based /Site point search / grid based optimization	Force field
HOOK	Miranker, Karplus	[172]	Fragment based	Empirical
Ligin	Sobolev <i>et al.</i>	[173]	Geometry based / surface complementarity	Force field
Ludi	Bohm	[95]	Fragment based / Interaction site matching	Force field

receptor-ligand complexes and used with caution as a general strategy. The alternative is to choose binding free energy based methods [183-193], which are rigorous and reliable, but extremely demanding computationally. Developing binding free energies of a ligand for a biomolecular target could again be based on a single configuration of the complex (obtained from X-ray, NMR or molecular modeling). This method neglects the thermodynamic (enthalpic and entropic) contributions associated with the thermal effects, flexibility of the target site and the ligand in particular. This is useful for qualitative estimates of binding free energies and ranking candidates with some training. The alternative is to develop molecular dynamics trajectories (or an ensemble of structures consistent with Boltzmann distribution as for instance with Metropolis Monte Carlo simulations) with an explicit atomic-level description of the solutes and solvent medium for the unbound and bound states and estimate free energies.

The statistical mechanical theory [194, 195] of binding affinities in aqueous media is presented below.

$$[P]_{aq} + [L]_{aq} = [P^*L^*]_{aq} \quad (1)$$

P and L are the reactants and P\*L\* is the product of binding in aqueous medium. The superscript "\*" denotes structural changes accompanying binding. At equilibrium

$$\mu_{P,aq} + \mu_{L,aq} = \mu_{P^*L^*,aq} \quad (2)$$

$\mu_{P,aq}$  is the chemical potential of species P in the solvent medium (partial molar Gibbs free energy) and  $\mu_{P,aq}^o$  is its standard chemical potential i.e. under conditions of 1 bar in gaseous state and 1 molar (designated as  $C^o$ ) in liquid state.

$$\begin{aligned} & \mu_{P,aq} + RT \ln ( \gamma_P C_P / C^o ) + \mu_{L,aq} + RT \ln ( \gamma_L C_L / C^o ) \\ & = \mu_{P^*L^*,aq} + RT \ln ( \gamma_{P^*L^*} C_{P^*L^*} / C^o ) \end{aligned} \quad (3)$$

where  $\gamma_P$  is the activity coefficient of species P and  $C_P$  its concentration. The standard molar Gibbs free energy of the reaction (standard absolute molar Gibbs free energy of binding) is

$$\begin{aligned} G_{aq} & = \mu_{P^*L^*,aq} - (\mu_{P,aq} + \mu_{L,aq}) \\ & = - RT \ln [ \gamma_{P^*L^*} C_{P^*L^*} C^o / (\gamma_P C_P)(\gamma_L C_L) ] = - RT \ln K_{eq,aq} \end{aligned} \quad (4)$$

In terms of canonical partition functions (Q)

$$G_{aq} = A_{aq} + P V_{aq}^o = -RT \ln K_{eq,aq} = -RT \ln \left\{ \frac{Q_{P^*L^*,aq}}{(N_A Q_w)} / \left\{ \frac{Q_{P,aq}}{(N_A Q_w)} \left( \frac{Q_{L,aq}}{(N_A Q_w)} \right) \right\} \right\} + P V_{aq}^o \quad (5)$$

$$G_{aq} = -RT \ln \left\{ \frac{Q_{P^*L^*,aq}^{rot} Q_{P^*L^*,aq}^{int} Q_{P^*L^*,aq}^{el} N_A Q_w}{\left\{ \left( Q_{P,aq}^{rot} Q_{P,aq}^{int} Q_{P,aq}^{el} \right) \left( Q_{L,aq}^{rot} Q_{L,aq}^{int} Q_{L,aq}^{el} \right) \right\}} \right\} + P V_{aq}^o \quad (6)$$

is the standard Helmholtz free energy of the reaction.  $N_A$ , the Avogadro number in the above equation originates in expressing partition functions  $Q$  as molar partition functions and  $P V_{aq}^o$  is the pressure-volume correction to Helmholtz free energy in the solvent medium.  $Q_w$  denotes the partition function for pure solvent (water).  $Z^{int}$  is the configurational partition function. It includes contributions from vibrations and internal motions as well as solvation (hydration) effects. The translational and rotational terms have been separated out.

$$Z_{P,aq}^{int} = \dots \exp \{ (-E(\mathbf{X}_P^N, \mathbf{X}_W^M) / k_B T) \}$$

$$d\mathbf{X}_P^N d\mathbf{X}_W^M = \langle \exp (E(\mathbf{X}_P^N, \mathbf{X}_W^M) / k_B T) \rangle \quad (7)$$

$\mathbf{X}_P^N$  and  $\mathbf{X}_W^M$  represent the configurational space accessible to the solute  $P$  and solvent  $W$  respectively, in the presence of each other.  $E(\mathbf{X}_P^N, \mathbf{X}_W^M)$  denotes the total potential energy of the system describing non-idealities. It includes intramolecular interactions within the solute  $P$  and solvent  $W$  as well as intermolecular interactions between the solute and the solvent.  $k_B T$  is the product of Boltzmann constant and temperature (in Kelvin).

$$Q_P^{el} \sim 1 \text{ (for non-covalent associations)} \quad (8)$$

The standard free energy change accompanying binding may be written as a sum of external (translational and rotational) and internal (intramolecular, intermolecular and solvation) contributions.

$$G = RT \ln \left[ \frac{Q_{P^*L^*}^{tr} N_A / (Q_P^{tr} Q_L^{tr})}{(Q_P^{rot} Q_L^{rot})} \right] - RT \ln \left[ \frac{Q_{P^*L^*}^{rot}}{(Q_P^{rot} Q_L^{rot})} \right]$$

$$- RT \ln \left[ \frac{(Z_{P^*L^*,aq}^{int} Q_w)}{(Z_{P,aq}^{int} Z_{L,aq}^{int})} \right] + P V_{aq}^o \quad (9)$$

Eq. (9) is an exact expression for evaluating binding free energies for non-covalent associations in aqueous medium. The first two terms on the right hand side of eq. (9) can be computed analytically. The third term is accessible to free energy molecular simulations configured in the canonical ensemble such as the perturbation method, thermodynamic integration, potential of mean force method etc. [196], albeit they are computationally expensive even for a single ligand and not practical in a high through-put sense even on supercomputers.

In the following, we consider some simplifications to bring the binding free energy computations into feasibility domain. The molecular translational partition function of  $P$  is

$$q_P^{tr} = V / \left( \frac{h^2}{2 m_P k_B T} \right)^{3/2} \quad (10)$$

The molar partition function of  $P$  is

$$Q_P^{tr} = (q_P^{tr})^{N_A} \quad (11)$$

Note that the volume  $V$ , has been included in the translational part consistent with ideal gas statistical mechanics. This would require that the  $Z^{int}$  be divided by  $V$  to quantify non-idealities (excess free energies). The translational part of the free energy in eq (9) is now given by the Sackur-Tetrode [196] equivalent as

$$G_{tr} = -RT \ln \left[ \frac{(N_A/V) \left( \frac{3}{2} \frac{3}{2} \frac{3}{2} \frac{3}{2} \right)}{(h^2/2 k_B T)^{3/2} \{ m_{P^*L^*} / (m_P m_L) \}^{3/2}} \right] = -RT \ln \left[ \frac{(N_A/V)}{(h^2/2 k_B T)^{3/2} \{ m_{P^*L^*} / (m_P m_L) \}^{3/2}} \right] \quad (12)$$

The expression in the square brackets in eq (12) is dimensionless.  $(N_A/V)$  may be replaced by a concentration term ensuring that upon transfer to aqueous medium standard free energies are recovered with the reference state anchored to a molar concentration of unity. Note that this expression is the same whether in gas phase or liquid phase provided the translational and rotational motions of the solute are unaffected by the solvent. This will be true only in a continuum, friction-less solvent influencing the position dependent potential energy but not the velocity dependent kinetic energy of the solute. Hence in a transfer process (an experiment involving transfer of species  $P$  from one phase to another phase such as from gas phase to liquid phase or octanol to water etc.), this term cancels out. In binding processes however, no such cancellation occurs. Also if  $P$ ,  $L$  and  $P^*L^*$  could be seen as a collection of non-bonded mono-atomic particles, then again the translational partition function for each species could be written as a product of the individual partition functions of the constituent atoms and since the number of atoms is conserved during binding, these terms would cancel out. Again, this is not so for polyatomic species where the mass in translational partition function  $m_P (= \sum_i m_i)$  is evaluated as a sum of the masses of the constituent atoms. It is recommended that Sackur-Tetrode equation be applied not in aqueous medium directly where it is invalid but upon transfer to vacuum via a suitable thermodynamic cycle.

Similar arguments apply to the rotational partition functions. Separating the rotational part from internal motions implies working under rigid rotor approximation.

$$G_{rot}^o = -RT \ln \left[ \left( \frac{I_P}{I_{P^*L^*}} \right) \left( \frac{1}{8} \right) \left( \frac{h^2}{2 k_B T} \right)^{3/2} \times \left\{ \frac{(I_P^a I_P^b I_P^c)}{(I_{P^*L^*}^a I_{P^*L^*}^b I_{P^*L^*}^c)} \left( \frac{I_P^a I_P^b I_P^c}{I_L^a I_L^b I_L^c} \right)^{1/2} \right\} \right] \quad (13)$$

$I_P^a$ ,  $I_P^b$  and  $I_P^c$  are the components of moments of inertia of species  $P$  along the principal axes and  $\rho_P$  its symmetry number. Murray and Verdonk [197] brought out the importance of rotational and translational entropies lost by small molecules on binding to proteins.

$$G = G_{tr}^o + G_{rot}^o - RT \ln \left[ \frac{(Z_{P^*L^*,aq}^{int} Q_w)}{(Z_{P,aq}^{int} Z_{L,aq}^{int})} \right] + P V_{aq}^o \quad (14)$$

Free energy contributions from internal motions which are coupled to solvent are best handled via molecular simulations. Separating the two will amount to an approximation.

$$Z_{P,aq}^{int} = Z_P^{vib.conf} Z_P^{solvn.}$$

$$Z_{P,aq}^{int} = \dots \exp \{ (-E(\mathbf{X}_P^N, \mathbf{X}_W^M) / k_B T) \} d\mathbf{X}_P^N d\mathbf{X}_W^M$$

$$\left[ \dots \exp \{ (-E(\mathbf{X}_P^N) / k_B T) \} d\mathbf{X}_P^N \right] \times \left[ \dots \exp \{ (-E(\mathbf{X}_W^N) / k_B T) \} d\mathbf{X}_W^N \right] \quad (15)$$

Equations similar to (15) can be written for  $L$  and  $P^*L^*$  and converted to excess free energies. Such a separation allows

$$G = G_{tr}^o + G_{rot}^o + G_{int}^o + G_{solvn.}^o \quad (16)$$

Eq. (16) forms the theoretical basis for the additivity assumed in free energy computations as employed in master equation methods [198, 199]. The  $P V_{aq}^0$  term in eq. (9) is often neglected in liquid-state work. If eqs. (15) and (16) are employed for each structure generated according to Boltzmann distribution either *via* molecular dynamics or Metropolis Monte Carlo and averages computed with a suitably calibrated dielectric continuum solvent model for solvation energy for each structure, the results are expected to correspond to eq. (9) which is exact.

Recent advances in free energy methodology offer two attractive methods *viz.* the MMPBSA [192, 201, 202] and the MMGBSA [203-205], which utilize the structural information emanating from molecular dynamics simulations to develop estimates of binding free energies using eqs. (15) and (16) above, in a *post facto* analysis of the trajectories on each structure followed by energy component averaging. The essence is to generate structures with explicit solvent and transfer these to continuum solvent for energy evaluations thus rendering the free energy problem computationally tractable. A practical implementation of the above free energy methodology involves computation of average intramolecular energy (internal energy / enthalpy), corresponding entropies, solvation free energies of the solute along the MD trajectories of the free and bound protein and ligand.

$$G_{int}^0 = H_{int}^0 - T S_{int}^0 \quad (17)$$

$$H_{int} = H_{intermolecular} + H_{intramolecular}$$

$$H_{intermolecular} = E_{el} + E_{vdW} \\ < E_{intermolecular} > = < E_{el} > + < E_{vdW} > \quad (18)$$

$$H_{intramolecular} = < E_{intramolecular} > \quad (19)$$

In the above equations,  $E_{el}$  and  $E_{vdW}$  represent the electrostatic and van der Waals components of the intermolecular interaction energy between the protein and the inhibitor.  $E_{intramolecular}$  represents changes in both bonded and non-bonded contributions to the intramolecular energy of the protein and the inhibitor upon binding. All these quantities can be computed from a force field either for a fixed structure (from minimization studies) or for an ensemble of structures from MD simulations.

$$S_{int} = S_{vib,config} \quad (20)$$

$S_{vib,config}$  can be calculated by normal mode analysis for energy minimized structures ( $S_{vib}$ ) or by quasi-harmonic approximation introduced by Karplus and Kushick [206] and subsequently extended and adapted to MD simulations by Schlitter [207] and van Gunsteren [208].

To account for structural deformation upon binding, we include adaptation expense which accounts for changes in the intramolecular energetics explicitly in  $G_{int}$ . This is calculated as the difference in the free energies of the bound and unbound states of the protein and the inhibitor (steps I and II in the thermodynamic cycle) in the presence of the solvent.

In the MMGBSA or MMPBSA models, the solvation free energies are computed as

$$G_{solv} = G_{GBSA} = G_{GB} + G_{SA} \quad (21)$$

where  $G_{GB}$  refers to the electrostatic component of solvation while  $G_{SA}$  is the non-electrostatic contribution, called cavitation energy in literature [209]. The defining equation employed for evaluating the electrostatic contribution to the solvation free energy [203] with the MMGBSA model is:

$$G_{el,solv}^0 = -166 \left( 1 - \frac{1}{\epsilon} \right) \sum_{i=1}^n \sum_{j=1}^n q_i q_j / f_{m2GB} \quad (22)$$

$$G_{GB} = G_{el,solv}^0(\text{final state}) - G_{el,solv}^0(\text{initial state}) \quad (23)$$

Similar equations are formulated to deal with added salt effects at the Debye-Huckel level [210]. Small ions associated with the biomolecular target and the ligand to maintain electroneutrality are dealt with explicitly in simulations and processed as part of the solute. The non-electrostatic (nel) contributions to the solvation free energy [211] are computed as a function of the solvent accessible (SA) surface area [212]

$$G_{nel,solv}^0 = \gamma_{nel} A \quad (24)$$

$$G_{SA} = G_{nel,solv}^0(\text{final state}) - G_{nel,solv}^0(\text{initial state}) \quad (25)$$

The quantity  $\gamma_{nel}$  has been assigned a value of 7.2 cal/mol/Å<sup>2</sup> [175]. This may be considered [210] as a resultant of +47 cal/mol/Å<sup>2</sup> from the cavity term [213] and -39.8 cal/mol/Å<sup>2</sup> from van der Waals interactions of the solute and the solvent [214]. This separation is only for the purpose of interpretation and does not alter the free energy estimates. Thus, a combination of equations (9), (12), (13), (14), (17) and (21) yields the absolute binding free energies. The governing equation (16) for estimation of free energy change upon binding is

$$G = G_{tr}^0 + G_{rot}^0 + G_{int}^0 + G_{solv}^0$$

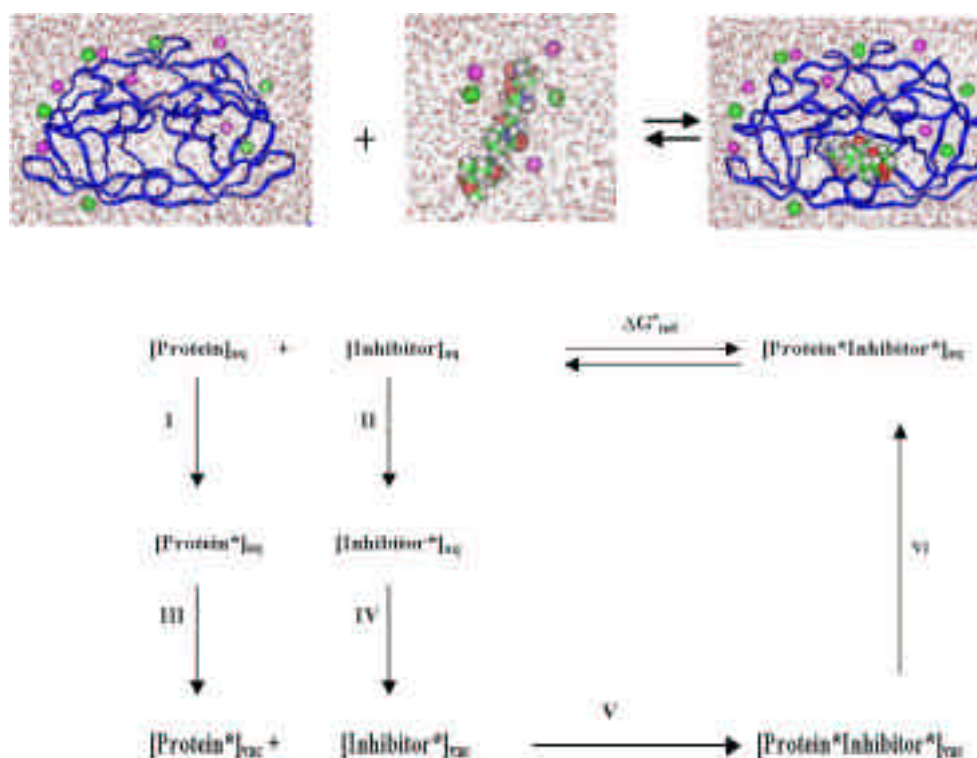
The internal and solvation energy components in eq. 16 can be further described as,

$$G_{int}^0 + G_{solv}^0 = G_{adapt}^0 + G_{vc}^0 + G_{el}^0 + G_{vdW}^0 + G_{solv,el}^0 + G_{solv,nel}^0 \quad (26)$$

with  $G_{adapt}^0$  and vibrational configurational entropy changes identified as  $G_{intramolecular}^0$  and  $G_{solv,nel}^0$  with  $G_{cav}^0$  and  $G_{net,el}^0$  as a sum of  $G_{el}^0$  and  $G_{solv,el}^0$ . These components can be computed via the MMGBSA or MMPBSA methodology. The thermodynamic cycle employed to construct the standard free energies of protein-inhibitor binding in solution is illustrated in (Fig. 5). Building on eqs. (16) and (26), the net binding process is decomposed into six steps and the corresponding binding free energy is calculated as a sum of five components:

$$G_{net}^0 = G_{vdw}^0 + G_{net,el}^0 + G_{cav}^0 + G_{adpt}^0 + G_{trvc}^0 \quad (27)$$

In a phenomenological view, eq. (16) may be rearranged (eq. 27) and the net binding free energy may be considered to be a sum of the free energy changes due to the following terms: (i) van der Waals interactions between the protein and the inhibitor indicating the influence of shape complementarities and packing effects; (ii) net electrostatics which includes interactions between partial or full charges, hydrogen bonds and electrostatics of desolvation upon binding and added salt effects, (iii) cavitation effects, which account for change in size and shape of solvent cavity on binding giving rise to water reorganization, a component of which, originating from nonpolar sources, is the hydrophobic



**Fig. (5).** A thermodynamic cycle for the computation of absolute binding free energies.

effect. Here the nonelectrostatics of desolvation of both polar and nonpolar atoms is accounted for in the cavitation term; (iv) the deformation expense (i.e. the intramolecular contributions due to structural variations upon complexation); (v) translational, rotational and vibrational, configurational entropy losses.

#### **Qualitative Estimates of Binding Free Energies and Segregation of Potentially Good Candidates**

The theory described above enables energy analysis of docked and energy optimized complexes *via* a thermodynamic cycle (Fig. 5). The protein and the inhibitor are separated from the complex, taken through the cycle and the free energy components are computed using eq. (27) except  $G_{\text{adpt}}$  which is amenable to estimation if structures of the native (unbound) protein and the ligand are available. These qualitative estimates provide a quick means to check whether or not to pursue with a candidate molecule. Noskov and Lim [215] following a small variant of the cycle calculated the free energy of the binding of antibody to lysozyme which was in agreement with experimental data.

#### **Semi-Quantitative Estimates of Binding Free Energies**

Reliable estimates of binding free energies are obtained by performing molecular dynamics simulations on the protein, ligand and the complex and taking each of them through the thermodynamic cycle (Fig. 5) to compute free energy components given by eq. (27). This computationally intensive step can be reserved for potentially good candidates.

In a nutshell, computation of absolute binding free energies from atomic level descriptions of the systems is a formidable task. The theory and methodology described above is an attempt to link structure with thermodynamics and to elicit the binding free energies in a computationally expeditious manner without compromising the rigors of statistical mechanics. The computational protocols described above for binding free energy estimates have been tested and validated in a series of studies on protein-drug [216-219], DNA-drug [220], and protein-DNA complexes [221, 222]. Most notably, the results obtained in *post facto* analyses of MD trajectories are found to be converged, stable and satisfactory (i.e. within a statistical uncertainty of  $\sim 2$  kcal/mol in most cases) in relation to experimental binding constant data. The theory presented also points to further systematic improvements to attain better accuracies.

#### **G. MD Simulations and Post Facto Binding Free Energy Analyses**

To account for flexibility of the candidate and the target and to deal with solvent and salt effects in binding in a rigorous way, all atom molecular mechanical simulations form the current choice. Molecular dynamics simulations can be configured on the bound (protein-candidate molecule docked complex) and the unbound species (free protein and free ligand) with explicit solvent and small ions under ambient temperature and pressure conditions [223-228]. To obtain converged net binding free energies with MMGBSA and MMPBSA methods, at least 100 structures or more of the free protein, the drug and the complex from molecular dynamics simulations (typically of length 2 nanoseconds or

longer) with explicit solvent are required. In addition, a force field compatible continuum solvent model is necessary for estimating the electrostatic component of the solvation free energy of each structure. Besides this, MD trajectory analysis programs for calculating intermolecular as well as intramolecular interaction energies and programs for calculating accessible areas, translational, rotational and vibrational / configurational entropies are required. In relation to the proposed pathway, this is one of the most computationally intensive steps and practical only for a select few promising candidates, which could be identified in the previous step *via* qualitative estimates of free energies on the docked complexes with a multitude of ligands.

## H. Mutational Studies for Better Leads

The overall protocol proposed envisages an assessment and selection of potentially good candidates based on qualitative estimates of binding affinities on docked and energy optimized complexes of candidates with the target in the first step. MD simulations are envisioned for improved binding affinity estimates as a subsequent step. Chemical modifications on the candidates are attempted with a view to further enhancing the affinities as the third step. This can be followed by a rapid computational assay of the binding affinity as mentioned in section II F. MD simulations with explicit solvent and *post facto* binding free energy analyses can be conducted in promising cases to ensure that the mutations introduced indeed result in better binders. We considered aspirin as the original lead and APHS as the mutated lead with better binding attributes [229]. Results of these studies are presented and discussed below.

## III. Case Studies

### COX-2

Preliminary studies conducted on several small molecules comprising some known NSAIDs and non-NSAIDs for COX-2 target demonstrated that the protocol could segregate drugs from non-drugs [230]. Jorgensen and coworkers investigated the binding of celecoxib and analogs to COX-2 via free energy perturbation method [231] as well as extended linear response formulation of the free energy based on Monte Carlo simulations [232]. The calculated free energy differences agree with experiment to within a kcal. Interaction energies of amino acid residues in the active site of cyclooxygenases interacting with indoprofen and NS-398 have also been quantified based on molecular dynamics trajectories [233].

Encouraged by the success of the free energy based computational methods, we undertook a mapping of the complete comprehensive computational pathway (christened '*Sanjeevini*') from steps A to H (Fig. 2) for aspirin and APHS. We generated these molecules from templates as shown in (Table 2 and Fig. 3), passed them through empirical filters to assess their drug-likeness. We optimized their geometries, derived charges and assigned force field parameters. We docked candidates in the active site of the target, performed molecular mechanics calculations on each candidate in the active site in several orientations and selected the minimum most energy complex (Fig. 4) for further molecular dynamics processing. All atom explicit

solvent MD simulations were then performed on COX-2, aspirin, APHS, complex of COX-2 with aspirin and complex of COX-2 with APHS and binding free energies were computed using eq. (27) as discussed in section II (F).

Convergence plots for the free energies of binding of COX-2 with aspirin and APHS are shown in (Fig. 6). Beyond one nanosecond, the computed binding free energies seem stable.

Results on the binding free energies as well the diverse phenomenological energy contributors are shown in (Fig. 7). Also shown in the figure are binding free energy results from energy minimized structures (single points in configuration space) as well as experimental values. Binding free energies computed in a *post facto* analysis of the MD trajectories are in close correspondence with experiment. Additionally the observed preferential binding of APHS over aspirin to COX-2 is well reproduced by the computational pathway.

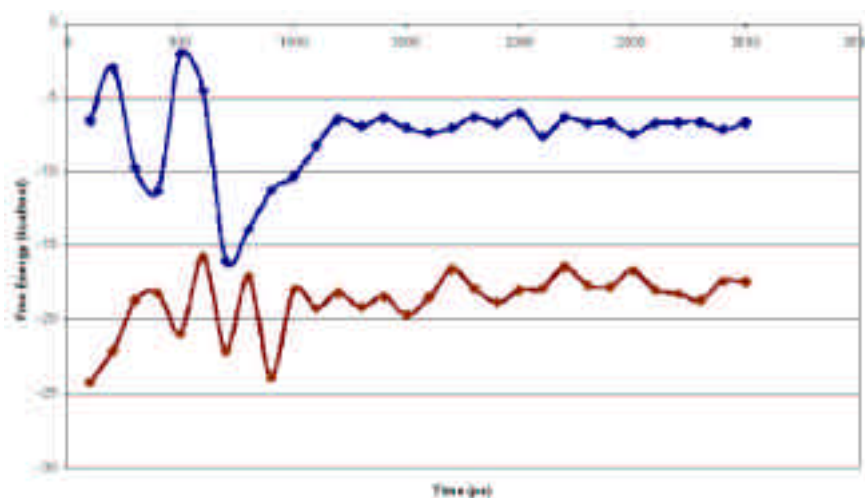
As expected, the free energies from minimized structures tend to exaggerate the interaction strength (van der Waals to a small extent and electrostatic component to a large extent). Accounting for solvent, thermal effects brings these in line with experiment. In our experience, the MMGBSA and MMPBSA if parameterized and implemented on MD structures seem to work quite well. From a methodological point of view, Boltzmann averaging and a consideration of entropy seem essential for semi-quantitative estimates.

On the lead design front, drugs and non-drugs for COX-2 separate out in most cases [230] by virtue of the contribution of electrostatics to binding free energy and this is particularly the case for aspirin, a small molecule for the available cavity size in COX-2. APHS suggests the alternative theme in that it carries no formal negative charge yet binds better due mostly to van der Waals as expected from its extended hydrophobic chain. Binding free energy computations and energy component analyses thus provide an extremely useful grip on the lead design problem. Although the computational protocol presented is a plausible route to lead design targeted to binding sites on proteins, the methodology is equally applicable to nucleic acid targets. Quite reassuringly most of the steps in the pathway (Fig. 2) starting from candidate generation leading to binding affinities are amenable to automation.

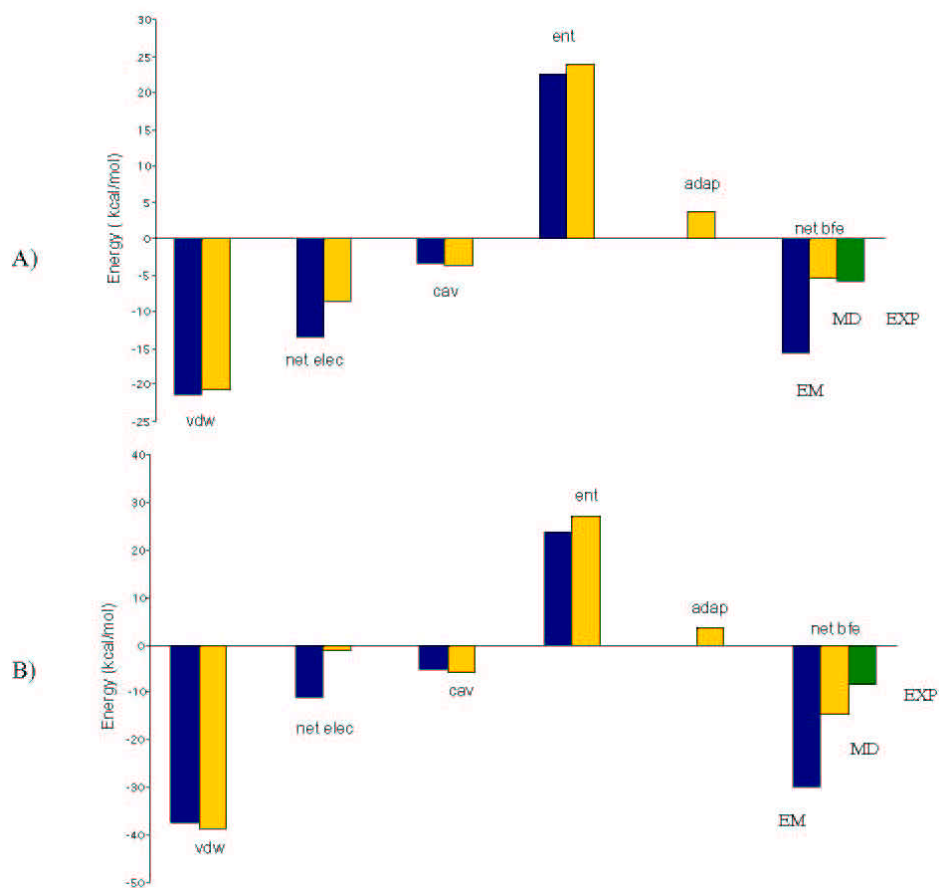
### HIV-1 Protease

Illustrative calculations i.e. MD simulations followed by free energy analysis, were performed on the binding of 4hvp and 8hvp (Fig. 8) with HIV-1 protease target [218]. Both the inhibitors are known experimentally to bind to the target site with nanomolar dissociation constants with 8hvp reported as a tight binder. The binding free energy calculations not only capture the energetics semi-quantitatively but also reproduce the known relative affinities. The HIV-1 computations further illustrate that concerns connected with the size of the inhibitors and the differential ionization states of the active residues are under control.

Contribution of structural water within the active site of HIV-1 to binding affinity has been analysed recently by Fornabaio *et al.* [234]. A method is proposed by Zoete *et al.* to calculate the binding energetics between different ligands



**Fig. (6).** Convergence plots of the calculated standard absolute binding free energies versus molecular dynamics run length for (A) COX-2 with aspirin (blue) (B) COX-2 with APHS (red).



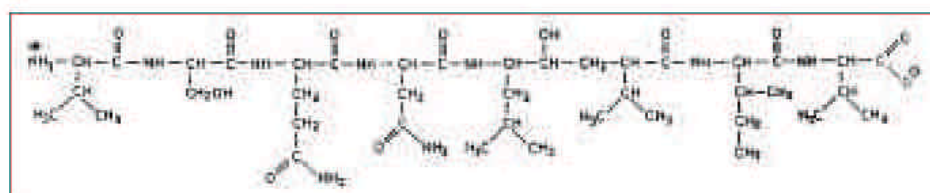
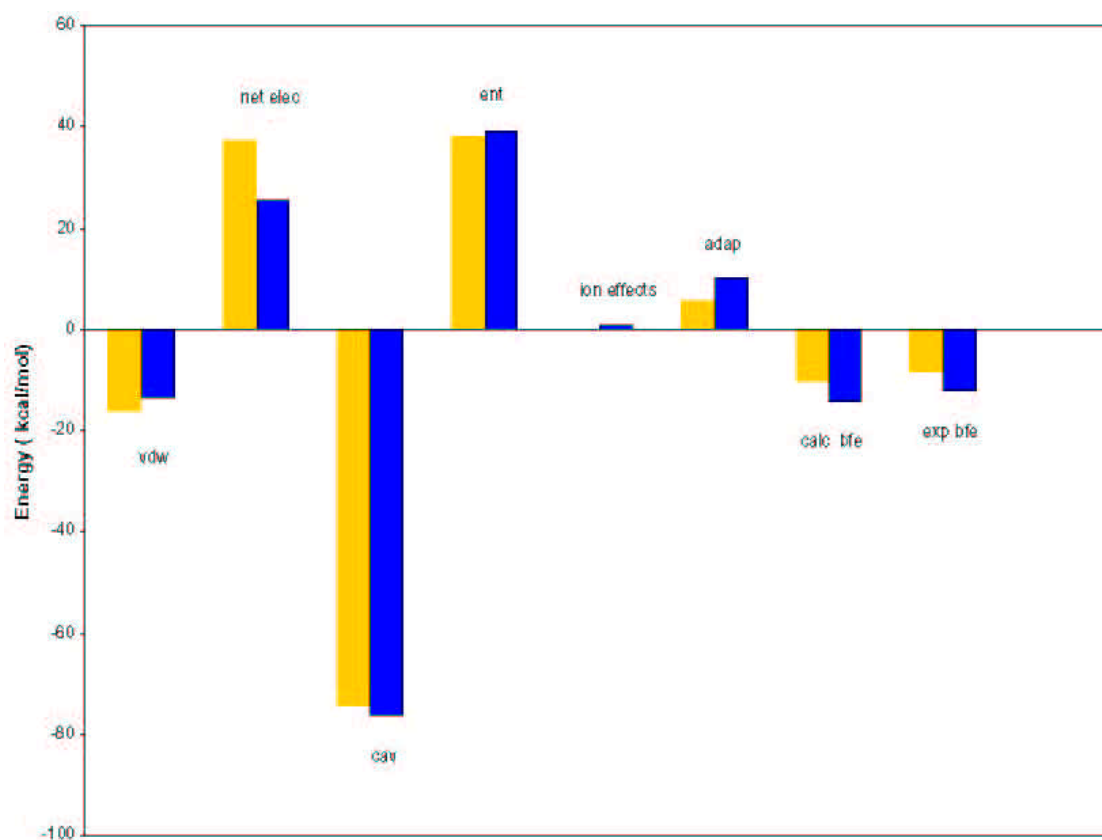
**Fig. (7).** A free energy component analysis of binding of COX-2 to A) Aspirin and B) APHS. (vdw: van der Waals, net elec: net electrostatics (Coulomb interactions and electrostatic component of desolvation, small ion effects), cav: cavitation, ent: entropy, adap: adaptation, net bfe: net standard binding free energy, EM – energy minimization, MD - molecular dynamics, EXP - experimental binding free energy).

and HIV-1 protease that provides a satisfactory correlation between the calculated binding affinities with the experimental  $IC_{50}$  values [235].

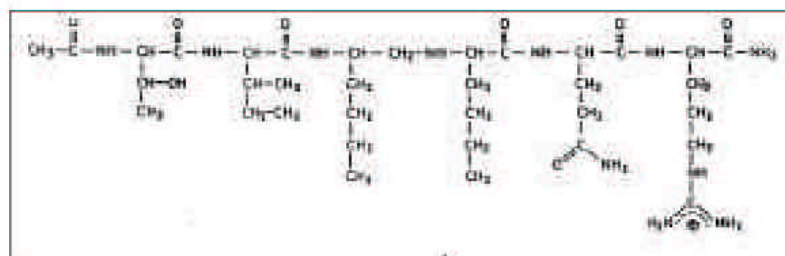
### FKBP12

McCammon and coworkers [202] investigated the binding of 4-hydroxy-2 butanone (BUT) to FKBP12.





4hvp



8hvp

**Fig. (8).** A free energy component analysis of the binding of HIV-1 protease with 4-hvp and 8-hvp. (vdw: van der Waals, net elec: net electrostatics (Coulomb interactions and electrostatic component of desolvation), cav: cavitation, ent: entropy, adap: adaptation, calc. bfe: calculated standard binding free energy).

Calculated binding free energies from *post facto* analyses of MD trajectories were within 10 kJ of the experimental value.

#### Aspartyl-tRNA Synthetase

Karplus and coworkers [188] analyzed the preferential binding of the negatively charged substrate Asp over the neutral Asn to aspartyl-tRNA synthetase *via* free energy

simulation studies and *post facto* analyses of the MD trajectories. The results are in accord with experiment.

#### Cyclin-Dependent Kinase

Sims *et al.* [236] applied a free energy methodology to study the interactions between cyclin-dependent kinase-2 (CDK2) and analog of the clinically tested anticancer agent flavopiridol. The computed free energies showed that

flavopiridol derivatives were significantly more potent than the lead compound flavopiridol.

### **Galectins**

Galectin binds to beta-galactosides with very high specificity. The calculated binding free energies agree [237] quite well with the experimental data. Also the ranking of binding affinities is well reproduced.

### **RNA Aptamer**

Free energy computations and the binding of theophylline to an RNA aptamer by Gouda *et al.* [238] using MMPBSA approach yielded a reasonable absolute binding free energy.

### **Class-I MHC Proteins**

A simple and fast free energy scoring function has been developed to predict the binding free energy of peptides to class I major histocompatibility (MHC) proteins. The average error reported with some test peptides was 3.1 kJ/mol. [239]. Another methodology developed by Froloff *et al.* [240] to calculate the binding free energy of a class-I MHC protein-peptide showed promising result when compared with experimental binding affinities.

## ***IV. Areas for Further Research/Improvements***

The case studies reported from diverse laboratories are clear indicators of the maturity of the binding affinity based *in silico* drug design methodology [188] and pave the way for conceiving a computational pipeline for drug design. Despite the rigor introduced in the comprehensive computational pathway (Fig. 2), further attention is called for in dealing with two issues immediately. One is the computational times involved and second is the statistical uncertainties in the binding affinity estimates.

It is estimated that most extant lead design programs can handle one molecule per processor per minute [150], which remains an achievable, but distant goal for the procedure outlined here (Fig. 2). The main computationally demanding areas in computer aided drug design protocols in general are, (i) generation of billions of compounds with the attendant time and storage issues – active site information must be fed at the design stage to truncate the sample space; (ii) quantum mechanical derivation of partial charges and parameter assignment – some transferability rules have to be evolved for organic molecules; (iii) docking including conformational flexibility and (iv) molecular dynamics with explicit solvent. The pathway examined in itself is robust and appears to hold well for non-covalent associations. Any bond formation will necessitate quantum treatment increasing in turn the computational demands.

Secondly, the statistical noise associated with simulation work, particularly in the development of absolute rather than relative binding free energies ties the estimates to semi-quantitative levels at present. Correlation with IC<sub>50</sub> values needs some further parametric and procedural improvements. Other considerations include location of active sites among multiple alternatives spaced closely on energy scale, adequacy of Monte Carlo search and molecular dynamics in probing fully the conformational space of the candidate and

the target [150]. The protocol reviewed refers to design of inhibitors, which may be lead-like. Whether *in silico* or *in vitro* studies can really mimic conditions *in vivo* remains another question.

## ***V. Promises and Perspectives***

The key driving forces for current day *in silico* drug design endeavors are the availability of structural information of the targets due mainly to advances in proteomics, emergence of reliable energy functions in the form of second generation force fields and force field compatible solvation treatments, as well as free energy methodologies and accessibility of high-end computing clusters. The overview presented here demonstrates a culmination of these developments as an integrated computational pathway all the way from chemical templates to lead discovery - a long haul of unprecedented magnitude. The pathway discussed in this review for *in silico* lead selection / design combines in a natural way basic concepts in chemical bonding (generation of candidate molecules from templates), quantum mechanics (geometry optimization and charge derivation), classical mechanics (molecular mechanics and dynamics), statistical mechanics (configurational / Boltzmann averaging) and thermodynamics (standard free energies of complex formation). The protocol out-lined can be fine-tuned at each stage to improve accuracies. The successes seen are indicators of the current state of computational chemistry and molecular theoretical biophysics in service of biology.

The worldwide efforts on genomics and proteomics have given a significant boost to both experimental and computational methods to march towards personalized medicine with minimal side effects. Automated lead design *in silico* seems a realizable dream in the near future.

## **ACKNOWLEDGEMENTS**

Funding from the Department of Biotechnology, Govt. of India, is gratefully acknowledged. Ms. N. Latha acknowledges the support from Sri Venkateswara College, Univ. of Delhi. The authors wish to thank Mr. Tarun Jain and Mr. Pankaj Sharma for the help received in preparing the manuscript. Thanks are also due to Dr. N. D. Kurur and Dr. N. Pant for their critical reading of the manuscript.

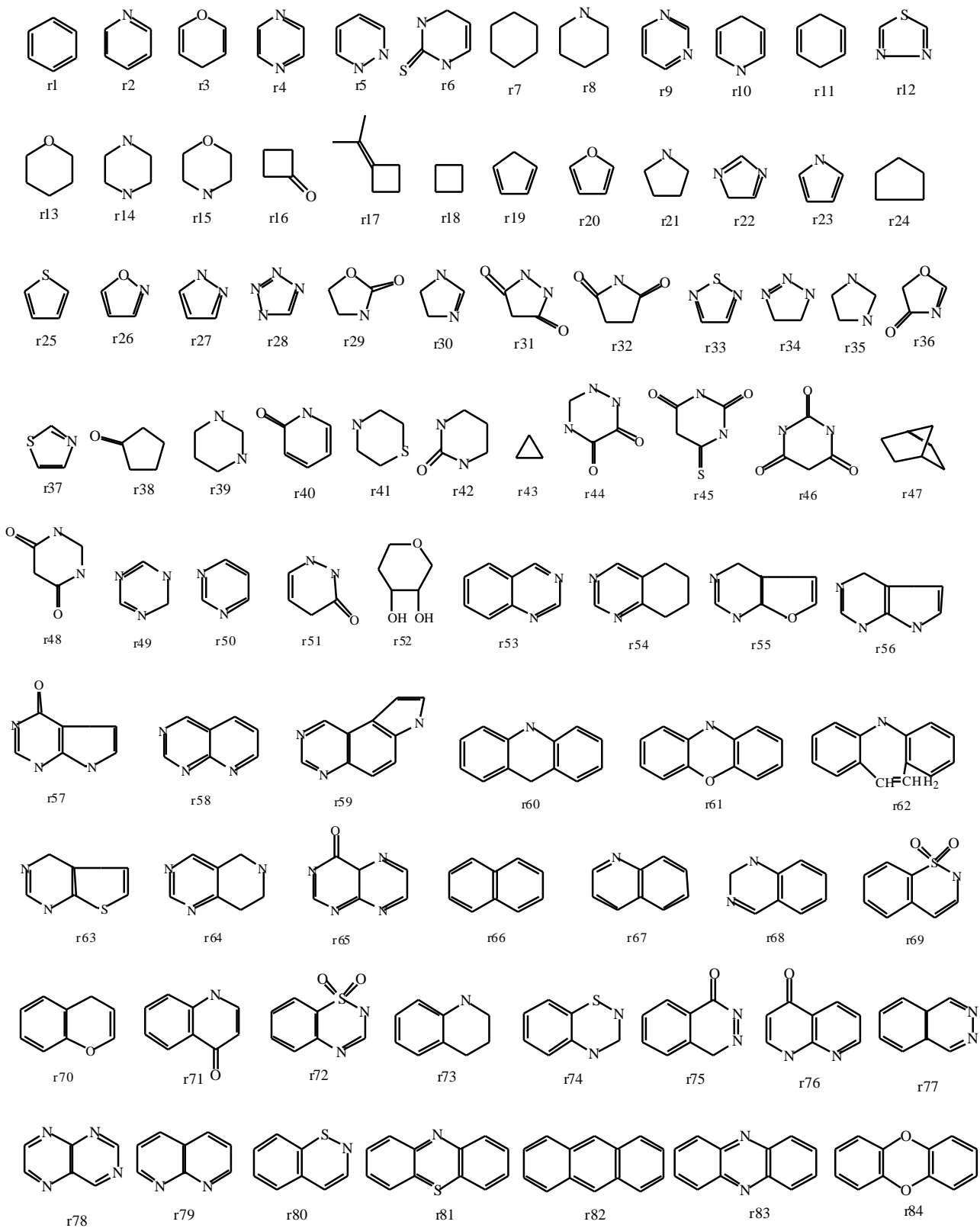
## **ABBREVIATIONS**

GAMESS	=	General atomic and molecular electronic structure system
RESP	=	Restrained electrostatic potential fit
MMGBSA	=	Molecular mechanics generalized born solvent accessibility
PB	=	Poisson boltzmann
ADME	=	Absorption, distribution, metabolism and excretion
MD	=	Molecular dynamics
RMSD	=	Root mean square deviation
COX	=	Cyclooxygenase enzyme

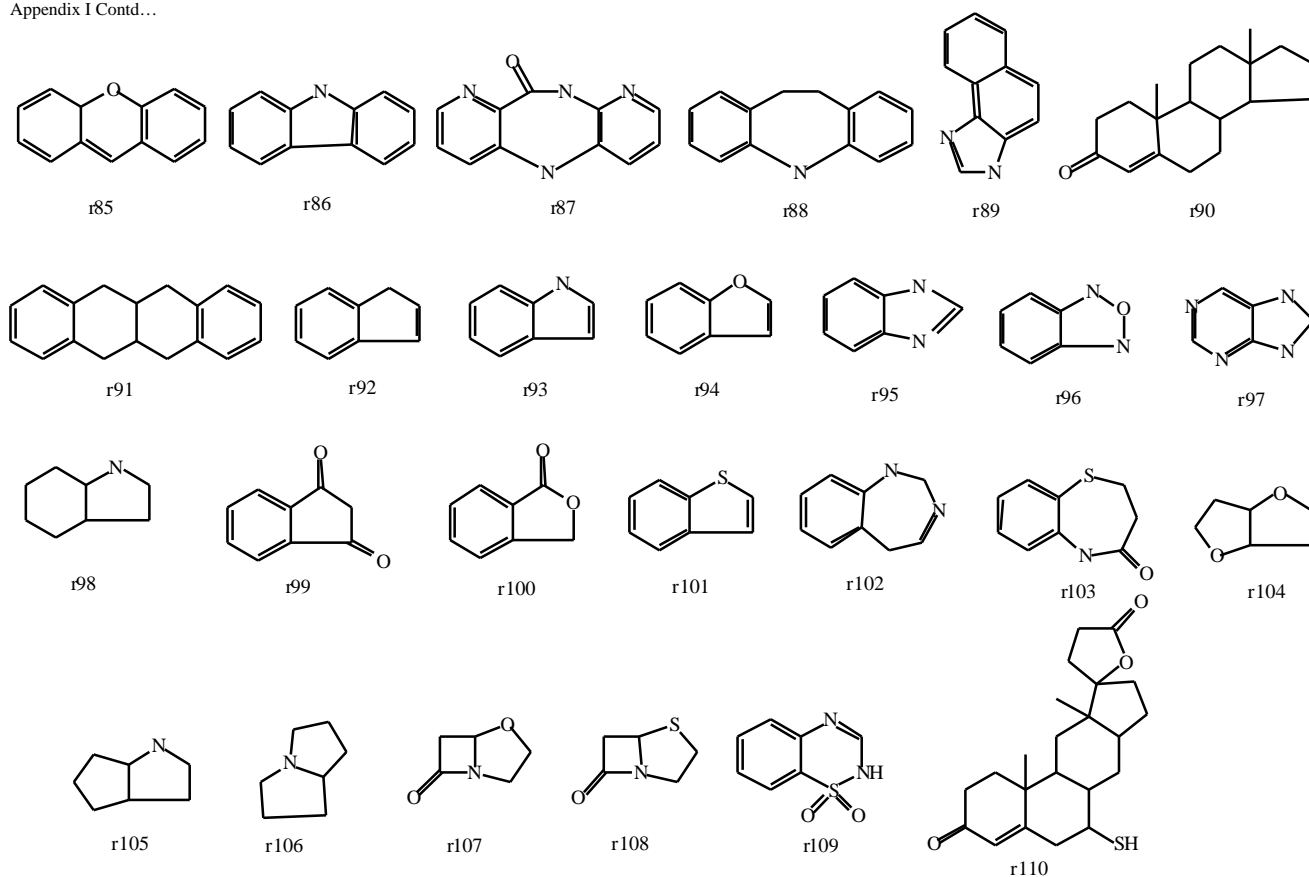
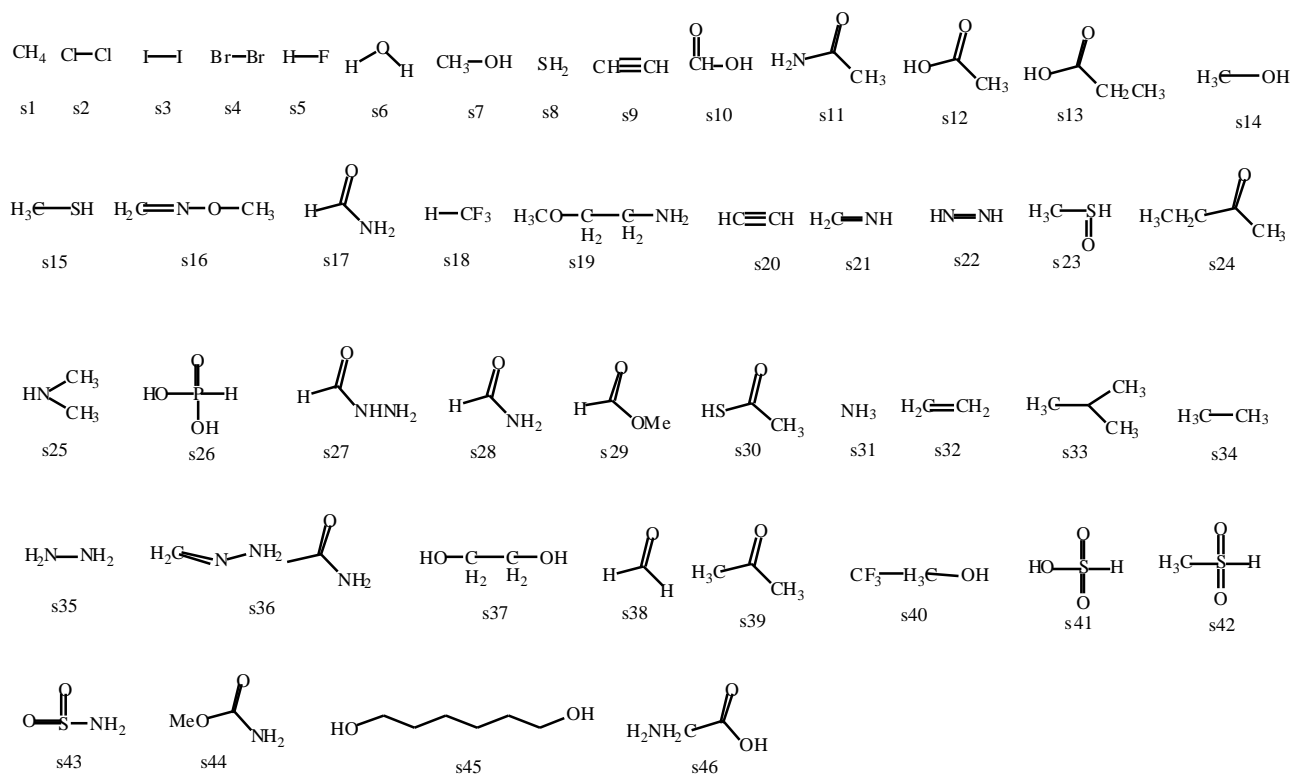
## Appendix

## Template Library

## Rings



Appendix I Contd...

**Side Chain and Linkers**

## REFERENCES

- [1] Walters, P.W.; Stahl, M.T.; Murcko, M.A. *Drug Discovery Today* **1988**, *3*, 160.
- [2] Comprehensive Medicinal Chemistry (CMC), Molecular Design Limited, San Leandro, CA, <http://www.mdli.com>.
- [3] Drews, J. *Nat. Biotech.* **1996**, *14*, 1516.
- [4] Joseph-McCarthy, D. *Pharmacol. Therap.* **1999**, *84*, 179.
- [5] Klebe, G. *J. Mol. Med.* **2000**, *7*, 141.
- [6] Neamati, N.; Barchi, J. J. *Cur. Top. Med. Chem.* **2002**, *2*, 211.
- [7] Konrad, H. B.; Hans-Joachim, B.; Klaus, M.; Alexander, I.A. *Nat. Rev. Drug Disc.* **2003**, *2*, 369.
- [8] Kuntz, I.D.; Meng, E.C.; Shoichet, B.K. *Acc. Chem. Res.* **1994**, *27*, 117.
- [9] Whittle, P.J.; Blundell, T.L. *Ann. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 349.
- [10] Marrone, T.J.; Briggs, J.M.; McCammon, J.A. *Ann. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71.
- [11] Wlodawer, A. *Ann. Rev. Med.* **2002**, *53*, 595.
- [12] Gane, P.J.; Dean, P.M. *Curr. Med. Chem.* **2000**, *10*, 401.
- [13] Farber, G. K. *Pharmacol. Therap.* **1999**, *84*, 327.
- [14] Whittle, P. J.; Blundell, T.L. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 349.
- [15] Lesyng, B.; McCammon, J.A. *Pharmacol. Therap.* **1993**, *60*, 149.
- [16] Pritchard, J.F.; Jurima-Romet, M.; Mark, L.J.R.; Mortimer, E.; Rolfe, B.; Cayen, M.N. *Nat. Rev. Drug Disc.* **2003**, *2*, 542.
- [17] Murray, C.W. *J. Comput. Aided Mol. Des.* **1997**, *11*, 193.
- [18] Bohacek, R.S.; McMartin, C.; Guida, W.C. *Med. Res. Rev.* **1996**, *16*, 3.
- [19] Augen, J. *Drug Discovery Today* **2002**, *7*, 315.
- [20] Norman, P. *Drug News Perspect.* **2001**, *14*, 123.
- [21] Jackson, R.C. *Curr. Opin. Biotechnol.* **1995**, *6*, 646.
- [22] Dean, P. M. *Bioessays* **1994**, *16*, 683.
- [23] Whittaker, P. A. *Trends Pharmacol. Sci.* **2003**, *24*, 434.
- [24] Ooms, F. *Curr. Med. Chem.* **2000**, *7*, 141.
- [25] Terstappen, G.; Reggiani, A. *Trends Pharmacol. Sci.* **2001**, *22*, 23.
- [26] Davis, A.M.; Teague, S. J.; Kleywegt, G. *J. Angew. Chem. Int. Ed.* **2003**, *42*, 2718.
- [27] Schneider, G.; Bohm, H.J. *Drug Discovery Today* **2002**, *7*, 64.
- [28] Bursavich, M.G.; Rich, D.H. *J. Med. Chem.* **2002**, *45*, 541.
- [29] Babine, R.E.; Bender, S.L. *Chem. Rev.* **1997**, *97*, 1359.
- [30] Kingsbury, D.T. *Drug Dev. Res.*, **1997**, *41*, 120.
- [31] Verlinde, C.L.M.J.; Hol, W.G.J. *Structure* **1994**, *2*, 577.
- [32] Kubinyi, H. *Drug Discovery Today* **1997**, *2*, 457.
- [33] Garg, R.; Kurup, A.; Mekapati, S.B.; Hansch, C. *Chem. Rev.* **2003**, *103*, 703.
- [34] Desiraju, G.R.; Gopalakrishnan, B.; Jetti, R.K.R.; Raveendra, D.; Sarma, J.A.R.P.; Subramanya, H.S. *Molecules* **2000**, *5*, 945.
- [35] Chavatte, P.; Yous, S.; Marot, C.; Baurin, N.; Leiseur, D. *J. Med. Chem.* **2001**, *44*, 3223.
- [36] Hansch, C.; Leo A. *Exploring QSAR Fundamentals and Applications in Chemistry and Biology*, Oxford University Press, USA. **1995**.
- [37] Kubinyi, I. *Burger's Medicinal Chemistry*, John Wiley & Sons, New York. **1995**.
- [38] Cramer, R.D.III.; Patterson, D.E.; Bunce, J.D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
- [39] Ohlstein, E. H.; Ruffolo, R. R. Jr.; Elliott, J. D. *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 177.
- [40] Jones, D. A.; Fitzpatrick, F. A. *Curr. Opin. Chem. Biol.* **1999**, *3*, 71.
- [41] Meyer, E. F.; Swanson, S. M.; Williams, J. A. *Pharmacol. Ther.* **2000**, *85*, 113.
- [42] Bharatam, P. V. "Pharmacoinformatics: IT solutions for Drug Discovery and development", CRIPS, **2003**, *4*, 2.
- [43] Bharatam, P. V. "Molecular Modelling Studies on Anti-diabetic Agents: A review", CRIPS, **2002**, *3*, 2.
- [44] Xue, L.; Bajorath, J. *Comb. Chem. High Throughput Screen* **2000**, *3*, 363.
- [45] Jana, S.S.; Bharati, D.J.; Mani, P.; Maitra, A.; Gupta, C.M.; Sarkar, D.P. *FEBS Lett.* **2002**, *515*, 184.
- [46] Narasimhamurthy, J.; Raghu, R. R. A.; Narahari Sastry, G. *Current Medicinal Chemistry - Anti-Cancer Agents* **2004**, *4*, 523.
- [47] Saxena, M.; Agarwal, S. K.; Patnaik, G. K.; Saxena, A. K. *J. Med. Chem.* **1990**, *33*, 2970.
- [48] Reddy, M. R.; Parill, A. In ACS Symposium Series 719. *Rational Drug Design: Novel Methodology and Practical applications*, A. Parill and M. Rami Reddy, Eds. American Chemical Society, Washington, **1999**.
- [49] Hillisch, A.; Hilgenfeld, R. (Eds.) *Modern Methods of Drug Discovery*. Birkhauser Verlag, Basel, **2003**.
- [50] Thomas J. Perun (Ed.) *Computer-aided drug design: Methods and applications*. Marcel Dekker, **2001**
- [51] Charifson, P. S. (Ed.) *Practical applications of computer-aided drug design*. Marcel Dekker, **1997**.
- [52] Vinter, A. G. (Ed.) *Molecular modeling and drug design*. CRC Press, **1994**.
- [53] Lengauer, T. (Ed.) *From Genomes to drugs*. VCH Verlag, **2001**.
- [54] Testa, B.; van de Waterbeemd, H.; Folkers, G.; Guy, R. (Eds.) *Pharmacokinetic Optimization in drug research*. VCH Verlag **2002**.
- [55] van de Waterbeemd, H.; Lennernas, H.; Artursson, P.; Mannhold, R.; Kubinyi, H.; Folkers, G.(Eds.) *Drug Bioavailability: Estimation of solubility, permeability, absorption and Bioavailability*. VCH Verlag, **2003**.
- [56] Poupaert, J.H. (Ed.) *Design of drugs: Basic principle and applications*. Marcel Dekker, **2002**.
- [57] Veerapandian, P. (Ed.) *Structure based drug design*. Marcel Dekker, **1997**.
- [58] Flower, D.R. (Ed.) *Drug design: Cutting Edge Approaches*. Royal Society of Chemistry, **2003**.
- [59] Gubernator, K.; Böhm, H. -J. (Eds.) *Structure based ligand design*. VCH Publishing, **1998**.
- [60] Böhm, H.-J.; Schneider, G.(Eds.) *Protein-ligand interactions: From molecular recognition to drug design*. John Wiley & Sons, **2003**.
- [61] van de Waterbeemd, H.; Testa, B.; Folkers, G. (Eds.) *Computer assisted lead finding and Optimization: Current Tools for Medicinal chemistry*. John Wiley & Sons, **1997**.
- [62] Merlot, C.; Domine, D.; Cleva, C.; Church, D.J. *Drug Discovery Today* **2003**, *8*, 594.
- [63] Ramström, O.; Lehn, J.M. *Nat. Rev. Drug Disc.* **2002**, *1*, 26.
- [64] Lloyd, D.G.; Buenemann, C.L.; Todorov, N.P.; Manalack, D.T.; Dean, P.M. *J. Med. Chem.* **2004**, *47*, 493.
- [65] Bajorath, J. *Nat. Rev. Drug Disc.* **2002**, *1*, 882.
- [66] Lamb, M.L.; Burdick, K.W.; Toba, S.; Young, M.M.A.; Skillman, G.; Zou, X.; Arnold, J.R.; Kuntz, I.D. *Proteins Struct. Funct. & Genet.* **2001**, *42*, 296.
- [67] Pickett, S.D.; Mclay, I.M.; Clark, D.E. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 263.
- [68] Zauhar, R.J.; Moyna, G.; Tian, L.; Li, Z.; Welsh, W.J. *J. Med. Chem.* **2003**, *46*, 5674.
- [69] Stahl, M.; Todorov, N.P.; James, T.; Mauser, H.; Boehm, H.J.; Dean, P.M. *J. Comput. Aided Mol. Des.* **2002**, *16*, 459.
- [70] Teague, S.J. *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 3743.
- [71] Viswanadhan, V.N. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 400.
- [72] Blundell, T.M.; Jhoti, H.; Abell, C. *Nat. Rev. Drug Disco.* **2002**, *1*, 45.
- [73] Schneider, G. *J. Comput. Aided Mol. Des.* **2000**, *14*, 487.
- [74] Langer, T.; Krovat, E.M. *Curr. Opin. Drug Discov. Devel.* **2003**, *6*, 370.
- [75] Bohm, H.J. *Curr. Opin. Biotechnol.* **1996**, *7*, 433.
- [76] Miranker, A.; Karplus, M. *Proteins* **1995**, *23*, 472.
- [77] Leach, A.R.; Kilmington, S.R. *J. Comput. Aided Mol. Des.* **1994**, *8*, 283.
- [78] Bohm, H.J.; Stahl, M. *Curr. Opin. Chem. Bio.* **2000**, *4*, 283.
- [79] Ghose, A.K.; Viswanadhan, V.N.; Wendoloski, J. J. *J. Comb. Chem.* **1999**, *1*, 55.
- [80] Olender, R.; Rosenfeld, R. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 731.
- [81] Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C. J. *Med. Chem.* **1999**, *42*, 3251.
- [82] Mason, J. S.; Good, A. C.; Martin, E. J. *Curr. Pharm. Des.* **2001**, *7*, 567.
- [83] Güner, O.F.(Eds.) *Pharmacophore perception, development and use in drug design*. International University Line, **2000**.
- [84] Gordon, E.M.; Kerwin, J.F. (Eds.) *Combinatorial chemistry and molecular diversity in drug design*. Wiley-Liss, **1998**.
- [85] Ghose, A.K.; Viswanadhan, V.N. (Eds.) *Combinatorial library design and evaluation. Principles, software, tools and applications in drug discovery*. Marcel Dekker, **2001**.
- [86] Fattori, D. *Drug Discovery Today* **2004**, *9*, 5.
- [87] Bemis, G.W.; Murcko, M.A. *J. Med. Chem.* **1996**, *39*, 2887.
- [88] Bemis, G.W.; Murcko, M.A. *J. Med Chem.* **1999**, *42*, 5095.

- [89] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic Acids Res.* **2000**, *28*, 235.
- [90] Schmidt, M.W.; Baldridge, K.K.; Boatz, J.A. Elbert, S.T.; Gordon, M.S.; Jensen, J.S.; Koseki, S.; Nguyen, K.A.; Su, S.; Windus, T.L.; Dupuis, M.; Montgomery, J.A. *GAMESS J. Comp. Chem.* **1993**, *14*, 1347.
- [91] Pearlman, D.A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham III, T.E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.
- [92] Case, D.A.; Pearlman, D.A.; Caldwell, J.W.; Cheatham III, T.E.; Ross, W.S.; Simmerling, C.L.; Darden, T.A.; Merz, K.M.; Stanton, R.V.; Cheng, A.L.; Vincent, J.J.; Crowley, M.; Tsui, V.; Radmer, R.J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G.L.; Singh, U.C.; Weiner, P.K.; Kollman, P. A. AMBER 6, University of California, San Francisco. **1999**.
- [93] Bayly, C.I.; Cieplak, P.; Cornell, W.D.; Kollman, P.A. *J. Phys. Chem.* **1993**, *97*, 10269.
- [94] Bamborough, P.; Cohen, F.E. *Curr. Opin. Struct. Biol.* **1996**, *6*, 236.
- [95] Bohm, H.J. *J. Comput. Aided Mol. Des.* **1992**, *6*, 539.
- [96] Gillet, V.; Johnson, A.P.; Mata, P.; Sike, S.; Zsolodos, Z.; Williams, P.; *J. Comput. Aided Mol. Des.* **1993**, *7*, 127.
- [97] Tschinke, V.; Cohen, N.C. *J. Med. Chem.* **1993**, *36*, 3863.
- [98] Cambridge Crystallographic Data Centre, (CCDC), <http://www.ccdc.cam.ac.uk>
- [99] Available Chemicals Directory (ACD), Molecular Design Limited, San Leandro, CA, <http://www.mdli.com>
- [100] MACCS-II Drug Data Report (MDDR), Molecular Design Limited, San Leandro, CA, <http://www.mdli.com>
- [101] World Drug Index (WDI), Derwent Information, London, UK, <http://www.derwent.com>.
- [102] Lipinsky, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Adv. Drug Del. Rev.* **1997**, *23*, 3.
- [103] Wang, J.; Ramnarayan, K. *J. Comb. Chem.* **1999**, *1*, 524.
- [104] Walters, P.W.; Murcko, M.A. *Adv. Drug Del. Rev.* **2002**, *54*, 255.
- [105] Walters, P.W.; Ajay, X.; Murcko, M.A. *Curr. Opin. Chem. Biol.* **1999**, *3*, 384.
- [106] Clark, D.E.; Pickett, S.D. *Drug Discovery Today* **2000**, *5*, 49.
- [107] Walters, P.W.; Stahl, M.T.; Murcko, M.A. *Drug Discovery Today* **1998**, *3*, 160.
- [108] Sadowski, J.; Kubinyi, H. *J. Med. Chem.* **1998**, *41*, 3325.
- [109] Wildman, S.A.; Crippen, G.M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868.
- [110] Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. *J. Med. Chem.* **2002**, *45*, 2615.
- [111] Bergstrom, C.A.S.; Strafford, M.; Lazarova, L.; Avdeef, A.; Luthman, K.; Artursson, P. *J. Med. Chem.* **2003**, *46*, 558.
- [112] Mitchell, T.; Showell, G. A. *Curr. Opin. Drug Discov. Del.* **2001**, *4*, 314.
- [113] Muegge, I.; Heald, S.L.; Britelli, D. *J. Med. Chem.* **2001**, *44*, 1841.
- [114] Navia, M.A.; Chaturvedi, P.R. *Drug Discovery Today* **1996**, *1*, 179.
- [115] Brustle, M.; Beck, B.; Schindler, T.; King, W.; Mitchell, T.; Clark, T. *J. Med. Chem.* **2002**, *45*, 3345.
- [116] Hann, M.M.; Leach, A.R.; Harper, G. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856.
- [117] Blake, J.K. *Curr. Opin. Biotechnol.* **2000**, *11*, 104.
- [118] Oprea, T.I. *J. Comput. Aided Mol. Des.* **2000**, *14*, 251.
- [119] Oprea, T.I.; Davis, A.M.; Teague, S.J.; Leeson, P.D. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308.
- [120] Butina, D.; Segall, M.D.; Frankcombe, K. *Drug Discovery Today* **2002**, *7*, 83.
- [121] Waterbeemd, H.V.; Gifford, E. *Nat. Rev. Drug Disc.* **2003**, *2*, 192.
- [122] Hodgson, J. *Nat. Biotechnol.* **2000**, *19*, 722.
- [123] Tarbit, M.; Berman, H. *Curr. Opin. Chem. Biol.* **1998**, *2*, 411.
- [124] Stouch, T.R.; Kenyon, J.R.; Johnson, S.R.; Chen, X.Q.; Doweyko, A.; Li, Y. *J. Comp. Aided Mol. Des.* **2003**, *17*, 83.
- [125] Beresford, A.P.; Segall, M.; Tarbit, M.H. *Curr. Opin. Drug Discov. Del.*, **2004**, *7*, 36.
- [126] Modi, S. *Drug Discovery Today*, **2004**, *1*, 14.
- [127] Ghombar, V.K.; Silver, I.S.; Zhao, Z. *Curr. Top. Med. Chem.* **2003**, *3*, 1205.
- [128] Etkins, S.; Rose, J. *J. Mol. Graph. Model.* **2002**, *20*, 305.
- [129] Boobis, A.; Gundert-Remy, U.; Kremers, P.; Macheras, P.; Pelkonen, O. *Eur. J. Pharm. Sci.* **2002**, *17*, 183.
- [130] Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 251.
- [131] Clark, D. E.; Grootenhuus, D. J. *Curr. Opin. Drug. Disc. Dev.* **2002**, *5*, 382.
- [132] Wang, J.; Lai, L.; Tang, Y. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1173.
- [133] Etkins, S. *Biochem. Soc. Trans.* **2003**, *31*, 611.
- [134] Greene, N. *Adv. Drug Del. Rev.* **2002**, *54*, 417.
- [135] Cronin, M. T. D. *Pharm Pharmacol. Commun.* **1998**, *4*, 157.
- [136] Benigni, R.; Richard, A. M. *Meth. Enzymol.* **1998**, *14*, 264.
- [137] Gillet, V.J.; Myatt, G.J.; Johnson, A.P. *Persp. Drug Discovery Des.* **1995**, *3*, 34.
- [138] Long, A.K.; Kappos, J. C. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 915.
- [139] E. Lewars (Ed.) *Computational Chemistry –Introduction to the theory and applications of Molecular and Quantum Mechanics*, Kluwer Academic Publishers, Boston **2003**.
- [140] Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. *J. Comput. Chem.* **2004**, *25*, 1157.
- [141] Frisch, M.J.; Head-Gordon, M.; Trucks, G.W.; Foresman, J.; Schlegel, B.; Raghavachari, K.; Robb, M.A.; Binkley, J.S.; Gonzalez, C.; Defrees, D.J.; Fox, D.J.; Whiteside, R.A.; Seger, R.; Melius, C.F.; Baker, J.; Martin, L.R.; Kahn, L.R.; Stewart, J.J.P.; Topiol, S.; Pople, J.A. *Gaussian 90*, Gaussian Inc, Pittsburgh, PA, **1990**.
- [142] Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- [143] Mackerell, Jr. A.D. *J. Comput. Chem.* **2004**, *25*, 1584.
- [144] Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- [145] Wang, J.M.; Cieplak, P.; Kollman, P.A. *J. Comp. Chem.* **2000**, *21*, 1049.
- [146] Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. *Nat. Rev. Drug Disc.* **2004**, *3*, 935.
- [147] Gorlick, R.; Goker, E.; Trippett, T.; Waltham, M.; Banerjee, D.; Bertino, J.R.N. *Eng. J. Med.* **1996**, *335*, 1041.
- [148] Wang, R.; Lu, Y.; Wang, S. *J. Med. Chem.* **2003**, *46*, 2287.
- [149] Kuntz, I.D. *Science* **1992**, *257*, 1078.
- [150] Broijmans, N.; Kuntz, I.D. *Ann. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335.
- [151] McConkey, B.J.; Sobolev, B.; Edelman, M. *Curr. Sci.* **2002**, *83*, 7.
- [152] Kuntz, I.D.; Meng, E.C.; Shoichet, B.K. *Acc. Chem. Res.* **1994**, *27*, 117.
- [153] Cherfils, J.; Janin, J. *Curr. Opin. Struct. Biol.* **1993**, *3*, 265.
- [154] Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759.
- [155] Halperin, I.; Ma, B.; Nussinov, R. *Proteins*, **2002**, *47*, 409.
- [156] Leach, A.R. *J. Mol. Biol.* **1994**, *235*, 345.
- [157] Abagyan, R.; Totrov, M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375.
- [158] Taylor, R.D.; Jewsbury, P.J.; Essex, J.W. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151.
- [159] Goodsell, D.S.; Olson, A.J. *Proteins* **1990**, *8*, 195.
- [160] Bartlett, P.A.; Shea, G.T.; Telfer, S.T.; Waterman, S. *CAVEAT: A Program to Facilitate the Structure-derived Design of Biologically Active Molecules*; Roberts, S. M., Ed.; Royal Society of Chemistry, London, **1989**.
- [161] Lawrence, M. C.; Davis, P. C. *Pro. Struct. Fun. Genet.* **1992**, *12*, 31.
- [162] Kick, E.K.; Roe, D.C.(co-first authors); Skillman, A.G.; Liu, G.; Ewing, T.J.; Sun, Y.; Kuntz, I.D.; Ellman, J.A. *Chem. Biol.* **1997**, *4*, 297.
- [163] (a) Meng, E.C.; Shoichet, B.K.; Kuntz, I.D. *J. Comput. Chem.* **1992**, *13*, 505; (b) Lewis, R.A.; Roe, D.C.; Huang, C.; Ferrin, T.E.; Langridge, R.; Kuntz, I.D. *J. Mol. Graph.* **1992**, *10*, 66.
- [164] Clark, K.P.; Ajay, J. *J. Comput. Chem.* **1995**, *16*, 1210.
- [165] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
- [166] Miller, M.D.; Kearsley, S.K.; Underwood, D.J.; Sheridan, R.P. *J. Comput. Aided Mol. Des.* **1994**, *8*, 153.
- [167] Gabb, J.; Jackson, R.M.; Sternberg, M.J.E. *J. Mol. Biol.* **1994**, *272*, 106.
- [168] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
- [169] Moon, J.B.; Howe, J.W. *Proteins* **1991**, *11*, 314.
- [170] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. *J. Med. Chem.* **2004**, *47*, 1739.
- [171] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry,



- J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. *J. Med. Chem.* **2004**, *47*, 1750.
- [172] Miranker, A.; Karplus, M. *Proteins* **1991**, *11*, 29.
- [173] Sobolev, V.; Wade, R.C.; Vriend, G.; Edelman, M. *Proteins* **1996**, *25*, 120.
- [174] Das, A.; Jayaram, B. *J. Mol. Liq.* **1998**, *77*, 157.
- [175] McCammon, J.A.; Harvey, S.C. *Dynamics of proteins and nucleic acids*. Cambridge University Press, Cambridge. **1987**.
- [176] Northrop, S.H.; Boles, J.O.; Reynolds, J.C.L. *J. Phys. Chem.* **1987**, *91*, 5991.
- [177] Teague, S.J. *Nat. Rev. Drug Disc.* **2003**, *2*, 527.
- [178] Carlson, H.A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447.
- [179] Carlson, H.A.; McCammon, J. *Mol. Pharmacol.* **2000**, *57*, 213.
- [180] Wang, J.; Kollman, P.A.; Kuntz, I.D. *Proteins* **1999**, *36*, 1.
- [181] Roche, O.; Kiyama, R.; Brooks, C.L.III. *J. Med. Chem.* **2001**, *44*, 3592.
- [182] Arora, N.; Jayaram, B. *J. Phys. Chem.* **1998**, *102*, 6139.
- [183] Kollman, P.A. *Chem. Rev.* **1993**, *93*, 2395.
- [184] Beveridge, D.L.; Mezei, M.; Ravishanker, G.; Jayaram, B. *Int. J. Quantum. Chem.* **1986**, *29*, 1513.
- [185] Beveridge, D.L.; DiCapua, F.M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
- [186] van Gunsteren, W. F. *Computer Simulations of Biomolecular Systems. ESCOM*: Leiden, The Netherlands. **1989**.
- [187] McCammon, J.A. *Curr. Opin. Struct. Biol.* **1991**, *1*, 196.
- [188] Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430.
- [189] Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P.A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211.
- [190] Bash, H.A.; Singh, U.C.; Langridge, R.; Kollman, P.A. *Science*, **1987**, *236*, 564.
- [191] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T.E. *Acc. Chem. Res.* **2000**, *33*, 889.
- [192] Reddy, M.R.; Singh, U. C.; Erion, M. C. *J. Am. Chem. Soc.* **2004**, *126*, 6224.
- [193] Jorgensen, W.L. *Acc. Chem. Res.* **1989**, *22*, 184.
- [194] McQuarrie, D.A. *Statistical Mechanics*, Harper and Row Publishers, New York. **1976**.
- [195] Gilson, M.K.; Given, V.; Bush, B.L.; McCammon, J.A. *Biophys. J.* **1997**, *72*, 1047.
- [196] Atkins, P.W.; Paula, J. *Physical Chemistry* Oxford University Press, New York. **2002**.
- [197] Murray, C. W.; Verdonk, M. L. *J. Comput. Aided. Mol. Des.* **2002**, *16*, 741.
- [198] Ajay; Murcko, M.A. *J. Med. Chem.* **1995**, *38*, 4953.
- [199] Aqvist, J. *J. Comp. Chem.* **1996**, *17*, 1587.
- [200] Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
- [201] Srinivasan, J.; Cheatham III, T.E.; Cieplak, P.; Kollman, P.A.; Case, D.A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.
- [202] Swanson, J.M.J.; Henchman, R.H.; McCammon, J.A. *Biophys. J.* **2004**, *86*, 67.
- [203] Jayaram, B.; Sprous, D.; Beveridge, D.L. *J. Phys. Chem. B.* **1998**, *102*, 9571.
- [204] Bashford, D.; Case, D.A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
- [205] Zou, X.; Sun, Y.; Kuntz, I.D. *J. Am. Chem. Soc.* **1999**, *121*, 8033.
- [206] Karplus, M.; Kushick, J.N. *Macromolecules*. **1981**, *14*, 325.
- [207] Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.
- [208] Scahfer, H.; Mark, A.E.; van Gunsteren. *J. Chem. Phys.* **2000**, *113*, 7809.
- [209] Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T.J. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
- [210] Jayaram, B.; McConnell, K.J.; Dixit, S.B.; Beveridge, D.L. *J. Comput. Phys.* **1999**, *151*, 333.
- [211] Jayaram, B.; Liu, Y.; Beveridge, D.L. *J. Chem. Phys.* **1998**, *109*, 1465.
- [212] Lee, K.; Richards, F.M. *J. Mol. Biol.* **1971**, *55*, 379.
- [213] Sharp, K.A.; Nicholls, A.; Fine, R.F.; Honig, B. *Science* **1991**, *252*, 106.
- [214] Ohtaki, H.; Fukushima. N. *J. Sol. Chem.* **1992**, *21*, 23.
- [215] Noskov, S.Y.; Lim, C. *Biophys. J.* **2001**, *81*, 737.
- [216] Kalra, P.; Das, A.; Dixit, S.B.; Jayaram, B. *Ind. J. Chemistry A.* **2000**, *39A*, 262.
- [217] Kalra, P.; Das, A.; Jayaram, B. *Appl. Biochem. Biotech.* **2001**, *96*, 93.
- [218] Kalra, P.; Reddy, V.; Jayaram, B. *J. Med. Chem.* **2001**, *44*, 4325.
- [219] Madhusudhan, M. S.; Vishveshwara, S.; Das, A.; Kalra, P.; Jayaram, B. *Ind. J. Biophys. Biochem.* **2001**, *38*, 27.
- [220] Shaikh, S. A.; Ahmed, S. R.; Jayaram, B.; *Arch. Biochem. Biophys.* **2004**, *429*, 81.
- [221] Jayaram, B.; McConnell, K.J.; Dixit, S.B.; Das, A.; Beveridge, D.L. *J. Comput. Chem.* **2002**, *23*, 1.
- [222] Kombo, D.C.; Jayaram, B.; McConnell, K.J.; Beveridge, D.L. *Molecular Simulation* **2002**, *28*, 187.
- [223] Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*, Oxford University Press: Oxford, **1987**.
- [224] Leach, A. R. *Molecular Modelling*, Addison Wesley Longman Limited: Singapore, **1996**.
- [225] Goodfellow, J.M.; Moss, D.S. *Computer Modelling of Biomolecular Processes*. Ellis Horwood Series in Molecular Biology, Chichester, **1993**.
- [226] Lipkowitz, K.B.; Boyd, D.B. *Reviews in Computational Chemistry*, Volume 11, Wiley-VCH, New York, **1997**.
- [227] Goodfellow, J.M. *Molecular Dynamics: An Overview of Applications in Molecular Biology* CRC Press, LLC, U.S.A. **1991**.
- [228] Warshel, A. *Computer simulation of chemical reactions in enzymes and solutions*. John Wiley and Sons, New York. **1991**.
- [229] Kalgutkar, A. S.; Crews, B. C.; Rowlinson, S. W.; Garner, C.; Seibert, K.; Marnett, L. J.; *Science* **1998**, *280*, 1268.
- [230] Latha, N.; Jain, T.; Sharma, P.; Jayaram, B. *J. Biomol. Struct. Dyn.* **2004**, *21*, 791.
- [231] Melissa, L.; Price, P.; Jorgensen, W.L. *J. Am. Chem. Soc.* **2000**, *122*, 9455.
- [232] Wesolowski, S.S.; Jorgensen, W.L. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 267.
- [233] Kothekar, V.; Sahi, S.; Srinivasan, M. *J. Biomol. Struct. Dyn.* **1999**, *16*, 901.
- [234] Fornabai, M.; Spyrikis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. *J. Med. Chem.* **2004**, *47*, 4507.
- [235] Zoete, V.; Michielin, O.; Karplus, M. *J. Comput. Aided Mol. Des.* **2003**, *12*, 861.
- [236] Sims, P. A.; Wong, C. F.; McCammon, J. A. *J. Med. Chem.* **2003**, *46*, 3314.
- [237] Mandal, T. K.; Mukhopadhyay, C. *Protein Eng.* **2002**, *12*, 979.
- [238] Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. *Biopolymers.* **2003**, *68*, 16.
- [239] Rognan, D.; Lauemoller, S.L.; Holm, A.; Buus, S.; Tschinke, V. *J. Med. Chem.* **1999**, *42*, 4650.
- [240] Froloff, N.; Windemuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293.