The Challenges of Sciences. A Tribute to Carlos Chagas
Pontifical Academy of Sciences, *Scripta Varia* 103, Vatican City 2002
www.pas.va/content/dam/accademia/pdf/sv103/sv103-janssen.pdf

HOW TO FIND BETTER DRUGS FOR THE TREATMENT OF AIDS

PAUL A.J. JANSSEN

The Center for Molecular Design (CMD), which became operational in April 1996, was conceived as a think-tank. The objective of the CMD is to provide ideas far the synthesis of chemical compounds by means of molecular modeling and using crystal structures of proteins of therapeutic interest. The design team is multi-disciplinary and comprises biochemists, chemists, a physicist, mathematicians and a computer engineer. It relies heavily on the use of dedicated computers and modeling software.

Current targets for molecular design are HIV-reverse transcriptase, HIV-protease and Influenza-neuraminidase. Most of the effort of the CMD is presently concentrated on the design of nonnucleoside inhibitors of HIV-reverse transcriptase (NNRTI). The immediate goal of the CMD is to design compounds using a 'de novo' approach, i.e. from first principles using only the three-dimensional structure of a target protein and the rules of chemistry.

Molecular modeling at the CMD proceeds in three steps. First, a proposed ligand is 'docked' into the binding sit a of the target protein, using the ligand's minimal energy conformation. This operation can be regarded as a test for geometrical fit of a ligand. After successful docking, the change in free energy between the ligand in solution and the ligand-protein complex is determined. In reverse transcriptase, most of the interactions between the ligand and its environment are governed by the polar and lipophilic properties of the ligand. It is also possible to identify the specific amino acid residues of the binding site, which account for the largest part of the ligand-protein interaction. Finally, using computed binding energies and observed antiviral activities of known compounds, a prediction can be made for the activity of a newly designed compound.

The 'do novo' approach of the CMD makes use of docking, binding and prediction of activity in an automatic design process which mimics natural evolution and which is, therefore, referred to as 'genetic algorithm'. At the start of this process one has to define a tractable initial 'population' of chemical structures. Each of these is assigned a 'fitness' value, according to its predicted biological activity, using the molecular modeling procedure described above. 'Parent' structures are selected far 'breeding' according to their fitness, i.e. structures with high fitness stand a better chance of being selected than those with low fitness. Diversification of chemical structure in the 'offspring' is obtained by exchanging structural fragments from the parents, which is equivalent to 'cross-over' in natural reproduction. 'Child' structures can be diversified further by adding and removing fragments, which corresponds to 'mutation'. Finally, the fitness of the offspring is evaluated and the best-performing ones may replace structures in the population that are less fit. At this point one 'generation' is completed. After several generations, the process of variation and survival of the fittest will breed a population of chemically diverse structures with high-predicted activity.

Molecular modeling and 'de novo' design are by no means substitutes for chemical imagination and serendipity. Rather, the role of the chemist is strengthened by the need to evaluate a large quantity of computer-generated compounds with respect to the ease with which these can be synthesized, as well as their stability and solubility. The approach has led to the design of a novel class of highly active anti-HIV compounds (substituted dianilino pyrimidines or DAPY's) which are presently being evaluated in clinical trials.