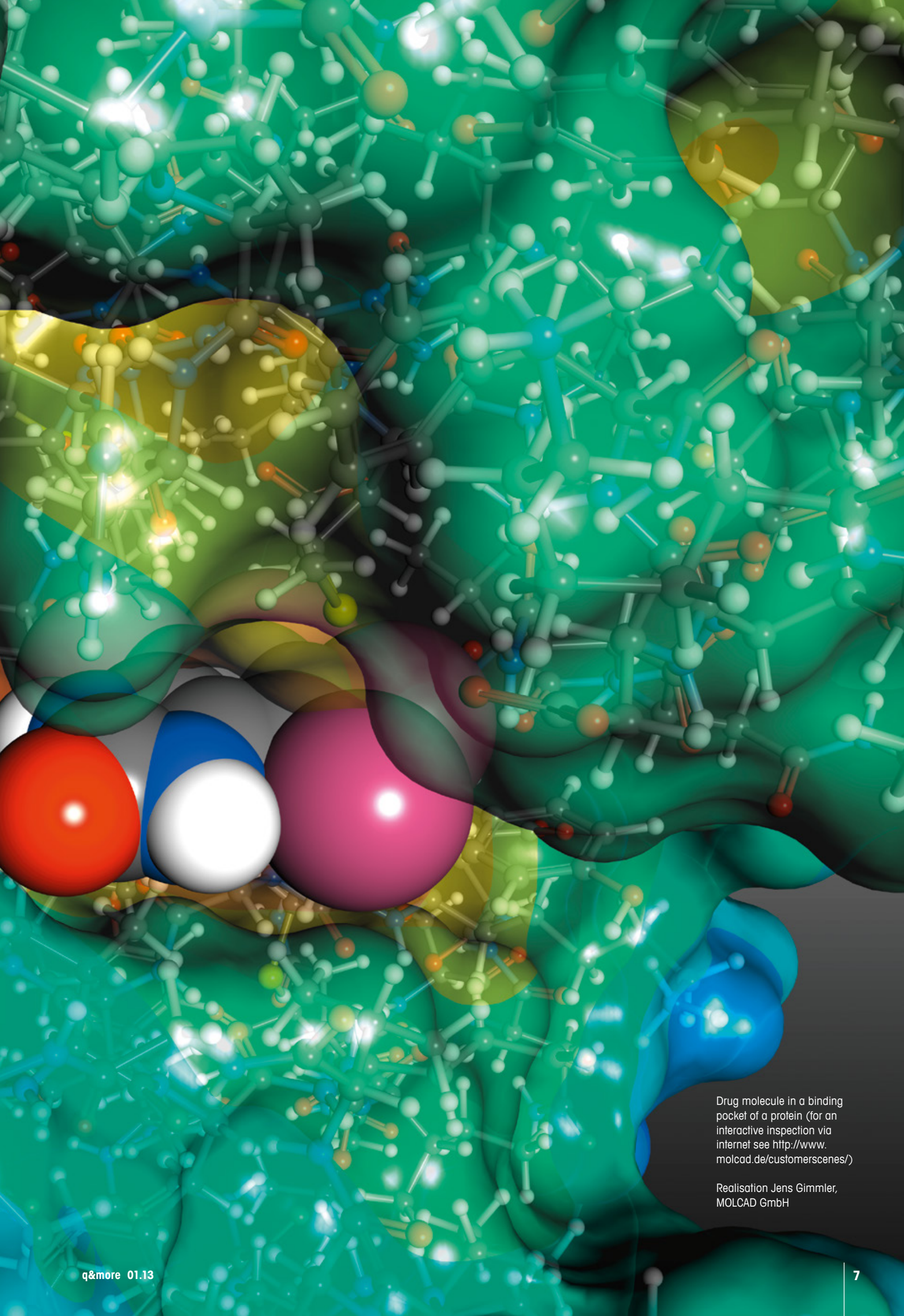
New Trends in Computer-Aided Drug Design

On-line Integration of Super-Computers

Prof. Dr. Jürgen Brickmann,
Physical Chemistry i, Technical University Darmstadt, Germany

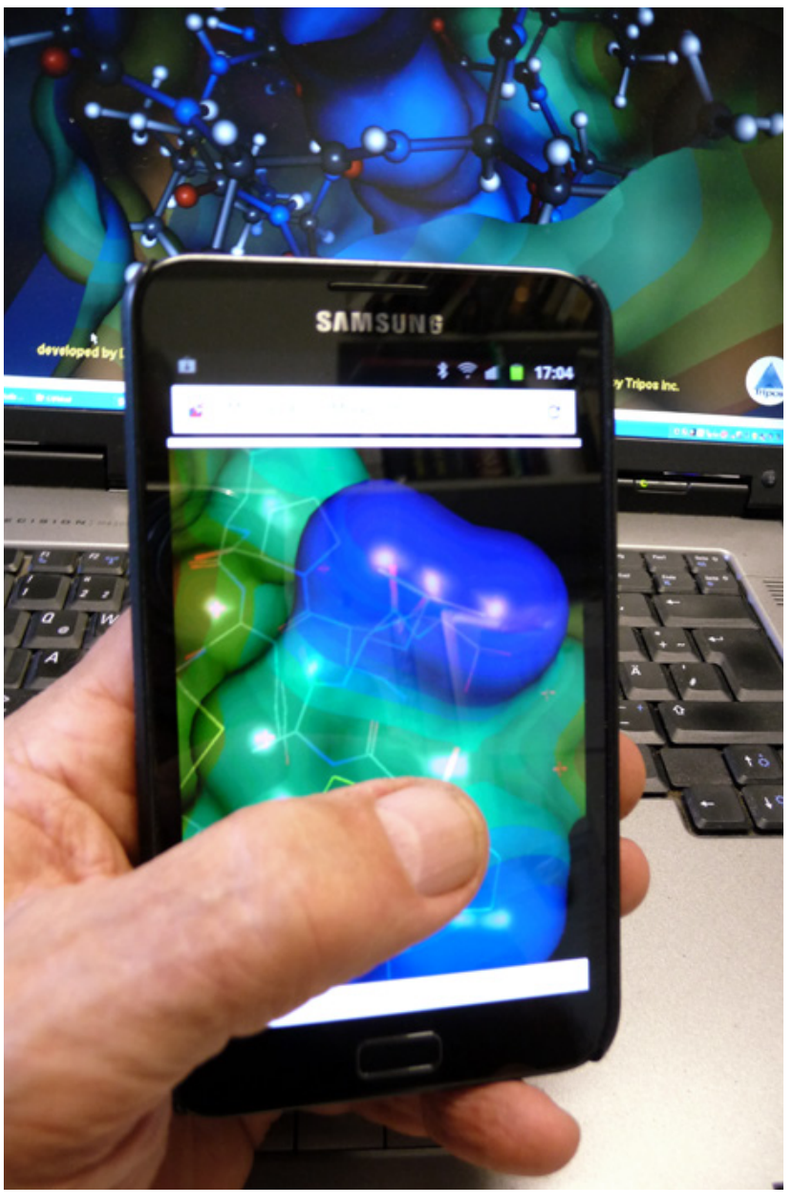
Prof. Dr. Timothy Clark,
Computer-Chemie-Centrum, University of Erlangen-Nuremberg, Germany



Drug molecule in a binding pocket of a protein (for an interactive inspection via internet see http://www.molcad.de/customerscenes/)

Realisation Jens Gimmler, MOLCAD GmbH

Computer-Aided Drug design (CADD) is not new. The Journal of Computer-Aided Molecular Design (Springer) was founded in 1987, when computers in the worldwide top 500 were slower than today's smart phones. This makes the field a quarter of a century old. Generally, scientific disciplines of this age have become mature, the major developments have been made and procedures have become routine. Superficially, this is also the case for CADD. However, the environment for all computer-based scientific disciplines has changed rapidly and continuously in the past quarter century. Our phones and automobiles have more CPU power that supercomputers 25 years ago and importantly, can also speak to each other more easily. This situation is exciting. We can do things that we couldn't dream of in 1987. This should mean that CADD is also a rapidly expanding field in which new compute-intensive techniques are being introduced continuously to improve performance and applicability. Sadly, this is not the case. CADD has not used the possibilities open to it. However, there are signs that things are beginning to change.



The actual output of number crunching results can be interactively inspected via the internet even with smart phones and tablets

In the traditional CADD scenario the number crunching was decoupled from the interactive data processing and knowledge generation process. Visualization was the domain of specialized graphics computers that were able to handle the extreme demands of visualizing proteins and their properties. That was before the games industry. Top quality molecular graphics are now at the bottom end of the demands placed on graphics hardware. At the risk of being boring, even smart phones can produce interactive 3D-molecular graphics that would have been beyond the first generation of raster-graphics machines. New web-based technology, such as Molcad's web3d molecule visualizer can make high-quality interactive 3D molecular graphics available anywhere at any time. This is only the beginning; interactive 3D graphics that can be manipulated in real time can be shared by anyone with a suitable web browser. Moreover, the new Web based technologies open completely new direction in the simulation scenario: Number crunching can be directly integrated in the real time treatment of CADD-activities, even when these processes are performed at locations thousands of miles apart.

The in many ways unimaginable increase in the speed of computations opens possibilities of using theoretical techniques that were unthinkable in the early days of CADD. Sadly, very little use has been made of these possibilities. The vast majority of CADD-techniques today rely on atomistic classical mechanical models that were designed in the 1970's for the hard- and software of the time. One of the most important interactions between molecules (electrostatics) is treated in this model by assigning a single charge at the center of each atom. This model is completely unsuitable for elements like

chlorine, bromine and iodine, which appear negative from some directions and positive from others. These elements are generally assigned a negative charge, meaning that they repel other negative atoms such as the oxygens in the backbone chains of proteins in the model. This is the reverse of the true situation. In the correct orientation, backbone oxygen atoms attract atoms such as chlorine, bromine and iodin.

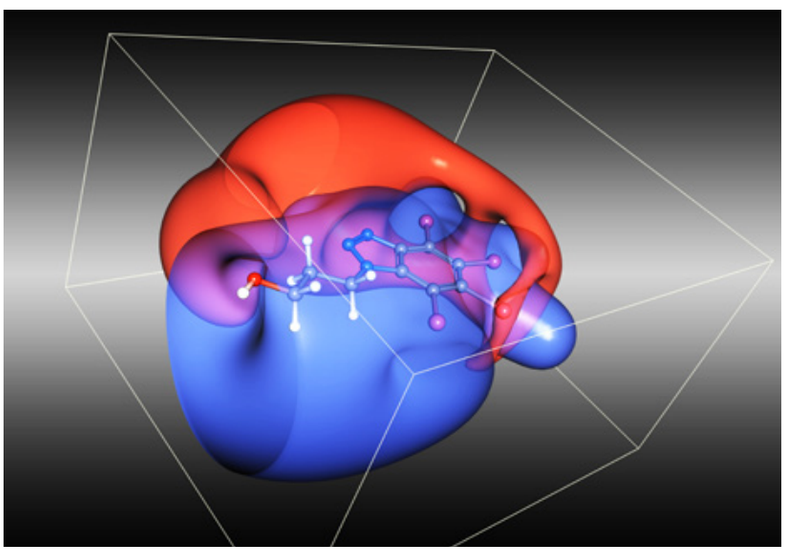
We are thus confronted with the situation that the models that we have used for CADD for a quarter century give the wrong sign for an important interaction (very many current drugs contain chlorine). How can this happen? The answer is probably that we have confused our models with reality. "Everybody knows" that chlorines in drug molecules "are negative". That is true if you approach them perpendicular to the carbon-chlorine bond, but chlorine atoms "are positive" if you approach them from opposite the carbon to which they are bound. Currently, many attempts are being published to correct this situation by adding an extra positive charge to chlorine in the model.

This will fix the problem – but what about the next one? One of Thomas Kuhn's signs of a failing paradigm is that it needs increasing numbers of ad hoc fixes. So why stick to classical mechanical models? We have a variety of quantum mechanical methods that all reproduce the chlorine-oxygen interaction correctly. Why not use them? The answer, sadly, is that they used far too much computer time in the 1970's and 1980's. Today's hard- and software are easily capable of using standard quantum mechanical calculations to treat pharmaceutical databases of hundreds of thousands of molecule or even to calculate an entire protein target. Our challenge is therefore to combine modern theoretical techniques from other branches of chemistry with high-performance hard- and software to improve the performance and reliability of CADD.

Why do we need to improve such a well established field? After all, CADD is used every day in dozens of pharmaceutical companies, apparently with good results. Well, another consequence of a paradigm at the end of its days is that it is defended strongly in the face of hard evidence that it does not work. A published test of docking and scoring algorithms used to estimate the binding affinity of drugs to their protein targets concluded with a table in which ten different docking-and-scoring techniques were used on datasets for ten different targets. The table reported the correlation coefficients between the experimental and calculated binding affinities. Of one hundred entries in the table, 64 were negative (i.e. they predicted the reverse trend to that found experimentally). Of the remaining 36, the highest was approximately 0.2. This is hardly acceptable performance but nonetheless, pharmaceutical companies rely on docking and scoring and dozens of papers using the technique without experimental confirmation are published every day. The whole situation is reminiscent of "the emperor's new clothes".

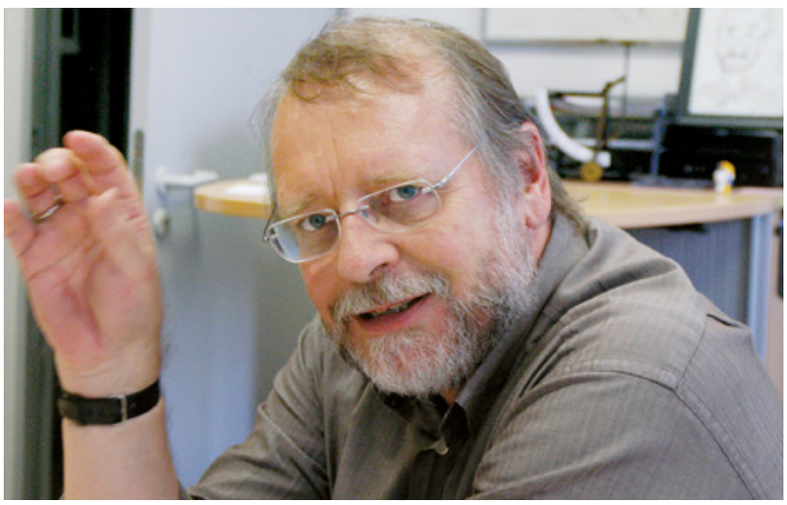
So what needs to be done? We need to rethink the way that we do CADD. Over the last decade, the emphasis has been placed on using existing techniques for increasing numbers of compounds. Given the results discussed above, this amounts to collecting even larger numbers of wrong predictions. Of course, the likelihood of a few predictions being correct increases with the number of predictions made. What we have been doing is buying more and more lottery tickets.

Is there a "system" that allows CADD to "win" and become predictive? We don't know. What we do know is that there are theories of intermolecular interactions (which are what really interest us) that are far more accurate and above all general than the ones we currently use in CADD. A variety of quantum mechanical methods range from semiempirical MO-theory (which can calculate hundreds of thousands of molecules or complete proteins) to high-level ab initio theory (which for small molecules is usually more accurate than experiment). Density-functional theory (DFT) has become the workhorse of computational quantum chemistry and can easily calculate drug-sized molecules within a few minutes on a modern multicore node. It has been parameterized extensively in the last decade so that the newest functionals are very accurate for everyday molecules like drugs.



Electrostatic potential (isopotential surfaces for negative and positive potentials) around a drug molecule containing chlorine atoms. (for an interactive inspection via internet see http://www.molcad.de/customerscenes/)
Realisation Jens Gimmler, MOLCAD GmbH

Jürgen Brickmann, born 1939 in Schwerin, studied physics 1959-1965 (Uni Munich, Uni Innsbruck, Tu Munich) and obtained his Ph.D from the TU Munich in 1967. In 1973 he became a Private Docent for physical chemistry at the Uni Freiburg, from 1974 to 1978 he was Professor for chemical dynamics at the Uni Konstanz. From 1979 to 2004 hold a chair for physical chemistry at Darmstadt University of Technology. He was guest scientist at different Universities (FU Berlin, Tel Aviv University, Hebrew University, Jerusalem). Presently he is principal owner and CEO of MOLCAD GmbH as well scientific director of SUCCIDIA AG, a publication and communication company in Darmstadt (Germany).



Tim Clark was born in 1949 in England and obtained his Ph.D. from the Queen's University Belfast in 1973. He is Director of the Computer-Chemie-Centrum in Erlangen and the Centre for Molecular Design at the University of Portsmouth, UK. He develops and applies modelling and simulation techniques in chemistry, materials science and biology. He is the author of 300 research articles and two books and is the founding editor of the Journal of Molecular Modeling. In 2009, he was awarded the Klaus-Wilhelm von der Lieth Medal.

Treating intermolecular interactions properly is only half the battle. Biological systems are flexible and dynamic at physiological temperatures. This means that we must consider all the conformations that are present in the living system. This is the so-called conformational sampling problem, which means that we must use molecular-dynamics simulations to allow the molecules to move and to adopt all possible conformations open to them. This is a major computational task, for which special dedicated hardware has been built. Considering conformational sampling adequately involves a true paradigm shift in CADD from thinking about static single structures to considering what the real moving molecules are doing.

Finally, to make things even worse, biology happens in aqueous solution that is full of ions, small molecules, proteins, nucleic acids etc. Simulating the effect of the solvent water is very expensive computationally. It can be done by simulating a system that includes the water solvent explicitly but a liter of water contains more than 3×10^{25} molecules. We don't need to simulate a liter but there are still going to be a lot of water molecules. One solution would be to represent the solvent as a continuous medium or continuum. This would be fine and very efficient if it were accurate. The problem is that water is a very complicated molecule and doesn't look much like a simple continuum. We therefore also need new calculational models in which the water solvent is treated implicitly and accurately.

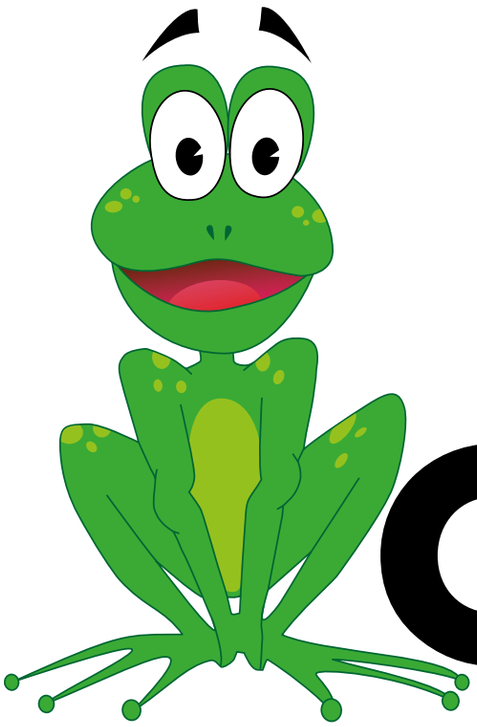
Are we doing anything about improving CADD and using the capabilities of modern computers? The answer is a cautious "yes". The hpCADD project (www.hpcadd.com) involves partners from academia (the Universities of Erlangen-Nurnberg and Dortmund) and industry (Sanofi, Frankfurt) and is funded by the German Ministry of Education and Research. The € 1.5 million, three year project involves computational and theoretical chemists, computer scientists who specialize in high-performance computing and a pharmaceutical company to test and validate new methods in real life. The aim of the project is exactly that outlined above – to drag CADD into the 21st century by allowing it to make use of modern high performance highly parallel computers and real time manipulation of the simulation scenarios via internet and modern graphical user interfaces.

We like to thank Jens Gimmler (Molcad GmbH) and Matthias Hennemann (Cepos InSilico) for computational assistance.

■ **brickmann@molcad.de**
■ **tim.clark@chemie.uni-erlangen.de**

References

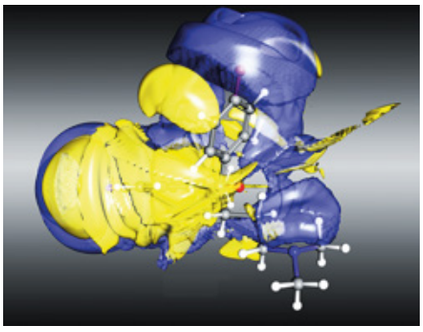
[1] Warren et al., J. Am. Chem. Soc., 2006, 49, 5912.



DO IT ONLINE

CADDLE.

SOFTWARE FOR MOLECULAR PROFESSIONALS



CADDLEs are a revolutionary new way to model drug molecules, their activity and ADME-properties. CADDLEs are completely web-based apps that require only a suitable browser on the client machine. They require absolutely no software installation apart from on the central server but nevertheless provide molecular modeling and visualization on the highest level. They can be used on any hardware with a suitable browser (desktop, tablet, smart phone).

Further details
www.ceposinsilico.com
www.molcad.de

Made by

CEPOS

MOLCAD