J Comput Aided Mol Des (2007) 21:591–601
DOI 10.1007/s10822-007-9142-y

Computer-aided drug design: the next 20 years

John H. Van Drie

Received: 27 September 2007 / Accepted: 18 October 2007 / Published online: 8 November 2007

Abstract This perspectives article has been taken from a talk the author gave at the symposium in honor of Yvonne C. Martin's retirement, held at the American Chemical Society spring meeting in Chicago on March 25, 2007. The talk was intended as a somewhat lighthearted attempt to gaze into the future; inevitably, in print, things will come across more seriously than was intended. As we all know—the past is rarely predictive of the future.

Keywords Thermodynamics · Pharmacophores · CoMFA · ADME models · Protein–protein interactions

Introduction

Despite the title of this article, the intent is not to actually make a prediction of the future 20 years out[1]—that is foolhardy. "Making predictions is hard, especially about the future" (a quote variously attributed to Yogi Berra, Niels Bohr, and others). The aim in this Perspectives is to provoke some thinking about what may lie ahead. Almost any time one looks backward 20 years or more, it is actually somewhat depressing to see how slowly we make progress. One of the most memorable conferences in our field was that organized by Peter Goodford in Erice, Sicily in 1989, on "3D Molecular Structure and Drug Action". Intriguingly, Peter summarized at the end of the conference his perspectives on where our field stood at the time, and where it needed to go (Fig. 1). This list is stunning to look at almost 20 years later, in that it looks so 'modern'. He highlights, for example:

- "we must deal properly with water"
- "we must remember that conformation depends on structure and environment"
- "we must combine theory and experiment"
- "we must predict solubility"
- "we must improve homology modeling."

In part, one must credit Peter's prescience in composing this list. One of the areas where significant progress has been made since 1989 is in homology modeling—the progress here has been stunning (e.g. [2]), building upon both our improvements in conformational analysis of proteins and on the rise of genomics data to assist obtaining excellent sequence alignments. But on topics like predicting solubility, we're not much farther along than we were in 1989.

The evolution of technology

In 1993, J. Bezdek proposed his now-famous curve [3] showing how most technologies progress (Fig. 2). Time is shown on the *x*-axis, and expectations are shown on the *y*-axis. Briefly, he proposed that all technologies initially begin with a great burst of enthusiasm—the "naïve euphoria" phase. At some point, a "peak of hype" is reached, and expectations begin to diminish. Expectations steadily fall, an overreaction to immature technology. That phase finally comes to an end, as the "depths of cynicism"

J. H. Van Drie (✉)
John H Van Drie Research LLC, 34 Stinson Rd.,
Andover, MA 01810, USA
e-mail: johnvandrie@mindspring.com
URL: www.vandrieresearch.com

[1] A 20 year timeframe was chosen to commemorate events which took place 20 years ago.

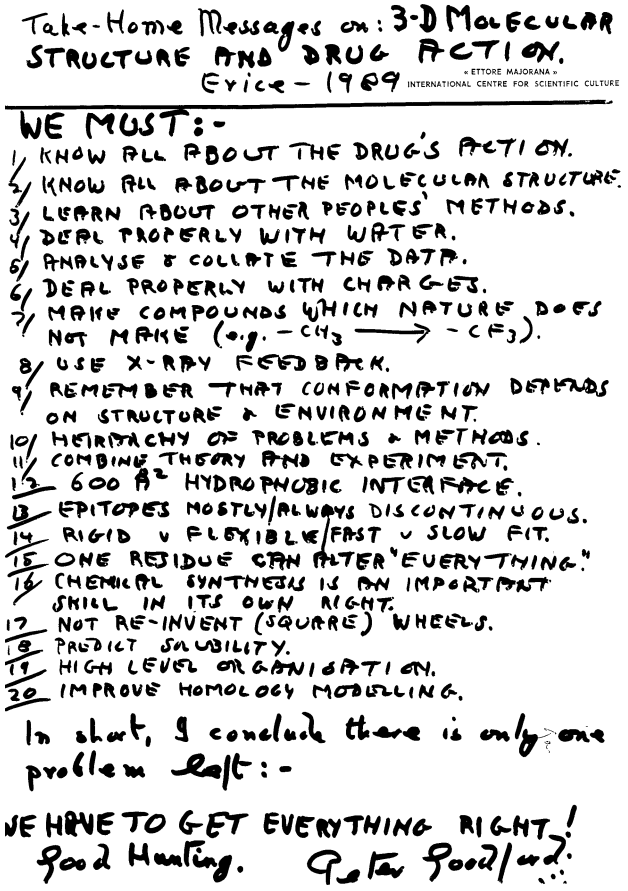
Take-Home Messages on: 3-D Molecular Structure and Drug Action.
« ETTORE MAJORANA » INTERNATIONAL CENTRE FOR SCIENTIFIC CULTURE
Erice – 1989
WE MUST:-
1/ KNOW ALL ABOUT THE DRUG'S ACTION.
2/ KNOW ALL ABOUT THE MOLECULAR STRUCTURE.
3/ LEARN ABOUT OTHER PEOPLES' METHODS.
4/ DEAL PROPERLY WITH WATER.
5/ ANALYSE & COLLATE THE DATA.
6/ DEAL PROPERLY WITH CHARGES.
7/ MAKE COMPOUNDS WHICH NATURE DOES NOT MAKE (e.g. $-CH_3 \longrightarrow -CF_3$).
8/ USE X-RAY FEEDBACK.
9/ REMEMBER THAT CONFORMATION DEPENDS ON STRUCTURE & ENVIRONMENT.
10/ HIERARCHY OF PROBLEMS & METHODS.
11/ COMBINE THEORY AND EXPERIMENT.
12/ 600 Å² HYDROPHOBIC INTERFACE.
13/ EPITOPES MOSTLY/ALWAYS DISCONTINUOUS.
14/ RIGID v FLEXIBLE/FAST v SLOW FIT.
15/ ONE RESIDUE CAN ALTER "EVERYTHING."
16/ CHEMICAL SYNTHESIS IS AN IMPORTANT SKILL IN ITS OWN RIGHT.
17/ NOT RE-INVENT (SQUARE) WHEELS.
18/ PREDICT SOLUBILITY.
19/ HIGH LEVEL ORGANISATION.
20/ IMPROVE HOMOLOGY MODELLING.
In short, I conclude there is only one problem left:-
WE HAVE TO GET EVERYTHING RIGHT!
Good Hunting. Peter Goodford

Fig. 1 Peter Goodford's concluding slide from his 1989 conference in Erice, Italy, "3D molecular structure and drug action". Note especially his final comments at the bottom "we have to get everything right"—the drug designer's lament. Used with kind permission of P. Goodford

are reached. According to Bezdek, only then do we see a steady march towards "true user benefits", with expectations approaching a positive plateau—the "asymptote of reality".

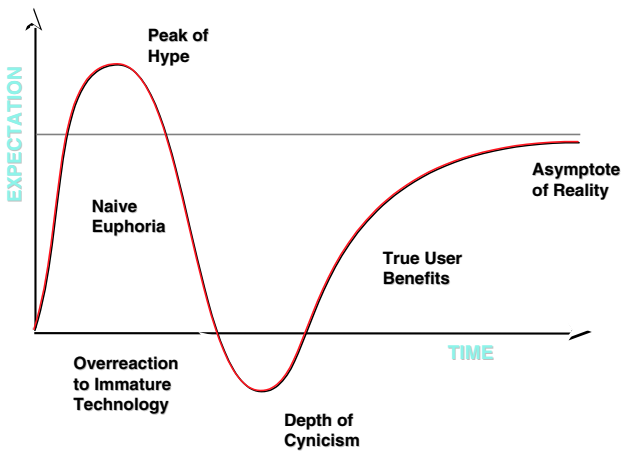

Fig. 2 Bezdek's 'evolution of technology' [2]. Image adapted from his original by J. D. Baker

One must take care not to take Fig. 2 too seriously, but nonetheless there is some wisdom contained in it. In particular, based on this author's observation and experience, it holds reasonably well for the evolution of CADD technology over the past 25 years; in fact, one can put dates on each of the key milestones (Fig. 3).

Naïve euphoria

There was an explosion of interest in the potential for CADD in the pharmaceutical industry in the early 1980s; some observers trace this to a cover article in *Fortune* magazine dated October 5, 1981, entitled "The Next Industrial Revolution: designing drugs by computer at Merck" (this was preceded by a scientific overview in *Science* [4]). Whether it was this particular event which was the fuse which lit the explosion of interest in CADD in the early 1980s may be debated, but beyond doubt the industry was in a naïve euphoria phase throughout the 1980s; enormous investment took place in the pharmaceutical industry in this era, with millions of dollars in investment in hardware and software and waves of hiring establishing groups of scientists specialized in this area.

Peak of hype

In the timeframe 1989–1991, a particular phrase became popular: "we can design drugs atom-by-atom" (out of respect for the promulgators of this, no reference will be made here, but the origins of this phrase can easily be retrieved via a Google search). Especially in hindsight, this attitude is risible, but, in 1989–1991, people actually bought this—quite a comment on the tenor of the times, stemming possibly from a mystical belief that computers could do anything. There were the beginnings of some real successes in the application of CADD, but these successes tended to get drowned out by these overripe claims.

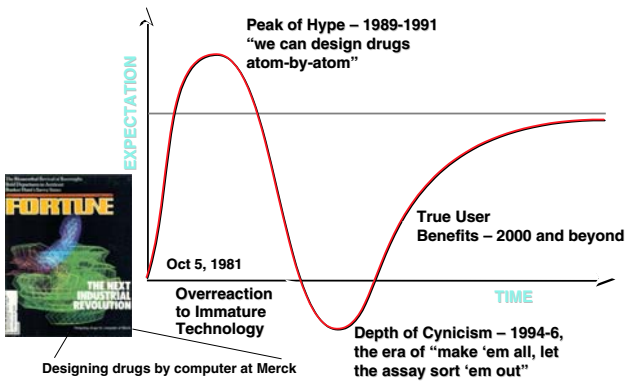

Fig. 3 In CADD, one can date each of the points on Bezdek's curve

Overreaction to immature technology, reaching the depths of despair

Two competing technologies began their own 'naïve euphoria' phase in the early to mid-1990s, puncturing the bubble of belief around computer-aided drug design. High-throughput screening and combinatorial chemistry painted a view of the world at the opposite extreme of designing drugs atom-by-atom—the new mantra was 'make billions of molecules with little or no design, and screen them all to find the right molecule'. In the era 1992–1995, hiring of CADD scientists in the pharmaceutical industry fell to almost nil, and the questions of the day became reminiscent of medieval clerics arguing about how many angels fit on the head of a pin: fierce arguments occurred over how large the total space of molecules was, what would be the ideal 'diverse' library to properly sample that space, etc.

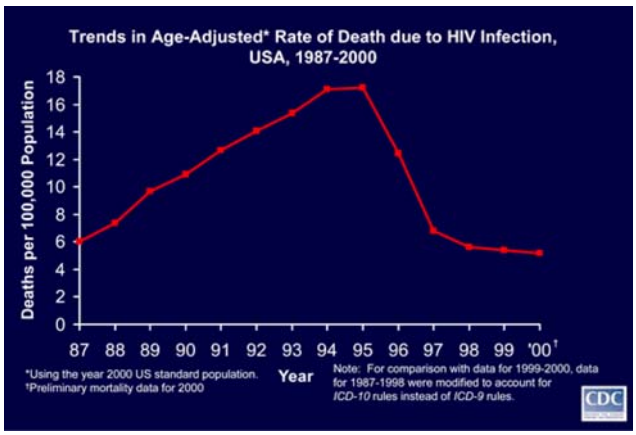

Fig. 4 Deaths due to AIDS in the US were rising rapidly until the introduction of HIV protease inhibitors and HIV reverse transcriptase inhibitors in the mid-1990s. This represents a triumph for the entire enterprise of pharmaceutical R&D, one in which CADD played an important role

True user benefits—approaching the asymptote of reality

By the late 1990s, hiring of CADD scientists in the pharmaceutical industry regained momentum, as people looked around and identified some of the real successes that had occurred during the previous 15 years. Structure-based design had played a key role in the discovery of the HIV protease inhibitors (ritonavir, indinavir, saquinavir) which hit the market in the mid-1990s, and played a key role in reversing an rapid growth in deaths due to AIDS in the US (Fig. 4). Virtual screening was beginning to show real promise; numerous discovery projects in the industry traced the roots of their series to a lead from virtual screening, especially in CNS targets and targets which had failed in high-throughput screening (e.g. phosphatases). Furthermore, the first drug candidate whose origins were in a pharmacophore-based virtual screening lead [5] was approved in 1999: tirofiban, a fibrinogen antagonist. The first report of a true success with de novo design was reported by scientists at Agouron in 1995 [6]. Chris Lipinski et al. introduced amazingly simple rules to guide one towards molecules with drug-like properties [7]. The expectations for CADD were now becoming more realistic, and the technology was maturing, and starting to deliver on the promise.

Yvonne Martin's experiences prior to 1980

Yvonne might look at Fig. 3, and observe that she'd been active for a dozen years prior to that already. In Fig. 5, the *x*-axis is backed up a couple of decades, to record the earlier but less prominent surge in activity in QSAR. The key events highlighted are:

- Early 1960s: Corwin Hansch devises QSAR, and discovers the surprisingly consistent relationship between a molecule's in vivo biological activity, and a simple physicochemical measure, the log of the octonol/water partition coefficient (logP)
- 1968: Yvonne Martin begins working with Corwin Hansch on QSAR
- 1975: The QSAR Gordon Conference is initiated
- 1978: Yvonne's seminal work on QSAR appears, *Quantitative Drug Design*

This earlier period is stressed, to show that the evolution of CADD has had at least two eras of growth.

The future (?)

The second reason for highlighting the earlier, smaller surge in interest in QSAR prior to the main technology evolution curve is to use this as a springboard for the fundamental prediction for the next 20 years of CADD (Fig. 6): *this has only been a warmup*! The transition from QSAR to more general molecular modeling led to a jump in expectations in what drug discovery scientists expected from the computational scientists. Similarly, it seems reasonable to anticipate another quantum leap in expectations in the near future, although this time it won't be triggered by science evolving on such a narrow front, but instead it will reflect a multi-dimensional evolution of the computational science and technology. As this author envisions it, the leap in expectations will be prompted by the evolution

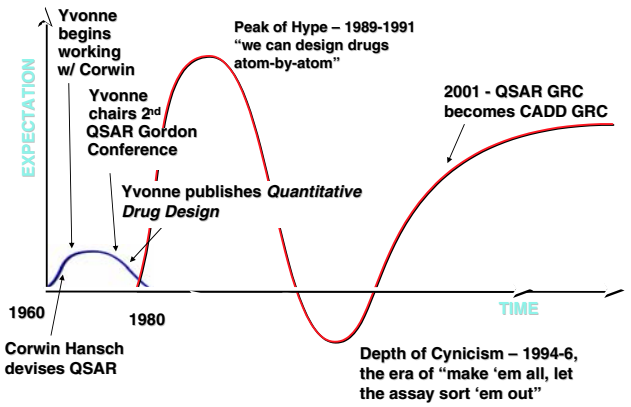

Fig. 5 Yvonne Martin and Corwin Hansch were at work on QSAR long before the curve in Fig. 3 starts

of the science along many lines, as will be elaborated momentarily.

It is important to realize that no specificity is being given, regarding *when* this leap in expectations will take place, and what specific event(s) will *trigger* this projected new phase of rapidly-rising expectations. Nonetheless, it is clear that the underlying forces driving this evolution reflect the growing impact of CADD in the academic and industrial research labs doing pharmaceutical discovery. In some senses, CADD is emerging as a distinct discipline in its own right, an off-shoot of computational chemistry, much as computational chemistry emerged in the 1960–1970s as a discipline branching off from its origins in physical chemistry. CADD distinguishes itself in one key regard relative to both physical and computational chemistry: while physical/computational chemistry is mainly focused on providing atomic/molecular-level explanations of chemical phenomena, CADD has in addition the aspiration to not merely explain, but also to answer the fundamental question of medicinal chemistry:

What molecule(s) should be made next?

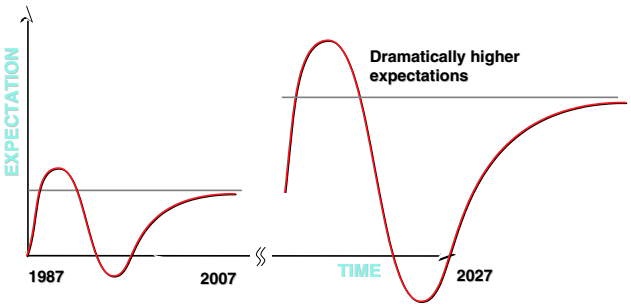

Fig. 6 The main projection to the future: what we've experienced thus far has only been a warmup. It is difficult to imagine precisely *when* the new wave will begin, nor what will be the stimulus to kick it off

In other words, CADD aims to make *prospective* predictions about what molecules should be synthesized, to achieve the desired set of biological properties (potency, etc.). This leads to computational methods like virtual screening, virtual library design, de novo design, all things integral to CADD but heresy to most academic computational chemists.

Another, more recent element of CADD also distinguishes itself from traditional academic computational chemistry, which arises from the appreciation that having a potent ligand is a necessary *but not sufficient* condition for having an effective drug. As E. H. Cordes has observed (personal communication) "it's relatively easy to discovery a potent ligand—it's damn tough to discover a drug". As will be explained momentarily, CADD is becoming increasingly involved in understanding the differences between potent molecules and drug candidates (ADMET—absorption, distribution, metabolism, excretion, toxicity)—this is already fueling a rise in impact and expectations for CADD.

Seven specific outlooks on the future of CADD

As this author looks across the landscape of CADD, evolution in seven areas appears poised to drive the future:

1. Computational thermodynamics will flower.
2. We'll learn to turn potent ligands into drug candidates: ADMET, etc.
3. We'll face new classes of drug targets which will challenge our competencies.
4. We'll encounter novel molecular mechanisms for drug efficacy, e.g. self-assembling drugs.
5. Rather than thinking about inhibiting a single target, we'll learn to model an entire signal transduction pathway, and use that understanding to better select drug targets.
6. Today's sophisticated CADD tools only in the hands of experts will be on the desktops of medicinal chemists tomorrow. The technology will disperse.
7. Virtual screening will become routine.

Let us consider each of these in turn.

Outlook 1: computational thermodynamics will flower

When molecular dynamics was first applied to macromolecules in the late 1970s [8], there was excitement at two levels: (1) computation now had the potential to display the conformational flexibility inherent in proteins, unlike the static pictures of X-ray crystallography, and (2) the

potential to calculate thermodynamic properties was inherent, as was later demonstrated with the evolution of free-energy perturbation (FEP) [9]. In general, as applied to drug discovery we've fallen short of the unrealistically high hopes created for molecular dynamics; in particular, while some practitioners of FEP methods claimed to be able to calculate experimental ΔG's of binding of ligands to proteins to within 1 or 2 kcal/mol, in actual application in the design of ligands for drug discovery this author is not aware of any example where a key new design idea emerged prospectively from the use of such methods. They have been reputed to be helpful in retrospective explanations.

However, lately a number of novel techniques for performing high-quality thermodynamic calculations on small-molecule/protein interactions have arisen in the arena of small start-up companies, which anecdotally do indeed appear to be having significant impact on molecular design. First among these was Locus Pharmaceuticals, whose scientific founder Guarnieri built upon the academic work of Gibbs' ensembles of Mezei [10] while at the Mt Sinai medical school and the Sarnoff Institute. The key breakthrough there was the recognition that molecular dynamics was insufficiently sampling thermodynamic phase space, and that a more exhaustive sampling of phase space was necessary to obtain good thermodynamic averages. Also, that group tended to focus on the application of these methods to very small, rigid molecules, in effect developing the computational equivalent of fragment-based screening or "SAR by NMR" [11] (the sampling problem should be less severe for conformationally-constrained molecules). Locus' success has spawned a number of imitators: BioLeap, Vitae, SolMap and Verseon. Unfortunately, very little is published by these groups, making it extremely difficult to get an accurate read on what progress they're making. This author was initially skeptical, but based on unpublished reports of their achievements, now tends to believe real progress is being made, in terms of developing novel designs for molecules, and reducing the number of iterations of ligand design/optimization in achieving the potency required for a drug candidate. The success of these efforts has been facilitated both novel developments in the computational science, as well as huge increases in computational horsepower, as such methods are exceptionally compute-intensive.

Another development where improved sampling appears to be leading to a better estimate of thermodynamic properties is through internal-coordinate conformational sampling [12]. For example, Jacobson et al. are using these methods to better understand the process of transport of a drug molecule across a membrane at the molecular level [13].

Another area where a recent breakthrough holds promise for computer-aided drug design is in the area of irreversible thermodynamics (the thermodynamics of processes like diffusion, versus equilibrium processes like ligand-protein binding). The fundamental science of irreversible thermodynamics has been available for decades (e.g. the work of Onsager, Prigogine, and others [14]), but it has largely remained inaccessible and resistant to practical applications. However, some recent work by Phillips and Dill [15] interpreting an inscrutable 30-year-old result of E. T. Jaynes, the "principle of maximum caliper", appears to hold great promise. Their applications thus far have only been to fluid flow in microfluidic devices, but the application to the dynamical processes inherent in ligand–protein interactions, phenomena involving transport across membranes, etc. appears to be on the horizon. This may be the theoretical formalism that we've been waiting for, able to achieve on the unfulfilled promise of molecular dynamics.

Outlook 2: we'll learn to turn potent ligands into drug candidates: ADMET, etc.

As enticing the prospect of accurately calculating ligand/protein ΔG's may be, it must be kept in mind that it is not enough to have a potent ligand—this is only one of many properties a successful drug candidate must have. The Rule-of-5 of Lipinski et al. [7] swept like wildfire through our business, in part due to its simplicity and intuitive appeal, and in part because people in the CADD community had been so blind to the molecular attributes conferring good pharmacokinetic properties. But the Rule-of-5 is only a start. As shown in Fig. 7 [16], the rate of attrition of drug candidates in the clinic is still extremely high—less than one molecule in 10 survives the process of initiation of a phase I trial through approval by the FDA. It is astonishing to realize that 90% of all the intellectual

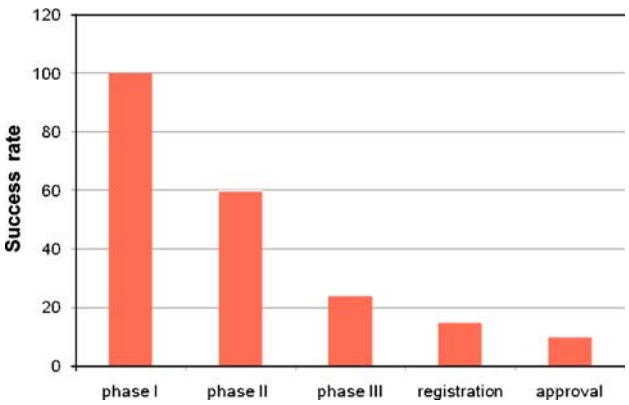

Fig. 7 The rates of attrition in the clinic are staggering—only one out of ten candidates entering the clinic make it to the market. Figure adapted from Ref. [16]. The grand challenge for CADD in the next 20 years is to improve the success rates of candidates in the clinic

effort and creativity that goes into the pharmaceutical research process, in effect, goes down the drain—a prodigious waste of scientific talent. The more we understand *why* molecules fail in the clinic, the more we'll be able to design molecules that *don't* fail. For this reason, it is this author's contention that

The grand challenge of drug design in the next 20 years will be to learn how to design drug candidates with a diminished likelihood of attrition in the clinic.

One can imagine innovations on a number of CADD fronts which will facilitate these advances in ADMET modeling (absorption, distribution, metabolism, excretion, toxicity):

Outlook 2.1: we'll have increased molecular understanding of the mediators of unwanted biological effects

The "long QT" syndrome is a drug-induced cardiac arrhythmia, mediated by a specific ion channel, the hERG channel [17]. A number of marketed drugs were withdrawn in the 1990s, due to an appreciation of those drugs' ability to induce the long QT syndrome. A group at Aventis [18] was able to construct a homology model of that receptor, and used it to explain the structure-activity relationship around hERG binding. Other cell-surface receptors are increasingly being appreciated as mediators of off-target effects. A natural role for CADD is to use models of these receptors, either homology models or pharmacophore models, to help design out such receptor-mediated off-target effects.

Outlook 2.2: our methods for computationally learning from data will improve

This is an easy prediction to make, given that our current techniques for learning from data, e.g. regression, recursive partitioning, partial least-squares, etc., perform so poorly. Techniques that perform admirably in other domains, e.g. image processing, simply do not function well in the world of structure–activity relationships. This author introduced the term "Kubinyi Paradox" to a behavior observed by a number of workers in the field, most prominently Hugo Kubinyi: the models that fit the best retrospectively tend to predict the *worst* prospectively [19]. There was a flurry of excitement at the start of this decade for a new machine learning technique, Support Vector Machines (SVMs) [20], which theoretically held the promise to handle the Kubinyi Paradox. Alas, the practical experience of SVMs in CADD applications has not lived up to those expectations. Nonetheless, computational learning techniques are a very fruitful area for future research, with immediate application to the problem of learning empirically how to design drug candidates devoid of attributes which may cause it to die in the clinic.

There is a tsunami of data already in the pharmaceutical companies of in vitro parameters correlating with in vivo ADME (e.g. cytochrome P450 binding data), with comparatively little in the way of CADD help in interpreting and understanding the data. This must and will change.

Outlook 2.3: we'll get much better at building empirical 3D models; something will emerge to replace CoMFA/ CoMSiA

CoMFA (Comparative Molecular Field Analysis) [21] is now over 20 years old, and is a standard technique for constructing 3D models in the absence of direct structural data of the target. One notable variation of that technique emerged about 10 years ago, CoMSiA (Comparative Molecular Similarity Analysis) [22], but with that exception there has been surprisingly little innovation in this area of empirical 3D model construction. The vicissitudes of CoMFA are well-acknowledged [23], and the need for empirical 3D models in the area of ADMET are strong—cases like hERG where one can readily construct homology models are the exception, not the rule. Just as the invention of CoMFA was catalyzed by the invention of partial-least-squares as a data analysis tool, advances in empirical 3D model construction may be tied to Outlook 2.2, an improvement in techniques for learning from data.

Outlook 2.4: use of pharmacophores will grow

Pharmacophore techniques are central to projects for which no direct experimental structural data is available, and where homology models are unreliable [24]. A pharmacophore is the spatial arrangement of functional groups on a ligand which is essential for biological activity—this may be used to construct overlays of molecules for visualization or for input to more refined procedures like CoMFA, or may be used to scan 3D databases to find novel scaffolds with similar biological activity. The simplest way of building empirical 3D models of off-target effect is by building a pharmacophore; the hERG model of Cavalli et al. [25] is such an example, though that particular model is less useful when the hERG homology model is available. But, referring to the tsunami of in vitro ADMET data mentioned a few paragraphs earlier, it would be a straightforward but tedious task for any CADD group to build pharmacophore models of each in vitro endpoint—the science is well-established (though admittedly what is

encoded today in most vendor software falls short of what one needs in this arena).

Outlook 2.5: we'll be able to find the data we need

This topic refers to the fact that our current chemical database systems are still woefully inadequate. When one compares the ease with which data on any topic can now be retrieved from home via simple Google searches, to the pain and agony one must undergo to retrieve important proprietary data tied to chemical structures at most pharmaceutical companies, one realizes that there are no technological limitations—things must get better. In fact, as one looks around, one sees increasingly that companies are building their own in-house custom software for chemical databases, despite the trend in all other areas of preferring buying over building (examples of which this author is aware include Lilly, Pfizer, Johnson and Johnson, Novartis, and Vertex). Chemical databases are prosaic, but very important: one key piece of overlooked data may be the datapoint that shoots down a cherished hypothesis; the right data will often be more persuasive than the most sophisticated models, theories, etc. It is a constant amazement to see how often a certain "lore" about a structure–activity relationship will be driving the analog design on a drug discovery project, only to have a new team member join the project, whose diligent searching of the chemical/biological database reveals that there are many examples where the lore has already been invalidated—this is a testament to the hurdles inherent in finding the right data.

Outlook 3: we'll face new classes of drug targets which will challenge our competencies

Through the 1970s, the majority of drug targets were cell-surface receptors (e.g. what became labeled in the 1980s as "G-protein-coupled receptors", or GPCRs); these targets were typically discovered empirically to be the mediator of the in vivo activity of a drug molecule. Beginning in the mid-1970s, enzymes were begun to be studied as drug targets. The first effort to bear fruit along these lines were the ACE inhibitors for hypertension (captopril, enalapril); indeed captopril is often considered the first drug whose design was facilitated by the presence of an X-ray structure [26]. However, captopril's origins were still the empirical observation that a polypeptide present in the venom of a snake resulted in a precipitous drop in blood pressure. The great promise of enzymes as drug targets, epitomized by the approach of Merck under the direction of R. Vagelos, was that biochemical knowledge of a metabolic pathway could be exploited to identify enzymes whose inhibition would lead to the desired therapeutic effect. This has been brilliantly exploited by the pursuit of the enzyme HMG-CoA, which led to the statins for control of cholesterol, one of which is the most successful drug in history: atorvastatin.

It has been an article of faith in the drug industry for the past 15 years that genomics in general and the Human Genome Project in particular would similarly revolutionize our outlook on drug targets. Thus far, the results from that has fallen short of expectations. Multiple interpretations can be given, regarding why this is so. One reason may be that we've all anticipated the Human Genome Project would tee up new targets that looked a lot like our old targets, allowing us to apply our tried-and-true approaches to drug discovery to this new targets. But at least two new classes have risen to prominence in the modern genomics era that are untraditional as targets, and hence extremely challenging to our competencies to prosecute them:

- Protein/protein interactions (PPIs) [27]
- Nucleic acid/protein interactions, e.g. transcription factors [28]

PPIs are often considered to be an impossible challenge for drug design, as the protein/protein interface is a surface often in excess of 1500 $Å^2$. Traditional high-throughput screening for PPI inhibitors frequently identifies no true positives in a sea of artifactual hits. Yet, progress is being made. The two paradigmatic examples of PPIs where progress is being made are the IAPs, proteins which regulate apoptosis (programmed cell death) [29], and p53/mdm2, a key protein pair involved in DNA repair and the checkpoint for the cell cycle progression [30]. In both cases, some 'out-of-the-box' thinking was necessary to make progress.

And, sadly, too often the CADD perspective on such target classes are that they are impossible, citing the thermodynamic impossibility of a small molecule disrupting a 1500 $Å^2$ interaction, etc. These targets are challenging our competencies, and we in CADD must rise to the challenge, e.g. to help identify allosteric sites away from the huge protein/protein interface which could serve to regulate that protein pair, or to devise novel virtual screening strategies to respond to the failure of HTS against this target class. PPIs aren't going away as a drug target class: if one calculates that there are approximately 30,000 genes encoding for proteins in the genome, there are potentially $30{,}000 \times 30{,}000 \sim 1$ billion protein pairs possible in the genome. In other words, once we start looking for PPIs as potential points of intervention, the odds of us stumbling across them are much higher than a traditional single-gene target.

Drugs regulating transcription are relatively uncommon (e.g. the 'glitazones' targeting the PPARs), yet our knowledge of biology would suggest that this is a natural, but extremely challenging, intervention point for therapeutics. We must start thinking more about nucleic acids in partnership with proteins as points of intervention. Until recently, we've tended only to look at the parts of the genome which code for proteins (which is only ~20% of the genome). Hence, the shock that accompanied the recent report by the ENCODE consortium, which studied 1% of the genome in great detail, and observed that 80% of the genome is indeed transcribed, suggesting that various forms of RNA play a much greater functional role that has been hitherto expected [31]. We know a lot about ligand/protein interactions, which we use to design drugs. We know comparatively little about ligand/RNA interactions—if one were asked to design a drug to specifically interact with a given RNA, one would hardly know where to begin. However, there has been some work along the lines of designing molecules to interact with a specific sequence of DNA [32]. This ENCODE result also suggests that we're still on the early part of the learning curve with the Human Genome, and maybe it was premature (naively euphoric?) to expect the consequences of that to pay out in drug discovery so quickly.

Outlook 4: we'll encounter novel molecular mechanisms for drug efficacy, e.g. self-assembling drugs

This next outlook is admittedly a wild idea, but it is one which has some precedents established already. If one needs to deliver a relatively large molecule, e.g. 1200 mol wt, to a target to properly modulate that target, it is typically thought that oral delivery is impossible, as the fundamental biophysics of absorption across the gut makes this a low likelihood (unless one gets lucky with active transport). But, yet, if one could orally deliver two smaller pieces, of 600 mol wt each, that spontaneously came together non-covalently at the site of action, one could achieve this goal. Obviously, this is a subtle, tricky thing, but at least two examples are already known, one a drug on the market, another a drug candidate undergoing clinical testing.

Exjade is a recently-approved drug for iron overload, which works by chelating the iron in the blood and dragging it out of solution [33]. As shown in Fig. 8, it works by a 2:1 stochiometry, with each Exjade molecule forming a bidentate interaction with the iron, resulting ultimately in a tetradentate complex.

More outlandish is the recent work of Stupp et al. [34], where small orally available molecules spontaneously form a cylindrical micelle around nerve fibers, to promote growth of damaged nerves. These molecules are being tested in diseases associated with damaged nerves. The stoichiometry here is thousands of molecules per nerve fiber, and the process is a subtle self-assembly at the site of action. These molecules were designed based on principles of self-complementarity, which like the Chinese yin/yang symbol, is a necessary property of such self-assembling molecules.

Fig. 8 Exjade s a new iron chelation therapy which binds to iron in a 2:1 stoichiometry. It is the simplest example of a 'self-assembling' drug

It remains to be shown whether these are two fortuitous examples, of whether this is a harbinger of a long-term trend.

Outlook 5: rather than thinking about inhibiting a single target, we'll learn to model an entire signal transduction pathway, and use that understanding to better select drug targets

It is the nature of our scientific enterprise to be 'reductionist', i.e. to understand something by breaking things down into steadily smaller pieces. Increasingly, people are realizing that something becomes lost in this process, that, for example, there may be aspects of cell function that are only apparent as one looks at how all the pieces work together as a whole. People have been promoting this idea for many years, but only recently is it becoming clear that there is a direct relevance to drug discovery. And it is this new relevance that will likely propel it forward into our consciousness.

For example, in the field of kinase drug discovery, three kinase targets were viewed as equally valid oncology targets, all residing along one signaling pathway, the MAP kinase pathway crucial in the control of cell division [35]:

$$\text{RAF} \rightarrow \text{MEK} \rightarrow \text{ERK}$$

In our standard linear thinking that RAF activates MEK, which in turn activates ERK, which in turn controls downstream transcriptional events, it was initially anticipated that the cellular phenotype of inhibiting RAF would be the same as the cellular phenotype of inhibiting MEK, which would be the same as that for inhibiting ERK. But,

the cellular phenotypes of each has been strikingly different, pointing to some being better drug targets than the others [36].

Studying an entire network of biological signaling molecules, rather than looking at one single enzyme, has become quite fashionable recently under the label 'systems biology' [37]. This has enormous potential to transform the field of CADD in the next 20 years, as people devise mathematical approaches to model the behavior of networks of proteins.

The most striking example of where this type of systems analysis has had an impact on drug discovery is the work of Perelson et al. in the mid-1990s, studying the dynamics of HIV infection [38]. Until that work appeared, it was a puzzle how that virus was able to wipe out the immune system of the host. With that analysis, an empirical model that looked at the dynamics of HIV infection in a human, it was then clear.

More recent work of Perelson represents a more typical approach to modeling biological networks [39]. A biological network is modeled as a set of coupled ordinary differential equations, with rate constants determined experimentally for each discrete step. This set of differential equations is solved numerically to make predictions about the behavior of the whole.

A more sophisticated approach is exemplified by the work of Aguda [40], where by modeling the network of reactions controlling the cell cycle he is able to identify regimes of parameter space which give rise to different types of cell cycle behavior. And, this type of modeling is not constrained to look simply at a network of enzymes; by modeling the behavior of a network of neurons, key insights emerged into the nature of Parkinson's disease, suggesting that the rhythms of firing were the key to understanding the pathology [41].

It is the nature of the modern scientific enterprise in general and biology in particular to reduce our understanding to ever smaller functioning units, i.e. to be 'reductionist'. But, in doing so, the 'emergent' behavior of how these units interact can become lost, interfering with our ability to understand how to modulate the system via small molecule drugs. This is a tremendous opportunity for those of us in the modeling field, to use mathematical modeling to achieve that type of understanding.

Outlook 6: today's sophisticated CADD tools only in the hands of experts will be on the desktops of medicinal chemists tomorrow. The technology will disperse

Twenty-five years ago, modelers worked with million-dollar room-sized computers with 3D display systems half the size of a refrigerator. Today, the computer which sits on my lap is far more powerful, both in computation speed and in 3D display capabilities. Twenty-five years ago, the software running on those computers was arcane, with incomprehensible user interfaces; much of the function of modelers in those days was to serve as a user-friendly interface to that software, and their assistance was often duly noted in manuscripts, if not as a co-author then as a footnote. Today, scientists of all backgrounds routinely festoon their publications with the output of molecular graphics software, running on their desktop/laptop machines with slick easy-to-use graphical user interfaces, e.g. Pymol [42].

This is a trend that will accelerate. Things that seem sophisticated and difficult-to-use, but are truly useful, will in 20 years be routinely available on desktop/laptop machines (and even laptops may be displaced by palmtops, multi-functional cellphones, etc.). Too many modelers are still in the business of being 'docking slaves' for their experimental collaborators (i.e. the experimentalist asks the modeler 'please dock my new idea for a molecule', and waits for the result to see if it confirms their design); this will ultimately disappear, as that type of routine task will be handled by more sophisticated user interfaces to current docking algorithms, e.g. the software from Molsoft is well on its way to fill such a role [43] (Fig. 9). Whereas the 'information retrieval specialists' that once populated corporate libraries have disappeared, replaced by desktop Google searches, this trend of modeling-to-the-desktop should not be a source of job insecurity for CADD scientists—this will free us up from the routine 'docking slave' tasks to focus our energies on higher-valued-added work.

As a rule, things today that seem finicky and fiddly to use (e.g. de novo design software), or things that take large amount of computer resources (e.g. thermodynamic calculations, or a docking run on the full corporate database) are things that one can easily imagine will in the future sit on the desktops of chemists, used by them with minimal intervention by CADD scientists.

Outlook 7: virtual screening will become routine

Virtual screening is the process of sifting through an electronic compound database for molecules which will be submitted for experimental testing. Such searches may be done with docking software, with pharmacophore searching software, or with simple similarity-searching software. Virtual screening contrasts with random, or high-throughput, screening (HTS), in which all molecules from a collection are submitted for experimental testing. It is a constant source of surprise to this author that, while the technology of virtual screening has become refined and

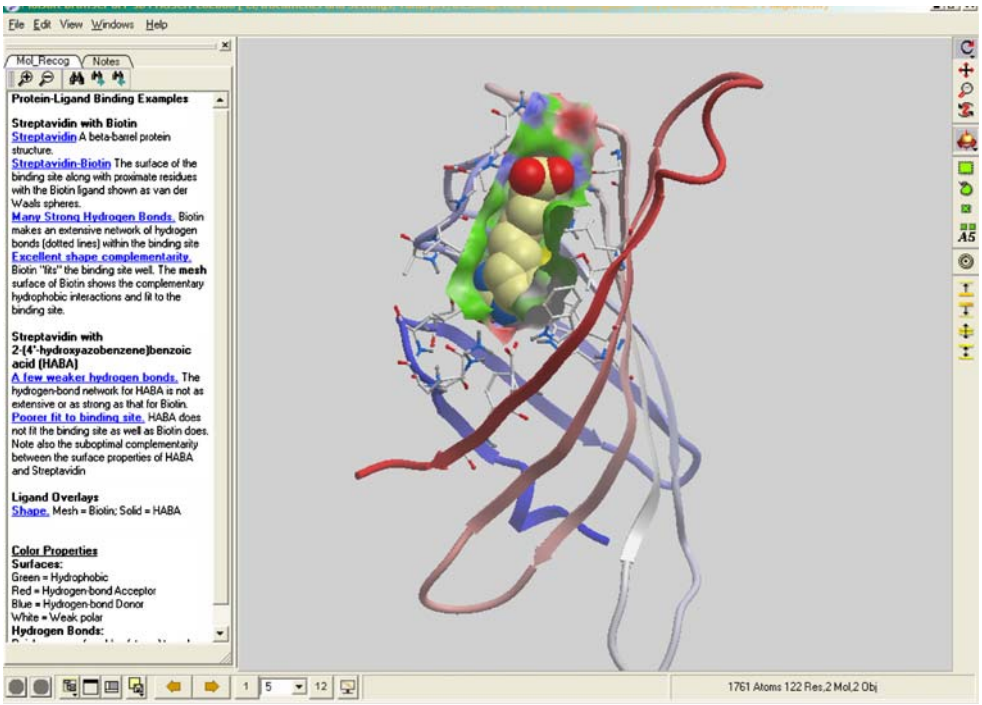

Fig. 9 The software from Molsoft provides an unusually simple interface, with customized hyperlinks (as on web pages) allowing the user to direct the actions of the software. Shown is the binding of biotin to streptavidin

straightforward to use, and furthermore that it has been proven to be useful beyond all doubt in a variety of settings [44], that it still occupies a small role in drug discovery in major pharmaceutical companies, certainly by comparison to HTS. As mentioned at the outset, the success of virtual screening was first demonstrated 20 years ago.

The obstacles to the broader adoption of virtual screening lie neither in the technology, nor the underlying science. One can only surmise that it reflects sociological issues, e.g. the inherent inertia of large organizations, or the relative power of the groups controlling high-throughput screening compared to that of the groups responsible for virtual screening (sometimes CADD groups, sometimes not). But, when looking out at a time horizon 20 years distant, it is easy to imagine that these obstacles will fade.

It is not widely appreciated that the aspect of virtual screening that makes it so compelling is the *speed* with which novel chemical starting points can be identified. When an X-ray structure is available for docking, or when SAR is available to develop a pharmacophore for pharmacophore-based 3D database searching, or when one active analog is available to seed a similarity search, one typically need only wait a few months until chemists can begin making analogs of active compounds. By contrast, with an HTS campaign, while the actual screening of plated compounds of a corporate collection takes only weeks with modern automation equipment, the great delay occurs in transforming a low-throughput assay into one formatted for high-throughput screening. HTS has the virtue of being exhaustive, while virtual screening has the virtue of speed. *Both* forms of screening are plagued by the potential for false positives and false negatives (the latter tends to be especially neglected when thinking about HTS, since it is rare that control experiments are done to highlight those). The costs associated with HTS (reagents, machines, etc.) are minimal compared to the costs associated with the delay of initiating chemistry on a drug discovery project. The fact that virtual screening can so quickly jump-start a project is the reason for the confidence in this prediction that the use of virtual screening will grow in the coming decades.

Summary and conclusion

Our general assertion is that, looking ahead 20 years into the future, the outlook for the growth of CADD is bright. Seven specific outlooks were presented which bolster that claim:

1. Computational thermodynamics will flower.
2. We'll learn to turn potent ligands into drug candidates: ADMET, etc.
3. We'll face new classes of drug targets which will challenge our competencies.
4. We'll encounter novel molecular mechanisms for drug efficacy, e.g. self-assembling drugs.

5. Rather than thinking about inhibiting a single target, we'll learn to model an entire signal transduction pathway, and use that understanding to better select drug targets.
6. Today's sophisticated CADD tools only in the hands of experts will be on the desktops of medicinal chemists tomorrow. The technology will disperse.
7. Virtual screening will become routine.

Implicit in all of these outlooks is that the driving force for the evolution of CADD is our close association with the actual practice of drug discovery in general, and medicinal chemistry in particular. The need for high-quality drug candidates has never been higher, and the pressure to discover those drug candidates is already high, and growing. This presents a historic opportunity for our field—the advances that have been foretold above will not happen from simple momentum and proceeding down our current paths, putting one foot in front of the other. This future will only be realized if we as a scientific discipline are able to boldly rise to the challenges, by discovering novel science, by developing novel technology, and by more successfully applying that science and technology to drug discovery.

References

1. Martin YC (1992) J Med Chem 35:2145; Van Drie JH, Weininger D, Martin YC (1989) J Comput Aided Mol Des 3:255; Kebabian JW et al (1990) Am J Hypertens 3:40S (the events described therein took place in the summer of 1987)
2. Rong SB et al (2002) J Med Chem 45:4140. Compare that prediction with the experimental structure published later, Davis MI (2005) Proc Natl Acad Sci USA 102:5981
3. Bezdek J (1993) IEEE Fuzzy Syst 1:1
4. Gund P et al (1980) Science 208:1425
5. Hartman GD et al (1992) J Med Chem, 35:4640
6. Gehlhaar DK et al (1995) J Med Chem 38:466
7. Lipinski CA et al (1998) Adv Drug Deliv Rev, 23:3
8. McCammon JA, Gelin BR, Karplus M (1977) Nature 267:585
9. McCammon JA (1987) Science 238:486; van Gunsteren WF, Berendsen HJ (1987) J Comput Aided Mol Des 1:171
10. Guarnieri F, Mezei M (1996) J Am Chem Soc 118:8493
11. Shuker SB et al (1996) Science 274:1531
12. Abagyan R et al (1997) Proteins Suppl 1:29; Jacobson MP et al (2004) Proteins 55:351
13. Rezai T et al (2006) J Am Chem Soc 128:14073
14. deGroot SR, Mazur P (1962) Non-equilibrium thermodynamics. North Holland Pub. Co., Amsterdam
15. Dill K et al (2006) Am J Phys 74:123
16. Kola I, Landis J (2004) Nat Rev Drug Discov 3:711
17. Sanguinetti MC, Tristani-Firouzi M (2006) Nature 440:463
18. Pearlstein R et al (2003) Bioorg Med Chem Lett 13:1829–1835
19. Van Drie JH (2003) In: Bultinck P, De Winter H, Langenaeker W, Tollenaere J (eds) Computational medicinal chemistry and drug discovery. Dekker, New York, pp 437–460
20. Warmuth MK et al (2003) J Chem Inf Comput Sci 43:667
21. Cramer RD, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959
22. Klebe G et al (1994) J Med Chem 37:4130
23. Cramer RD, Wendt B (2007) J Comput Aided Mol Des 21:23
24. Van Drie JH (2003) Curr Pharm Des 19:1649
25. Cavalli A et al (2002) J Med Chem 45:3844
26. Ondetti MA, Cushman DW, Rubin B (1983) Captopril. In: Bindra JS, Lednicer D (eds) Chronicles of drug discovery. Wiley, NY, pp 1–31
27. Keskin O et al (2007) Curr Top Med Chem 7:943
28. Welsh SJ et al (2006) Semin Oncol 33:486
29. Sharma SK, Straub C, Zawel L (2006) Int J Pept Res Ther 12:21
30. Tovar C et al (2006) Proc Natl Acad Sci USA 103:1888
31. ENCODE Project Consortium (2007) Nature 447:799
32. Viger A, Dervan PB (2006) Bioorg Med Chem 14:8539
33. Kontoghiorghes GJ (2006) Expert Opin Emerg Drugs 11:1
34. Silva GA et al (2004) Science 303:1352
35. Peyssonnaux CC (2001) Biol Cell 93:53
36. Ohren JF et al (2004) Nat Struct Mol Biol 11:1192
37. Kirschner MW (2005) Cell 121:503
38. Perelson AS et al (1996) Science 271:1582
39. Dahari H et al (2007) J Theor Biol 247:371
40. Aguda BD (1999) Proc Natl Acad Sci USA 96:11352
41. Terman D et al (2002) J Neurosci 22:2963
42. http://www.pymol.org
43. Abagyan R (2006) Trends Biochem Sci 31:76
44. Alvarez J, Shoichet B (2005) Virtual screening. CRC Press, Boca Raton