Modelling and Simulation: A Computational Perspective in Anticancer Drug Discovery

Federico Gago*

Departamento de Farmacologia, Facultad de Medicina, Universidad de Alcala, E-28871 Alcala de Henares, Madrid, Spain

Abstract: The availability of high-quality molecular graphics tools in the public domain is changing the way macromolecular structure is perceived by researchers, educators and students alike. Computational methods have become increasingly important in a number of areas such as comparative or homology modelling, functional site location, characterization of ligand-binding sites in proteins, docking of small molecules into protein binding sites, protein-protein docking, and molecular dynamics simulations. The results obtained yield information that sometimes is beyond current experimental possibilities and can be used to guide and improve a vast array of experiments. On the basis of our improved level of understanding of molecular recognition and the widespread availability of target structures, it is reasonable to assume that computational methods will continue aiding not only in the design and interpretation of hypothesis-driven experiments in the field of cancer research but also in the rapid generation of new hypotheses.

INTRODUCTION

Cancer is usually viewed as a result of cumulative genetic damage in susceptible cells that makes them escape the normal constraints to unscheduled proliferation. The ever increasing list of both gain-of-function mutations that convert protooncogenes to oncogenes and loss-of-function mutations that inactivate tumour-suppressor genes [1] is providing a wealth of putative targets for anticancer drug discovery other than the DNA molecule itself, which has been one of the prototypical targets for decades [2-4]. Genes "simply" code for protein sequences but the functionality emerges through many protein-protein, protein-DNA, protein-RNA, and protein-ligand interactions in both cells and whole organisms. These interactions are in many ways akin to computations (but alas using an operating system – the 'logic of life' [5] – that is far from being completely understood), and are organized in networks or pathways [6, 7], some of which are recurrently found mutated in human tumours. Interestingly, in tumour cells harbouring a number of cancer-causing mutations, it has been shown that correction of even a single defect can impair cellular proliferation and/or induce apoptosis [8]. Conceptually, this improved knowledge brings us closer to the paradigm that the same abnormalities that lead to the emergence of cancer cells may also eventually serve, through properly targeted pharmacological intervention, as the basis for their demise. To increase the therapeutic index of novel anticancer agents, however, it is crucial to identify both the changes that are continuously required for maintenance of the transformed phenotype [9, 10] and one or more target gene products that may not be essential in normal cells but are essential for survival in cancer cells that lack the function of some other gene(s) [8]. Experiments with genetically modified animals can help in delineating the true association between gene expression and disease [11, 12], as well as in validating a given gene product as a therapeutic target. The phenotype a given gene produces in a cell or organism depends not only on the biochemical function of the protein it encodes but also on the cellular function this protein performs as part of an assemblage or complex with other molecules. For this reason, one of the first steps in the process of converting sequence information (as provided by the genome projects) into functional information is the interpretation of how amino acid (and less often nucleotide) sequences dictate the three-dimensional structures into which they fold.

The availability of high-quality molecular visualization programs from the public domain (e.g. RasMol [13], Swiss-PdbViewer - a.k.a. DeepView - [14], ViewerLite [15], VMD [16], etc), and also of extremely effective plug-ins (e.g. Chime [17]) that can be installed on one's favourite navigator for free, is changing the way molecular structure is perceived by researchers, educators and students alike. Molecular graphics and computer simulations are frequently used today to guide and improve a vast array of experiments, to interpret and present some of the results, to pose novel relevant questions, and to extract structural and energetic information that is usually beyond current experimental possibilities [18]. Apart from being an integral part of X-ray crystallography and NMR spectroscopy methods that are commonly used in structure determination and structure-based screening [19, 20], a variety of computational methodologies have become increasingly important in a number of areas such as (i) comparative or homology modelling [21, 22], (ii) functional site location and characterization of ligand-binding sites in proteins [23], (iii) docking of small molecules into protein binding sites [24, 25], protein-protein docking [26, 27], and (v) molecular dynamics (MD) simulations (Fig. **1**).

*Address correspondence to this author at the Departamento de Farmacologia, Facultad de Medicina, Universidad de Alcala, E-28871 Alcala de Henares, Madrid, Spain; Tel: (+34) 918 854 514; Fax: (+34) 918 854 591; E-mail: federico.gago@uah.es

1568-0118/04 $45.00+.00

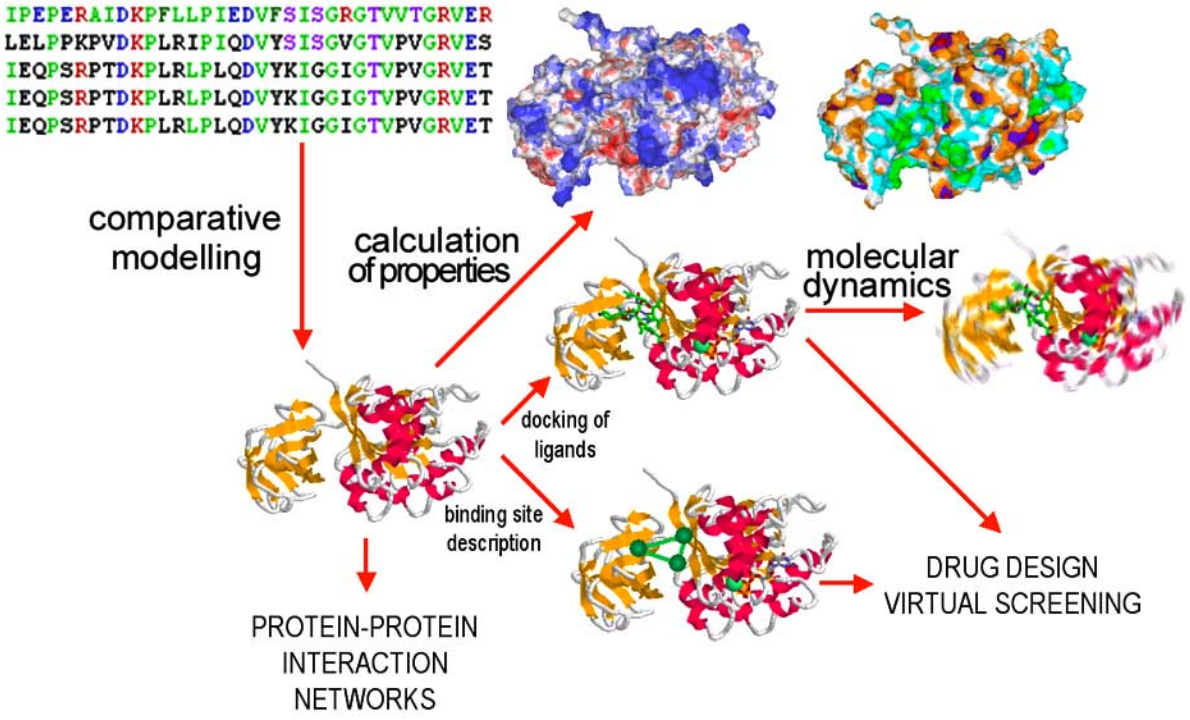

Fig. (1). Pictorial representation of major uses of computational applications in protein target building and characterization, ligand and protein docking, and molecular dynamics simulations that can be of value in the design of novel anticancer drugs.

The initial feeling that theoretical calculations were a waste of time [28] (not least because of the near impossibility of treating such complex systems as macromolecules in living cells with the level of detail that was thought to be necessary to add something of importance to our knowledge) is giving way to a renewed wave of confidence in simulation methodologies. Ongoing improvements in molecular mechanics force fields [29, 30], the almost systematic incorporation of solvent molecules and counterions into the models [31], and a reliable treatment of long-range electrostatics using particle mesh Ewald methods [32] are making it possible to simulate the dynamics of relatively large systems (including such highly charged ones as nucleic acids [33, 34]) over considerably long times, sometimes approaching the microsecond time scale [35], without the instabilities that plagued earlier simulations. Moreover, replacement of explicit solvent by the much faster Generalized Born (GB) implicit model [36] in some applications is extending the length of these MD simulations even further. These methodologies are now well suited to study conformational changes that are coupled to function [37] and also, in conjunction with quantum mechanical methods, the course of enzymatic reactions [38]. A number of tools ('targeted' or 'steered' [39, 40] MD) also exist that can 'pump' some extra energy into the system to accelerate the crossing of local barriers thus allowing the study of processes that would normally occur too slowly in a standard MD run based only on random thermal fluctuations. The repertoire of macromolecules relevant to cancer that can be studied by these methods has also expanded enormously in the last 20 years, from just a handful of 'classical' targets (e.g. short DNA oligonucleotides, dihydrofolate reductase or thymidylate synthase) to DNA-protein complexes [41], RNA [42], and whole protein families (e.g. growth factor receptors [43], kinases [44], and phosphatases [45]), both alone and in complex with ligands.

Researchers arguably have on their laptops today more computing power than was available on most mainframe computers when MD simulations of macromolecular systems started at the end of the 1970s [46]. Moreover, a realistic alternative to large-scale 'supercomputers' these days is to use massively grid-distributed computing [47]. Thus, the power of dozens or hundreds of inexpensive personal computers (PCs) can be harnessed toward a common goal such as the folding of a protein [48] or the docking of potential ligands into a protein binding site. An extension of this concept to literally millions of PCs linked on the Internet using their idle time is the CAN-DDO cancer screening project [49] sponsored by the National Foundation for Cancer Research (NFCR) and scientifically coordinated from the NFCR Centre for Computational Design [50]. The software currently used [51] builds 3D coordinates for drug-like derivatives of both commercially available molecules and compounds from combinatorial chemistry libraries, generates conformers, docks into selected protein targets, and ranks the hits using a scoring function [52]. The aim of this virtual screening (VS) approach, which represents a typical compromise between speed and accuracy, is to

eliminate inactive compounds from testing and increase the chemical diversity of putative hits [53].

The obvious need for experiment and serendipity in the search for new anticancer agents will continue in the near future. Nonetheless, based on the current level of understanding of molecular recognition and the widespread availability of target structures, it is reasonable to assume that computational methods will become even more useful. Due to their fastness, they will aid not only in the design and interpretation of hypothesis-driven experiments but also in the rapid generation of new hypotheses. This latter approach, sometimes derisively called 'ignorance-based' [54], can also be of value: in the words of Nobel Prize winner Linus Pauling, to have a good idea you need to have lots of ideas.

REFERENCES

[1] The Cancer Genome Project, The Wellcome Trust Sanger Institute. http://www.sanger.ac.uk/genetics/CGP/

[2] Hurley, L. H.; Boyd, F. L. *Trends Pharmacol. Sci*., **1988**, *9*, 402–407.

[3] Hurley, L. H. *J. Med. Chem*., **1989**, *32*, 2027–2033.

[4] Hurley, L. H. *Nat. Rev. Cancer*, **2002**, *2*, 188–200.

[5] Noble, D.; Levin, J.; Scott, W. *Drug Discov. Today*, **1999**, *4*, 10–16.

[6] Bader, G. D.; Betel, D.; Hogue, C. W. *Nucleic Acids Res*., **2003**, *31*, 248–250 (http://www.blueprint.org/bind/bind.php).

[7] Babu1, M. M.; Luscombe, N. M.; Aravind, L.; Gerstein, M.; Teichmann, S. A. *Curr. Opin. Struct. Biol*., **2004**, *14*, 283–291.

[8] Reddy, A.; Kaelin, W. G. *Curr. Opin. Pharmacol*., **2002**, *2*, 366–373.

[9] Scherf, U.; Ross, D. T.; Waltham, M.; Smith, L. H.; Lee, J. K.; Tanabe, L.; Kohn, K. W.; Reinhold, W. C.; Myers, T. G.; Andrews, D. T.; Scudiero, D. A.; Eisen, M. B.; Sausville, E. A.; Pommier, Y.; Botstein, D.; Brown, P. O.; Weinstein, J. N. *Nat. Genet*., **2000**, *24*, 236–244.

[10] Hughes, T. R.; Marton, M. J.; Jones, A. R.; Roberts, C. J.; Stoughton, R.; Armour, C. D.; Bennett, H. A.; Coffey, E.; Dai, H.; He, Y. D.; Kidd, M. J.; King, A. M.; Meyer, M. R.; Slade, D.; Lum, P. Y.; Stepaniants, S. B.; Shoemaker, D. D.; Gachotte, D.; Chakraburtty, K.; Simon, J.; Bard, M.; Friend, S. H. *Cell*, **2000**, *102*, 109–126.

[11] Zambrowicz, B. P.; Sands, A. T. *Nat. Rev. Drug Discov*., **2003**, *2*, 38–51.

[12] Zambrowicz, B. P.; Turner, C. A.; Sands, A. T. *Curr. Opin. Pharmacol*., **2003**, *3*, 1–8.

[13] http://www.umass.edu/microbio/rasmol/

[14] http://www.expasy.org/spdbv/

[15] http://www.accelrys.com/viewer/

[16] http://www.ks.uiuc.edu/Research/vmd/

[17] http://www.mdlchime.com/chime/

[18] Kazlauskas R. *Science*, **2001**, *293*, 2277–2279.

[19] van Dongen, M.; Weigelt, J.; Uppenberg, J.; Schultz, J.; Wikström, M. *Drug Discov. Today*, **2002**, *7*, 471-478.

[20] Coles, M.; Heller, M.; Kessler, H. *Drug Discov. Today*, **2003**, *8*, 803–810.

[21] Sali, A.; Kuriyan, J. *Trends Cell Biol.*, **1999**, *9*, M20–M24.

[22] Marti-Renom, M. A., Stuart, A. C.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. *Ann. Rev. Biophys*., **2000**, *29*, 291–325.

[23] Campbell, S. J.; Gold, N. D.; Jackson, R. M.; Westheady, D. R. *Curr. Opin. Struct. Biol*., **2003**, *13*, 389–395.

[24] Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput. Aid. Mol. Des*., **2002**, *16*, 151–166.

[25] Glen, R. C.; Allen, S. C. *Curr. Med. Chem*., **2003**, *10*, 763–777.

[26] Smith, G. R.; Sternberg, M. J. E. *Curr. Opin. Struct. Biol*., **2002**, *12*, 28–35.

[27] Wodak, S. J.; Méndez, R. *Curr. Opin. Struct. Biol*., **2004**, *14*, 242–249.

[28] Karplus, M. *Acc. Chem. Res*., **2002**, *35*, 321–322.

[29] Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn*., **1999**, *16*, 845–862.

[30] MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem*., **2000**, *21*, 105–120.

[31] Cheatham, T. E.; Kollman, P. A. *Annu. Rev. Phys. Chem*., **2000**, *51*, 435–471.

[32] Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct*., **1999**, *28*, 155–179.

[33] Giudice, E.; Lavery, R. *Acc. Chem. Res*., **2002**, 35, 350–357.

[34] Cheatham, T. E. *Curr. Opin. Struct. Biol*., **2004**, *14*, 360–367.

[35] Duan, Y.; Kollman, P. A. *Science*, **1998**, *282*, 740–744.

[36] Bashford, D.; Case, D. *Annu. Rev. Phys. Chem*., **2000**, *51*, 129–152.

[37] Young, M. A.; Gonfloni, S.; Superti-Furga, G.; Roux, B.; Kurigan, J. *Cell*, **2001**, *105*, 115–126.

[38] Mendieta, J.; Martín-Santamaría, S.; Priego, E.-M.; Balzarini, J.; Camarasa, M.-J.; Pérez-Pérez, M.-J.; Gago, F. *Biochemistry*, **2004**, *43*, 405–414.

[39] Isralewitz, B.; Gao, M.; Schulten, K. *Curr. Opin. Struct. Biol*., **2001**, *11*, 224–230.

[40] Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. USA*, **1997**, *94*, 11905–11910.

[41] Marco, E. ; García-Nieto, R.; Gago, F. *J. Mol. Biol*., **2003**, *328*, 9–32.

[42] Zaman, G. J. R.; Michiels, P. J.; van Boeckel, C. A. *Drug Discov. Today*, **2003**, *8*, 297–306.

[43] Hunter, T. *Cell*, **2000**, *100*, 113–127.

[44] Hubbard, S. R. *Curr. Opin. Struct. Biol*., **2002**, *12*, 735–741.

[45] Klumpp, S.; Krieglstein, J. *Curr. Opin. Pharmacol*., **2002**, *2*, 458–462.

[46] McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature*, **1977**, *267*, 585–590.

[47] Chien, A.; Foster, I.; Goddette, D. *Drug Discov. Today*, **2002**, *15*, S176–S180.

[48] Folding@home project: http://www.stanford.edu/group/pandegroup/folding/

[49] http://www.chem.ox.ac.uk

[50] http://www.chem.ox.ac.uk/ccdd/ccdd.html

[51] Davies, E.K.; Glick, M.; Harrison, K.N.; Richards, W.G. *J. Comput. Chem*., **2002**, *23*, 1544–1550.

[52] Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. *J. Comput.-Aided Mol. Des*., **1997**, *11*, 425-445.

[53] Davies, E. K.; Richards, W. G. *Drug Discov. Today*, **2002**, *7*, S99–S103.

[54] Lazo, J.S.; Ducruet, A.P.; Koldamova, R. P. *Mol. Pharmacol*., **2003**, *64*, 199–201.