International Journal of Pharmacy Research and Technology
2012, Volume 2, Issue 2, 01-05.
ISSN 2250 – 0944 (Online)
ISSN 2250 – 1150 (Print)

Review Article

Druggability of Genome

H. B. Patel*, S. K. Mody, A. B. Chukewar, C. M. Modi, G. B. Dudhatra, Avinash Kumar and Ratn DeepSingh
Department of Pharmacology & Toxicology, College of Veterinary Science & Animal Husbandry, Sardarkrushinagar Dantiwada Agricultural University, Sardarkrushinagar-385506, Gujarat, India
***Corresponding author:** E-mail: drhitesh2002@rediffmail.com*
Received: 21/2/2012, Revised: 19/05/2012 Accepted: 25/05/2012

ABSTRACT
The process of drug discovery in pre-genomic era was considered as very complex process requiring great deal of time and money with high degree of unpredictability. The genomic research revolution has brought strategic shift in the concept of drug discovery and development with availability of tool for early identification of molecular targets based on human genomics. This approach has led to reduced rate of compound attrition and rejection, making process of drug discovery more viable economically and to some extent predictable with respect to time frame. The criteria of compound which make the compound fit for passing the process of drug discovery are studied and gene based targets are indentified, validated and interpreted. This approach no doubt is at its infantile stage but has tremendous future potential to play key role in drug discovery process. The present review describes the interrelationship between the druggability of genome and its impact on drug discovery process.

Keywords: Druggability, Drug Discovery, Drug Targets, Genome

INTRODUCTION

In the beginning of 20th century, the pharmaceutical industry witnessed that too many compounds were terminated in clinical development because of unsatisfactory pharmacokinetic (PK) profile. The scenario is somewhat still continues causing economical loss due to unreasonable rate of drug failure. Medicinal chemists need to address these problems during lead optimization and therefore novel tools are in urgent need to assess the relationship between structure of compound and pharmacokinetic prospect. Search for an understanding of what is responsible for compound attrition has led to the development of criteria which are characteristic for compounds that successfully pass through the development process [1]. Such compounds have been called 'druggable' or 'drug-like' [2, 3].

The introduction of Lipinski's 'Rule of Five' (RO5) has initiated a profound shift in the thinking paradigm of medicinal chemists. Understanding the difference between biologically active small molecules and drugs became a priority in the drug discovery process, and the importance of addressing pharmacokinetic properties early during lead optimization is a clear result. These concepts of 'drug-likeness' and 'druggability' have been extended to proteins and genes for target identification and selection [24, 25]. How should these concepts be integrated practically into the drug discovery process? This review summarizes the recent advancement in the field and examines the usefulness of druggability of compound and genome and its prospects in drug discovery and development process.

Whole sequence of the human genome with high-quality annotation is available; it is a good time to have a fresh look at the druggable genome. Analysis suggests a druggable gene count can be between 2000 and 3000, in general agreement with previous estimates (3000). However, there is evidence for a significant shift in the contribution of the major target families, with fewer rhodopsin-like GPCR (G-protein Coupled Receptor) and protein kinases, and more proteases than expected. The statistics of molecular drug targets and percentage of different molecular targets of current drugs are shown in Table 1 and Figure 1 respectively.

What is druggability?

The druggability is defined as a presence of protein folds that favour interactions with drug-like chemical compounds [4]. An extension of Lipinski's 'Rule of Five' to protein targets that can bind, such 'drug-like' compounds and therefore are thought to be amenable to modulation by compounds with oral bioavailability has led to the terms like 'druggable protein' [2] and 'druggable genome' [4].

Proteins lacking these structural features might have interesting biological properties, but are unlikely to be readily amenable to pharmaceutical modulation. The problem can be visualised as doors and keyholes. A door (protein) might control access to an interesting pathway, but if it does not have the appropriate keyhole (druggable domain) it cannot be opened. Analysis of druggability, so to speak, is the analysis of keyholes [5].

Table: 1. Molecular drug targets [26, 27]

Class of drug target	No. of molecular targets
Targets of approved NMEs (Human and anti-infectives)	301
Human Targets of approved NMEs	238
Human Targets of approved NCEs	170
Targets of approved biological	59
Human Targets of approved antibodies	15

HISTORY

In 2002, Hopkins and Groom introduced the concept of the 'druggable genome' [4]. Their purpose was to identify the limited set of molecular targets for which commercially viable, oral compounds can be developed. Because such targets are expected to bind RO5-compliant compounds, they analyzed databases and used

computational methods to identify all proteins belonging to families which have at least one member that has successfully been targeted by drug-like molecules. Assuming that druggability is shared among protein families and taking into account that, by necessity, a drug target needs to have the potential to be disease-modifying, they estimated that the human genome encodes 600–1500 targets for small-molecule drugs [5].

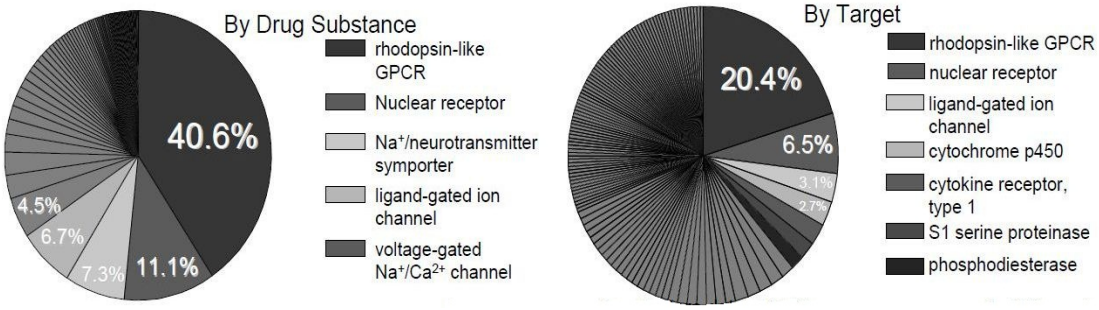

Figure: 1. Molecular target of current drugs [26, 27]

Features of druggability

(1) It is also important to keep in mind the distinction between druggability of a protein, and its actual qualities as a drug target [4].
(2) Many proteins are druggable according to their structure, but modulating their biological function may not provide any therapeutic benefit; not every door with a keyhole leads to a desirable place.
(3) The actual drug targets are the subset of druggable proteins which possess both structural and functional features of druggability, and their only real validation comes with successful clinical use.
(4) The numbers of protein target for commercially viable compounds are considerably smaller than originally thought.
(5) The size of the Pharmaceutically Tractable Genome, a term that was suggested for genes that can be targeted by small molecular weight compounds, antibodies and recombinant therapeutic proteins [6] is considerably larger.
(6) Multimeric protein complexes and successful promiscuous drugs acting on more than one target do not allow to immediately generalizing from gene to drug target. When discussing the merits of a druggable genome annotation, it is very important to keep all these limitations in mind.

Steps in druggability

A first step towards a more reliable way to assess the druggability of individual proteins is the identification of binding sites for drug-like molecules [7]. Kellenberger and co-workers have created an annotated database of ligand binding sites extracted from several experimental structure databases [8]. Such data can be used to derive rules or training sets for the computational identification of binding pockets. Many algorithms [9] are available for this purpose and in general they have been successfully applied in identifying true ligand binding pockets on the surface of proteins [8].

The second step of quantitatively assessing the druggability of the identified pockets is more challenging. An obvious approach is to screen a large library of drug-like compounds pharmaceutical compound collection or a chemical genetics library [10] and assess the resulting hits. Unfortunately, this approach has three significant drawbacks: it is very expensive, applied rather late in the drug discovery process and it produces a large number of false positives and promiscuous hits which complicate the analysis.

Another method has recently been developed by researchers from Abbott. Using 2-D heteronuclear-NMR they studied the interactions of 10000 lead-like or fragment- like compounds with protein surfaces. This approach has the advantage that it samples a large fraction of chemical space (even though the size of the library is small) and yields more reliable data than conventional high-throughput screening. Furthermore, such an NMR based analysis of druggability could be performed with limited resources and relatively early in the drug discovery process. Most importantly, an analysis of the NMR data has led to the development of 'druggability-indices' that can be used for the computational assessment of proteins with known structure [7].

Table: 2. Druggable genome predictions [26, 27]

Druggability prediction method	No. of molecular targets
Targets of approved NCEs	170
Sequence homology to NCE drug targets	945
Targets of chemical leads with activities (binding affinities) below 10 µM	707
Targets of Ro5 chemical leads with activities (binding affinities <= 10 µM)	587
Sequence homology to targets with chemical leads	2921
Feature-based druggability sequence probability prediction	2325
Structured-based prediction	427
Sequence homology to proteins predicted druggable by structure-based method	3541
Predicted Druggbable Genome (high confidence)	3505
Human Genome	23000

How precisely can we define the druggable genome?

Over time, the estimated numbers of potential targets are now converging at around 3000 druggable loci [4, 11, 12].

Main parameters influencing the count are the coverage of sequence databases, the tools used for sequence annotation, structural information about tractable folds, bioinformatics tools, and biological information about protein function. One key parameter, the genome sequence itself, has stabilised now. While the first working draft sequence covered only 90% of the genome, the preliminary final assembly covers 99% of the euchromatic genome in high quality sequence, with only 341 gaps remaining. Thus, current estimates should only miss 1% of potential targets due to lack of database representation. The current status of druggable genome prediction is depicted in Table 2 and Figure 2, 3.

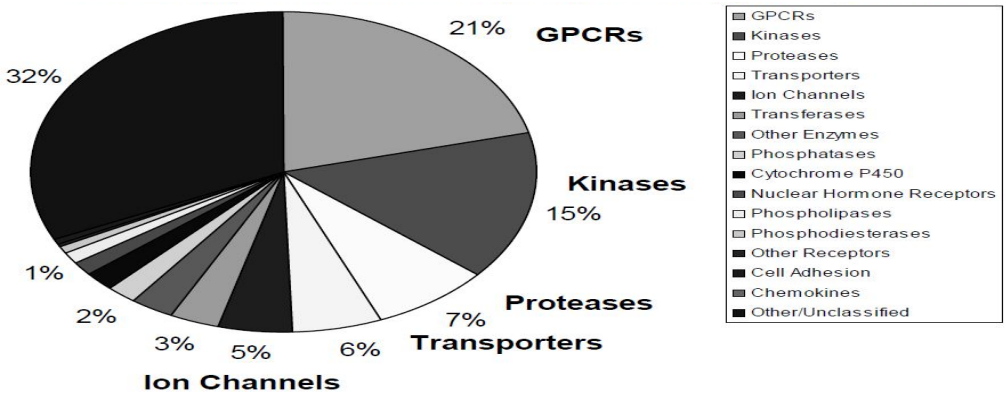

Figure: 2. Gene family distributions of predicted druggable genome [27]

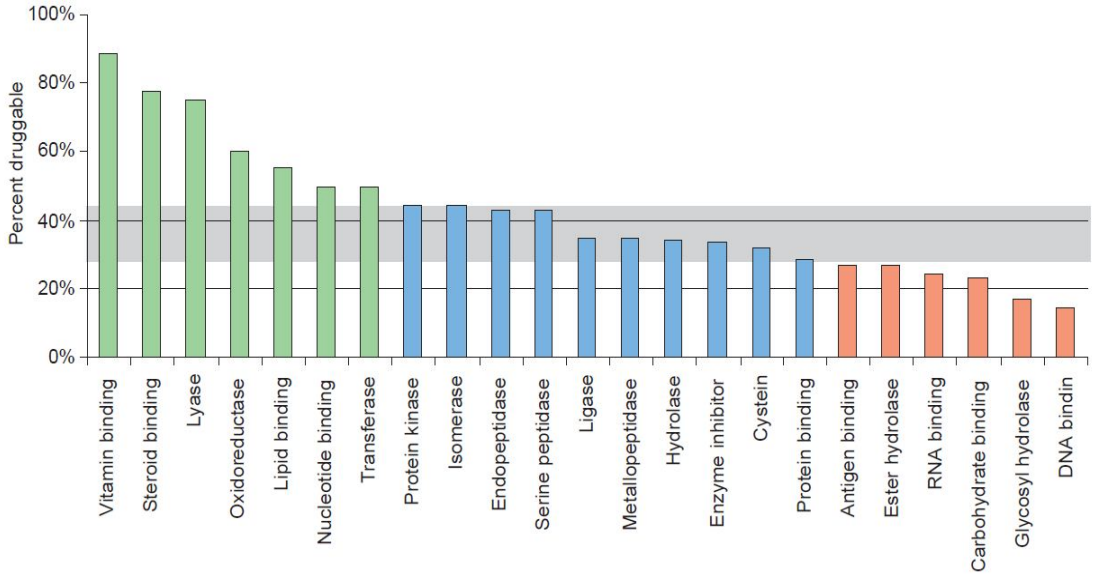

Figure: 3. Predicting protein druggability [7]

The results of druggability indices derived from 1096 Non-redundant human proteins, in which the percentage of members of selected target classes that contain a druggable binding site ('percent druggable') is plotted. In this preliminary analysis, no attempt was made to differentiate the known active or ligand-binding site, nor was conformational flexibility taken into account. Overall, 35% of the targets in this set contained at least one druggable binding site. Target classes depicted in green have a higher than average percentage of members with druggable binding sites, whereas those depicted in red have a lower than average percentage. The horizontal gray bar represents the average and standard deviation for the entire dataset

Is the power shifting between the ruling families?

The two largest families, protein kinases and Rhodopsin-like GPCR, are still topping the league table, but are smaller than expected. IPRO and PFAM predictions are very close to put the count for kinases at 480-500, down more than 20% from earlier predictions and in close agreement with the detailed annotation of the "kinome" (518) [13]. Table 3 is showing comparison of the druggable genomes of selected eukaryotes.

Obtaining an exact count of rhodopsin-like GPCR drug targets is cumbersome, as hundreds of closely related sensory receptors inflate the automated domain count, but are not likely to have any therapeutic potential. Near about three hundred non-sensory members of the family have been identified, this is in agreement with the IUPHAR receptor database [14] and other recent publications [15]. Both predicting under 300 druggable members of this family.

Table: 3. Comparison of the druggable genomes of selected eukaryotes [4]

	Homo sapiens	***Drosophila melanogaster***	***Caenorhabditis elegans***	***Saccharomyces cerevisiae***
Total number of predicted genes	~30,000	13,601	18,424	6,241
Number of proteins in proteome	21,688	13,849	17,946	6,127
Number of estimated druggable targets	3,051	1,714	2,267	508
Percentage that are predicted druggable targets	~10–14%	12%	12%	8%

These receptors, very many of them linked to disease-relevant physiology, are still most likely to be the richest opportunity for drug discovery in the short- to mid-term, but the lower number suggests that the area might soon be saturated by drug discovery efforts. The counts for non-rhodopsin GPCR, ion channels, transporters, nuclear hormone receptors and cytochromes are largely unchanged, indicating that these smaller families are annotated in great detail already. The most important area that consistently appears larger than previously expected is the proteases. Even the most stringent of counts (PFAM-CCDS), which is likely to be an underestimate, yields ~230 putative proteins. Applying the IPRO models increases the count to ~380. The higher numbers are consistent with recent curated counts [16]. Which predict 553 proteases and related proteins (although they include additional domain signatures that are not defined as druggable). It appears that the druggable protease space might be at least as large as the rhodopsin-like receptors. The heterogeneous group of druggable enzyme families seems to be of similar size as previously predicted, but due to its complexity a detailed validation of this observation is very tedious. If it is assumed that the numbers for kinases, GPCR and smaller target families will remain stable in the optimistic scenario, the larger number of proteases compensates for the reduction of other families, and arrives at just over 3000 targets, the same total as Hopkins and Groom in 2002. The conservative count uses the validated protease subfamilies and high-stringency predictions for enzymes and other target classes only, yielding a total of around 2200 druggable genes.

The available evidence suggests that qualitative druggability arguments are useful strategic tools; however, more accurate, quantitative assessments are needed, especially for proteins with borderline druggability. Moreover, the breakthroughs in small molecule disruption of protein–protein interactions (*e.g.* p53–MDM2 [19] and Bcl-2–Bax [20]) might lead to further expansion of the 'druggable' genome. Since its discovery, nuclear magnetic resonance (NMR) has become the single most powerful form of spectroscopy in both chemistry and structural biology. The dramatic technical advances over the past 10–15 years, which continue apace, have markedly increased the range of applications for NMR in the study of protein–ligand interactions. These form the basis for its most exciting uses in the drug discovery process, which range from the simple identification of whether a compound (or a component of a mixture) binds to a given protein, through to the determination of the full three-dimensional structure of the complex, with all the information this yields for structure-based drug design. This is a very attractive approach that can be used in the early stages of drug discovery and provides a solid basis for computational druggability estimations [21, 22, 23].

CONCLUSIONS

What do these numbers tell us, and how stable can we expect them to be? As discussed above, this approach to annotating the druggable genome estimates the potential maximum size of the playing field for current small molecule drug design. It does not consider biological (where the rules are harder to define), RNA (which is not based on protein structure), or future breakthroughs in medicinal chemistry or biology (no crystal ball available).

It is to be keep in mind that these simple sequence comparisons do not allow any immediate far-reaching conclusions about the biological function of a protein. However, sequence-based protein structure prediction certainly pinpoints areas of research that will be highly enriched for "real" drug targets. Investigating the function of orphan receptors has been very successful in identifying unexpected novel players in important physiological pathways [17, 18] and the druggable genome will further guide experimental efforts to understand the biology of other potential drug targets.

The RO5 and its extensions have been useful tools to generate awareness about the importance of PK parameters for development. In addition, this concept has led to the realization that there may be whole families of proteins for which it is extremely challenging to design compounds with good oral bioavailability.

REFERENCES

1. C A Lipinski, Filtering in drug discovery, Ann. Rep. Comp. Chem. 1, 155-168 (2005).
2. A L Hopkins and C R Groom, Target analysis: a priori assessment of druggability, Ernst. Schering Res. Found Workshop. 42, 11-17 (2003).
3. A P Orth, S Batalov, M Perrone, S K Chanda, The promise of genomics to identify novel therapeutic targets, Expert Opin. Ther. Targets. 8 (6), 587-596 (2004).
4. A L Hopkins and C R Groom, The druggable genome, Nat. Rev. Drug Discov. 1 (9), 727-730 (2002).
5. G Muller, Medicinal chemistry of target family-directed master keys, Drug Discov. Today. 8 (15), 681-691 (2003).
6. R Morphy, C Kay and Z Rankovic, From magic bullets to designed multiple ligands, Drug Discov. Today. 9 (15), 641-651 (2004).
7. P J Hajduk, J R Huth, and C Tse, Predicting protein druggability, Drug Discov. Today. 10 (23-24) 1675-1682 (2005).
8. E Kellenberger, P Muller, C Schalon, G Bret, N Foata and D Rognan, Sc-PDB: an annotated database of druggable binding sites from the protein data bank, J. Chem. Inf. Model. 46, 717-727 (2006).
9. J An, M Totrov and R Abagyan, Comprehensive identification of druggable protein ligand binding sites, Genome Inform. Ser. Workshop Genome Inform. 15, 31-41 (2004).

10. F Darvas, G Dorman, L G Puskas, A Bucsai and L Urge, Chemical genomics for fast and integrated target identification and lead optimization, Med. Chem. Res. 13, 643-659 (2004).
11. J Drews and S Ryser, The role of innovation in drug development, Nat. Biotechnol. 15 (13), 1318-1319 (1997).
12. J Drews, Drug discovery: a historical perspective, Science. 287 (5460), 1960-1964 (2000).
13. G Manning, D B Whyte, R Martinez, T Hunter and S Sudarsanam, The protein kinase complement of the human genome, Science. 298 (5600), 1912-1934 (2002).
14. S M Foord, T I Bonner, R R Neubig, *et al*., International Union of Pharmacology. XLVI. G Protein-Coupled Receptor List, Pharmacol. Rev. 57, 279-288 (2005).
15. R Fredriksson, M C Lagerstorm, L G Lundin and H B Schioth, The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints, Mol. Pharmacol. 63 (6), 1256-1272 (2003).
16. X S Puente, L M Sanchez, C M Overall, C Lopez-Otin, Human and mouse proteases: a comparative genomic approach, Nat. Rev. Genet. 4 (7), 544-558 (2003).
17. S B Seminara, S Massager, E E Chatzidaki, *et al*., The GPR54 gene as a regulator of puberty, N. Engl. J. Med. 349 (17), 1614-1627 (2003).
18. S Messager, E E Chatzidaki, D Ma, *et al*., Kisspeptin directly stimulates gonadotropin-releasing hormone release via G protein-coupled receptor 54, Proc. Natl. Acad. Sci. USA. 102 (5), 1761-1766 (2005).
19. L T Vassilev, *et al., In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2, Science. 303, 844-848 (2004).
20. T Oltersdorf, *et al.,* An inhibitor of Bcl-2 family proteins induces regression of solid tumours, Nature. 435, 677-681 (2005).
21. P J Hajduk, J R Huth and S W Fesik, Druggability indices for protein targets derived from NMR-based screening data, J. Med. Chem. 48, 2518-2525 (2005).
22. G C K Roberts, Applications of NMR in drug discovery, Drug Discovery Today. 5 (6), 230-240 (2000).
23. M Betz, K Saxena and H Schwalbe, Biomolecular NMR: a chaperone to drug discovery, Current Opinion in Chemical Biology. 10, 219-225 (2006).
24. T H Keller, A Pichota and Z Yin, A practical view of druggability, Current Opinion in Chemical Biology. 10, 357-361 (2006).
25. C A Lipinski, F Lombardo, B W Dominy and P J Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 46, 3-26 (2001).
26. J P Overington, B Al-Lazikani and A L Hopkins, How many drug targets are there? Nature Reviews. 5, 993-996 (2006).
27. A L Hopkins, The druggable genome: how do we deliver on the promise of new drug targets? An International perspective on pharmacogenetics: the intersection between innovation, regulation and health delivery, Rome, Italy (2005).