ISSN NO. 2320-5407

*Journal homepage: http://www.journalijar.com*

*INTERNATIONAL JOURNAL OF ADVANCED RESEARCH*

**REVIEW ARTICLE**

# *In silico* approaches to drug discovery –A review

**Rama Adiga and Indrani Karunasagar**
Faculty of Biomedical Sciences, Nitte University Centre for Science Education and Research, NUCSER, Nitte University, Mangalore

## *Manuscript Info*

***Manuscript History:***

Received: 14 May 2015
Final Accepted: 22 June 2015
Published Online: July 2015

***Key words:***

drug discovery, lead, optimization, toxicity

***Corresponding Author***

**Rama Adiga**

## *Abstract*

The study of the origin and mechanisms of disease in human and other vertebrates has been the main reason leading to the discovery and development of new drugs. Drug candidates sometimes fail at late stages of clinical trials resulting in wasteful expenditure for pharmaceutical companies. Other major public health concerns are side-effects of drugs or adverse drug reactions which cause the withdrawal of the drug even after they have reached the market. Drug guidelines for potential risks like carcinogenicity, genotoxicity and reproductive toxicity are useful in assessing the risk profile of any drug in the development stage. Safety guidelines issued by the Committee for Medicinal Products for Human Use (CHMP) have also stressed on reduction of animals for drug testing. It emphasizes the use of *in silico* approaches as part of a tiered non-clinical testing. Some of the *in silico* methods for determining the toxicity of drugs in drug discovery are briefly reviewed.



## INTRODUCTION

Drugs induce perturbations to biological systems which encompass various molecular interactions such as protein-protein interactions, metabolic pathways and signal transduction pathways, leading to the observed side-effects. Non-specific interaction with non-target molecules with varying affinities causes unfavourable interaction observed as side effects.

Screening procedures involve several steps to test for toxicity of new chemicals which have the potential to be formulated as a drug. Physicochemical profiling form an essential part of the drug discovery process generating drug lead compounds. The use of *in silico* estimation of toxicity helps remove costly animal testing and speed up the drug discovery process.

### Chemicals as drugs

Drugs are chemicals which interact specifically in the body bringing about a specific effect. The targets for drugs are proteins (mainly enzymes, receptors and transport proteins) and nucleic acids (DNA and RNA). Drugs react with the binding site and become permanently attached via a covalent bond that has bond strength of 200–400 kJ mol. However, most drugs interact through weaker forms of interaction known as intermolecular bonds. In the body, the drug has to travel through an aqueous environment in order to reach its target. To interact favourably, the water molecules surrounding the drug and the target binding site have to be removed. This requires energy and if the energy required to desolvate is greater than the energy gained by the binding interactions, then the drug may be ineffective. Pharmacodynamics is the study of how drugs interact with their targets through binding interactions (Otagiri et al,1999). Pharmacokinetics is the kinetics of how the drug is Absorbed, Distributed, Metabolized and Excreted (ADME).

Xenobiotics cause many types of toxicity by varied mechanisms. Some chemicals which are toxic are the undegraded "parent" compounds. Others must be metabolized before they cause toxicity. Many xenobiotics often

affect only specific target organs. Others, however, on contact can damage any cell or tissue. The target organs affected may vary depending on route of exposure and dosage.
Toxic effects are generally classified according to the site of the toxic effect. In some cases, the effect may occur at only one site or specific target organ. In other cases, toxic effects may occur at multiple sites. This is referred as systemic toxicity. Some of the types of systemic toxicity are acute toxicity, chronic toxicity, developmental toxicity, genetic toxicity (somatic cells) and carcinogenicity. The scientific and ethical committee recommends predictive toxicity testing which represents a challenge to those who would like to see animal testing replaced by non-animal tests and testing strategies. The usefulness of animal studies for predicting long-term target-organ and target-system effects in humans is limited by inter-species differences as is in the case of carcinogenicity testing and reproductive toxicity testing.

**Target Identification and Validation**

Identification of disease associated pathways furthermore allows to link pathway-specific biomarkers and drug targets.A validated target should have a clear biological function, have an essential role for the growth or survival of the organism, should be expressed during the relevant life stages , be druggable and should have the possibility of being screened in a biochemical or cellular assay.
Common target validation methods are RNA interference, antisense RNA, and antibody-mediated inhibition experiments. Alternatively, the validation of a drug target is performed using chemical compounds. In such cases, experimental compounds with well-understood modes of action are tested directly and screened. With positive results, it is inferred that the phenotypic effect is due to the interaction of the chemical compound with its known target.
Data mining refers to the use of bioinformatics approach especially to extract or filter valuable targets cited in literature database or other biomedical data such as microarray data have given a boost to target identification and discovery (Yang et al, 2009). No single data mining approach is sufficient for understanding the cellular mechanisms and reconstructing the biological networks. Some of the search engines include Chilibot, Textpresso, and PreBIND. Literature searches on model organism and subject-specific articles are available, in collaboration with databases like Flybase, Wormbase and The Arabidopsis Information Resource (TAIR) (Cohen et al,2008). Normalization of gene and protein names mentioned in biomedical texts has become an important step in many datamining pipelines, which is being done by the Gnat library (Hakenberg et al, 2011).
Proteomic data mining are emerging datamining approaches in the post-genomic era. High-throughput mass spectrometry analysis has emerged and its data mining would help to analyze and extract information from massive dataset. A study of disease-related networks by Krauthammer et al., 2004 was developed into a mining tool called GeneWays (Krauthammer et al., 2004). It searches by automatically extracting, analyzing, visualizing and integrating molecular pathway data from full-text research articles and predicting the physical interactions among candidate disease genes and understanding cell functions.
Pathway approach: A novel therapeutic target in various human cancers. This conserved signaling pathway precisely regulates self-renewal and terminal differentiation in embryonic development, but is typically silenced in adult tissues, with reactivation usually only during tissue repair. For example, in malaria, a shift from blood stage to liver stage for high-throughput phenotypic drug screens where a number of genes and proteins are expressed only during the liver stage,, thus represent a likely stage-specific drug target (Jamal et al,2013).
Molecular imaging has increasingly been used in drug development. The identification of imaging targets usually cell surface or membrane-bound proteins,help in selecting candidates for drug development and understanding of drug activity and disease ( Pospisil et al,2006; Willmann et al, 2008).
The data are collected from a variety of sources which include publications and patent information, gene expression data, proteomics data, transgenic phenotyping and compound profiling data. Identification approaches also include examiningmRNA/protein levels to determine whether they are expressed in disease and if they are correlated with disease exacerbation or progression. Biological targets have to be prioritized by the uniqueness of their roles and integrated and analyzed across many different disciplines to retrieve meaningful targets (Yang et al, 2009). Defects in expression or function or deregulation of apoptosis may cause diseases, such as cancers, immune diseases, and neurodegenerative disorders (Jiang et al, 2004). Apoptotic proteins leading to programmed cell death involved in the pathogenesis or progression of several diseases, including their biochemical mechanisms and three-dimensional structures, have provided a wealth of information in drug discovery. One such database is the Gene Expression Omnibus (NCBI) (http://www.ncbi.nlm.nih.gov/geo/which provide extensive datasets for expression levels in diseased statesfor analysis.

**Prioritizing genomic drug targets**

The identified drug target must fulfill a variety of criteria to allow the progression to the next stage. The drug target must be essential to the pathogen's growth. The homology between target and host must be low or nonexistent, which would prevent host toxicity. The activity of the target in the diseased conditions must not be compromised with minimal or no interaction with the host and host flora but high binding specificity for the pathogen. Targets with high sequence similarity to its host and host flora would be eliminated from the list (Hasan et al, 2006). The prioritized genes in application to the problem of TB drug development is listed in Table 1. The selection criteria can be found easily by querying publicly available bioinformatics resources, including metabolic pathway databases such as KEGG (Kyoto encyclopedia of genes and genomes), protein classification sets such as COGs (clusters of orthologous groups) , and databases of "druggable" proteins. A recent assessment system is the Druggable Protein-protein Interaction Assessment System' (Dr. PIAS)(Sugaya et al, 2011). Information on tertiary structures, drugs and chemicals, and biological function associated with PPIs retrieved from public databases are stored as a database. Select candidates of druggable PPIs that are more promising as drug targets in a research area (e.g., disease, pathway, or protein family) are available at http://www.drpias.net.

**Hit Discovery**

Hit is the starting point for small molecule drug discovery. A hit is usually defined as a molecule which binds to the target, which has been identified to be important in the disease of interest. Until recently, hits were identified by randomly screening millions of compounds against a target. In practice, this has yielded fewer hits than expected so a new tailored approach has emerged (Hughes et al, 2011). A workflow depicting the stages of generation of lead compound to the final consumer is shown in Fig1.

The compound library can then be screened to provide hits. Each research site has the infrastructure to quickly perform medium throughput screens (MTS, approximately 40,000 compounds) and UCB has a global HTS facility (100,000 - 1,000,000 compounds).

**Identify Leads**

Hit to lead molecule generation involves

1. Knowledge of 3D structure of target site
2. Calculating the target affinity for the hit (ligand)
3. Virtual screening of the hits

The screening comprises docking small molecule ligand (hit) available from a library or *de novo* construction of ligand which can bind to active site of target. In this way after a series of iterations the molecules which bind better, subsequently become 'hits'. In stage two docking studies are used to check for binding specificity which become suitable to be passed on to stage three (Pitt et al,2013). Stage three include ADMET profiling to be labeled as 'leads' (Moroy et al, 2012). ACD/ADME Suite from ACD/Labs (http://www.acdlabs.com/products/pc_admet/adme/adme/) is a software for the prediction of ADME properties from chemical structure which provides predictions relating to the pharmacokinetic profiling of compounds and predicts ADME properties, P-glycoprotein specificity, oral bioavailability, passive absorption, blood brain barrier permeation, distribution and P450 inhibitors.

The identification of small molecule modulators of protein and the process of transforming these into high-content lead series are key activities in modern drug discovery.

**1. Strategies for Hit and Lead generation**

In situations where the specific enzymes responsible for the disease are unknown, it is essential to investigate complex phenotypes in living cells. Phenotypic screens and target-based screens are generally employed for the purpose.The former also called *in vivo* screens in live cells or intact organisms, looks at the disease-modifying effects, or phenotypes, that compounds induce in cells, tissues or whole organisms. This would help find out the target or targets that are being "hit" by the candidate molecule, causing the effect.The target-based screens measures the effect of compounds on a purified target protein via *in vitro* assays where an enzymatic or binding reaction with purified protein - is assayed (Burbaum et al, 1997,Crews and Splittgerber, 1999, Yeh and Crews, 2003, Boppana et al, 2009). The response of an intact organism to a drug is often dependent on interactions between various cell types and tissues that are not possible to predict based on the results of a pure protein high-throughput screen. Phenotype-based

compound screening was used in zebrafish models which has the potential for identifying drug leads. Phenotypic high-throughput screens have identified compounds that are active in various cell culture-based or small animal models of cardiovascular disease.

In situations where the target protein is known, knowledge-based screening or focused screening is employed. It involves selecting from the chemical library smaller subsets of molecules that are likely to have activity at the target protein based on knowledge of the target protein eg. literature or patent precedents for the chemical classes likely to have activity at the drug target (Boppana et al, 2009). This type of knowledge has given rise, to early discovery paradigms using pharmacophores and molecular modeling to conduct virtual screens of compound databases (McInnes, 2007). Pharmacophore modeling is a powerful means to generate and use 3D information to search for novel active compounds, particularly when no receptor geometry is available. Currently, various automated pharmacophore generators have been developed, including commercially available software HipHop (Barnum et al, 1996), HypoGen (Li et al, 1999) (Accelrys Inc., http://www.accelrys.com), DISCO (Martin, 2000), GASP (Jones and Willet, 2000), GALAHAD (Tripos Inc., http://www.tripos.com), PHASE (Dixon et al, 2006)(Schro¨dinger Inc., http://www.schrodinger.com) and MOE (Chemical Computing Group, http://www.chemcomp.com)

Ligand profiling is an emerging computational method for predicting the most likely targets of a bioactive compound and therefore anticipating adverse reactions, side effects and drug repurposing. A few encouraging successes have already been reported using ligand 2-D similarity searches and protein–ligand docking (Ghemtio et al, 2012). Other types are receptor–ligand-derived pharmacophore searches as a tool to link ligands to putative targets

Fragment screening involves the generation of libraries having small molecular weight compounds. These molecules can be grown efficiently or have evolved and then optimised (Law et al,2009,Hughes et al,2011). Finally, a specialized physiological screening is employed which is a tissue-based approach looking at *in vivo* effects and screening them.

## 2. Physicochemical profiling:

Physicochemical properties at an early phase of drug discovery and development is crucial to reduce attrition rates due to poor biopharmaceutical properties. With the profiling data and risk assessment of chemicals, it is possible to enhance the predictive power of *in vitro* tools (Wang and Skolnik, 2009). Among these properties, ionization, lipophilicity, solubility and permeability are mandatory to predict the pharmacokinetic behavior of new chemical entities. Pharmacokinetic processes often referred to as ADME, determine the drug concentration, their distribution and metabolism in the body when the medicines are prescribed. Drugs of different nature show different types of ADME's. For example most of the drugs are usually excreted from the body after the earlier liberation, absorption, distribution and metabolism but in rare cases, some drugs irreversibly accumulate in body tissue.

## 3. Toxicity estimation

The aim of reducing toxicities in lead optimization that lack appropriate experimental test can be achieved by *in silico* tools which provide a means of assessing toxicity. Chemicals and its potential breakdown products experimentally tested or untested, are evaluated using TOxicity Prediction by Komputer Assisted Technology TOPKAT (Prival, 2001), ADMET Predictor and Toxicity Estimation Software Tool (T.E.S.T) (Bakhtyari et al,2013). Public databases for genotoxicity and carcinogenicity: The Carcinogenic Potency Database (CPDB) has been designed to obtain data which would give the best estimates of carcinogenic potency. Information on the time of death and tumor pathology for each animal were available from a set of National Cancer Institute (NCI) bioassays, NCI bioassays on aromatic amines (Russfield, 1973) , and from tests in nonhuman primates by the NCI Laboratory of Chemical Pathology through chemical carcinogenesis studies (Thorgeirsson et al, 1994). Potency values in the CPDB are calculated using analysis. Both the CPDB and the online NTP database have been "chemically-indexed" in the DSSTox (Distributed Structure-searchable Toxicity) database and the National Center for Computational Toxicology (NCCT).The **E**uropean chemical **S**ubstances **I**nformation **S**ystem (ESIS) is a freely accessible data via the JRC ex-ECB website EXCHEM (http://dra4.nihs.go.jp/) was developed by the Chemicals Investigation Promoting Council, Japan and was supervised by Office of Chemicals Safety Evaluation and Licensing Bureau Pharmaceutical and Food Safety Bureau Ministry of Health, Labour and Welfare, Japan. The Genetic Activity Profile Database (GAP) developed by US EPA and The International Agency for Research on cancer (IARC), is now maintained by ILS (http://www.ils-inc.com). ToxRefDB and

TOXNET is maintained by the US National Library of Medicine (NLM). The Organisation for Economic Cooperation and Development, OECD Toolbox, is connected with genotoxicity and carcinogenicity database. TOPKAT and Toxtree are used to assess ecotoxicity and mutagenicity, and were created using bacterial mutagenicity data. These test methods for chemical substances were reviewed by the Institute for Health and Consumer Protection of the European Commission (Bakhtyari et al, 2013).

## CONCLUSION

Significant advances in Drug chemistry and ADMET profiling tools have helped decreased the attrition rate of compounds entering the clinical phase. Identifying more phenotypic biomarkers may help bring out better drugs with reduced toxicity.

| **Interpro Domain of *M. tuberculosis*** | | |
|---|---|---|
| | AminoacyltRNAsynthetase | Short chain dehydrogenase/ reductase |
| | ATP binding region | FAD dependent pyridine nucleotide Disulphideoxidoreductases |
| | Peptidase eukaryotic cysteine | S - adenosyl Methionine |
| | Peptidase active site | IMP dehydrogenase |
| | Carboxyl transferase | Peptidase S1 and S6 chymotrypsin |
| | Phosphoribosyltransferase | Glyceraldehyde 3 phosphate dehydrogenase |
| | Aldehyde dehydrogenase | Dihydropteroate synthase |
| | Cytochrome P450 | ABC transporter |
| | DNA topoisomerase | Carbohydrate kinase |
| | Zn containing alcohol dH | Ribonucleotidereductase (large subunit) |
| | Glycosyltransferase | Rhodopsin like GPCR family |
| | Aldo/ketoreductase | Peptidase M14 |

Table 1: List of prioritized genes in application to the problem of TB drug development (Hasan et al, 2006)

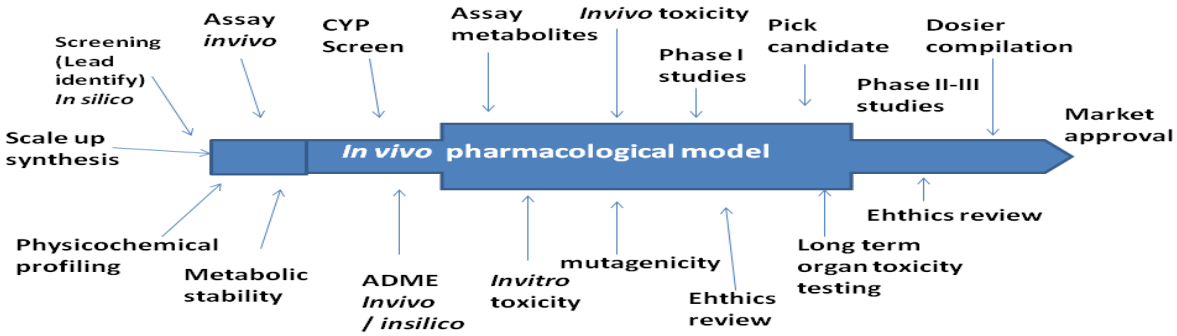


Fig 1: Workflow: Pharmacodynamic studies for generation of a lead compound into a drug (adapted from Hughes et al, 2011)

## ACKNOWLEDGEMENT

**We gratefully acknowledge the support received from the faculty of Biomedical Sciences, Nitte University, Mangalore towards carrying out this study.**

## REFERENCES

1. Bakhtyari, N.G., Raitano, G., Benfenati, E., Martin, T. and Young, D.(2013). Comparison of *in silico* models for prediction of mutagenicity. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 31:45-66.

2. Burbaum, J.J. and Sigal, N.H.(1997). New technologies for high-throughput screening. Curr Opin Chem Biol. 1:72–78.

3. Barnum, D., Greene, J., Smellie, A. and Sprague, P.(1996). Identification of common functional configurations among molecules. J. Chem. Inf. Comput. Sci. 36: 563–571.

4. Boppana, K., Dubey, P.K., Jagarlapudi, S.A.R.P., Vadivelan, S., Rambabu, G. (2009). Knowledge based identification of MAO-B selective inhibitors using pharmacophore and structure based virtual screening models. Eur J Med Chem. 44: 3584–3590.

5. Crews, C.M. and Splittgerber, U.(1999). Chemical genetics: exploring and controlling cellular processes with chemical probes. Trends Biochem Sci. 24:317–320.

6. Cohen, K.B., Hunter, L. (2008). Getting started in text mining. PLoS Comput Biol .4: e20

7. Dixon, S.L.,Smondyrev, A.M., Knoll, E.H, Rao, S.N., Shaw, D.E., Friesner, R.A.(2006). PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening. 1. Methodology and preliminary results. J. Comput. Aid. Mol. Des. 20: 647–671.

8. Ghemtio, L., Perez-Nueno, V.I., Leroux, V., Asses, Y., Souchet, M., Mavridis, L., Maigret, B., Ritchie, D.W. (2012). Recent trends and applications in 3D virtual screening. Comb Chem High Throughput Screen. 15:749-69.

9. Hasan, S., Daugelat, S., Rao, P.S., Schreiber M.(2006). Prioritizing genomic drug targets in pathogens: application to *Mycobacterium tuberculosis*. PLoS Comput Biol 2:e61

10. Hakenberg, J., Gerner, M., Haeussler, M., Solt, I., Plake, C., Schroeder, M., Gonzalez, G., Nenadic, G., Bergman, C.M.(2011). The GNAT library for local and remote gene mention normalization. Bioinformatics. 27: 2769-71.

11. Hughes, J.P., Rees, S., Kalindjian, S.B, Philpott, K.L.(2011).Principles of early drug discovery.Br J Pharmacol. 162: 1239–1249.

12. Jamal, S., Periwal, V.(2013). Open Source Drug Discovery Consortium, Scaria V. Predictive modeling of anti-malarial molecules inhibiting apicoplast formation.BMC Bioinformatics. 14:55.

13. Jiang, X. and Wang, X.(2004). Cytochrome C-mediated apoptosis. Annu Rev Biochem. 73:87-106.

14. Jones, G., Willet, P. GASP: genetic algorithm superimposition program. In Pharmacophore Perception, Development, and Use in Drug Design International University Line (O. F. Guner, Ed.) IUL, La Jolla, CA, 2000 pp.85–106.

15. Krauthammer, M., Kaufmann, C.A., Gilliam, T.C. and Rzhetsky.,A.(2004). Molecular triangulation: bridging linkage and molecular-network information for identifying candidate genes in Alzheimer's disease. Proc Natl Acad Sci USA. 101: 15148-15153.

16. Law, R., Barker, O., Barker, J.J., Hesterkamp, T., Godemann, R., Andersen, O., Fryatt, T., Courtney, S., Hallett, D. and Whittaker, M.(2009). The multiple roles of computational chemistry in fragment-based drug design. J Comput Aided Mol Des. 23:459-73.

17. Li, H., Sutter, J., and Hoffmann, R. Hypogen: An automated system for generating 3-D predictive pharmacophore models. Pharmacophore Perception, Development and Use in Drug Design (O. F. Guner, Ed.), IUL, La Jolla, CA 1999, pp. 171–189.

18. Martin, Y.C. DISCO: what we did right and what we missed. In Pharmacophore Perception, Development, and Use in Drug Design (O. F. Guner, Ed.) IUL, La Jolla, CA 2000, pp.49–68.

19. McInnes, C.(2007). Virtual screening strategies in drug discovery. Curr Opin Chem Biol. 11:494–502.

20. Moroy, G., Martiny, V.Y., Vayer, P., Villoutreix, B.O., Miteva, M.A.(2012). Toward in silico structure-based ADMET prediction in drug discovery. Drug Discov Today. 17:44-55.

21. Otagiri, M., Imai, T., Fukuhara., A. (1999). Improving the pharmacokinetic and pharmacodynamic properties of a drug by chemical conversion to a chimera drug. J Control Release. 62:223-9.

22. Pitt, W.R., Calmiano, M.D., Kroeplien, B., Taylor, R.D., Turner, J.P. and King, M.A. (2013).Structure-based virtual screening for novel ligands. Methods Mol Biol. 1008:501-19.

23. Pospisil, P., Iyer, L.K., Adelstein, S.J., Kassis, A.I.(2006). Acombined approach to data mining of textual and structured data to identify cancer-related targets. BMC Bioinformatics. 7:354.

24. Prival, MJ.(2001). Evaluation of the TOPKAT system for predicting the carcinogenicity of chemicals. Environ Mol Mutagen. 37:55-69.

25. Russfield, A.B., Homburger, F., Boger, E, van Dongen, C.G., Weisburger, E.K., and Weisburger, J.H. Carcinogenicity of Chemicals in Man's Environment. Final Report, Contract No. NIH-NCI-E-68-1311. Cambridge, MA: Bio-research Consultants, Inc., 1973.

26. Sugaya, N., Furuya, T.(2011). Dr. PIAS: an integrative system for assessing the druggability of protein-protein interactions. BMC Bioinformatics 12:50.

27. Thorgeirsson, U. P., Dalgard, D. W., Reeves, J., and Adamson, R. H. Tumor incidence in a chemical carcinogenesis study of nonhuman primates. Regul. Toxicol. Pharmacol.1994; 19, 130–151.

28. Wang, J.(2009). Skolnik S.Recent advances in physicochemical and ADMET profiling in drug discovery.Chem Biodivers. 6:1887-99.

29. Willmann J.K., Bruggen N.V., Dinkelborg, L.M. and Gambhir, S.S. (2008). Molecular imaging in drug development. Nature Reviews Drug Discovery. 7: 591-607 doi:10.1038/nrd2290

30. Yang, Y., Adelstein, S.J. and Kassis, A.I.(2009). Target discovery from data mining approaches. Drug Discov Today. 14:147-54.

31. Yeh, J.R.J. and Crews, C.M.(2003). Chemical genetics: adding to the developmental biology toolbox. Dev Cell. 5:11–15.