Rafael Depetris
Principal Scientist
Kadmon Corporation

Rafael obtained his Ph.D. from NYU in 2008, working with Stevan Hubbard in the field of X-Ray crystallography - specifically focusing on the structural analysis of proteins that interact with the insulin receptor. His post-graduate work was done at Weill Cornell Medical College, where he focused in the structural analysis of the HIV-1 envelope proteins as well as of the HIV-1 neutralizing antibodies.

Rafael went on to assume a position of a Senior Scientist at Pfizer CTI in New York, where he initiated his work focused on modeling of targets and therapeutic antibody candidates. He was also commanding the efforts for the humanization of the lead antibody candidates based on the modeling analyses.

At Kadmon he is combining his expertise in structural biology with in silico approaches to drive forward discovery projects. Here he shares with us his insights into computational drug development at Kadmon, challenges faced and possible solutions, as well as his thoughts on the future of the field. Read the full interview below.

For those who are unfamiliar, could you briefly describe your current research focus and involvement in the computational modelling field?

I like to define myself as a structural biologist. This is a fascinating discipline that can be very helpful in the rational design of therapeutic agents, which comprehends a body of theory that hasn't been thoroughly described yet. It is a field that is in constant growth, as the amount of protein structures deposited keeps increasing. It is deeply related to computational modelling given that modelling tries to predict the structural features of proteins for which there is no experimental data (usually, X-ray crystallography).

You gave a fabulous presentation at the inaugural CDD for Biologics meeting where you detailed the comparative use of X-Ray crystallography for the structural analysis of therapeutic antibodies and computational model validation. How has this work continued to progress in the months since?

We have recently concluded the validation of the model that last years talk revolved around. When putting together the previous talk last year we didn't have full confirmation of the accuracy of the model, but there were structural features that were arranged in a coherent manner that supported the idea that the model was right. We were able to confirm the mechanism of action of our antibody by mutagenesis and binding assays.

> "The current challenge for CADD is to develop a method to evaluate the different outputs of a modelling process, particularly how to get a valid score parameter to be able to rank them"

AI and machine learning are quickly becoming additive tools for in silico analysis of biologics. Given your interest in the application of structural analyses in the improvement of artificial intelligence, what do you consider to be current key advantages and limitations of these approaches for assessing and improving biologics design and developability?

The approaches involving machine learning methods can be very influential. Most importantly, they will tilt the balance towards the use of computational methods for the design of therapeutic agents, and lead scientists to abandon more classical approaches like high throughput screening. The major impact will be on the development of small molecule agents; there are fewer intents of applying these algorithms in the development of therapeutic antibodies.

The efficacy of these methods relies heavily on the quality of the data used to train the algorithms, and that's why I like to say that discussion of the applicability of AI methods is more a data science subject than anything else. One advantage for structural biologists is that the quality of the data is very good given that the structures that are published or deposited have been validated or can be validated in many ways.

The current limitation then stays as to whether the size of the available data is large enough to be able to make the most accurate predictions.

What do you consider the single greatest un-met challenge for the use of AI and CADD in biotherapeutic development?

It is definitely too early to have a diagnostic to tell where AI is standing and what the shortcomings will be. It is a strategy that needs to be further evaluated and which may be useful only for some specific steps in the drug development process.

The current challenge for CADD is to develop a method to evaluate the different outputs of a modelling process (poses); particularly how to get a valid score parameter to be able to rank them. There are specific ways in which AI can help to overcome this limitation and many ongoing efforts in this regard.

Another challenge for both AI and CADD is a cultural one: whether the scientists will trust the output of computational methods accordingly and whether will be prone to take decisions on a project based on that output. The introduction of AI is raising the bar for a discussion that still is at a much more basic level.

How should we seek to overcome this as an industry?

We need successful stories about the application of AI; that may come next year. Stories in which people have been reportedly using AI methods and can show to the community how the application of these methods translated in savings of time and money, which are the main drivers on the decision-making process.

There has to be, however, a predisposition of the scientists to listen and rely on those successful examples. The latter is more difficult, and is more a cultural change I would say.

> "I like to say that discussion of the applicability of AI methods is more a data science subject than anything else"

Over the next five years, where does the greatest opportunity lay AI and machine learning in for computational aided drug design and development?

It can be very helpful to solve the scoring problem mentioned before. The current computational speed and the methods used for in silico modelling are very good but they do provide many outputs; and the scientists need to rely on either their trained eye or on scoring methods that are still being improved. On a more general basis, AI will render any in silico prediction more reliable and accurate.

Have you and Kadmon begun to see an impact from utilizing AI and CADD to your R&D pipeline?

Fortunately yes, and we are now able to fully understand the mechanism of action of every therapeutic agent in our pipeline. In addition to that we can also design our therapeutic agents on a rational basis.

What are you most looking forward to at CDD for Biologics 2017 and can you give us a preview of some of the key points you will be addressing?

I hope there will be a thorough discussion of the role of computational biology on therapeutic antibody development much like last year. I hope we can get an update on how our colleagues are overcoming the barriers I just described and how deep AI is penetrating this specific field. For my presentation, I will describe the current methods available to describe structural features and generate fingerprints that can feed the machine learning algorithms.

> "We need successful stories about the application of AI"

A huge thank you to Rafael Depetris for taking the time share his insights with us.

If you enjoyed reading about Rafael's experiences, there will also be an opportunity to meet him at the CDD for Biologics Summit (December 5-7, Boston).

For more information about the event, visit the website at www.cdd-biologics.com