Jacques Loeb Centre
for the History and Philosophy
of the Life Sciences
Prof. Ute Deichmann, Director

Ben-Gurion University of the Negev
P.O. Box 653, Beer-Sheva, 8410501, Israel
Building 39, Room-114 | Phone: +972-8-6472258
Email: jloebcentre@post.bgu.ac.il | Web: in.bgu.ac.il/en/loeb

Interview by Ute Deichmann with Hans Lehrach

Max Planck Institute for Molecular Genetics, 18 January 2015

Hans Lehrach obtained his Ph.D. at the Max Planck Institute for Experimental Medicine and the Max Planck Institute for Biophysical Chemistry in 1974. As a postdoctoral fellow at Harvard University, from 1974 - 1978, he carried out one of the first cDNA cloning experiments. From 1978 - 1987, as a group leader at the EMBL, Heidelberg, he was among the first to conduct positional cloning experiments in mouse (Brachyury) and man (e.g. Huntington's disease and Cystic Fibrosis). He was one of the scientists who initiated the human genome project. As head of the Genome Analysis Department at the Imperial Cancer Research Fund in London (1987-1994), he developed new structural and functional genome analysis technologies (e.g. the first array robotics in 1987). Since 1994 he has been Director and Scientific Member at the Max Planck Institute for Molecular Genetics in Berlin, focusing on genetics, genomics and systems biology. He has co-founded several biotechnology companies such as Sequana Therapeutics, GPC Biotech, Scienion, Prot@gen, PSF Biotech, Atlas Biolabs, Alacris Theranostics, as well as the Dahlem Centre for Genome Research and Medical Systems Biology.

Early career and research

UD: You studied chemistry at the University of Vienna and did your PhD in Berlin -

HL: Actually I did my diploma and started to work on my PhD at the Max Planck Institute for Experimental Medicine in Göttingen. Then I moved together with the head of the group I was working in to the

Max Planck Institute for Biophysical Chemistry which was then opening.

UD: Why did you move from chemistry to biology?

HL: I liked biophysics. For a long time, I couldn't decide if I wanted to do physics or chemistry and then started with chemistry because it was easier to switch from chemistry to physics than from physics to chemistry because you had all the labs in chemistry. So I was very interested in trying to work on some biophysics type of project. Max Planck was just a very attractive science organization particularly because the situation in Vienna, as far as the support of science went, was pretty miserable.

My wife to be and I tried to get a PhD project in Munich and failed. But we kept trying and eventually got PhD positions at the Max Planck.

I started working on a project which was, in retrospect, completely insane. I was trying to analyze the molecular mechanism of the enzyme nucleotide phosphorylase, which was very complex. It is easily attacked by proteases, so I had to basically make it myself and I had to build this fluorescence start-flow machine – design one and hook it up to a computer (PDP-11/45) and acquire and analyze the data. Maybe an interesting aspect is that what I ended up doing with completely inadequate computation power was a sort of systems biophysics. So I was doing something which is very similar to what we do now. You can measure the fluorescence, which is generated in the release of fluorescent ribonucleosidediphosphates, but you can also measure the kinetics of the binding of the enzyme to the polymer; and so there are many things you can measure with the system from milliseconds to hours. And then, based on the possible mechanisms, formulate systems of differential equations which can be solved numerically and then did parameter optimization using non-linear regression, and all that with a computer that had 64 kilobytes of memory and a 120 kilobyte hard disk. So the computing power has

gone up but we are still trying now to revolutionize medicine by somewhat similar concepts. It just has gotten a bit more complicated in between.

So this was very difficult, and anybody in their right mind doesn't do biophysics on anything which they cannot simply buy off the shelf. So if you do biophysics, you take an enzyme which you can buy easily and then at least there are a whole lot of things which cannot go wrong any more. If you have to make your own enzyme, which is tricky (I started with kilos of E. coli), you have to make your own machine and then design all your measurements and your whole analysis – programs and all that – with the computing power available then, it was just a bit on the ambitious side.

UD: If I understand correctly, you were always interested in developing new techniques, is this right?

HL: Yes, because I always came up with completely crazy questions which nobody else was thinking about. And then, obviously, the technology wasn't there to answer them, so we had to think about how to do it. So when I went to Harvard…

UD: How come you went to Harvard? It was still not common for postdocs from Germany or Austria to go to Harvard.

HL: To some extent it was a stroke of luck, because in the same department we had a group of very interesting post-docs …

UD: That was the Eigen Institute in Göttingen?

HL: It was the Eigen Institute, but basically I was working in the department of Tom Jovin, and lots of young Americans applied to Eigen and some of them he then pointed to the department of Tom Jovin I was in. So we had some really interesting guys in that department that got us started on all sorts of things from watching American football games once a week at the Institute, to ordering a shipment of true American ice cream from Wisconsin in a dry ice package. They would always talk about how American science was,

so we applied to Harvard because that was what they recommended. In fact, one of them got an Assistant Professorship in Harvard, so we had local contacts. The others got Assistant Professorships in Philadelphia and in Princeton.

I got the Jane Coffin Childs Fellowship because as Austrian citizens we couldn't get German money to go to the U.S. It's one of the private fellowships which is very competitive.

I went to work in the lab of Paul Doty, where my wife had already gotten a post-doc paid through the lab.

Paul was a very nice guy, and he had a very nice German wife, Helga Boedtker-Doty who basically ran the lab while he spent much of his time with politics – there is an institute at Harvard for disarmament questions, which he was heading. So in a sense he was leading, under Richard Nixon's presidency, an attempt by the institute to keep the disarmament question active. He had been an advisor to President Kennedy and spent a lot of time traveling and discussing with his Russian counterparts. It was the era of the Cold War. But he was really nice, so if I needed something I asked him to write a letter, and since he had a very good name, it was quite good.

UD: How did you get to cDNA cloning?

HL: It became obvious at that point that there was a revolution with the first cloning experiments going on. I was working with Helga on the question of what the so-called hnRNA really was. We had at that point a few paradoxes. One was a question of why the genomes were so large, far exceeding any reasonable size estimated by the number of genes, plus the size of the mRNA which people knew by then from the size of mRNAs. The master/slave hypothesis was postulated by Charles Thomas (my first postdoctoral advisor) - that there was one master gene and lots of slave genes which were basically just identical copies.

The other question was why you supposedly got huge RNA by the separation techniques available. And the conclusion was, on one hand, that a lot of this supposedly very large RNA was the result of aggregation. So I worked on aggregation in different denaturing RNA gel electrophoresis systems, a paper which is still pretty heavily quoted. The other explanation is splicing; something, which, at that point nobody knew about. But it was pretty clear that it would be much easier to work on these questions, if there were clean sequences of specific genes, basically cDNA clones. So there was a lot of work going on in Harvard, I was talking a lot to Wally Gilbert's people - and others. Resources were pretty scant in our lab, so we established a bartering (lab) economy. For instance, Will McClure, an Assistant Professor at the time, and I would swap radioactive triphosphates for enzymes, which we made ourselves. After the Asilomar moratorium, I was waiting to access a P3 lab – an extremely politically charged topic in Cambridge at the time. I'm sure you know this whole story, Major Velucci, George Wald, Ruth Hubbard, Jonathan Beck etc. – it was all nonsense, but basically this led to the fact that there was an absolutely over-specified P3 lab in Harvard, finally approved months after anything anybody wanted to do was downgraded to P2.

This in the end was one of the more expensive bicycle sheds in the U.S. for some time. I think it cost at that point more than a million dollars to build and was, as far as I know, never used for cloning experiments. Wally Gilbert ended up going to Basel to clone Insulin cDNA, which was sensible because of the economic implications, and I went to Cold Spring Harbor to generate chicken collagen cDNA clones. I used a P3 facility in Cold Spring Harbor for one of the first cloning experiments after the Asilomar moratorium.

UD: The moratorium was in '75, if I remember correctly.

HL: Yes, '75 sounds right.

There were two strategies that were then being developed – linkers and poly-A tailing - and I tried both and one of them worked fantastically well. I had very little time. The first experiment failed because of a critical tube breaking at some point, but the second one worked very well. I got 4.3 kilobase clones at a time when this was just unheard of.

Becoming a pioneer of genomics

So that basically got me into the cDNA cloning and it got me into sequencing. Wally Gilbert's people were doing the Maxam-Gilbert sequencing, so I got this technology. When I went to EMBL, one of the things I started was to try to propose to convert EMBL into the first worldwide genome center, since its official aim had been to do for molecular biology what CERN had done for physics. To have a center, which would have been able to do things which were too expensive for the individual countries. And sequencing the genome of *E. coli* seemed to me a pretty good description of that job. I still think it would have been a brilliant idea, but it was politically not do-able at the time. So I was involved in the discussions on the EMBL sequence library work because it was just heating up. I went to some of the meetings.

The first head of the sequence library was Greg Hamm who had been helping me write programs to analyze sequences and stuff like that. And then we had a meeting in Schönau in which the EMBL sequence library was accepted as a goal for EMBL and my proposal of sequencing the *E. coli* genome was discussed. But the idea was not approved. Some thought it was too ambitious some not enough. In fact, Charles Weissman who was at that meeting thought that the project was not ambitious enough and EMBL should start sequencing chromosome 21 instead, which I thought was a good idea.

UD: You were one of the pioneers of genomic research. Why were you so convinced that it was important? As you said, many biologists

at the time did not want it – developmental biologists and also geneticists.

HL: My basic philosophy has always been that logic helps.

It was to me completely obvious that whatever the hypothesis in research is, it is fine to find some possible mechanisms, but you can never prove a hypothesis. Because proving a hypothesis requires disproving all alternative hypotheses. And that is at best possible if you know all relevant facts about the experimental system. So, for example, if you have an observation that knocking out a certain gene increases the frequency of cancer of some type and you don't know that knocking out many other genes has exactly the same effect, I think you haven't contributed that much to the state of knowledge.

To identify genes for inherited diseases or phenotypes in mice, we developed all of this positional cloning instrumentarium in trying to find the gene for Huntington's disease or for Brachyury; it just required completely different concepts and approaches. And ultimately there are things you can do if you have the whole genome. If you have all genes of a human cell tested functionally, you know their activities, Kms, interaction partners, downstream targets, you are in a much better position to arrive at the truth. Typically, the result of hypothesis-driven research will give you one explanation (and maybe a high-impact paper), but often with moderate benefit ultimately to the state of research. So I think it was completely clear that we would be able to work much faster and much more efficiently and understand the human biology much faster if we simply worked on a genome-wide basis – genome/transcript/proteome-wide basis. I couldn't see why there were technologies in the lab which allowed you to learn a specific aspect of the function of one gene or protein – which couldn't be automated and then applied systematically to all genes of the organism.

Hypothesis-driven science

UD: Do you want to replace hypothesis-driven science by correlations using big data?

HL: It's not correlations, it's basically the same experiments you do in hypothesis-driven experiments, but you do them on essentially all genes of the organism in parallel. So you do not just have one gene which phosphorylates a protein. You have all genes which phosphorylate that protein; then you can understand the mechanism of action much better. So I think this is understanding mechanisms, it's just on a genome-wide basis. Many things we can only study on the level of the cell or the organism. Therefore, if you try to connect the experiments we do one protein at a time in a culture with some phenotype of the organism it takes a hell of a lot of hand waving to try to argue why what you have seen in the test-tube actually has anything to do with the phenotype you are talking about.

UD: I understand. In a way, you keep hypotheses; you just transfer them to the systems-level?

HL: Yes, sure.

UD: The creative scientist is still there – it is not just somebody who adores machines and algorithms.

HL: Sure. My view of biology is that the system is really likely to be so complex that the only thing we can ultimately say is that we can model our concept of what happens and then the closest we can come to the truth is to show that our model predicts correctly everything we can measure about the organism. Ideally we can define a range of models. There are always going to be more complex models which will also be able to describe a system. We will not be able to do mathematical proofs in biology because it's not a binary process; it's much too complicated. So the only thing we can do is have a model, compute the consequences, and make sure that the predicted consequences of our model are not negated by the

experiments. Which is obviously not what we scientists are used to doing or have been trained to do. It's also not something which fits in very well to the usual rewards for scientists or their personal career planning, but it's probably the only thing which is ultimately going to help the patients to benefit from the whole enterprise.

We really need to keep in mind that we are given tax payers money to fund our research - we should be using this money to solve society's pressing problems in the most efficient way we can.

UD: Eric Davidson emphasized the necessity of hypothesis in his systems approach in contrast to scientists pursuing Discovery Science or ENCODE scientists, who have the hypothesis created by the computer and don't test them against reality.

HL: Yes, but then they publish the data and then every scientist can formulate hypotheses on the basis of the data. I think that we have to define a goal. And once we define the goal, then everything else is a tool to reach that goal. Hypothesis-driven research is a tool which is appropriate in some circumstances – effective in some circumstances – and ineffective in other circumstances.

UD: OK, but also to understand development and other non-medical questions.

HL: Yes, I think that you can divide science into hypothesis-driven and data-driven science, while in reality most of science is impact-point driven.

UD: So you think the future scientist is different?

HL: I would hope so. You know, we have spent a hell of a lot of money with very little impact on, for example, cancer treatment. Not very little, but I think we could do much more good for the money we are getting, if that was a high priority. As long as the reward mechanisms are such that for every individual, science is only to think about his/her career and not to think about how to help other people, then that's what we get. If we have a reward system in which the bankers

get rewarded for destroying the economy, then that's what's going to happen.

I had a few discussions with Benno Müller-Hill on the merits of hypothesis-driven research. Hypothesis-driven research is clearly better than random research. So we fully agree on that. But it makes sense based on the specific trade-off between the cost of generating data and the value you get out of it. If the only thing you can ever afford to do is analyze one protein, then hypothesis-driven research is the only thing you would be able to do. If you can sequence the entire genome and therefore start hypothesis-free to look at all hypotheses compatible with the much larger dataset you have available, then it may not be appropriate to continue with narrowly defined hypothesis-driven research.

UD: In his last paper Eric Davidson said that in the systems approach which he uses, in order to come up with this entire network model that he created, you need to have a hypothesis at every stage because it is so complicated. Hypotheses are very important; otherwise you will not get anywhere.

HL: You go hypothesis-free wherever you can and then where hypothesis-free approaches don't work anymore, you have to start with hypotheses because, for example you can analyze all proteins in principle, but you will have a harder time to analyze all potential binary complexes between those proteins – triple complexes, quadruple complexes. So you are caught at some point in a combinatorial explosion which enforces, in a sense, the hypothesis-driven approach. So the question is only how far you can push it – hypothesis-free – and where do we have to continue with hypothesis-driven.

UD: But the whole idea of a hierarchical gene regulatory network is based on the hypothesis that there is a hierarchy of genes; with some genes having a high impact on others and others just involved in coding for proteins.

HL: There are so-called trivial hypotheses. It's a bit like the hypothesis that sequencing the genome could be interesting. Sequencing the genome requires a hypothesis that sequencing the genome is interesting, but it's very different from hypothesis-driven research in the classical sense.

Personalized medicine

UD: One of your goals in genome sequencing is to contribute to personalized medicine. How realistic is this aim?

HL: Extremely realistic. That's what we are doing right now. We take the tumors from patients, in fact patients who come to us or are sent to us – in a spin-out company, as Max Planck is obviously not the right place to do something like that, where the tumors are analyzed. Typically, we analyze the exome of the tumor, transcriptome of the tumor, exome of the patient. But we would like to feed in as much information as we can get – proteome, metabolome, anything we can in principle measure, ideally spatially resolved to one tumor cell at a time, and then make a model of the tumor in its complexity, ideally a model of the liver because of the pharmacogenomics and models of the normal tissues which are possibly killed by the drug faster than the tumor is killed. Also, ideally a model of the immune system of the individual patient. So you have a virtual patient which allows you to make mistakes not on a real patient, but first of all to test everything on a virtual patient.

UD: But you don't have all the data of the patient.

HL: We have a lot more than we had 10 years ago. We have the genome, transcriptome, we have proteome data, and we are working on this spatial resolved sequencing. Also there are technologies around which are going to make it simpler. And I am convinced that with all the information that we can collect at a reasonable cost from every patient, combined with the increasing computing information we have

available, we can easily beat the 25% success rate which the physicians have, for example, in cancer treatment in the hospital. And in fact, I would argue that those colleagues who don't believe that, are, in a sense, arguing, that science is practically not very useful making you wonder, if the money should not have been used to improve public transportation etc. instead of basic research because it helps the patients much more! I'm sure that the S-Bahn in Berlin, for example, could use the money!

So you cannot argue, on one hand, that science will solve all problems, and on the other hand, argue that even for a comparatively simple problem we are much too far away from knowing everything after spending maybe a trillion dollars on cancer research since the early 1970s. In a sense, you just have to think logically about the available facts.

Big-data driven research

UD: What do you think about these new developments – big-data driven discovery science or big data in the Google collaborations for example with Stanford University or the Broad Institute?

HL: I think it will work in some situations but not for others. It's OK if you are working with extremely complex processes for which you have no molecular mechanisms. If you want to understand which people buy i-Phones or Samsung, it's not going to help very much to sequence their genomes any time soon. If you want to understand which people get Type-1 diabetes and how to treat them, we have molecular mechanisms to model the process and then to try out on all the models which drugs would work best.

To me, for medicine nothing will work which also doesn't work to predict the weather, because it's a very similar type of problem. The weather of every day is different and you cannot do statistics, or stratification. A stratified weather forecast – one for the summer and

one for the winter, would be only moderately useful, but that's more or less what we have in medicine often.

There are many examples of processes which have to be understood in enormous details to make the right predictions. The weather is one of them. Virtual crash testing is one of them, where it might depend on the strength of one particular part of the car - a sheet of metal to predict if the driver will survive or die. There are many things which you have to know in exceeding detail to have a chance to make the right predictions and so things cannot be predicted by big data or by looking it up in the literature because exactly that case has never been described, just as much as the weather of today has never been observed so you cannot simply say, "OK, I will look up in my records when the 70 terabytes of starting information which I have measured about the weather of today have exactly occurred previously." They have never occurred and they never will occur again. Similarly, the molecular biology of a tumor has never occurred and will never occur again.

I think Google genomics is going to help to learn something about the basic rules, but data quality is key - there is a big danger of "garbage in, garbage out." If you know a lot of patients you know very little about, you are mostly measuring noise. What you really should do is characterize fewer patients very accurately, and then try to build a model which you can test against reality. Not necessarily in the hospital. We are doing a lot with xenografts with 3D cell models, but you then develop and adjust your models, for example, finding the regions in the parameters base which are compatible with life, in a sense. Compatible with the data by comparing your predictions to reality. And I would predict that within a relatively short time, we will generate more information about human biology by comparing the predictions of models to reality than by the entire basic research enterprise.

UD: I understand that many projects are just about correlating genomic data. They are not trying to find out about individual data of individual patients or mechanisms, but just take their genomes and -

HL: Yes, it will maybe tell us something, but it will never be very good for predicting how to treat an individual patient.

UD: From what I understand, this purely big data driven science doesn't find out about mechanisms.

HL: It can create leads into mechanisms. It creates crucial information to identify mechanisms that you don't know already.

UD: But in order to identity those mechanisms you need to know a lot more than just the sequences.

HL: Sure, as I said, I think it can be a case of "garbage in, garbage out," but depending on how well those things are done, they could identify things which don't fit with our current model, which then can be used to improve the model.

But as I said it's just a tool. You should think about at which point you generate more insight by formulating a hypothesis first and then designing the experiment or where you are better off generating all the data you can and then testing all possible hypotheses on the dataset. It's data generation and hypothesis; the question is what comes first?

UD: But the hypotheses are created by the computer.

HL: You can generate the hypothesis by the computer, but sooner or later the human brain is still required, at least until the artificial intelligence guys make lots of progress.

UD: Did you cooperate with Craig Venter during the time of genome sequencing?

HL: Yes, we have actually one paper together, about sequencing a region of the Chorea Huntington region when we were trying to find the

Huntington gene. I think running into each other in the Genome Project we were in opposite camps.

If he hadn't set up his private genome sequencing attempt, it could be that the public project would have taken a lot longer...

Modeling

With all these things, the only chance we have is to make accurate models and then see what happens to move the experimentation and the errors associated with the experimentation from the real domain into the computer. That's what we do everywhere. We don't build skyscrapers just to watch them fall down in the first storm. We model them and make sure that they survive a Pacific typhoon (even if it would be a bit unlikely in the middle of Berlin). But basically, if something is important then you try your best to make your mistakes on a computer model which is accurate enough to predict what is likely to work.

UD: From what you said, I understand that efficient modeling has to be based on experimentation.

HL: There are two components which are important. One is the knowledge of the mechanisms. For example, in the weather forecast, the knowledge of the gas equations and spinning of the earth's Coriolis forces and all those things, and the other one is a detailed characterization of the starting point of the model. What is the weather of today with ideally a meter resolution? At the moment it's probably a kilometer resolution, but it would be a lot better if it had a meter resolution in our forecasts.

In my talks I have a quote from a website for weather forecasts that a model is only as good as the input data. And that is completely true for medicine just as much as for the weather forecast. If your only information about your patient is blood pressure, you will not come up with an individualized therapy, just as much as you will not come up

with an accurate weather forecast if the only thing you know about the weather of today is the temperature in front of your door.

UD: There is a tendency in bioinformatics that people make computerized models which they don't intend to test and correct through experiments.

HL: Sure, that's pretty useless as well. But that's, again, the reward structures that promote the publish or perish mentality, rather than the actual impact on patients.

UD: I read in a paper of yours on reverse engineering of gene regulatory networks that you used mainly artificial data. Are there plans to create experimental data?

HL: Sure, we have huge amounts of experimental data from the RNA-Seq experiments. But it's still a difficult problem to model the transcription regulation because there are so many transcription factors which interact with so many promoters in different combinations and modification states. It just takes more work.

UD: In 2012, Eric Davidson and his group completed a computational model of a complete developmental gene regulatory network. Is this of any value to you?

HL: Sure, definitely. I had a guy working in my department who tried to work with Eric on the sea urchin. In effect, Eric took over many of our technologies. I think he even bought our robots – robots which we had developed to do exactly that type of analysis. Large-scale, systematic filter hybridizations to find out how all genes are expressed – so Eric bought into the concept completely.

Genome and environment

UD: Eric was convinced that development is completely determined by the genome. At least early development. Later on other factors such as

epigenetic ones become also important. You didn't mention epigenetics. Is it because it is determined by the genome?

HL: That's another discussion which I keep having. There are two possibilities. Either epigenetics is completely determined by the genome or it's unimportant as you can see in monozygotic twins.

UD: I agree with you, I just wanted -

HL: I had a discussion with a woman who was a historian of science and somehow there is this ideological anti-genetic undercurrent -

UD: Anti-genetic-determination.

HL: Which is nonsense! At the outset we are not genetically completely determined. If you don’t feed us we are dead within a month, so the environment has some effect on phenotype.

UD: Epigenetics is also used to claim the existence of a Lamarckian kind of evolution.

HL: Maybe there is some experimental evidence that effects of the environment can be translated across generations. There are some possibilities, but they are still fairly limited. As I said, epigenetics is an important mechanism in the translation of the information from the genome and from the environment into the phenotype. We know perfectly well there are things which are not genome-determined. We are analyzing here the immune system of identical twins where one has type-1 diabetes or lupus and the other doesn't. So it's obvious that the environment, and possibly infections, shape the immune system even if you have the same genome. So nobody, in their right mind, is arguing that the environment is unimportant. But everything is just a mechanism in translating those two inputs. These mechanisms – DNA methylation, microRNAs, non-coding RNAs – all those things still just translate signals accurately enough to make sure that monozygotic twins are still pretty similar at 80 years of age.

UD: Ellen Rothenberg told me that the histone modifications in the development of the immune system are determined by transcription factors.

HL: Yes, it's the interplay of mechanisms which we don't understand.

Ethics

UD: Coming to questions of ethics. Genomic sequencing makes it possible to compare genomes not only of individuals but also of human groups or of what people call races.

HL: We know a lot about that. We participated in the 1000 Genomes Project which is, after all, exactly that.

UD: There are some scholars, for example the author of "The Myth of Race," which appeared in 2014 at Harvard University Press, who think that a science that establishes genetic differences between human groups is racist.

HL: That's completely nonsense. A science is either correct or not. That's the only criterion you should judge it by. And it's also nonsense anyway because the only thing you show is that everybody is different. That you find some groups which are a little bit more similar than other groups. But the inherent message is everybody is different. There are, if you want, seven billion races on the earth.

UD: Right, but then you have groups which are different. And you can find statistically relevant differences between certain human groups.

HL: They are statistically significant, but they are not relevant because ultimately the only thing which is relevant is the individual. You should treat, for example, patients not by the race they belong to based on the fact that this ‘group’ has a higher frequency of some variant which causes them to get some particular type of diabetes. We should characterize every individual and treat him/her optimally.

UD: Right, but so far that is not possible because it is too expensive.

HL: But in a sense it's not true.

UD: I have read that this is why some medical scientists use as a rough indicator what they call race parameters. And the people – the minorities – even ask for it.

HL: Sure, if that's all you have, it's better if you have a slightly better chance of responding to your drug based on your skin color. But it's a lot worse than if it's based on your individual genome and transcriptome. Ultimately it’s best for every individual, irrespective of which race, to be treated as an individual. And that is the position that we should get at as quickly as possible.

Sequencing costs have dropped from billions to a thousand dollars per genome and further competition in sequencing technologies will help to drive prices down further. There are things on the horizon which could give us a sequence of the genome for thirty dollars.

So we have to plan ahead now for a future in which anybody is treated optimally as an individual; not stratified - as I said, stratified weather forecasts is maybe an example - but individually. And that's already now cost-effective for cancer patients.

One of the things I'm doing is trying to convince the EU to maybe invest 1 billion into building up a much more personalized medicine and prevention-based healthcare system in Europe. The EU alone spends 4 billion euros every day on healthcare. Compared to that, the costs of the Greek debt crisis are pretty trivial.

A lot of that money goes for drugs, which are expensive, and then either make the patients sicker, or – statistically, healthier. For example, many expensive cancer drugs on average extend the life of patients by 30 days. Simply because patients are not treated individually. Medicine is truly individualized in, for example, surgery. If you break your left arm, you are very unlikely to get a cast on the right arm because more people in a clinical trial had broken their right

arm so they look similar, and the best treatment for this group of patients happened to get a cast on the right arm. And that's exactly what we do with drug treatment.

It's completely obvious and understood that personalized medicine is the standard in surgery because we look at what happens if you try to fix the broken bone not of the average patient but of the individual patient as well as you can. It's because you have lots of data, imaging, and the brain of the doctor is well equipped to draw the right conclusions. With drug treatment, we have no data and no conclusions. That's what we're trying to put in. The data is mostly molecular, but sensor and imaging data can also be very useful. Optimisation of drug treatment will come from mechanistic models on which you can try all possible treatments and then select the one which doesn't kill the patient. But the fact that everybody accepts that surgery has to be individualized and everybody seems to be absolutely sure that drug treatment is called personalized (but isn't) and the best we can do is the stratification is just not good enough.

We should also appreciate that the way we approach the development of cars and planes etc., training pilots or making hand mixers – which is to model everything, make sure that we don't do something stupid, and then start building something will also have to be applied in medicine.

UD: But the systems approach is not so new anymore.

HL: Yes, but the systems approach – you have to be careful. The systems approach is sometimes very ill-defined. It can be anything from just generating lots of data on whatever you can generate data on – systematic, large-scale, two-hybrid, and whatever. I'm not quite sure where Leroy's four Ps, for example, stand because you really have to have a very clearly-defined way of proceeding. [According to Dr. Leroy Hood, the medicine of the future should be based on prediction, prevention, personalization, and participation] And to me, it's very simple. What we need is not something defined as a systems

approach, but we need the capability of making our mistakes on a computer model. At least I expect that will be the only thing which works. There is always the possibility that at some point we will have some brilliant insight and understand how things work. But it hasn't worked for a long time. I think the only thing will be to tediously try everything you are able to try on every individual patient, and it's clear that this would kill the patient very quickly so you basically have to try it on a computer model. Now that is, to me, the definition of systems medicine and systems biology and I think that's the only thing which will work.

I think you need a very clear definition of what the problem is and the problem is that every patient is different, should get ideally a different therapy, should get a different prevention, and should get a different fitness studio advice and the only way to achieve that is by virtual models which accompany the patient, in a sense, from birth to death as a sort of forecast of whatever happens to everyone should be first tested on the computer model of that individual, not only for immediate ill effects but also for long term ill effects. So everybody should have his virtual twin as a sort of guardian angel to protect him

UD: Coming back to the scholars according to whom science is racist when it establishes genetic differences between human groups – do you think science can be racist at all or establish values, or do you follow Max Weber who says that science is value free and it's only scientists or the society which establish values?

HL: Scientists should be value-free otherwise it's too dangerous that we fool ourselves. Humans have too much of a tendency to like to fool themselves, for example in the case of homeopathy. I think the more we look at every individual and the more we can handle every individual on his own, then those idiotic concepts like race will go out the window. They are just completely irrelevant for the things which are necessary, which is how to treat the individual patient. I think it is very unlikely that we will find a classification in which we will say,

"OK, those people are genetically so stupid that we shouldn't send them to school" or something like that.

UD: No, but there are significant differences in intelligence tests.

HL: I don't think that's our problem. If you look at the guys who blow themselves up among hordes of Shi'ite children, there is no race question involved. I don't see any case in which those problems are serious compared to the good we can do and cannot be taken care of by legal mechanisms. You know, if you are afraid that genome information could be used to affect the price of life insurance, that is something you can just solve legally. It's unethical to let cancer patients die prematurely for not having to pass a law forbidding misuse of genetic information. Just because data protection is so important that you cannot treat every patient the way he should be treated. This is OK for rich societies which have no problems to worry about or believe that. But in a situation in which every second person is likely to get cancer, we have other problems to worry about.

And I would really focus on the objectively serious problems and not on this stupid nonsense which people come up with if they get bored. If you are too afraid that you might be grabbed by a dinosaur when you cross the street, then you are more likely to be hit by a bus! Rational actions and, in fact, statistics could be a good basis for optimizing the happiness of everybody. But statistics is obviously something people do have some serious problems with. So I think to some extent the educational system could help to just teach people to think for themselves, to teach them basic statistics. Tell them what is a proof, what isn't a proof. I think you could do that in primary school.

UD: Teach them Mendel.

HL: I'm very interested in any counter-arguments and would welcome more of them in general.

Basic research

UD: Don't you think that there should be also a place for basic research, to find out, for example, about causes and mechanisms of animal development before thinking about patients? I think that not all biology should be related to medical questions. There are still a lot of basic biological questions which have not yet been solved. And science is also a cultural institution. Looking at the cultural impact science has had from its conception in early modern times or, let's say, 19th-century biology, I find it really, really important that -

HL: I think there is – we just have to be honest –science as a cultural activity competing with the money spent on the opera and various other cultural institutions. As an educational activity because you have a lot of people teaching students and it keeps the system somewhat honest if they somehow know the things they teach have been generated – a fact which might actually change in interesting ways if you take these internet-based courses with hundreds of thousands of students into account. And you have a science funding component which is driven by promising to solve problems of society.

I think as long as we are honest, in arguing for money for science, everything is fine. We can say, "OK, this is completely useless, but it is important for our self-understanding." We can say, "OK, we know that those guys are not going to get the next Nobel Prize, but it's better if the students, or the doctors and the medical research have some clue about how research works and therefore we let them play around and the money spent there is well invested." And we have the problem of, "we are going to solve cancer or global warming." We should not have scientists arguing, that they are going to cure cancer patients but forget this as soon as they have the money in the bank. I think that dishonesty is always a problem.

It's one of the major driving forces between lots of organizations. I wouldn't expect that either the military or the banks or the churches are any more honest about the goals of an individual when they are

going out to get money for what they are doing. So it's probably, to some extent, unavoidable what they have to say. I'm a bit disappointed that scientists are not inherently honest about everything. If you are lucky, they are honest about the data they generate, but that's pretty much it. And even that's not always a given.

I also think the effectiveness of science even as a way of generating knowledge could be improved a lot. I don't think Fred Sanger would have a very good chance of surviving in the current high-impact journal environment. People joked, that he got two papers and two Nobel prizes. But I don't think he would be a good candidate for a job in places with the tendency to look at certain high-impact factors. This science-as-a-competition might be fun, but I think it's cheaper to let the people run around the block to compete with each other than to spend taxpayers' money.

UD: Yes, but when you go back to the time of Sanger, you had these scientists who worked alone or with one or two people and built their hypotheses and tried to find out about mechanisms. This picture is more and more being forgotten.

HL: It was also an easier time. It was easier for drug companies to find new drugs when you just stumbled into things that happened to work, like quinine for malaria. Many of the simple things which we could solve with the technology at that time probably have been found. Now the technologies have changed and we have to adjust the way of proceeding. Now, as I said, ideology (like only classical hypothesis driven research) is always bad. It stops you from doing pragmatically whatever makes the most sense in an individual situation.

I've heard that good science has to be hypothesis driven. Homeopathy has a huge advantage in that -

UD: It doesn't do any harm; it may help psychologically.

HL: In some cases, it’s a reasonable, pragmatic response to the needs of the patient, but I wouldn’t use it for cancer like Steve Jobs [an American businessman and inventor who unsuccessfully tried to fight his cancer disease using pseudo-medical treatments]. It just depends on the exact circumstances. But that's too complicated for most people. They want to have one particular way of proceeding no matter what the situation is.

UD: If you want to find out about the genetic control of organ formation in animals, you use, in part, hypothesis driven science.

HL: Yes, as I said, it's just a tool. It's a trade-off between starting with a hypothesis and starting with data, and in some cases if you start with data it’s not bad because then it's much easier to formulate your hypothesis on a database. If it's a choice between spending three years of testing one hypothesis or spending two out of the three years generating the dataset and then you have one year to test thousands of hypothesis on your dataset, then the second form is not bad. Most hypotheses are inherently wrong. People tend to think that every hypothesis has a 50% chance, at least, of being right. That's not true.

UD: In the past, too, a lot of hypotheses were based on a lot of data, for example the hypothesis of the DNA double helix. I don't see a fundamental difference in principle in that, only in the question of whether or not you later test your hypotheses.

HL: First of all, that was a time when it was very expensive to do the experimentation, so every minute spent on thinking about how the few data actually could fit together was very worthwhile. And it's also a sort of unique problem – there's only one double helix. But to extend this to the question that 1 out of 100,000 proteins could be involved in Alzheimer's is something different.

UD: Yes, of course. I am just saying that so-called hypothesis-driven science is also based on data. You don't start from nowhere.

HL: Yes, if you want to arrive at the truth you have to eliminate all hypotheses but one. Good old Sherlock Holmes!
But it's something that sometimes is not sufficiently considered because hypothesis-driven research just fits very well into the organizational structure. You have one guy getting a PhD for testing one particular hypothesis. It doesn't matter how plausible it is as long as he gets money!

The hypothesis that on the far side of the moon there is a sheep governing developmental processes is also a hypothesis. It's a very unlikely hypothesis, but -. Actually, that's one interesting point. I would argue that good science is more or less based on statistics. You have a likelihood ratio which you generate by -

I've tried to discuss that with my local historians of science but they never rose to that challenge. A base in statistics really exemplifies what good science should be. At least in my field.

UD: But what about the discovery of CRISPR-cas? and other unexpected discoveries?

HL: The goal is obviously to combine hypothesis and data. Now science is very compartmentalized and if you compartmentalize it by the guy who generates the data and the guys who then formulate the hypothesis, or if you compartmentalize one gene at a time, I think the first version is more effective. Because you need only two groups, whereas with the other one you need 20,000 groups. Many of them in duplicate.

I think that is one of the problems in general with science. Science is a bit organized like an ant heap. Everybody decides on his/her optimal strategy based on looking around at what the neighbors do and what they don’t do. Which is fine to build an ant heap, but it's an inefficient way to reach the moon. If you want to reach the moon, it may be better trying to understand the problem from the goal downwards. You don’t just do whatever helps you survive in the system, but you say, "OK, we want to ultimately make sure that we

treat every patient optimally. What do we have to develop to get there?" Just as much as it is completely obvious to say, "OK, we ultimately want to understand human biology. What do we have to do to get there?

We have to be able to sequence the genomes; we have to be able to do a functional analysis on genes, transcripts, proteins (which we now do by hand) systematically on all genes; we have to build models which then can make predictions out of those extremely complex interactions; and we have to compare those predictions to reality. And as I said, by definition the best we can do is develop explanations which are able to predict correctly everything we can measure.

If it's a really important problem, you should forget about the organizational details but really start from the goal.

UD: Thank you very much for sharing with me your challenging insights.