History of Modern Biomedicine Research Group
School of History, Queen Mary University of London
Mile End Road, London E1 4NS
website: www.histmodbiomed.org

AUDIO INTERVIEW TRANSCRIPT

Rawlings, Chris: transcript of an audio interview (04-Aug-2016)

Interviewer: Tilli Tansey

Transcriber: Debra Gee

Editors: Tilli Tansey, Sarah Beanland

Date of publication: 25-May-2017

Date and place of interview: 04-Aug-2016; Queen Mary University of London

Publisher: Queen Mary University of London

Collection: History of Modern Biomedicine Interviews (Digital Collection)

Reference: e2017231

Number of pages: 27

DOI: 10.17636/01023369

Acknowledgments: The project management of Mr Adam Wilkinson and the technical support of Mr Alan Yabsley are gratefully acknowledged. The History of Modern Biomedicine Research Group is funded by the Wellcome Trust, which is a registered charity (no. 210183). The current interview has been funded by the Wellcome Trust Strategic Award entitled "Makers of modern biomedicine: testimonies and legacy" (2012-2017; awarded to Professor Tilli Tansey).

Citation: Tansey E M (intvr); Tansey E M, Beanland S (eds) (2017) *Rawlings, Chris: transcript of an audio interview (04-Aug-2016)*. History of Modern Biomedicine Interviews (Digital Collection), item e2017231. London: Queen Mary University of London.

Note: Audio interviews are conducted following standard oral history methodology, and have received ethical approval (reference QMREC 0642). Related material has been deposited in the Wellcome Library.

Rawlings, Chris: transcript of an audio interview (04-Aug-2016)*

Biography: Professor Chris Rawlings PhD (b. 1954) started his bioinformatics career at the Imperial Cancer Research Fund in 1982 during which time he was the Project Manager for the computing infrastructure needed for the Human Gene Mapping Workshops (10.5 and 11). From 1991 to 1996, he led a group that researched the application of advanced logic languages to genetic mapping and protein structure bioinformatics. In 1996 he moved to SmithKline Beecham, where he was responsible for the bioinformatics platforms supporting human genetics, comparative genomics, and gene expression. From 2000 to 2004, he was the Director of Bioinformatics at Oxagen Ltd, where his group worked on the identification and validation of genes and drug targets from human genetics and genomics technologies. He moved to Rothamsted Research in 2004, where he now leads the Department of Computational and Systems Biology, which comprises over 40 staff and students engaged in research into, and application of, bioinformatics, mathematical modelling, and statistics to problems from the agricultural sciences. His personal research interests are in the development and use of data integration systems for supporting systems biology and for candidate gene discovery from multi-omics datasets. He is a visiting Professor in the Department of Computing at Imperial College London, and was also one of the founding members and former Vice President of the International Society for Computational Biology.

TT: Tilli Tansey

CR: Chris Rawlings

TT: Can you tell me about your family background Chris, and how you got interested in science?

CR: I come from Warwickshire. My family was not a particularly scientific one. My father was an antique furniture restorer and my mother worked in the haematology lab and pathology lab in the good old NHS days. And I think probably my interest in biology and the natural sciences came from the dinner table conversations in which she was always talking about work in the lab. And so working in the lab and visiting her lab became something that was quite familiar to me. I then found myself doing well at sciences at school, and decided on perhaps some sort of career in science or medicine, because medicine was on the cards – there was a family history of being medical professionals, or paramedical scientists. So it didn't seem unnatural then to go on that route.

TT: So there was some science in your family, you were aware of science. Did you have any particularly inspirational school teachers, or was it just a question of knuckling down and getting through exams?

CR: I had an inspirational physics teacher, and he really made a big difference to my attitude towards science and gave me a love of physics, actually. Which is sort of how I ended up eventually in the career I'm in.

---

* Interview conducted by Professor Tilli Tansey, for the History of Modern Biomedicine Research Group, 04 August 2016, in the School of History, Queen Mary University of London. Transcribed by Mrs Debra Gee, and edited by Professor Tilli Tansey and Mrs Sarah Beanland.

Because having completed my A levels in physics, which I didn't do particularly well at examinations, I decided that I wanted to combine my interest in physics and my love of biology, and try and pursue a career in biophysics. I'd attempted to get into medical school, but as a result of a misspent youth I didn't really do as well in my exams as I could have done, and took a year out, and found a biophysical sciences degree at the University of East London; well, it was North East London Polytechnic in those days, not far from here, being in Stratford. So that was my route into a science career, through the polytechnic system.

TT: It was quite unusual then, because you're talking about the early 1970s.

CR: Well, it was called biophysical science. There was physics, there was some engineering, and there was biochemistry. But the focus of the physical science was physics and things. That was unusual. The only other biophysics degree at the time was at Kings, and that was the usual sort of competitive entry system. So the polytechnic at that stage gave me an opportunity as a not particularly successful A level candidate, an opportunity to get in.

TT: Could you just go back, and biophysics; again, that seems very unusual, because it certainly played to your strings in physics, your interest in physics, but very unusual for people to go to university to do biophysics.

CR: I think at the time I was thinking more of bioengineering. I think I hadn't made the distinction between physics and engineering quite so clearly in my head; I think I was just doing what I wanted to do. If I could, I found that there were courses in biophysics and I thought it's what I wanted to do. I'd had enough contact with biochemistry in my A level biology to know that I didn't really want to learn by rote quite so many of the biochemical pathways, which seemed to be a major part of teaching in biochemistry at the time, and I thought 'I love the physics bit and I like the experimental side of biophysics, and the physical techniques, playing with lasers, playing with electronics.' That was my hobby, playing with electronics, so it seemed like a natural place to end up where I could enjoy myself. I didn't quite know what I was going to do for a career, but I could at least have three years of enjoying something that might lead to something interesting.

TT: You just said something that clicked in my brain. Your hobby was electronics. Can you say a little bit more about that, because hobbies are so important?

CR: I had two hobbies, one of which was classical music, and the other one was electronics. I used to fiddle with lots of electronic things, and I used to make my own bits of electronic equipment for music making, so I dabbled with electronic music, and I dabbled with effects, building my own effects pedal for the electric guitar. It was only ever a hobby, and that gave me the practical and technical exposure to electronics that proved useful later, and still does actually.

TT: So you come to the polytechnic, you're going to do biophysical sciences. What did that encompass? It sounds an amazing degree.

CR: The first year was largely probably fairly similar to any modular biological sciences course. There was some biochemistry, some basic chemistry, some physics around optics, and some instrumentation; some lasers and maths. And it's going back a long time now so I don't quite remember all the options that I took, but I do remember the first year being fairly broad. I still ended up having to learn by rote quite a lot of biochemical pathways, so I didn't quite escape that, but it wasn't a major part of the course at that time. Crucially, in that first year, they took us on a one-week computer programming course, which actually we did in BASIC [Beginner's All-purpose Symbolic Instruction Code] at the time. And that week of computer programming was what sort of planted a seed in my mind about what I wanted to do in the future. That was the start of a change in what I thought I wanted to do. I hadn't quite realised it yet, but that was very important to me.

TT: Were you able to follow that interest into your second and third years at all?

CR: In my second year things got a little bit more serious in the course around mathematics and around population theory, and around the physics that we were doing experimentally. So we were making holograms, and we were doing quite a lot more electrochemistry, so there was a lot of the physics of excitable cells, and that sort of material that was being taught on the course. I found all that fascinating; I really enjoyed the fact that you could describe some of these behaviours with fairly simple electrical analogue circuits. In those days analogue computing was as important as digital computing at that time, and so playing on the idea of tissues and membranes and things as electrical components in an electrical circuit that resulted in a certain excitable behaviour in the cell. I found that very interesting. And then in the third year (it was a four-year course) we were given a one-year placement in industry or research; it depended on the luck of the draw. I tried to get into Roehampton at the time, because that was the physics of limbs and prosthetic engineering, and I thought I wanted to go more in the bioengineering area. But I didn't get that placement, or that placement didn't crop up. And the course tutor at the time put me in touch with the Institute of Cancer Research at Pollards Wood Research Station in Buckinghamshire; and I interviewed there, and managed to get a place doing something completely different, that I wasn't at all convinced that I wanted to do. But it was a placement, there wasn't a lot else going, and that was in DNA repair mechanisms in mammalian cells. As an interesting aside I also interviewed at Rothamsted Research, which is of course where I work now, as a sandwich student to work with a chemist on small molecule structures of, I think they were either insecticides or pesticides, and Mary Truter who'd moved her group out after the second world war, from UCL I think.

TT: It's an interesting place, Rothamsted.

CR: It's historical importance, apart from its experimental work, has been where R A Fisher was, and the statistics. Sir Walter Bodmer was R A Fisher's last PhD-student.

TT: Going back to the Institute of Cancer Research - was it a field station?

CR: It was the Institute of Cancer Research, so it was before Cancer Research UK merged the Institute of Cancer Research and the Imperial Cancer Research Fund. It was a chemical carcinogenesis group really, and it was a little breakout group from the Fulham Road laboratories, that had combined. There were three group leaders there; I'm trying to remember their names. John Roberts, who was my PhD supervisor and my sandwich year supervisor, who was the DNA repair scientist. There was Peter Brookes, who was an expert in chemical carcinogenesis. And there was a radiophysicist, radiochemical, looking at damaged DNA caused by radiation - the other senior scientsts at Pollards Wood at the time were: A R Crathorn, Ken Shooter, Stan Venitt, Phil Lawley. And they basically had created, they decided that they wanted to have somewhere separate because they found the conservatism in the Fulham Road laboratories too stifling, and they wanted to be out there on their own. How they negotiated to rent this old property in Pollards Wood Research Station, which had all sorts of history attached to it, I don't know. So they formed the relatively small group there, also joined by another scientist who was a guy called Stan Venitt who was also very well known in the assessment of chemical carcinogenesis using bacterial techniques. And they formed this core senior science group, but it was only four principal investigators, and PhD-students and technicians. It was very influential at the time in the whole area of chemical carcinogenesis, DNA repair, human mutagenesis, so it was quite an influential team.

TT: So it was quite a placement to get; even though you hadn't been thinking along those lines.

CR: Not at all. It completely shifted my view. First of all, I became convinced that I wanted a career in research, and I really enjoyed the laboratory experimental work. And John Roberts, who liked the work I was doing, and I'd fitted into the team well, said 'Well, if you get a decent degree, come back and do a PhD and I'll find some funding for you.' That's the way it worked in those days. So then when I went back into my fourth year, I knew I had something if I wanted it. I basically shook off the relatively laid back approach that I'd had to my education up until that point, frankly, and thought 'I'd better do well here.' Also, I'd then narrowed down my topics to ones that I really wanted to do. In the final year there was a general curriculum, but really you then opted to specialize in two topic areas. And I specialized in mathematical modelling of

biological systems and radiobiology. They linked in with my interests on the more mathematical and physical side to biology, and also on the radiobiology which had a close relationship with the sort of science I'd been doing in the Institute of Cancer Research.

TT: So when you graduated and you decided to do the PhD. Was this in Pollards Wood?

CR: I did the PhD. Yes. I went back to follow up on some of the work that I'd been doing.

TT: This was DNA repair, mammalian cells and cytotoxic agents?

CR: Yes. The thesis that I was pursuing was to investigate whether DNA repair mechanisms were responsible for anti-tumour drug resistance in mammalian cells. We got two cell lines, both coming from the original rat sarcoma line called Walker cells, and one was exquisitely sensitive to cytotoxic agents, and one was resistant. So, the question was, because originally one was selected from the other, had we selected the resistant cell line or ones which had a greater degree of repair capacity. And we were looking at a range of different cytotoxic drugs, and in particular a current anti-cancer drug called cisplatin. We were using sulphur mustard as the model alkylating agent, but then we were using cisplatin, which was reckoned to be operating by the same basic mechanism, which was cross-linking DNA.

TT: We did a whole Witness Seminar on platinum compounds. It's a fascinating story.

CR: It was a very interesting time. And, of course, John Roberts, who was my PhD supervisor, he'd started his career as a synthetic chemist, and was involved in the invention of chlorambucil as an early anti-cancer agent in the Fulham Road. In those days, the drug discovery was in the academic sector and then moved to Wellcome. And it was developed as a drug by Wellcome, I believe. And so he'd had that background as a chemist, and many of the people in the Pollards Wood Research Station were chemists first and foremost, and were more interested in the biological effects, the biological consequences of the chemistry, and what was going on in the cell.

TT: Were you bringing different techniques, different physical techniques?

CR: The challenge at the time was to try and study DNA damage at levels where you could actually follow cells in a relatively healthy state. For a lot of the early work done on alkylating agents and on platinum you had to use such horribly toxic doses that you never knew whether your results were the consequence of the death of the cell, and its response to the insult, or whether it was really the effect of the primary damage to DNA. There would have been a publication of a new technique for measuring interstrand crosslinks in DNA from a group in the States - Kurt Kohn as the inventor of this - and it involved using very special membranes, and allowing the DNA to filter through the membrane where the length of the DNA was proportional to the speed of it going through the membrane. The longer it was the slower it took going through the membrane. The idea was to establish that technique in the lab, and start to use it alongside the more traditional methods, which used alkaline sucrose gradients to measure molecular weight, and calibrate that new method, and then see if we could drop the dose down to a point where we could measure what was going on in these pretty sensitive cells, which would die at the slightest sniff of an alkylating agent. So my PhD was to establish this new technique, compare it to the older techniques to calibrate it, and then find ways of interpreting results and follow the behaviour of these two cell lines against a range of different alkylating agents.

TT: And this was all funded by the cancer research institute?

CR: Yes, it was all internally funded through the Institute of Cancer Research, some bursary scheme. In those days you didn't really care where the money came from. It was very much more straightforward to get funding than it is today, with all the different research agencies.

TT: Did you have much involvement, or were you out there in your own autonomous republic almost?

CR: We were out there in our autonomous republic. It became clear during the time that I was there that the days of the Pollards Research Station were numbered; it was not viable financially, I don't believe. I think there was a concern amongst the science management team that it was too isolated. And it was true; there were only two PhD-students there, myself and one other, so it wasn't a student experience at all like what some of our contemporaries were having in the university environment. But that didn't worry me; it was a good group of people and I enjoyed working there.

TT: And the computing, was this done there or in the background?

CR: I took it on myself to really build my computing skills. A lot of the techniques that I used produced a lot of data from using radioactive nucleotide bases that you incorporated into DNA. You then wanted to measure that in the scintillation counter - the scintillation counter would then produce a paper tape. You would read that paper tape into the computer and then you needed to do something with it. So I wrote a suite of programs building on some work that had already been done at the institute to handle data that came off scintillation counters. I took what I needed to do and wrote my own set of software in Fortran (I taught myself Fortran at the time), and we used very primitive technology by today's standards, where the institute was connected to Imperial College's computer centre. I had remote access to Imperial College from Pollards Wood by telephone, and an acoustic coupled modem, 110 bits per second connection to Imperial College, on telephone wires that went through the woods at Pollards Wood. So if there was a storm you could forget it. I would spend my evenings writing programs, playing with how I was going to handle this data, writing the programs to do it. And then, as I started to look at the data that I was getting, predominantly from the alkaline sucrose gradients, I realised there was no way in which I could actually relate that to the number of crosslinks that I was generating. So I started to do some work on a simulation program to see whether I could simulate the experimental technology, and use that simulation to come up with something that was close to a quantitation of the number of interstrand crosslinks.

TT: Were these very much your ideas, developed during your PhD, not something you were asked to do?

CR: That's right, it was very much a hobby. I enjoyed my time on the old teleprinter terminal, and I got a lot of satisfaction out of the programming, the creative nature of computer programming, and the satisfaction of being able to interpret the results in a way that hadn't been done before.

TT: How did your colleagues, and in particular John Roberts respond to this? Pleased? Mystified?

CR: Mystified was the biggest response I got. They understood the programming necessary to handle the data and produce nice plots, and get it charted out on fancy equipment at Imperial College, which would send out rolls of, in those days they were using *xy* line plotters, so you came back with very high quality prints from your data. But the whole simulation thing, and in the end I chose to include that as a chapter in my thesis, basically John said 'It looks good but I have no way of telling whether what you've done is any use, but it's interesting and you've written it so keep it in.' And my external examiner couldn't really interpret it either. So it was a bit on its own. But there was enough of the experimental traditional postdoc, postgraduate research thesis that it was probably not necessary.

TT: But you'd been thinking of a different mechanism, a different way.

CR: It was writing programs to simulate an experimental method in order to get some sort of mechanism, some sort of basis for getting a quantitative measure of the number of lesions in DNA, with all the usual assumptions that you have in these sorts of mathematical models and simulations. But it was an approach to science that I subsequently became more and more interested in.

TT: Were there people you modelled yourself on for doing that? Were there other groups, individuals? Could you read about people doing this sort of synthetic, or setting up these systems?

CR: Not really, it was partly my degree in biophysics. In my final year, let me just wind back a little bit. When I was at Pollards Wood in my undergraduate year, John Roberts had kindly allowed me to go off on a one-week mathematical modelling course at City University. He paid for my registration; I don't know where from. So I'd been exposed to people who were modelling biochemical processes; actually it was probably more modelling electrophysiological responses. And I'd got the idea that modelling was something that was quite interesting, and it was intellectually clean, if you like, as opposed to dirty old biology, so I liked that component of it. And then when I went back to university, to poly, to finish my final year, I'd got enough Fortran under my belt, and I'd read up about Monte Carlo simulation to know that in some circumstances you don't have to have the analytical equations to solve a problem, you could solve problems by simulating. So when I had final year coursework material on simulating epidemics, I played with two different modelling methods using the deterministic modelling method with a couple of simple differential equations, or a stochastic Monte Carlo method, and basically played with the fact that there were differences between those two as part of my coursework. And that gave me the confidence to think 'Some of this mathematics stuff is just dressed up in fancy language; when you turn it into computer programs it's not that big a challenge.' Of course, that was naive but at the time that's when I thought I could use this sort of simulation method to interpret the data from my experimental results in my PhD.

TT: So after your PhD you then stayed at ICRF, because you became a research fellow?

CR: This was my big break. When I finished my PhD, I thought I'd had enough of screwing the tops on plastic bottles and putting them in scintillation counters. It all seemed, the rate of progress seemed, quite frustrating to me. Basically your week, another round of experiments would take a week. At the end of the week it was satisfying to have your lab book filling up and filling up, but it almost seemed like this was too slow a pace for me. So I thought 'I'm going to have a go at trying to get a job in mathematical modelling in biology.' And an opportunity arose at the Imperial Cancer Research Fund for a research fellow to look at artificial intelligence [AI] methods in molecular biology. And the genesis of this opportunity really came from Sir Walter Bodmer, who was the director at the time. He'd been in Stanford University, and had interacted with people in the AI lab at Stanford. So when he took on Imperial Cancer Research Fund he created an opportunity to build a group looking at computing methods in biology. And he'd hired a guy called John Fox, who'd come from Carnegie Mellon University, and his interest was in medical applications of artificial intelligence, in particular decision-making and diagnosis. And it was in an era when expert systems were a very popular technology; there was a lot of hype, and he was basically setting up a new lab at Imperial Cancer Research Fund to look at expert systems in medicine. And he wanted to also explore what was possible in molecular biology. Bearing in mind that it was very early days in terms of DNA sequencing, there was very little concept of what we now call bioinformatics.

TT: We're talking about the late 1970s, early 1980s.

CR: Yes. I finished my PhD in 1982, and during my PhD they had sequenced the phi X 174 genome, which was the first complete organism genome that was sequenced. It would've been Sanger probably, Fred Sanger or somebody like that.

TT: Was that a worm or something?

CR: Well *Caenorhabditis elegans* was the worm, but this was predating that. These were small bacteriophage. Everybody was using these huge gel electrophoresis plates with P32-labelled nucleotides in order to generate the sequencing ladders that would then read to get the sequence data, and fragments of 100 bases at a time. This was a very slow laborious process, and it was happening in ICRF at the time. There were a number of molecular biology groups that were sequencing genes related to cancer, and so there was a cohort of people in the ICRF at that time who were interested in how they could use computers to analyse their data. So, fortuitously, the decision had been made by Walter or possibly his predecessor, I'm not sure, to invest in the same sort of computing equipment that was in place. So, yes, the ICRF had invested in what was called DEC computers at the time, digital equipment; a DECSYSTEM-20 computer, which happened to be the

same sort of technology as they were using in the Stanford AI lab. We could benefit from software which we could get for free from Stanford University, so we were playing with, we were one of the few places in the UK where you could run programs that were influential in the early days of artificial intelligence, running from the States. And included in that were medical diagnostic expert systems called MYCIN, which we could run locally and play with. And I was experimenting with a piece of software that had been developed in the Stanford AI lab, which was to handle molecular biology data as well, and there was a long history of, there would be a number of groups looking at AI techniques in Stanford, applied to both chemical and molecular biological processes. And one of those groups was led by a company called IntelliGenetics and IntelliCorp, which were very successful for a while in the boom era of AI in all sorts of different application areas, in the Stanford area.

TT: Was this the beginning of a boom in the UK? Because there's you and your boss, a guy who came from Carnegie Mellon.

CR: John Fox.

TT: Working in cancer research? This had been an idea of Walter's - Walter had had the vision to do this. Was this very unusual? Were there other people in the UK doing this sort of thing? Did you have people to talk to about this?

CR: There had been a report on AI research by a guy called Lighthill in, I don't know what year it was. I think he was a Cambridge mathematician or a Cambridge physicist. And he basically said that AI was bankrupt and the UK shouldn't invest in it. So there had been what we called an AI winter in the UK, and yet around the globe we had big announcements from the Japanese, who announced their fifth generation computing project, and a lot of noise from DARPA and other research agencies in the US, talking about AI. On the back of this the UK government decided to invest in a new programme of funded research called the Alvey programme, and that reawakened the interest in AI in the UK, because it had pretty much dwindled to three main centres, I would say: Edinburgh, at the Turing Institute and also at the University. Actually, the Turing Institute was Glasgow, wasn't it? So there was a strong grouping in Scotland, and there was another at Cambridge, mainly around natural language processing; this is Karen Sparck Jones. There was also a group at Imperial College, and Imperial College was focused around Bob Kowalski and the logic programming community. So there was a good tradition of new computing programming languages in the UK, and much of that had also been inspired by AI type problems, and the whole idea of logic programming was about representing human deduction in a declarative programming language. So for John and for myself, and there was another guy working on the medical side as well, a guy called Peter Alvey, we got a lot of our collegiate interactions with the group at Imperial College at the time, and so we were backwards and forwards. And Imperial Cancer Research Fund and Imperial College, the two Imperials, had quite a lot of exchange. We had bidirectional exchange with teams at Imperial, particularly with the team at Imperial that was more on the applied side of AI and expert systems. And that was a very useful place; I think John had done his PhD at Cambridge, so I think he had some links back to the Cambridge AI teams, and there was also a cybernetics group at Brunel if I remember. There was also some activity. So I might be slightly under representing what the UK community were like.

TT: In terms of medical application, were you pretty much alone?

CR: There was quite a lot of work around the medical applications of computing, but this whole idea of expert systems and diagnostics, there weren't many places doing that. There were other teams looking at other reasoning systems for medical diagnosis, but John was probably one of the more influential and certainly we were, because we had access to this software and these methods that had come from Stanford, it gave us quite a high visibility. And we were involved at the time in British Computer System, AI groups, and things, running, helping organize meetings.

TT: And did you go over there?

CR: I had a whale of a time. I went over to Stanford and I built on a system called the Unit Editor, based around a frame-based representation of knowledge, which had originally come from understanding, building natural language systems, where you represent concepts in a natural language in a unit or a frame of knowledge. We wouldn't use these terms now. They're more like they're concepts and entities and things. So they had this system, which they had then written a language around, which allowed you to manipulate these concepts as entities, and a language for manipulating the data that you stored in these data structures. I had extended that so that you could pull in data from what were then the early molecular sequence databases, the DNA data library that was running at the European Molecular Biology Lab in Heidelberg, which is where it was then. So I took my software that I built to extend their system, I took it over to the States, and then incorporated it into their platform because you could. It was a very useful interaction at the time, but it was a little bit unfortunate in that virtually everybody on the campus at the knowledge engineering lab, as it was called in those days, had gone off for the summer to do their consultancy work in the bay area. So there were all these startups happening, and all the graduate students basically had got consulting jobs in different companies, writing software or helping them build new systems. So there weren't a hell of a lot of people around to talk to, but there were a couple of the key people I was able to meet with, and throughout my career I've had a link with Doug Brutlag, who was somebody who was a biologist involved in that molecular biology computing revolution that was going on at the time.

TT: Tell me more about him.

CR: He was just one of those influential people with a lot of ideas, similar to those of Walter's, I think, where he felt that computers had a much bigger part to play in molecular biology than people had recognized, and that organizing this and reasoning over these datasets was going to be part of the future of how we worked as scientists. There were a number of people there who were opinion leaders in that area at the time. And I'm just trying to remember the other guy that was involved; there was Doug Brutlag and Larry Kedes and Peter Friedland, who founded this company called IntelliGenetics, which was an early bioinformatics company. And it was built around some of the ideas that had come out of their AI research lab, but they turned it into a more-user friendly tool. And that company was then either in a takeover or reverse takeover with an organized company called IntelliCorp, which I think had a bigger interest in the applications of AI. And they were also in the end subcontracted to run for a while some of the bioinformatics resources alongside the US national database, which was at that time subcontracted to Bolt, Beranek and Newman, which was a defence contractor. So this was before the National Centre for Biotechnology, Bethesda, was created. These things were all run by subcontracts by government departments. There was no real cohesive view. And they ran under an organization called Bionet. I think that's how it was organized; it's a bit hazy as to what all the American politics was; but it was still very much based on what we now call open-source software, and those routes of exchange of software between scientists and between labs, without any issues of intellectual property and licensing; this was all going on in the early 1980s. I was benefitting from that – my career was built on being able to access these tools.

TT: You briefly mentioned EMBL [the European Molecular Biology Laboratory]; did you have similar links with EMBL? Developing similar things in Europe?

CR: Europe had set up the European Data Library at Heidelberg, under the leadership of a guy called Greg Hamm, and he had basically built up a software and database team to gather from Europe the sequencing information that was being generated here. It really is a bit of a, although he was American, there was this… I don't know whether the then director of EMBL had given quite a lot of support to this, to make sure that there wasn't an imbalance in the contribution to this effort in the world, whereas we had then what was the GenBank database, of DNA sequences, which was the national database for the US. And GenBank was being run under contract. And then a few years later (I'd have to look back and remember when it was) the three - GenBank in the US, the databank, DNA database in Japan - started to build that relationship that persists to this day of an international collaboration. So I remember going to one meeting where the National Science Foundation and the European Union and the UK, and other government funding agencies were together in a workshop to decide the future and look at new funding methods.

TT: Could you put a date on that?

CR: It was probably around 1986, something like that.

TT: That's comparatively early, in a way. Things were just getting off the ground. You were really there at the right time, or it looks like that retrospectively.

CR: I have said for many years that I was fortunate in my career by having the right interests and being in the right place at the right time. When you end up in one of these unusual niches, you can make a lot out of it, and I wasn't particularly ambitious, I wasn't going around looking to see where I could hop to. I was quite comfortable being the strange computer geek in the corner at ICRF, and building relationships with people there. And doing things that are relatively low profile. As a science topic it was quite challenging, because there's a lot of this which is about engineering and tool-building, and it doesn't make for necessarily a lot of competitive science, so it was quite difficult in terms of a career as a scientist, because my CV was not that impressive by comparison to some of my colleagues who were working in labs, generating data, writing papers all the time. Having said that, it wasn't as competitive either, so I sort of survived because there weren't that many people, and when opportunities arose they could see that I had a set of skills which were unusual and were needed, even if I didn't tick all the boxes as a full blown academic. And I think I benefitted throughout my career on that basis.

TT: So you stayed at ICRF for four years and then?

CR: It got complicated. I think I was on a three-year fellowship to begin with, but, reflecting what I said about bringing over software and making it work for the people, I ended up doing quite a lot of work for other people when I was at ICRF. So basically then Walter and the other science administration team at ICRF said 'We'll give you another year on your fellowship to recognize that you've done a lot of work for other people during this time.' And I think that was helped because the visibility of, for want of a better term I'll call it bioinformatics activity, the visibility of ICRF had become much greater because we'd had a lucky break in terms of a discovery, from having got this software and having installed it on the computer, we were playing with some software from the US, which would allow you to compare a protein sequence with the database of proteins, that at that time was, the software had come from Russell Doolittle's lab, and the database had come from the Protein Information Resource in the US. And so I was working then with a visiting scientist who had come over from New Zealand, called Peter Stockwell, but he was based in one of the protein labs in ICRF. He was playing with some of their sequence data, and they were interested in human growth factors, and he was using as an example to test the software that I was installing and running for the centre, he was playing with this new sequence for platelet-derived growth factor. And what popped out of some of his early tests was a sequence similarity to a human oncogene, and so that was a very early identification of how you could use sequence comparison to say 'there's a functional relationship between these two proteins that we didn't know about before'. So it opened up a new area of biology because all of a sudden you had growth factors and oncogenes, and oncogenes were the new big thing; they were where most of the molecular biology at ICRF was happening. As a result of that the growth factor team at ICRF, whose group leader was, again it escapes me, was able to get a first publication before Russ Doolittle, who had also been working on this problem, and had given us this software, and we had accessed the database. It was a good PR coup for ICRF, but it did cause us some discomfort between Russ Doolittle and, it was Mike Waterfield was the head of the growth factor team, and the oncogenesis group at ICRF. That helped in the way that, because I was part of the background to how this happened, that contribution was recognized by letting me stay on for a bit longer. So then I was coming to the end of my 4 years and I really didn't want to move. This was the time when the Alvey research programme was getting up and going in the UK.

TT: Could you say a little bit about that?

CR: The Alvey programme was an attempt to revitalize the AI research community in the UK. It was very much linked with the industry. I'd been fortunate in that John Fox, who was the head of the lab at that time, had

been looking for funding opportunities outside of ICRF to do some of the other things he wanted to do, so we'd built a relationship with, we knew our way into the Alvey programme. I think maybe he'd been part of someone's steering group or something. Anyway, we applied for a one-year research assistant role, and got somebody to work on an application of logic programming in protein structure representation. And that was a collaboration with Professor Mike Sternberg, who was at Birkbeck University. So we had a three-way collaboration and that went very well. And around the same time I applied for what was then an Alvey-funded - the Science and Engineering Research Council [SERC] had an advance research fellows program - and I applied for one of these and got it, so I then had a five-year fellowship funded by SERC. And that was a bit unusual for ICRF, they didn't take external funds from research councils; in fact, I think they were forbidden from taking research council money by charter, or by agreement, which comes up later when we talk about human gene mapping. And I then had to find a home for my fellowship, and through work within other parts of the UK and the Expert Systems Group at the British Computer Society at the time I bumped into a guy called Tom Addis, who was at that time Professor of Computer Science at Brunel University, but actually transitioned to Reading University. I basically negotiated with him, and I got my fellowship based at Reading University, but I was seconded to ICRF at the time. So I didn't move; it was just a way of laundering the money through the system. And this was tolerated by everybody, and it worked out for Reading, because they got whatever I got as additional brownie points for the university for having an advanced fellowship for SERC, and I got to carry on working for ICRF. And, again, I was also collaborating with that group at Brunel, because by then we'd got a full Alvey-funded project to build some hardware that would accelerate database searching in molecular biology databases, which we were using to deal with some of the scale - we were hoping to deal with some of the scale problems that we were starting to see in the data coming from both protein and DNA sequencing and structure experiments.

TT: What was the remit of your advanced fellowship? Was it something different you had to do?

CR: I can't remember. It was basically building on these ideas of using logic programming languages to represent and reason about protein structure. An important part of that was actually being able to not just look at one protein at a time, but being able to look at a whole database of proteins that were available. And that was where we were running into problems in the ability of the software of the time to deal with the data volume, which, by today's standard, is trivial, but in those days it was still a challenge. So we built this collaboration with, under the Alvey programme, what was then the GEC Hirst Research Centre and Tom Addis at Reading, and Tom Addis had previously been involved in an ICL [International Computers Limited] project which was called CAFS, which was the Content Addressable File Store, which was an early attempt at, arguably, semantic computing, where basically you didn't treat your datastore as a series of bits and bytes, you had concepts in the datastore, and you merely asked for 'give me everything you can with this keyword in it'. So in the early days of ICL, they'd built some of these machines and they were selling them, and they were being used for applications such as the online telephone directories, as it was building then, because they needed to deal with some of the indexing problems. Basically it was a computer store with a built-in indexing system. Tom Addis had an interest in these big databases, so we'd met through a couple of workshops, at the time run by another computing pioneer called Simon Lavington, who had previously been at Manchester. And he had organized a series of workshops on the interaction between big databases and AI. I'd gone along to these meetings to try and say 'There are these biology problems, and we're doing AI things in biology, and we've got big data problems.' So we need this link between what databases could do for us, which is handle the big data, but we wanted the technologies that the AI community had for accessing that information, so that we could both represent and reason things in a much more AI style, but use database technology. So that project was in that middle ground there, and that led me into an interest in databases and what now we call "big data", but we didn't call it that then.

TT: When you were looking at the protein and DNA sequences, how did you get these databases? Did you just have a call out for people to provide their data? Or did you actually try to generate this yourself with colleagues?

CR: This was where I first came across the people at EMBL, and Greg Hamm's group. They were basically offering a database subscription where you just got on their list, and every six months or three months they

would send out another; it was a reel of tape in those days. And they started out as 10-inch spools and moved to the bigger spools, and then you'd start getting them in packs of three. I'd be getting them from EMBL and from the United States and one of the reasons I'd written my suite of software was that at that time they weren't coordinated. So you never knew whether to look in the EMBL database or the GenBank database, so what I wrote was some software that brought it all together, so you didn't have to worry where the data came from, you just had a knowledge base of all these sequences. And that's what I'd linked into the frame base systems that the Stanford guys had built.

TT: And people from around America or Europe were providing this information, with no IP (intellectual property), no sense of licensing?

CR: Not in those days. A lot of it was being culled from the literature, so at that time people were going to the literature and typing the sequences in by hand. But everybody could see this wasn't going to last. So increasingly there were discussions both in the US and in Europe, no doubt in Japan too, about the submission of these data directly at the time of publication. A lot of the work that the EMBL and the National Library of Medicine team, and I think probably what helped secure the National Centre for Biotechnology and Information as an important entity was that these data were going to come through the scientific publication process. It was a way of submitting supplementary material to a journal article, but it was going to be mandatory, and the editorial policy was, after much negotiation, the editorial policy of the journals was changed such that you had to have the identifiers for your DNA sequence, to show that you'd lodged it in a database before they would publish the paper. And that transformed the field; you can see the beginnings of open data and science data from those early days of changing editorial policy in the key molecular biology journals, *Nucleic Acids Research*, *etc.*

TT: Were you aware at the time of one particular person or organization driving that? Because it was incredibly forward-thinking wasn't it?

CR: Yes, it was, it was very important, and I would name Greg Hamm, but he possibly wasn't the only, almost certainly he wasn't the only person. There was another guy called Graham Cameron, who was Greg Hamm's collaborator at EMBL. Graham Cameron subsequently came over to the European Bio-informatics Institute at the Sanger campus, and was co-director there as that was being set up. And so these two, Graham Cameron and Greg Hamm, were part of this sort of political, cultural change movement, to make this happen. Whether they were the only ones in those days I don't know. I suspect there were also people from the funding agencies and things also working on that. I wasn't very close to what was going on there, but I do know it took a while and some persuasion. But in the end it was clear that the different organizations couldn't afford to keep paying people who were simply copying out of journals the sequence data to go into the database. There had to be a direct route. And the scientists would have to become more responsible for their own data. This is still an ongoing struggle, I have to say. We solved it with DNA sequences, but there are many areas of modern, even quite a high throughput science, where we haven't really yet solved this particular concept.

TT: So the model is non-transferable, you mean? Or there's just not a general agreement to do it?

CR: We've seen in the past decade a huge cultural change happening in the community; the whole open science movement, the whole open data, and open access to journals has changed. There's been push from every direction; the OECD [Organisation for Economic Co-operation and Development], funding agencies, national and European governments have all being saying 'science funded by public resources should be open and it should be freely accessible by all'. And I think what we saw in the early 1980s with the DNA databases was the beginning of that movement. I think the early efforts to build in software where the way in which everybody was circulating software for free was very influential in terms of the way the open source community developed as well, and became crucial to bioinformatics. So it was an interesting time, and it still is interesting, but definitely the early days of that were influential.

TT: And this is the late 1980s? You're still at ICRF?

CR: I was an advanced fellow based at ICRF through Reading.

TT: And is this when you start getting involved with the human gene mapping workshops? The first one, at least when you were talking in the Witness Seminar?

CR: Now we're talking about two years into my advanced fellowship; I was approached as one of the more computing-literate members, so ICRF had built a computing team. And the computing team was largely responsible for handling its clinical experiment data - I won't call them trials, because they weren't always trials - but they were doing quite a lot of work in extracting information from patient records, looking at their response to new combinations therapies, and there was a small team building databases and a set of interfaces to report this data, and being able to analyse it. And there were data managers in all the ICRF units, where they were doing this sort of clinical work, and there was a team back within the computing services department at ICRF, building databases. Originally in a system called 10/22, but later on in Oracle. And that team was in place from the early 1980s; that was one of the motivations, I think, for the computing infrastructure at ICRF, was that we had the clinical units doing the clinical trials. And there was a small statistics group as well, and a very big epidemiology and statistics group in ICRF, of which Jack Cusick was the head while I was there. But also a colleague at Leeds.

TT: Tim Bishop?

CR: Tim Bishop joined later, based in Leeds. So there was a big statistics group at ICRF as well. Not big, but there was a reasonable sized group of people looking at the epidemiology and some of the other aspects of cancer, and they were also using the computing infrastructure. There was a need to have the statisticians, the clinical scientists, and the database together. And that was supported through the computing services group at the time.

TT: We were coming up to human gene mapping workshops, and how you started.

CR: During my time at ICRF I'd also been involved in some MRC grant panels and in the broader community in the UK around bioinformatics, because it was a relatively small community. So I'd been asked to sit on the MRC panel.

TT: I wanted to ask you specifically about the panels, because you've been on a phenomenal number of scientific committees, which perhaps shows how unique and rare you are, that you're used so much on these or you're very amenable and say 'Yes.'

CR: Probably a bit of both.

TT: But we can talk about that later. Let's talk about the ICRF, and HGM [Human Gene Mapping (workshops)].

CR: Before HGM 10.5 happened, ICRF had agreed to organize these workshops in the UK. And I think Ellen Solomon had basically said we would do it, and Peter Goodfellow, and then Walter sort of came in alongside that and said 'ok, we'll do this'. And we also expected at that time, I think, that in taking on these workshops we were going to take on the need to supply the data infrastructure to do that, which largely meant taking on from the HGM 9.5 conference, which had been run in Yale by Ken Kidd, the dataset that he'd got. The idea was to take what he'd written for PC and turn it into something that would run on our VAX VMS and use Oracle. But at the same time, the Howard Hughes Medical Institute had been looking at what was going on in the human gene mapping project, and Peter Pearson and colleagues at Johns Hopkins had got this additional funding from Howard Hughes to set up a database of gene mapping information. Of course, that was on the back of OMIM [Online Mendelian Inheritance in Man], and the work that had gone on at Johns Hopkins for a number of years to capture the clinical information around human genetic defects. So when they'd secured the money from the Howard Hughes to rebuild OMIM and support that with Victor

McKusick, then it was a natural extension to think 'they want more of the genetic mapping data that's appearing from the human gene mapping conferences'. And for a while there was a bit of a standoff as to who was going to be doing what. It wasn't always an easy exchange of positions because ICRF were saying 'we're going to do this for the workshops and we're going to use Oracle,' and Howard Hughes and the guys at Johns Hopkins were saying 'we're going to do it and we're going to use Sybase.' There was a bit of a technology tussle, a bit of a science and technology leadership tussle. At that time I wasn't fully involved, people were asking my advice, and in particular as a result of my MRC links. There was a lady called Bronwyn Loder, who was if you like building, supporting Walter and others around this human gene mapping, in HUGO [Human Genome Organisation], in the human gene mapping organization. So Bronwyn was there, and she'd got caught up in helping with trying to resolve the political situation. In the end the agreement was that, because the amount of resources that the Johns Hopkins was getting from Howard Hughes to build a team, which was the genome database, the GDB team, and implement this national infrastructure with Sybase, we couldn't compete. We were not in a position to move that quickly, and ICRF wasn't really well placed to support the software engineering side of that. We just didn't really have the capability, whereas in the medical school they'd got a digital library. At Johns Hopkins, the William Welch Medical Library. It had quite a strong background in genome, and I think there was a guy called Dick Lucier, who was the head of the library, and had worked closely with some of these teams. He was the project manager for the GDB, and he was the one that was basically making sure things were happening, and had the grant from, it might have been Peter Pearson that had the grant, but he was Peter Pearson's chief operations officer, making sure everything was happening, very business-like and very organized, and professional and competent, quite a tough cookie. He was making sure that this GDB was going to be as ready as possible for the human gene mapping conferences. At that stage I was asked - because there wasn't enough capacity in the IT team within ICRF, I was asked if I could take on this particular project as a project leader at ICRF, and build a team that would support the workshops and do some of the database work. It seemed like it was a bit of a distraction from what my research programme was going to be, but it just seemed like one of those things that I should do. It was the right thing for ICRF, and it was an interesting different sort of thing for me to take on.

TT: So this was whilst you had your SERC fellowship?

CR: To be honest, I just carried on doing my SERC fellowship, and just did this as well. There really wasn't a lot of scrutiny about what SERC fellows were doing. I had a small team, and we were writing software and publishing, but they were having to get on with it while I was another floor up, with a different team. And so I was pretty much at that stage a bit detached from the research group that I'd got through as a result of some EU funding. I was building in an EU-funded project then as well. Again, there was some EU money and some SERC Alvey money that I built a small team around. In the midst of all this I'd basically moved on to another project and it was difficult, it wasn't an easy time, there was quite a lot of politics.

TT: You referred to that in the Witness Seminar, you had an interesting expression, 'A lot of political tussles in the background as to who was going to build this database, how it was going to be implemented, all the technical complexities, and who was controlling the whole thing.' You were already hinting at some of the tensions and, I wouldn't say disagreements but, disparities?

CR: It was; the tensions arose, I think, from where the ownership of the data and the database resided. And I think there was a, there were two players in this. There was the human gene mapping conferences as a loose federation of scientists, of whom Walter Bodmer had taken on the chairmanship of that conference for the HGM 10.5 and HGM 11, and therefore if you like was the custodian of that data on behalf of the human gene mapping conferences. And Walter felt that they should be the ones with that ownership, and have the scientific control over what was being done. But at the same time we had the Johns Hopkins people who had basically got the money and resources to say 'we can do that', and the last of the conferences was based in the States, and had been funded by Howard Hughes quite extensively in Ken Kidd's team in Yale, so the Howard Hughes also felt that they'd got a stake, and they wanted to see it turned into something that would complement OMIM and Victor McKusick's work. So I think there was a vision; this is more with hindsight. So I think the tension came from there, that this was actually about the scientific tension – that was the

major one – but then there was this minor tension over what database technology. And this reflected a bit of a US versus Europe view of how, what tools we would use. And it was, with hindsight, a bit of a silly squabble, but each team had an investment in a particular software platform. In the UK and ICRF in particular it was Oracle, and in the US it was Sybase. And the reason for Sybase being there over Oracle was as a result of two things: there had been a technical evaluation by a group of scientists, I think probably connected to the GenBank project, about which of the relational database vendors at the time produced the most performant technology that should be the basis for future genome project work. And so Sybase had come out on top of that competition. And then the National Science Foundation had negotiated with Sybase corporation a very substantial discount on what was very expensive software for use by US scientists, in fact used for scientists around - I think it was international. So anybody on a human genome project could potentially take advantage of this Sybase deal. Naturally that was the way in which the Johns Hopkins group was going to go, so they hired a bunch of people and consultants from Sybase, to make sure they could deliver on the Sybase platform. And it was a nice piece of software, I'm not denying that, and Oracle had probably rested on its laurels as a provider of software and database technology to the business sector, and Sybase was stealing a march on them from a science and technical position. These companies leapfrog one another all the time. But still that was also the tussle. So in the end we also had Sybase at ICRF, the clinical team continued using Oracle for their products and projects, and we in the genome mapping project were using Sybase. But we used it in a relatively minor way. It was agreed, given the resources, because I was trying to build up a team from nothing, because there hadn't really been any full allocation of any staff to the project, so I'd taken one member of the clinical database's team from the IT group, to start my project off, and then I had to hire a number of people to create a human gene mapping computing team at ICRF. I think at our max, we were about five at one stage. And by this time it was clear that the only thing we could really do was make sure that the workshop that was being organized in Oxford, HGM 10.5, would operate smoothly. What we did actually was be very pragmatic about it. We said 'we'll take the PC platform that Ken Kidd's team had got', and we would just tidy that up a bit and use that as the basis for collecting information from the different chromosome committees while they were in their preparatory phase, leading up to the conference. And it would be our job to make sure that data got into the relational database built by the Johns Hopkins so that they could then edit the live information when they got to Oxford. And I then didn't spend so much time worrying about the GDB archive database; I was much more concerned about running the relationship between the ICRF team and all the different editorial chairs of the different chromosome committees, because they were the ones that were basically being given a PC application where they could make updates to the then gene mapping information. And that was pretty much based on the work that Ken Kidd's team had done back a couple of years before. The revolution wasn't to come until we went to Oxford. That was sort of one half of my job, to make sure that everybody was getting, there was information flow between the people, there was data, we were sending out floppy disks to them that they could install on their computer, they had installers written, so it was much more a PC team that I had. So I stole a couple of people from the IT team who had got a bit more of the software.

TT: And how long did you have to do this?

CR: I think, from when I agreed to help out, it was probably only about 6 months to do this. So the other half of my job was to liaise with St John's College, Oxford.

TT: Wire that all up and put it in people's bedrooms.

CR: They basically said, we booked a chunk of the modern part of St John's College, and I then worked with some members of the Oxford University computing service, who were very helpful actually. Also whoever it was in charge of St John's College, I can't remember, and Ian Craig of course was at Oxford, and was able to make things happen for me on the ground. We would go up there and we looked at the room, but we had to bring in networking companies to cable the whole place. We also had to run networks between St John's College and Oxford University computing service, because what we did was we installed some computing equipment into their machine rooms so we could access the database. So this meant setting up in the short term what were in those days quite big pieces of computing kit, and we had to source all the computing technology to put equipment into people's studies.

TT: How many people are we talking about?

CR: Good question. So there was, I think it would have been about 70 people, something like that; 60-70 people. Because you got, for each chromosome, you had got a team of two or three who were doing the research.

TT: For each chromosome committee?

CR: Yes. So each chromosome committee would have a bedroom in this hall of residence, they would meet in this bedroom and do their editing. And what we did was to, on that occasion, equip each bedroom with a PC with simple terminal- emulator capability that would link back to a Unix machine running the genome database software in the Oxford University computing service. So I made use of some contacts that I'd got with Hewlett Packard, because I'd got a research project with Hewlett Packard, and they put us in touch with their PC team and they lent us a whole bunch of 40 odd PCs for nothing. So for the 10.5 we were able to beg, borrow, and steal enough computing equipment, but we still had to do a hell of a lot of rigging things up, and bits of cabling.

TT: It sounds a phenomenal amount of work for one meeting, and it did when you were talking about it at the Witness Seminar. Had that sort of thing ever been done before?

CR: They'd done it before at Yale, and I think we were a transition. After Yale they said 'this has to change', they had to put quite a lot of effort into building that first PC application. They'd set up the framework for collecting the data on floppy disks, and being something they could do electronically, rather than on paper, whereas previously they had done it all on paper. So that was the beginning of the move, and 10.5 and 11 were basically the end of the move. Because from then on it was clear that setting up this computer infrastructure for a meeting in any conference centre was too expensive; nobody wanted to fund it. And the future was in a live database which could be edited all the time. The conferences could then become just a meeting to finalize and agree the state of what was already in the database.

TT: Rather than actually to do the work.

CR: Now we'd call it a Hackathon or a Jamboree, and that model was very difficult to accommodate. I mean now with everybody having a laptop it wouldn't be such a big deal, we'd all be doing it in the Cloud, and we wouldn't have to set up this infrastructure, but in those days nobody had laptops.

TT: Who was funding 10.5?

CR: Predominantly ICRF agreed to pay the bill for 10.5, but we'd had a number of meetings with the MRC to secure a grant. So we'd also got funding from the Medical Research Council, but basically ICRF underwrote it.

TT: Why did you have it at St John's? Because 11 was at the ICRF.

CR: Eleven wasn't at the ICRF; it was just down the road in the Connaught Rooms, as ICRF didn't have any location where it was suitable, and so…

TT: So it had to be external?

CR: I think it was an Ian Craig connection to Oxford. Walter had been at Oxford before he'd come to ICRF; he was an Oxford guy. So it was natural that the Oxford genetics lab, which was probably still one of the most influential human genetics labs in the country, was a natural home for it. There was also Kay Davies there, and she was also very much involved. And I think it was just where the centres of politics and science were for the human genetics in the UK at the time.

TT: And that was 10.5; you were actually there overseeing everything?

CR: I was sort of like the Master of Ceremonies. It was my responsibility to do the work on the site, get everything ready. The Sybase guys and the Johns Hopkins team arrived; I think we ran the meeting just after a bank holiday, and we'd been in there the week before - it was a university holiday, so we'd been in there doing the cabling, and we'd got some of the servers, these big Sun microsystem servers, in place. The guys from Baltimore came over, they were installing their software, their database, getting it up and running. We were making sure all the PCs could connect to the database, making sure all the software was working. I'd also describe it like running a rock concert, because you had to get everything in and working.

TT: It does sound quite a phenomenal exercise.

CR: I was very lucky in that I had a couple of quite experienced computer networky guys helping out that I'd hired as consultants. They could turn their hands to anything, they were capable of making a cable to fit between the ports, and we were doing a lot of making stuff up by hand. So everybody was just running around doing everything we could. We set up, in the Junior Common Room, a sort of operations centre where we could fix things. The Genome Database team were in another bedroom where they were doing the back-end system database stuff, and then all the rest of the people were in their bedrooms editing chromosome information. So it was very frenetic, and it wasn't without its glitches, and things going wrong. But, on the whole, the technology stood up, because we were all very worried that, this was the first time that anything like this had gone live with anything like the number of users we were putting on it. It was a complete guess as to how big a server we needed, and how much data we were going to collect, and what the overheads were and things. It was very complicated at the time.

TT: So the success of that led to 11, which was fairly soon after? Had that been planned?

CR: Basically they were consecutive summers. We'd committed to 10.5 as part of our deal. So 10.5, the .5 conferences emerged just because of the amount of work that needed to be done as the number of new markers were appearing. And what we were talking about was an era where we were going away from gene-based markers to these anonymous DNA fragments. These were just bits of DNA markers that were more and more populating the literature, and as being linked to particular loci that were involved in disease. So that was vastly changing the number of markers that needed to be curated in the database, and we were seeing an exponential rise in the number of things that needed to be curated. And so 11 was the big one; 10.5 was only the chromosome chairs and their committees. Eleven was a full open conference, so we were running a full scientific conference, but just before that, or alongside that, I can't quite remember, we'd run another of these jamborees to collect the data and bring that up to date. I was working alongside another team at ICRF who were organizing the conference itself. So we needed a big venue like the Connaught Rooms, which has a huge number of rooms; it's about three different buildings bolted together. Do you know where the Masonic Temple is? It's between Lincolns Inn Fields and the Masonic Temple, the big Connaught rooms. Very smart now, I've been back since a couple of times. The déjà vu feeling is there; I remember walking underground round in the corridors, agreeing where the cabling was going to run. We had to do exactly the same thing again. But this time the computing requirements were put up a notch because the sophistication of the database was growing such that logging in only with a PC over a terminal server wasn't really ideal, so we were putting a SUN workstation, a UNIX workstation, a graphics workstation, into every room. So what we did was we put one SUN workstation and one PC terminal, or more than one PC terminal, that would then plug into the back of the SUN workstation, and then we had to run a full internet Ethernet connection throughout the building and back to a local server that we installed in one of the Connaught Rooms, with mobile air conditioning units to keep it cool, because it was blazing hot. And then we all sat to run a network between that centre and then the ICRF machine room, where we had another backup server, a big backup server. We were trying to mirror the database between the two sites so we never lost anything; it was a nightmare. And that was, there were people sleeping under the tables overnight in order to make sure we were there in case something went wrong. That was in a way more tense.

TT: An incredible investment. Time, energy, manpower, just the whole effort. So, after 11 that was when

people realised 'we can't carry on doing this'?

CR: Yes.

TT: And were you involved after 11?

CR: The product of HGM 11, as had been of HGM 10.5, was the printed publication that was coordinated with Karger Press. So, the other side in 10.5 of my job was the data collection, and then the provision of a camera-ready copy that could go to Karger for publication. That was also another piece of major coordination. For 10.5 we did it in the traditional way; we took the data from the data tables, and we put them into Word - well, we put them into something else actually - but we did it with a good old desktop publishing approach. I was a big fan of a desktop publishing system called Framemaker at the time, and I think we either used Word for Windows or Framemaker, I can't remember. At ICRF we were all Mackintosh-based, so I think Framemaker was the thing of choice. But we did it as a series of exchanging text files and things, and doing it that way. We didn't do anything fancy. But, of course, for HGM 11 all the data was in the database, so there was no longer a need, if you like, to go back to the original chromosome editors to make sure that we hadn't mis-typeset it, because the real data was in the database. Then the problem was to get the data out of the database and into a typesetting language. I got some experience of converting electronic database content using a markup language into what was then SGML, basically. It went into a markup language that Framemaker used, so I wrote all the software that took the contents of the database and turned them into a marked-up form, which could then go into Framemaker. Then it was almost a camera-ready copy from the database, and we just needed to send these things backwards and forwards, basically by fax, to make sure that we hadn't messed up something in the punctuation where it wasn't in the database, but you needed punctuation when you put it into printed copy. But that was a much bigger enterprise, because of course the volume of data had grown, but also the logistics with the generation of this thing as a body of text, was phenomenal. It was all fairly boring, because it was just tables and tables.

TT: But it was vital stuff.

CR: It was vital stuff, and that was the only way we could distribute it at the time. The next phase for the team at Johns Hopkins was to make the database open and accessible as a resource online, and that's where they were going with their results. But I had the responsibility for producing the printed copy, and collating. It's a full, big two-volume with text articles from each of the chromosome chairs as well, and also from the informatics committee, which nominally, I was co-leading with Dick Lucier, because we were responsible for parts of the data. So we had to write little articles and explain what we'd done. There was all the stuff that you needed the text around the data, so we basically had to put all that together as an editorial team as well. Once that was handed over, we disbanded the team, most of my team got jobs back in the ICRF, they went back to roles in the IT team, or they looked for jobs elsewhere. And so basically it was all wham-bam.

TT: What about you? We've concentrated a lot on HGM logics, and where you fit into the programme, but it's only a small fraction of your career.

CR: Two and a bit years of intense focus.

TT: What happened to your career after that? You went back to ICRF but you went back as head of bioinformatics because your SERC fellowship had now finished.

CR: The SERC fellowship ended, and I'm not quite sure, I can't remember how it happened, but I was basically offered a tenured position at ICRF to head up a bioinformatics and clinical trials database team, and it was called the biomedical computing unit. So the idea was to take the database developers from the IT support team, where they were a bit out on a limb anyway, because the rest of the guys were just looking after desktop PCs and doing networking and things, and there wasn't really a software culture in that team. To take that team and then put it alongside my group, which was this research group, still playing with AI methods in molecular biology, being run largely by a postdoc of mine, Dominic Clark, who was there at the

time. And then forming a new department, who would have a joint research and support activity. So I then had a small bioinformatics group providing support for molecular sequence databases and software, both the ones we ran on the servers and the one we ran on Macs; and also the database development team, which was working with the ICRF clinical units to support them in terms of curating their data. So that's when I got the tenured role at ICRF for that period of time.

TT: And in that role, how reactive and proactive were you? Were you developing new things and trying to push them out to the punters, or relying on people coming to you? Because you're mixing two things there, service and research.

CR: And I continued to do that. It can be quite a challenge. I've always argued that there are certain areas of science where the technology dominates, and bioinformatics is one of these; where there's a very thin line between providing a technical solution for somebody and the research that you need in order to develop the next generation of technology. And you need those close together, otherwise the service becomes too wound up in delivering service, and doesn't necessarily have the right people to look over the parapet to see what is coming, and they don't have time to do that. Whereas a research group is good at exploring lots of new opportunities, but are tied up with doing the research. And you need these two activities side by side, and that's what I wanted to create at ICRF, but it was actually quite difficult to do so, not least probably because the research that I was doing was largely funded. My research was entirely externally funded when I was at ICRF; I had a European grant, which was a joint grant with the European Computer Research Centre in Munich, and a number of other partners. And so we were pursuing the interests that I had had in how we might automate some of the protein structure prediction and protein structure data mining technologies that AI provided, working with these new technologies that were coming out of the logic programming community, and that wasn't really that close to what ICRF was interested in.

TT: They must have supported you getting the grants? The ethos was changing; you were allowed to have external funding?

CR: I was one of the few external grants. John Fox was head of his own; I don't remember what his department was called; it changed its name. So he also had quite a lot of external money, but we were quite unusual, and it was really because we were on the edge of the IT community; there were opportunities for us to apply for grants. And it clearly wasn't work that ICRF would necessarily fund; it was too close to computer science for them to justify spending charity money on. We understood that, and it was supported that we should go outside and get this, to do these things. Because it brought additional funding into ICRF. So we were doing that sort of thing, and that was quite good.

TT: You were saying there were still these tensions about balancing the research, because that's an obvious tension in your kind of role, people expecting you to come in and tell them at the end, rather than being part of the team. You've already got a research team and experience, so taking on this other role, was it very difficult to manage that, and also manage expectations of what people think you're going to be doing for them?

CR: If I think of a time in my life when I didn't handle the politics well it was then, in that I found myself spread too thinly between these responsibilities. I didn't look hard enough, or I was too naive, frankly, to see that I was putting myself out on a limb at ICRF, and basically the point came to - ICRF has this quinquennial review system - and so by the time I came up for quinquennial review, my research team hadn't really delivered enough for me to, I and my research team had not delivered enough in terms of publications and science, to do well at that quinquennial review. So when I was put up, they don't review one lab at a time, they review a set of labs, and my lab was noticeably different in its outputs, because nobody really counted software and databases, things like that, as scientific outputs, and still it's an issue for people in my position. There aren't the mechanisms for recognizing those outputs in the system to help them build their careers, and it remains a challenge.

TT: Yes. Things like that don't count. It's really difficult; how do these compare with papers in *Nature*

or *J. Physiol.*?

CR: The funding agencies have recognized this, but actually the evaluation systems don't really. It's improving each time we go through the REF [Research Excellence Framework], or we go through, as I work in a research council institute, so we have our own evaluation process. Each time they want this sort of thing, but actually it plays very much second order. And the whole model of how you measure the influence of a person's research career on their peers and on the community is very much changing as we see, these metrics and technologies that help us look at these other things.

TT: Going back to ICRF and the quinquennial review, your research was shaky by this stage?

CR: And I'd also run out of funding, so I was completely vulnerable. There were a number of different things that changed the context for me. My research had not really produced the number of scientific papers in the high enough impact journals. The culture again, in the computer science community, with which I was mixing, was more around grey literature conference papers, and my colleagues in the biology community just didn't recognize these as proper publications, even though they were going through an equally rigorous review process. So I wasn't publishing in *J. Mol. Biol.* or *Mutation Research*, or all the normal peer review processes. A lot of my papers had been published with computer scientists in conference proceedings. So it was, I didn't look good alongside their type of science. I put my hand up and say that line of research was not going to produce the sort of volume of papers that you would get out of a more traditional bioinformatics team probably. And, of course, because I was working in an area where I'd chosen a niche in protein structure bioinformatics, which was more exploring new technologies, so therefore I wasn't, and really proving that approaches would work or not, it wasn't really delivering into the biological sciences literature, it was delivering more into the computer sciences literature. So other biologists at ICRF weren't coming to me and saying 'can you apply your technique to my data', which is what was happening to some of my contemporaries in a more traditional protein informatics community, the likes of Janet Thornton, Mike Sternberg, and Willie Taylor, who were my contemporaries, who I was working with, but not necessarily at the sort of science level that they were operating. It just wasn't where I was working. And so the context at ICRF changed. I'd not done that well in my quinquennial, and it was fairly, the money had run out of my EU projects, there was nothing particularly obvious in the pipeline, and I'd also done something which, although it did my international profile a lot of good, it didn't do my local reputation any good, which was to get heavily involved in the foundation of the International Society for Computational Biology, ISCB, and that was on the back of co-organizing a new range of international conferences called intelligent systems for molecular biology. So this had started in a conference in Bethesda, at the National Library of Medicine, in the couple of years previously to my quinquennial review. There had been a meeting of people involved in AI and different aspects of molecular biology, who got together to say 'we're part of a new community, and we're interested in aspects of applied AI for molecular biology problems'. That was run by a guy called Larry Hunter, who was then at the National Library of Medicine. Then a second conference happened the year after at Stanford, again with the old guys back in Stanford research. And at that time, I thought 'we're doing some of this, we'll organize the next one'. So I organized with Dominic Clark, who was my postdoc at the time, to hold one at Cambridge, and we held that at Churchill College in 1995. And the combination of my relatively weak quinquennial review for my lab, plus the distraction of running this international conference on a shoestring, spending all my time on it, meant I'd really taken my eye off the ball in the local politics. The third strand of this, which had more impact on my career, was that I'd been asked to join a genomics advisory board by SmithKline Beecham by the then director of research who was George Poste. And myself and Kay Davies, actually, from the UK were travelling backwards and forwards to help them build their bioinformatics and genomic strategy for SmithKline.

TT: That was in?

CR: 1994, 1995. So we're going to Philadelphia. I think I started in 1994, doing that, and then all of a sudden they hired an old acquaintance of mine, David Searls, to be the Head of Bioinformatics, and I really liked this guy. We'd met at these different conferences and always got on, and we had a similar interest in AI, and approaches in bioinformatics, and I just said to him 'If there's ever an opportunity, I think I'd enjoy working

with you; is there any chance?' And I hadn't realised but they were recruiting for a UK director of bioinformatics, because up until that time they'd built their bioinformatics group in the US, and they'd really left their UK team behind, and they'd not succeeded in hiring anybody who they wanted to have as a head of bioinformatics. So they said 'Yes, give us a couple of months while we sort it out.' And basically I fell into a role at SmithKline Beecham in Harlow, which was very handy, being in the right place at the right time, with the right skills.

TT: And the right contacts.

CR: Yes. So I basically jumped before I was pushed from ICRF. And then completely changed my career, really. I went out with a great deal of trepidation, and went into the pharmaceutical sectors. But it was a time when the pharmaceutical sector was changing and opening up, and not being a fortress mentality any longer. It actually meant I could maintain a lot of my academic links still, so I still had some residual links back to colleagues, and people were joining SmithKline Beecham from ICRF. It wasn't long after I joined that Peter Goodfellow became Head of Discovery at SmithKline. And then he hired various other people who'd been geneticists and genomic scientists. Nigel Spurr was one, and I worked quite closely with him at SmithKline; he'd been very much a part of the organizing committee for the human gene mapping conferences, so it was very familiar. There was a big ICRF diaspora at SmithKline at the time, and they are all over the place anyway, and so that was my move, and then I stayed there from 1995 until 2000, basically.

TT: What were your main responsibilities there?

CR: I was asked to set up a research group and a service group, side by side. Because at that time, the support for bioinformatics at SmithKline was huge, and I think to a certain extent they were trying to outdo Glaxo Wellcome, who had been quite visible and vocal in supporting bioinformatics as a research discipline in the organization, as well as building the infrastructure within Glaxo to support the use of this new data resource for drug discovery, or for target identification which would lead into drug discovery. And SmithKline had made a decision to partner with a US company, HGenome Sciences, to have prior access to the human genome sciences EST [express sequence tags] database; this was the early days, this is when the intellectual property and patenting of genes boom was going on. And there was this huge scrabble to secure intellectual property around genes and the genome. And at the time there were these two companies, Human Genome Sciences and Celera. There was a lot of, you either went with one or the other, and SmithKline had done a deal with Human Genome Sciences, and the concept was that, as well as having access to the data for our own discovery, we would license on to other pharmaceutical companies in a pre-competitive collaboration, to allow them to mine the same dataset, and we would support them from a bioinformatics perspective. Most of the deals that SmithKline had down were with European pharmaceutical companies, the smaller ones. So Boehringer Mannheim - at the time there was also another French pharmaceutical company Synthélabo, and a German one, Merck, as well. There were three European-centred collaborations that were being run by SmithKline, but most of the support had to come from a European office. So I said 'we want you to set up that support team', and so I hired a number of people who would then act as the bioinformaticians that would help these external companies train their scientists in how to use this database, and the resources that we'd built, to give them access to be the contact point for a broader alliance in pretty competitive drug discovery, which was a sort of vision. It was called the ATG team, maybe AGT, the Alliance of Genome Technologies. I think it was ATG but I can't remember what the acronym stands for. So we also had a link to China because we'd also got involved in the early days of the Beijing Genome Institute that SmithKline had, and had started a Beijing laboratory. So there was also this – it was called Project Dragon at the time - at SmithKline. So there was this whole international dimension to what was happening, and I was very much part of the support network that made that happen. But still the engine of bioinformatics was largely based in Upper Merion, in Philadelphia. So I was very much, again a bit like I was when I was at Pollards Wood, at an outpost, doing what I could but without a lot of resource. I had resources in terms of the outward facing work, but the amount I could do for my UK colleagues was relatively limited. And as companies do, there were a number of different reorganizations around what responsibilities were done by whom in the team, and eventually it was agreed that I would focus on UK science, but SmithKline had a very strong trans-Atlantic ethos. They said 'It shouldn't matter what side of

the Atlantic you're on, you should have access to everything, you should have reporting relationships with your managers, it shouldn't matter where you are.' We were living on teleconferences, everything was transnational. So, as the department grew, and they reorganized, instead of being focused on the national support teams it was about areas of biology. And I took on the support for the Human Genetics Group, which, because of Peter Goodfellow's and Nigel Spurr's arrival, was based in the UK. There was also a genomics team looking at the early days of microchip transcriptomic datasets. And again there was quite a big team in the UK doing that as well. So it was agreed that I would take over the genetics and genomics activity, so anything to do with human genetics and the transcriptomic datasets I would lead for the entire of SmithKline. But that meant I then had a team in the US and a team in the UK, looking after that side of things. So then I started doing a lot of trans-Atlantic travel, and in some respects that was quite a satisfying time because I had much more of a close, I needed to be much more part of the team, I was no longer 'Oh, that group in the UK and Chris.' I was very much more part of the science team in SmithKline, in the bioinformatics group. And the department was round about 65 people at that time, and I probably was managing 10–15 of them. So that was a good time.

TT: This was all SmithKline Beecham. Were you still there when it merged?

CR: They had a similar sized team. So when it became GSK [GlaxoSmithKline] they were massive, and I wasn't really that closely involved. Once I left I didn't really hear that much.

TT: Why did you leave? Because you then went into another, Oxagen, a private company?

CR: Although I was enjoying my time at SmithKline there were a number of reasons why I left SmithKline Beecham. Really I'd found it quite frustrating to not be the master of my own destiny. And so I found, or somebody pointed out to me, that there was a job going in the spinout from Oxford University, where they were basically looking for the same set of skills. So the opportunity to not have to travel to the US on a regular interval, and be based at home, and be the head of an activity, and run things my own way, in a smaller outfit, was very appealing.

TT: The same kind of responsibilities in terms of research and service?

CR: It was pretty much all service. And in fact towards the end of my time at SmithKline most of the people involved in research had moved on. And it was decided that research for its own sake wasn't going to be how we contributed it. But of course a lot of what we were doing was forging new paths, and taking new approaches in bioinformatics.

TT: What was the main area of Oxagen?

CR: Oxagen was founded by a number of Oxford scientists, who had got large collections of disease families in chronic human disease. So it was a joint venture, spun out of the Wellcome Trust Centre for Human Genetics and the university. John Bell was one of the founders, but there were other people as well. And from that they decided that, by pooling all their disease families together, and their own genetics research, and giving DNA samples from that collection to a company, the opportunities to do the types of research which would help identify new causal genes in chronic human disease would be a significant commercial opportunity. So I joined Oxagen; it had probably been in existence for a couple of years when I joined it. It was pretty much in full swing, but they had really not built a bioinformatics capacity, they'd concentrated on the genetics. So when I moved in there was quite a strong team looking at genotyping. So they were doing high-density genetic markers of tandem nuclear repeat type markers, microsatellite markers. They were using standard panels of microsatellite markers, about 400 across the genome, and doing what you might call traditional linkage analysis, multipoint linkage analysis, for human disease, and we were looking at osteoporosis, coronary artery disease, autoimmune disease, psoriasis. It was pretty much driven by – and asthma – pretty much driven by the clinical interests of those people who had contributed into the partnership from the Wellcome Trust or from Oxford. But there were a few others as well, who had come alongside. And, of course, we were doing interesting science and we had a genetics team and a molecular

biology team, who were attempting to clone the genes that we were identifying through positional cloning techniques. So I was asked to come in and basically build a more substantial infrastructure that would enable them to move up to the next step, which was where we were going from just looking at our own genes that we were identifying, or picking out from these pretty large and broad QTL [Quantitative Trait Locus] regions in the genome, to mining databases that were in the public domain. Because we had a substantial injection of venture capital into the company, not long after I joined I had to grow the team quite quickly, because we also signed licence agreements with Celera, to mine their database. So we then had to put a really substantial bioinformatics pipeline in place to trawl the data for things that we were interested in. So I built a sequence-based bioinformatics team that would do that, but I'd also had a software development team building a database that would capture what we were doing within the company, linking it up to assets that we had identified in the database. Because in the end we were responsible to report any new discoveries, and any licensing, that we would do with third parties, that might have been built on the data that we had got from Celera. Basically, it was an intellectual property, we'd got licences, it was expensive enough, but the real value to Celera was if we discovered anything that subsequently became a drug target. So my main team was a genotyping team, looking at quality control, the genotyping team. I had a bioinformatics group looking at the sequences, a software development team, and a quantitative genetics team. So we were doing all of that stuff, and each one was about three or four people. So I had a team of about 20 looking after that in a company of about 80.

TT: And this was much more hands-on; you were much more directly involved?

CR: I was definitely much more hands-on. I was very much involved in the management team in the company. I was involved in some of the commercial development, in a minor way. We had a commercial director dealing with the new business partnerships, but I was much more involved in talking to potential new partners, explaining what we had built, how we could add value to another company, things like that. So it was without a doubt the most fun I had in my entire career. Because I had this, coming from the pharmaceutical sector, and seeing how drug discovery worked, and how medicinal chemistry and that engine worked, it was interesting to go to a small company. It's often the other way round. And then see how a small company dealt with trying to then sell its inventions into the big pharmaceutical sector. So although we were cash-rich for a while, we were having lots of ideas and pursuing all sorts of different avenues, it was very difficult, the pharmaceutical sector was starting to retrench at this stage. This is from 2000 to 2004 basically. So then it was a case of the small company having to take more and more of the risk in terms of the commercial development. So genetics, genomics, and bioinformatics were where Oxagen started. It then had to say 'Well, what have we discovered? What are we going to pin our hopes on in terms of new drug targets? Because companies don't want to sign a deal with us just because we've got a coronary artery disease collection.' The only company that signed up with us at that stage, because of our access to clinical samples, and the DNA and the genotyping, was AstraZeneca. So for a while I helped with that strategic partnership with AstraZeneca, and that was quite interesting. But nobody else was really signing up with a strategic partnership, whereas the business model originally for Oxagen was that we would have a number of these strategic partners, arranged around different disease, we would have the platform, the informatics, the genetics, and we would be delivering potential targets to the other companies. We could divide it up around disease sectors, and it would all make sense, and we would have these long-term deals. But, really, apart from AstraZeneca nothing came through.

TT: Was that part of why you left? You then went to Rothamsted?

CR: A number of different things happened at the same time - one of these "everything collapses at the same time". The dot.com bubble burst, and with that there was a whole retrenchment in the investment community. So then Oxagen basically had to prove that it had some generated value, in order to look for its next round of investment. So, whilst they were spending money hand over fist on agreements with Celera, and burning it doing the experimental work, we were building a set of genomics tools that would allow us to screen the tissues that we were looking at from the clinical collections, for more of the biology. We had gone from the genetics, which gave us an idea of where we should be looking in the genome. We identified from the genome those genes that we thought would be most likely to be both supported by genetic

evidence and because they were in gene families which were linked to known pharmaceutical target families, like G-protein coupled receptors, neuropeptide receptors, kinases, all these things that had been the targets of drugs before. We built that as a panel of candidate genes, which we then constructed as a bespoke microarray platform, and we started to look at the profiling of these things using gene transcriptomics approaches to help tease apart the biology more. So the company started to bring in more biologists, because we needed pursue the biology, we needed to have the functional assays to test whether we could, in order to then put them into screening platforms. So, the company basically shifted its sense of gravity away from the early discovery phase and into the preclinical target validation screening of compounds and the full drug discovery pipeline. As part of that they had to shift where they were investing, and the only way they could shift that investment was to say 'We've got what we've got from the genome, we have to just drop the genetics.' So they dropped the bioinformatics first, because we weren't going to mine the databases any longer. They dropped much of the genetics, and so I was there as we were doing the first waves of redundancies, losing a lot of people who had been there from the beginning. And then, with the agreement, I understood why I couldn't stay. They didn't need me any longer. I'd been involved in the way the company was emerging for that not to be a shock. So I left in December 2003, I think. Basically I was made redundant and started looking around for something else, and then a job appeared at Rothamsted, which happened to be quite close to where I lived. I knew Rothamsted of old, because I'd sat on a number of BBSRC [Biotechnology and Biological Sciences Research Council] panels where I'd reviewed the services that BBSRC had supported at Rothamsted, and it was familiar territory.

The reason why there was a position at Rothamsted was, in the regular reviews of science at institutes which happened roughly every five years there had been concern expressed by the previous review panel that the institutes as a whole were not investing in their bioinformatics and their computational sciences area, because the BBSRC was very strongly promoting systems biology, data-intensive biology, and informatics as the new way of working. And so they wanted to make sure that the institutes were moving with the times and not being, as they often feel we are, somewhat behind the times. The director had been encouraged to do something about it, and the BBSRC had created an opportunity under what was then, I don't know if you remember, the e-science initiative, which was funded by government. So this was the brainchild of the then chief scientist, I think, who considered that, again, in order to keep up with what was going on in the rest of the world, e-science and developing the infrastructure to both computational networking skills, that e-science needed its own special stream. It was recognized that often when put in competition with additional biological or other research disciplines, these interdisciplinary topics or these technology-dominated topics tended to do badly. So they said 'Well, we'll give each research council a chunk of money to do e-science research, and they'll come up with a method of funding how they're going to build their e-science capacity.' So the BBSRC basically enabled each research institute to bid for a person and some capacity to build their bioinformatics in some way. And so there was a role for a senior bioinformatician that became available. I applied and in fact somebody else good applied as well, so they hired me and they hired this other guy. But because of my experience the director said 'I can see that you've done more than just bioinformatics, and we've got a bit of an issue, because we have mathematical modellers, and statisticians, and they're splattered around the institute, and they often don't do very well in their careers because they're not very visible to their heads of department, or their senior postdocs, so they're often left a little isolated, and for some time I've wanted to know what to do with these guys. And I think with you there's an opportunity to pull them all together and create a new department.' And he said 'If I do that, are you willing to take it on?' And I said yes because I wanted to take a bit more on, I wanted to do more strategic involvement in the institute, I wasn't used to just being basically a group leader postdoc type post again. So at the end of 2004, that opportunity, the Head of Department role came out about October 2004, which also coincided with when the old Silsoe Research Institute was being closed down by BBSRC, or BBSRC funding had been withdrawn from Silsoe and so they were having to close down the BBSRC bits of that, and still trying to work out what to do with the rest of it. And Rothamsted had agreed to take a number of the scientists from there, and many of the good ones had come in with a modelling or statistics background, so I got two new group leaders from that as well. So all in all now there is a group of some 40 people odd - it goes up and down a little bit - and the areas of work within the department are statistics, bioinformatics, and mathematical modelling. We are about 10% of the institute, so a relatively small department, and we cover a wide territory of different activities.

TT: And again, is there tension between service and research?

CR: Yes. The mathematical modellers are generally free from any service role. Often they are, if you like, they are more theoretical biologists, and modelling is their research tool. But there aren't really people in the institute who want a modeller to come in and model something. Unless you see the need for that, it doesn't tend to happen. Whereas for both statistics and bioinformatics, there is a real tension between research and service. Indeed, when I started, the person who joined with me was recruited to build the bioinformatics research activity. Apparently, he was hired to build research and the service capacity in the institute, but actually only built the research. As directors come and go, and as BBSRC changed the way it supported the computing infrastructure for the Institute, the need arose within Rothamsted, we needed to have our own bioinformatics service team, which previously had been served by an organization called BITS, which is the Biological Information Technology Service, which used to be based at Rothamsted physically. With the change in government arrangements with the institutes it was no longer viable, so they whammed it down, and some of the people went out to the other institutes and found careers. But, basically, it went. With that really they were providing some bioinformatics, but it was very low key. It was the lowest common denominator. There was always a bit of bioinformatic support, in terms of collaborating, but it left lots of scientists who hadn't established that collaboration without anything. I continued to lead the research side of bioinformatics, and I've got a very capable group leader running the applied bioinformatics, and we're beginning to win grants again, so I think we're doing reasonably well. But it's pretty much applied research, it's new algorithm development, but it is about building resources and joining things up, and building tools that people can use, which is a good tradition in bioinformatics. Producing things that people want to use is often the most effective demonstration that we're doing something useful.

TT: You've been in a lot of different kinds of institutions, positioned between service and research. Are you unusual in that?

CR: There are also other groups that fall into this situation, and I know because I've talked to her about it; Kathryn Lilley at Cambridge runs a proteomics facility. She's caught between this where she is considered a scientist and is judged as a scientist, but she runs a service. The people under her also, their careers and career development within a service unit in an otherwise academic research organization, often have a hard time because they don't have the opportunities that other academics have in order to pursue, develop their own careers with funding applications.

TT: Do you think this tension has built throughout your career or were you aware of this at the beginning? Do you think there's still a tension there?

CR: I think it was there at the beginning, but I worked in research institutes, and I think there's a lot more flexibility in research institutes than there is in universities about this, so you might get a very different view from somebody working in a university, where they're very reluctant to fund service, because it's seen as a cost rather than a benefit. So within research institutes I go back to ICRF days, while you're doing something that benefits the organization - that will largely be acknowledged in some way, shape, or form. And I think institutes are much better at having these cross-disciplinary and mixed role positions in their science departments, because they're often not big enough to cover it as a technology only, in order to attract people with the talent to run the unit and make sure that it stays current with whatever science is emerging; they want somebody with a research background. It's not often that they hire somebody with a service background to run a service department. More likely it's something that a scientist will tag on to their career development. I have colleagues where they feel similarly to me that the career development and rewards system doesn't really reflect the mixed roles that they have. I think that is a challenge, and it's been recognized in the research councils that people do have a challenge. Because more and more parts of science have this high-tech component, and as soon as you build in high-tech, whether that's a research platform where you want people to come and use a facility, but that facility has to have really qualified personnel to run it, or in order to keep on top of it they have to keep innovating with the latest chemistry, or the latest

physics, or the latest imaging method; and then on the image method they need the data analysis method. Unless they're on top of what is essentially a pretty fast moving field, you don't remain competitive for the biology you do using that platform. When I started really, there were very few areas where that was the case. Now you see it in microscopy, where it's largely been a service but now there's research interest; you see it in spectroscopy, there's so many different types of spectroscopic equipment in terms of mass-specs and MR machines that are used in biology. You have all the imaging technologies where they all have the same problem. And now, as I say, we're seeing more and more in the environmental sciences; there are technology platforms that are making a huge difference to the way in which people are collecting data about the environment. There's another cadre of scientists that need to be aware of the sensor technologies and the earth observation technologies and what they can do to answer questions at the landscape scale, instead of just at the plant or pot scale. As you wish to move between levels of biological organization, you bring in other technologies to this. So we're seeing that at Rothamsted, for example, but I know other places have similar problems.

TT: Let me just then finally talk about these committees and your role on scientific committees, because you've been on a phenomenal number.

CR: Yes, I'm a serial offender, aren't I?

TT: Just about, yes. What a considerable role you've played in so many different kinds of committees, particularly for BBSRC and MRC. Would you like to say something first of all about how you first got onto this, and it's something you either enjoy or feel committed to or responsible for, because you just say yes.

CR: It all started with the MRC when I was still at ICRF. And I think that just, I can't quite remember how, I got invited onto the, I think it was around the time that, about the same time that discussions were happening about the human gene mapping workshops. There were a couple of committees that were sitting looking at training in bioinformatics type roles, postdoctoral fellowships in bioinformatics that they were setting up. I think because I had been engaged with the MRC as part of the ICRF contingent (and am local to London, so they do tend to know the Londoners probably better), I was asked if I would sit on one of these committees, so I said yes and served my time with the MRC. That continued until I moved to - I didn't really do very much when I moved into industry, because you do tend to disappear off the radar screen.

So when I moved to Oxford, the then Director of Science had been working with BBSRC, and he basically put my name forward to sit on a grant panel, which was then biochemistry and cell biology, BCB, which was the earlier committee for the BBSRC. Prior to that I'd sat on a couple of these strategy development panels for BBSRC bioinformatics, and among the e-science initiative, again, the number of people who had experience and had been in the business and were available was relatively limited. Because I was in industry I was very popular, because they wanted people who didn't have a vested interest, because I wasn't going to get any funding. So to be a bioinformatician, and in industry, and to be reasonably well known, was where it came from, I think.

TT: Do you think a lot of these are reflections on your rarity, your uniqueness, almost. There was a phenomenal range of committees.

CR: I think the people working in the research councils talk a lot to one another. They find people that will say yes, and who will commit the time, and are prepared to put in the effort, and they tend to be recycled quite a lot.

TT: You must enjoy being on the committees.

CR: I've enjoyed it a lot. Some more than others, but on the whole I like to be engaged with the community. It's a good opportunity to keep on top of what's going on in the field, and building your network. So it's been

valuable to me in my career, I think, that I've known how the system works and know people. I can pick up the phone and ask people directly rather than wait to be told. It's been a positive thing. It has come at a cost; you do pay out of your private life often in terms of how much you have to read and prepare for a meeting. But that's, I've always, it's always a pain at the time, but it's always satisfying afterwards, I think.

TT: Perhaps we'd better stop there. Thank you so much for your time Chris.

[END OF TRANSCRIPT]

Further related resources:

1. Jones E M, Tansey E M (eds) (2015) *Human Gene Mapping Workshops c.1973-c.1991*. Wellcome Witnesses to Contemporary Medicine, vol. 54. London: Queen Mary University of London.
2. Tansey E M (intvr); Tansey E M, Beanland S (eds) (2017) *Rawlings, Chris: transcript of a video interview (04-Aug-2016)*. History of Modern Biomedicine Interviews (Digital Collection), item e2017232. London: Queen Mary University of London.