

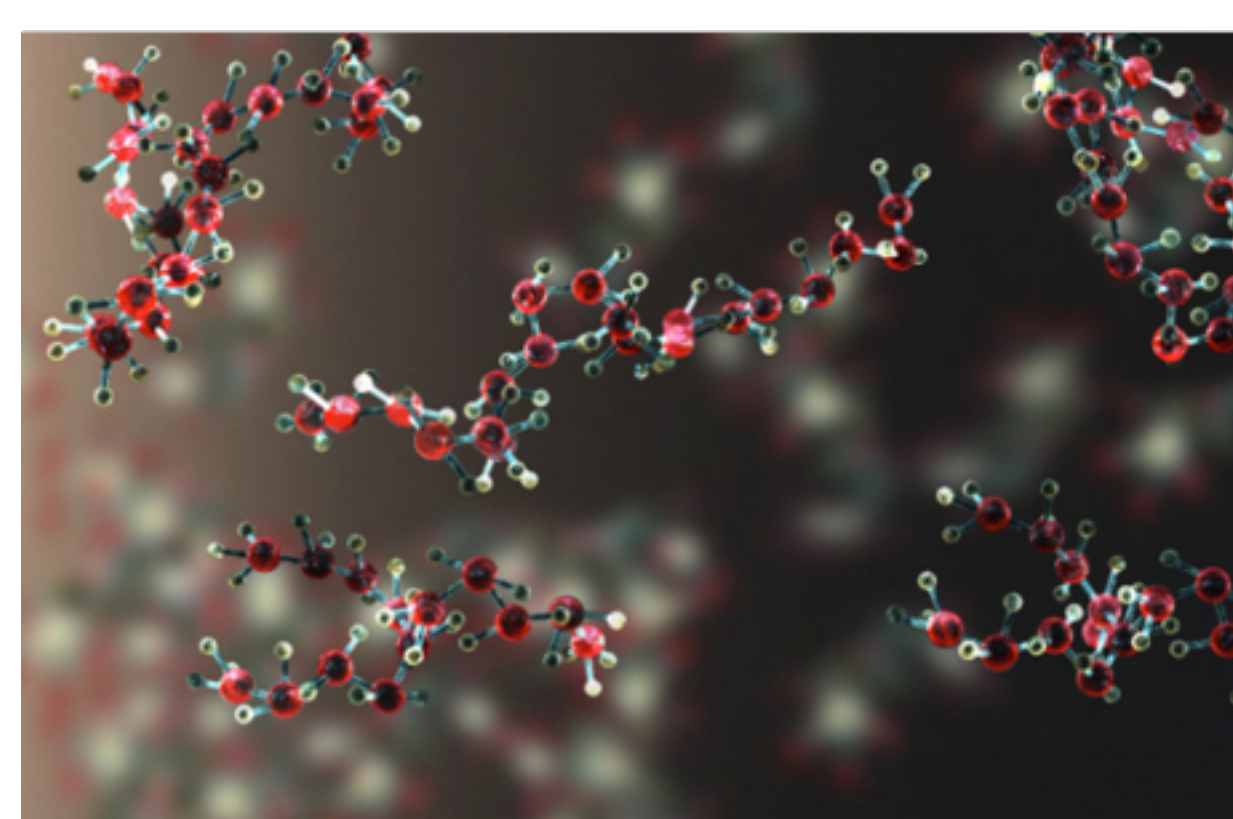
NMR_REDO: Analysing complex NMR data in the EGI Federated Cloud

Michel Drescher, EGI.eu • Matteo Turilli, OerC • Marco Verlato, INFN
Geerten W. Vuister, UoL • Wouter Touw, USR

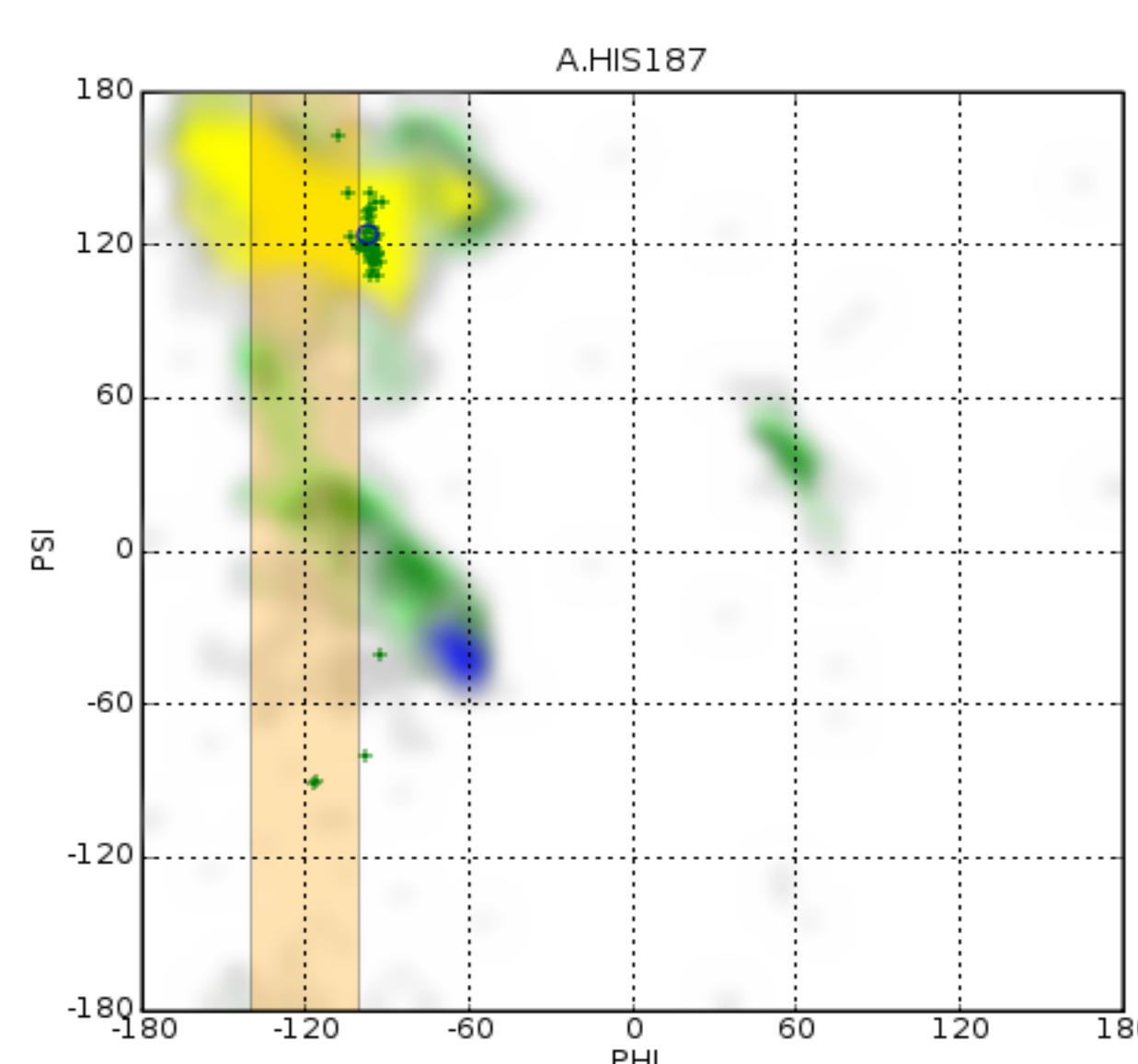
The Challenge

The WeNMR project is a global initiative focusing on the structural biology of proteins. Using NMR experiment data of live proteins, the project develops and improves **electronic models** of atomic protein structures that can be used for **simulations** and **protein visualisation**.

The NMR_REDO subproject works on the **validation** of protein models by analysing existing NMR data with the **EGI Federated Clouds** infrastructure.



Visualisation of bioactive proteins



Validation map of NMR data

Benefits using the EGI Cloud

Global accounting

Easy access to accounting figures across EGI providers – for free!

Federated Identity

Use your Grid credentials to access all supporting EGI providers.

Full application control

Deploy your applications when you want it, where you want it.

Elastic Computing

Cross-provider resources scaling, like a shrinking and expanding bubble.

NMR_REDO infrastructure

The NMR_REDO production infrastructure comprises three key components:

NMR Protein Database

This database holds the **raw NMR data** coming from the experiments, and the **validated protein structures**. It is at the same time the **central UI server** managing data, results, and vCing VM checks.

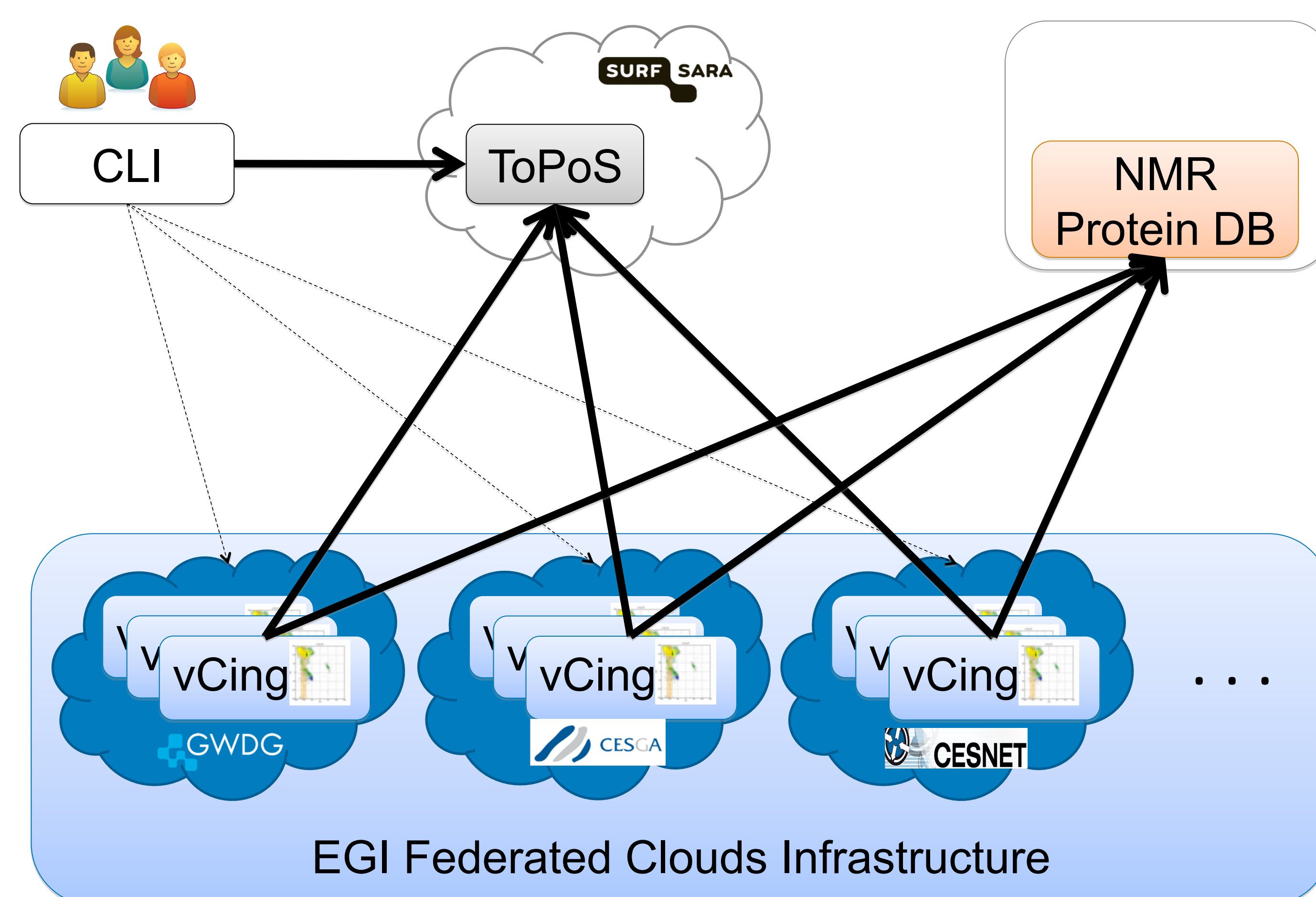
vCing VM images and instances

VirtualCing images encapsulate **~25 individual applications** for protein model validation. The code is maintained and updated through code.google.com. One image is instantiated **across all EGI Cloud providers**.

All vCing instances regularly contact the workload queue for new tokens. If no more tokens are available, the VM instance turns idle, until it is shutdown.

SURFsara ToPoS workload queue

All deployed vCing instances contact the token pool service for the next workload information. Each token encodes which data sets to fetch from the Protein DB, the specific protein, and which calculations to conduct.



NMR_REDO in numbers

5,000 protein recomputations

20 validation iterations

10 CPU hours per protein validation

114 CPU years in total, in a

3 year project

99% time saving using EGI Federated Clouds