

C. M. Ionescu^{1,2}, R. Svobodová Vařeková^{1,2,3}, D. Sehnal^{1,2}, O. Skřehota^{1,2,3}, J. Koča^{1,2}

¹Central European Institute of Technology, Brno, Czech Republic

²National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic

³ANF DATA, a Siemens company, Prague, Czech Republic

Introduction

Atomic charges can offer insight into chemical behaviour and reactivity and are part of force fields and structure-activity assessments. The *ab initio* approach to calculating atomic charges can be very precise, but too time and resource demanding for biomolecules. One alternative to fast and reasonably accurate charge calculation is EEM, which is based on the Density Functional Theory and Sanderson's Electronegativity Equalization Principle.

$$\begin{cases} X_1 = X_2 = \dots = X_N = \bar{X} \\ X_i = A_i + B_i \cdot q_i + k \cdot \sum_{j=1, j \neq i}^N \frac{q_j}{R_{i,j}} \\ \sum_{i=1}^N q_i = Q \end{cases} \Rightarrow \begin{pmatrix} B_1 & k & \dots & k & -1 \\ R_{1,2} & B_2 & \dots & k & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ R_{N,2} & R_{N,2} & \dots & B_N & -1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \\ \bar{X} \end{pmatrix} = \begin{pmatrix} -A_1 \\ -A_2 \\ \vdots \\ -A_N \\ Q \end{pmatrix} \Rightarrow \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix}$$

EEM can compute partial atomic charges with reasonable precision provided that proper parameters have been previously determined.

The objective of this study is to find a set of parameters k (A_i, B_i) so that EEM can be used to predict *ab initio* quality atomic charges for proteins.

Methods

Generally the process of parameterization involves several steps:

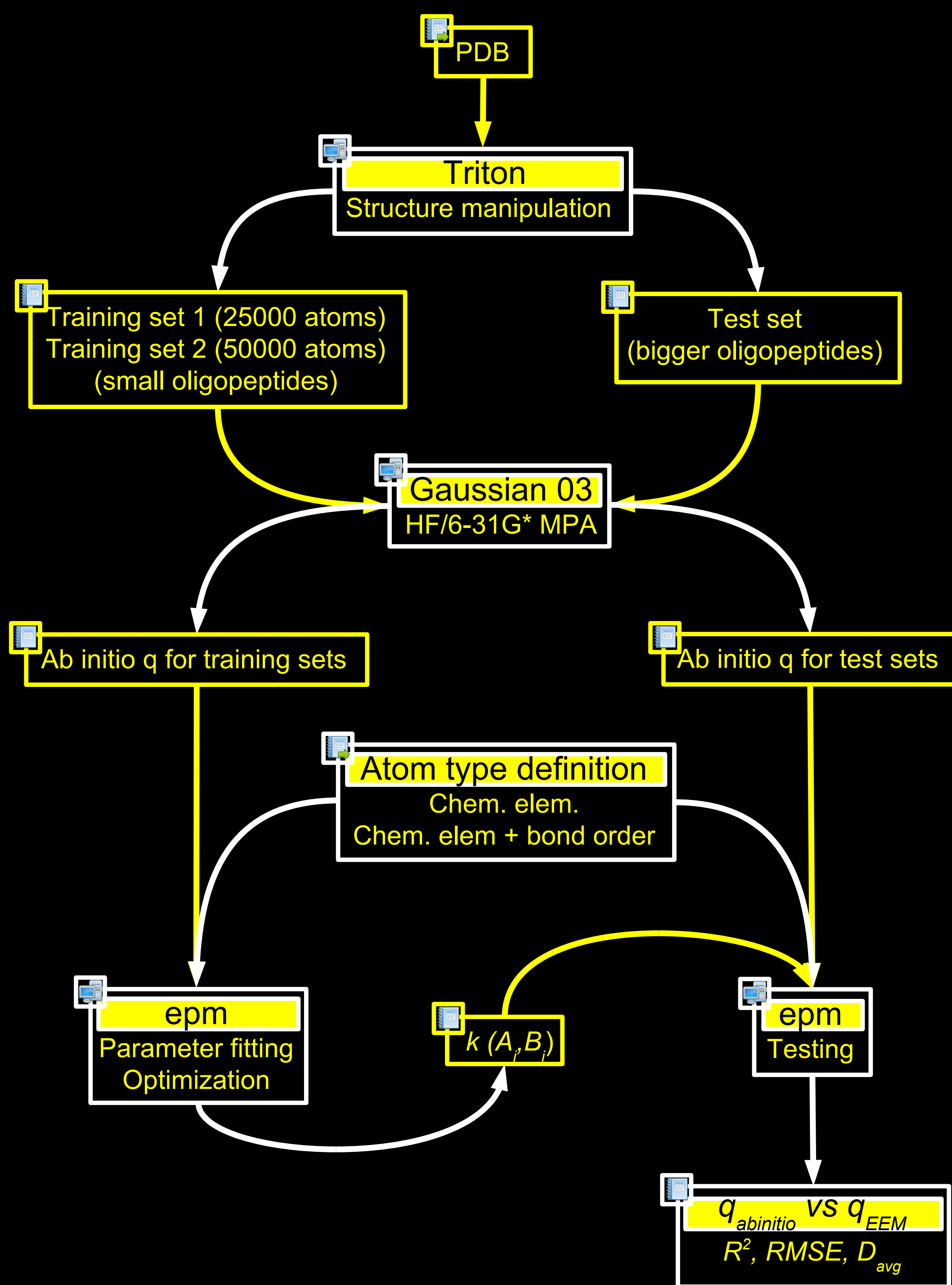
- Making up a training set of molecules.
- Computing *ab initio* charges for the atoms in the training set.
- Fitting the EEM parameters (A, B) to the *ab initio* data via least squares minimization for a range of k .

$$\begin{cases} A_i + B_i \cdot q_i = \bar{X} - k \cdot \sum_{j=1, j \neq i}^N \frac{q_j}{R_{i,j}} \\ \bar{X} = N \left(\sum_{i=1}^N \frac{1}{\lambda_i^0} \right)^{-1} \end{cases} \Rightarrow \begin{cases} A_i + B_i \cdot x_i = y_i \\ x_i = q_i \\ y_i = \bar{X} - k \cdot \sum_{j=1, j \neq i}^N \frac{q_j}{R_{i,j}} \end{cases} \Rightarrow (A_i, B_i) \text{ for each atom type for each value of } k$$

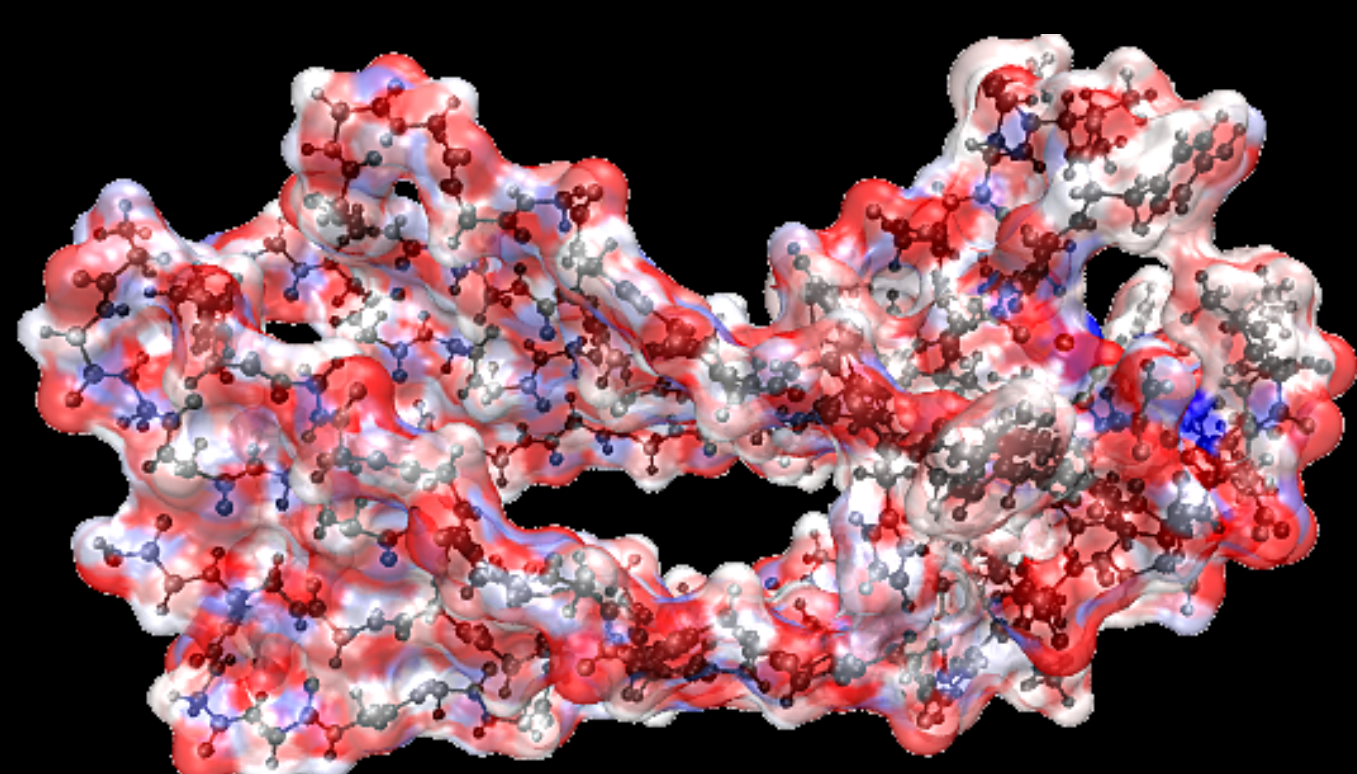
- Optimization to find the k for which parameters A and B perform the best.
- Making up a test set of molecules.
- Validating the parameters obtained in the previous steps (R^2 , $RMSE$, D_{avg})
- Monitoring the influence of the atom type definition and size of the training set.

Results

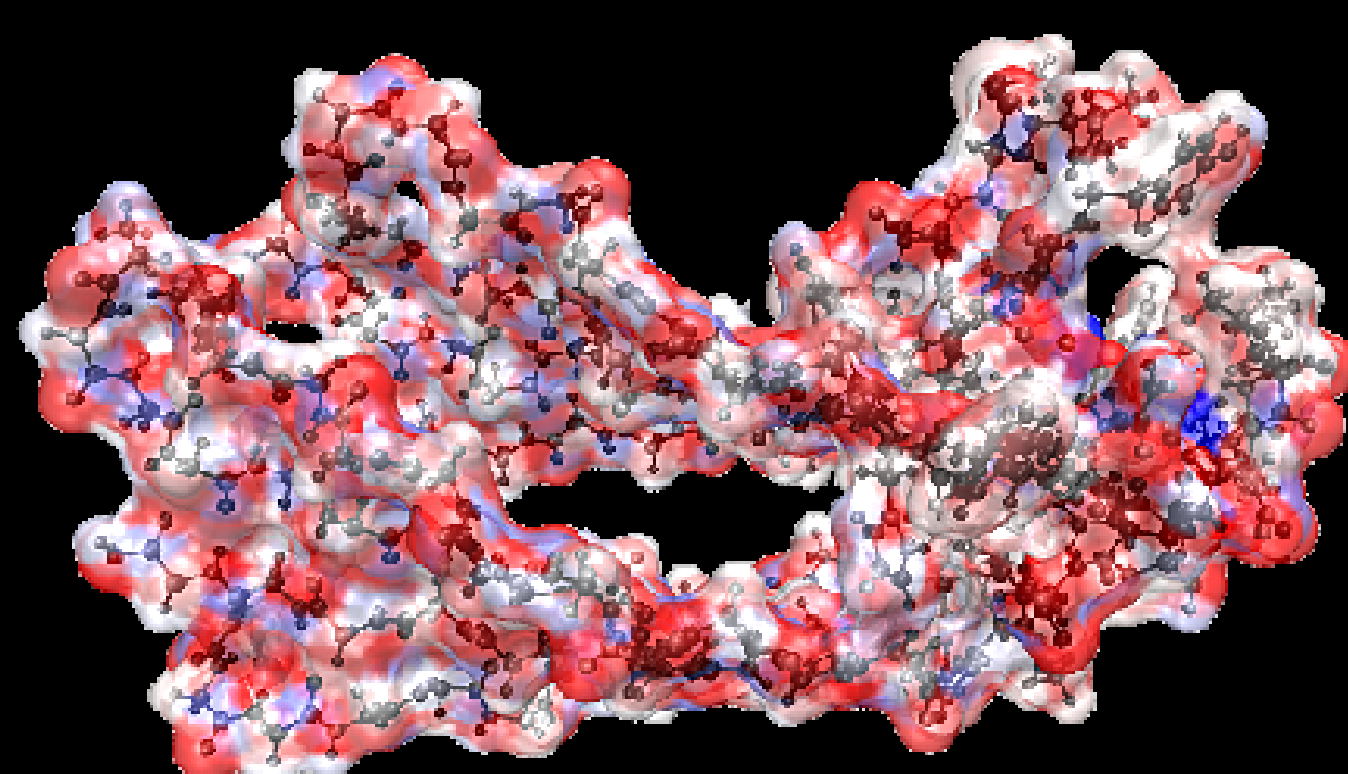
EEM vs *ab initio* correlation for test molecule 2 depending on the definition of the atom type and the size of the training set



Surface representation of test molecule 2 (Phospholipase fragment from 1BCI)

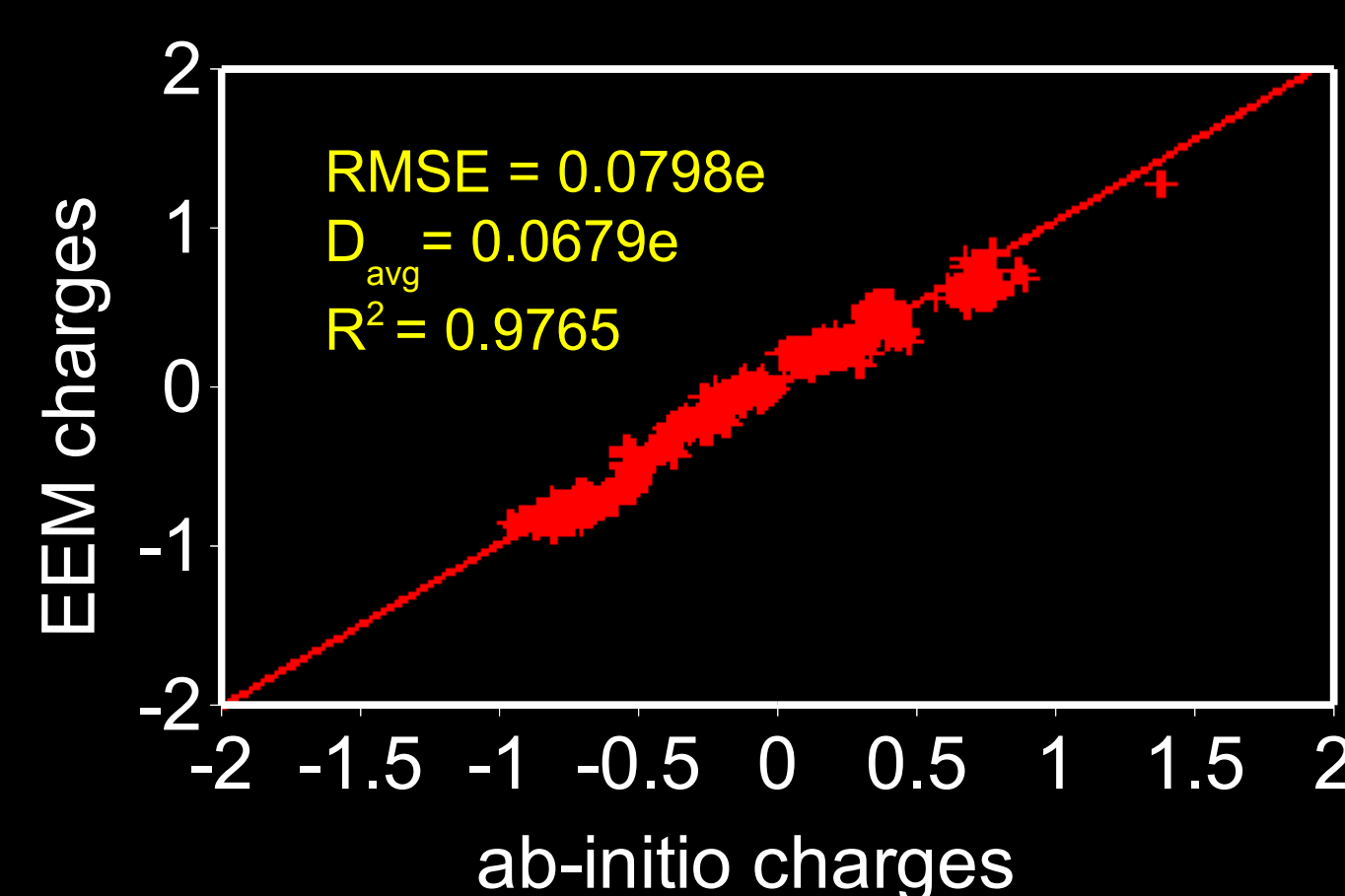


Method: HF/6-31G* MPA
Duration: 3.3 days

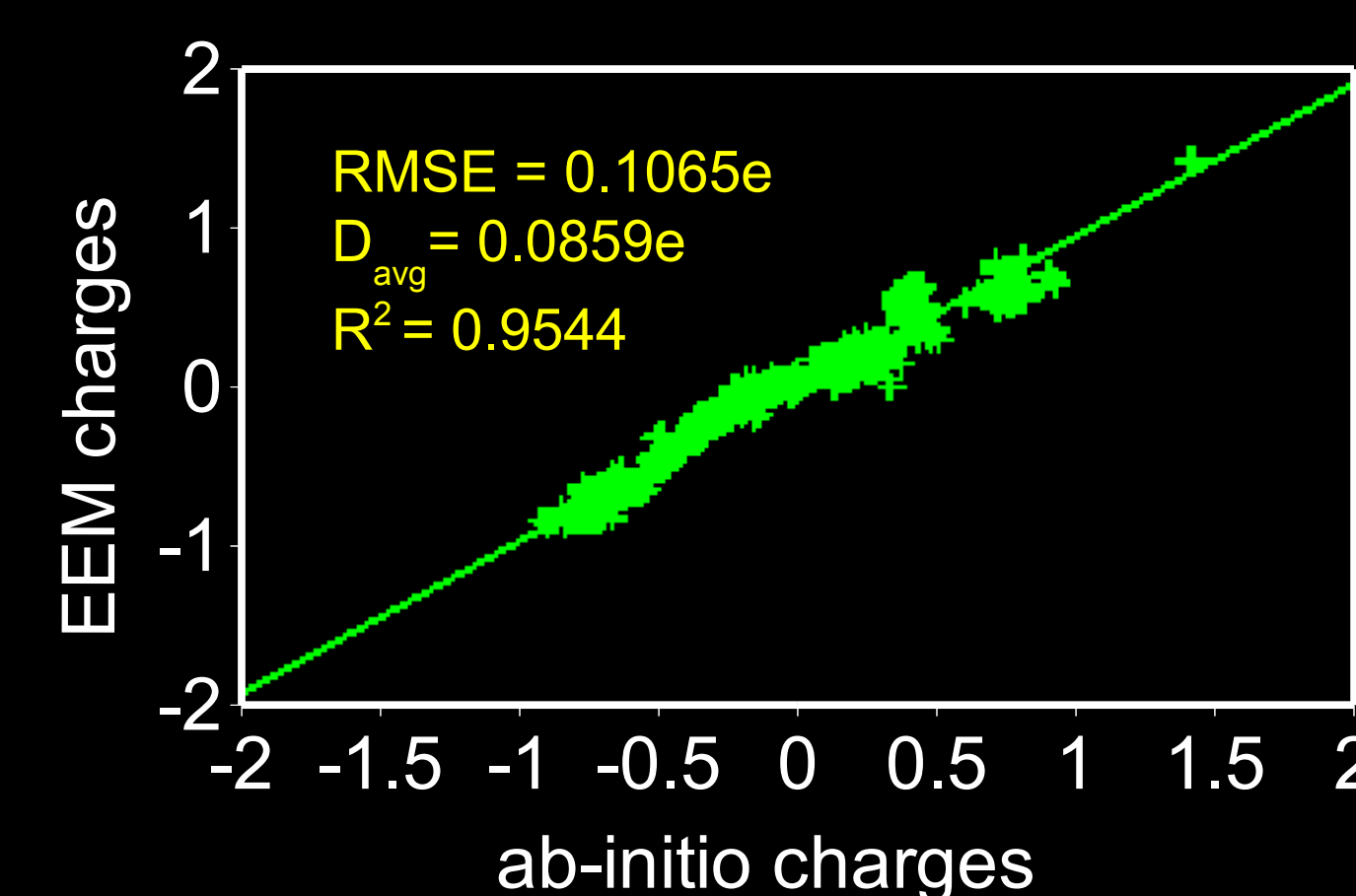


Method: EEM/TS1-E0
Duration: 5 seconds

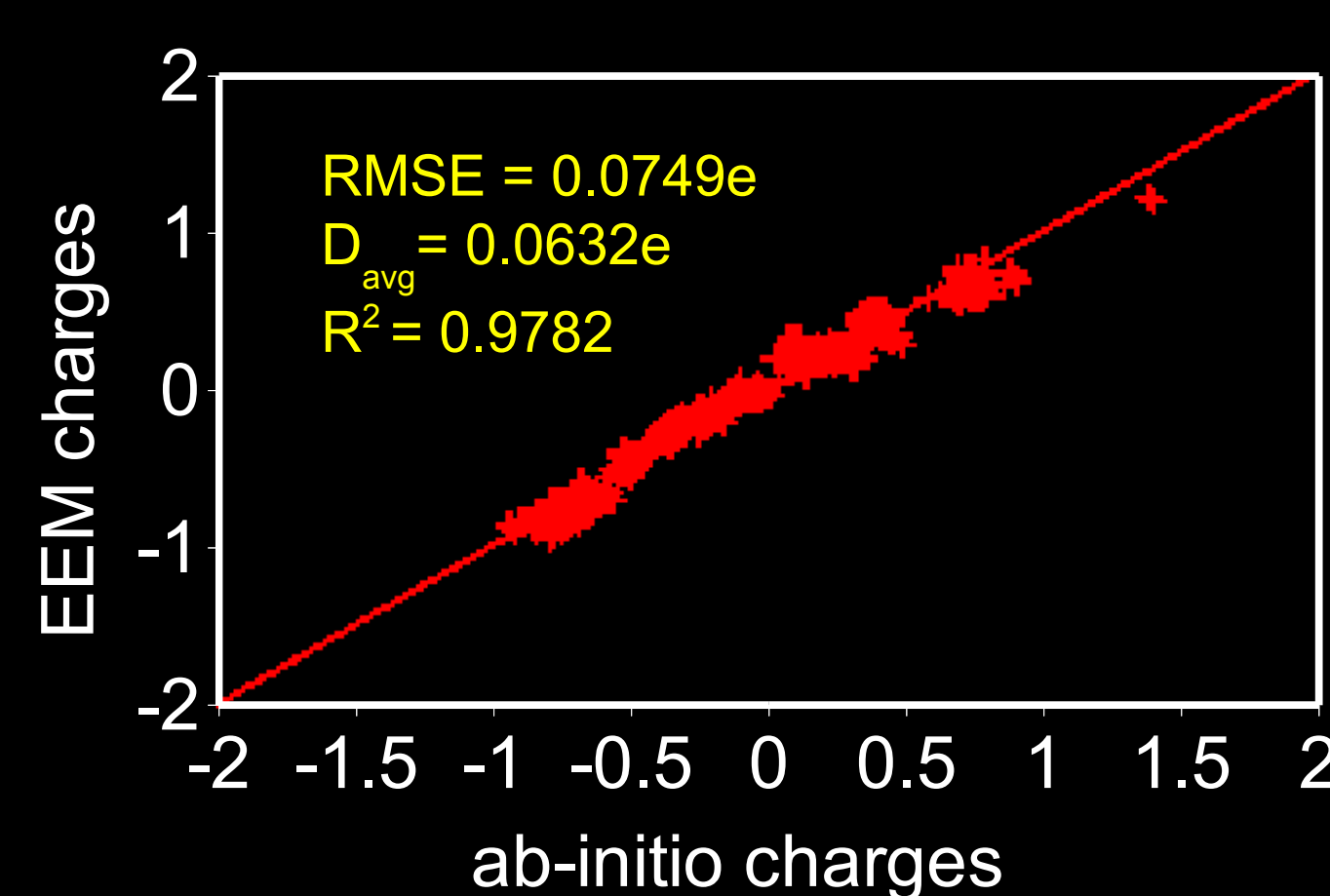
TS2-E0: training set 2 (~50000 atoms), atom type by element



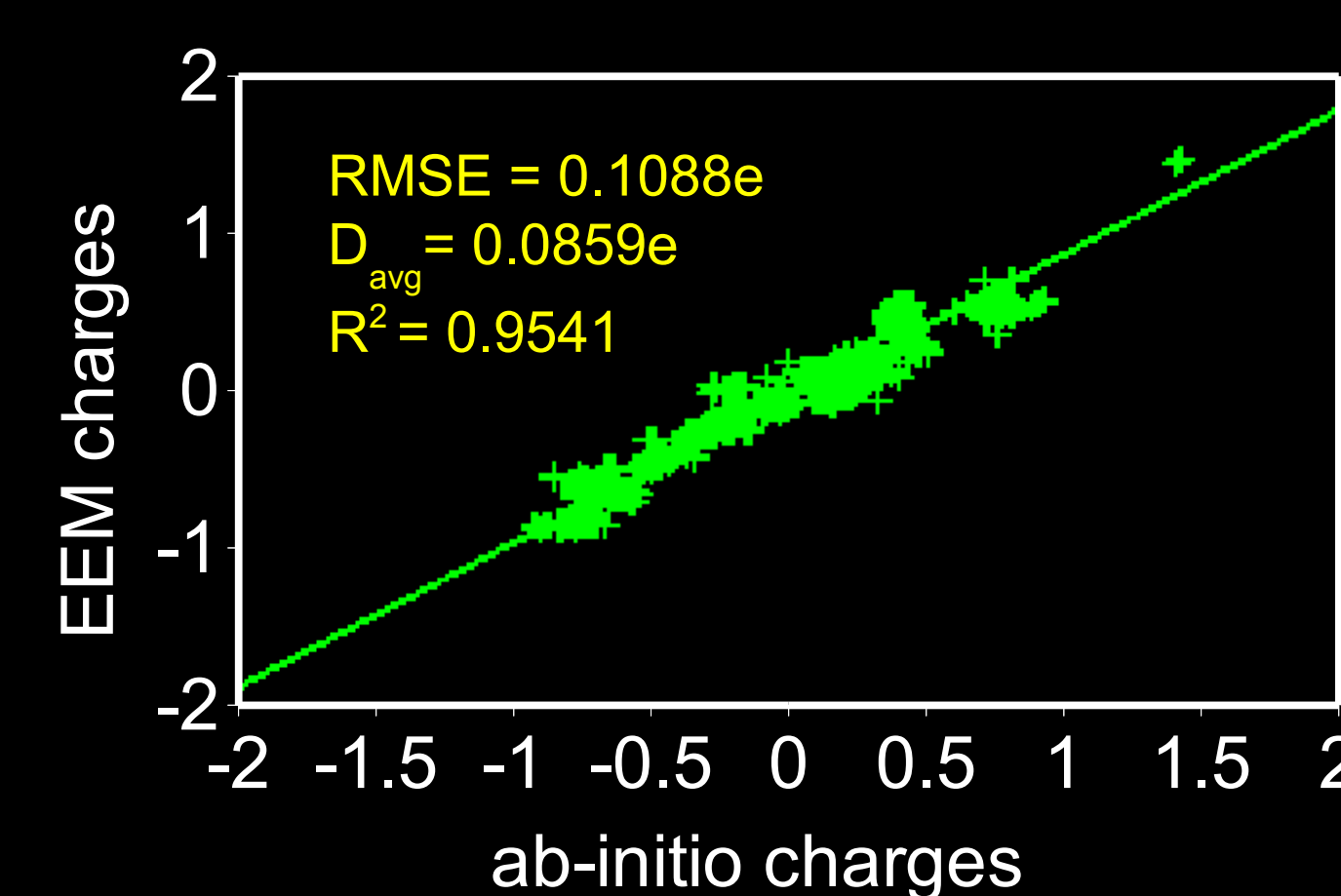
TS1-E0: training set 1 (~25000 atoms), atom type by element



TS2-EX: training set 2 (~50000 atoms), atom type by element and bond order



TS1-EX: training set 1 (~25000 atoms), atom type by element and bond order



Conclusion

We have obtained EEM parameters for the elements commonly found in proteins (C, H, N, O, S) and the Ca ion that may appear as a ligand. These parameters have been tested on significantly sized systems which are part of very large proteins.

EEM equipped with these parameters was able to predict HF/6-31G* Mulliken charges with reasonable accuracy both for the training sets and the test molecules ($R^2 \sim 0.98$, $RMSE \sim 0.075e$, $D_{avg} \sim 0.06e$). The main outliers were among the atoms on the surface of the peptides and the Ca ions. It is expected that solvating the structures would significantly reduce these errors.

We found that a more detailed differentiation of atomic types can improve the accuracy of EEM charges, but only if the training set from which the parameters were obtained was relatively large.

We conclude that the parameters we have obtained should enable EEM to predict partial atomic charges on full-sized protein systems to a good approximation.

Acknowledgements

This project is supported by funds from the Ministry of Education, Youth and Sports of the Czech Republic. Access to the computing resources was kindly provided by MetaCentrum, the supercomputing centre in Brno, Czech Republic. C.M.I. is grateful for the financial support provided to her by the Brno City Municipality through the program "Scholarships for Talented PhD Students", and to Mgr. Zora Štřelcová for the design idea.

References

- Jelfs, S.; Ertl, P.; Selzer, P., J. Chem. Inf. Model. 2007, 47, 450-459.
- Mortier, W. J.; Ghosh, S. K.; Shankar, S., J. Am. Chem. Soc. 1986, 108, 4315-4320.
- Rappe, A. K.; Goddard III, W. A., J. Phys. Chem. 1991, 95, 3358-3363.
- Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P., J. Phys. Chem. A. 2002, 106, 7895-7901.
- Svobodová Vařeková, R.; Koča, J., J. Comput. Chem. 2006, 27, 3, 396-405.
- Svobodová Vařeková, R.; Jiroušková, Z.; Vaněk, J.; Suchomel, S.; Koča, J., Int. J. Mol. Sci. 2007, 8, 572-582.