Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/2/11/e1600746/DC1

Supplementary Materials for

Prediction of interface structures and energies via virtual screening

Shin Kiyohara, Hiromi Oda, Tomohiro Miyata, Teruyasu Mizoguchi

Published 25 November 2016, *Sci. Adv.* **2**, e1600746 (2016)
DOI: 10.1126/sciadv.1600746

This PDF file includes:

section S1. The most straightforward method to determine the structure and energy of single grain boundary.

In the case of a coincidence site lattice (CSL) grain boundary of simple metal, three degrees of freedom, namely rigid body translation to x, y, and z directions, namely dx, dy, and dz are present. To determine the atomic structure of the single CSL grain boundary, the number of the atomic configuration to be considered often reaches several hundred ~ several tens of thousands.

For instance, about 18,000 atomic configurations have to be considered in case of Σ5 [001]/(210) CSL grain boundary of Cu. The grain boundary energy of those all possible configurations are shown in fig. S1. The most stable grain boundary, indicated by red arrow, is obtained by the rigid body translation by (dx, dy, dz) = (3.2 Å, 1.2 Å, 1.8 Å), and its grain boundary energy is 0.957 J/m^2.

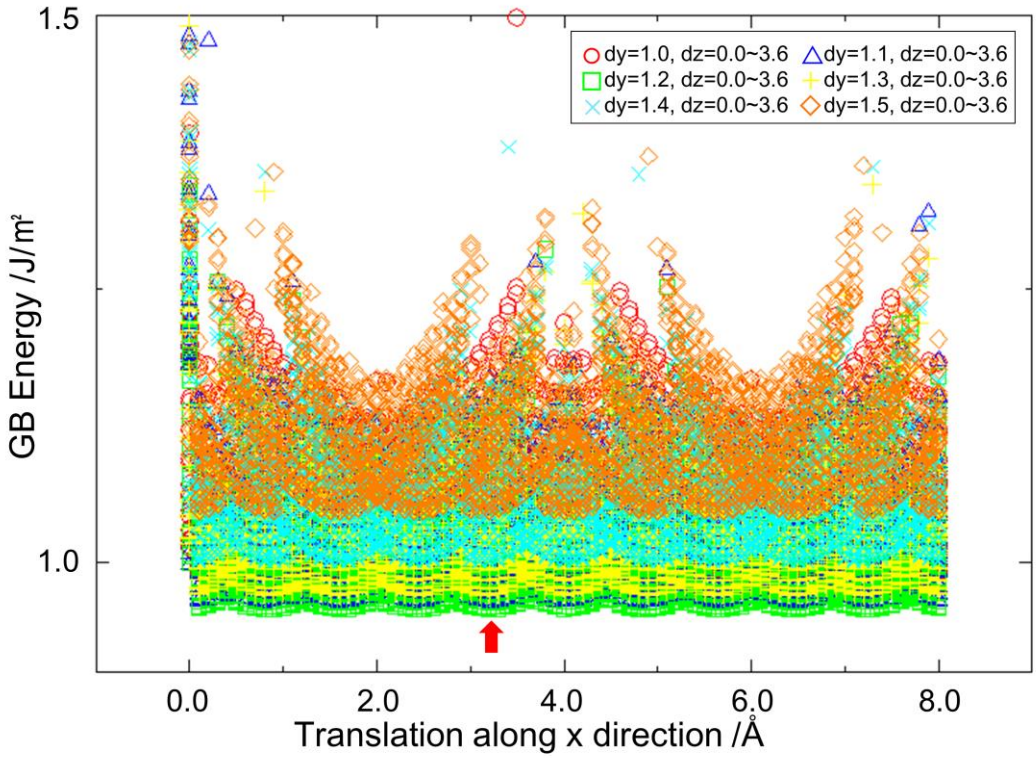

fig. S1. Plot of the calculated grain boundary energies by the all-candidate calculation method. Red arrow indicates the most stable configuration.

section S2. Descriptors used for the regression analysis in this study.

Figure S2 shows descriptors for the present SVR analysis. θ in tan(θ/2) and sin(θ/2) is misorientation angle between two crystals (Fig. 2). Number of shorter and longer bond length is estimated comparing with the optimized bulk's Cu-Cu bond length, 2.556Å. The Cu-Cu bond length of the bulk was obtained by the optimized FCC-Cu unit cell with the empirical potential described in the Methodology.

List of Descriptors
tan(θ/2)
sin(θ/2)
Atomic density around GB
Average 1st NN (Near Neighbor) bond length
Average 2nd NN bond length
Average 1st NN bond length around GB
Average 2nd NN bond length around GB
Number of shorter bond length
Number of longer bond length
Average shorter bond length
Average longer bond length
Shortest bond length
Number of dangling bond around GB
Relative translation distance along x direction
Relative translation distance along y direction
Relative translation distance along z direction

fig. S2. Descriptors for the SVR analysis.

section S3. The results obtained through the linear regression method.

The regression was performed by the linear regression with the same training and test data. As can be seen fig. S3(a), the regression itself was succeeded. However, the constructed model cannot predict the test data (fig. S3(b)).

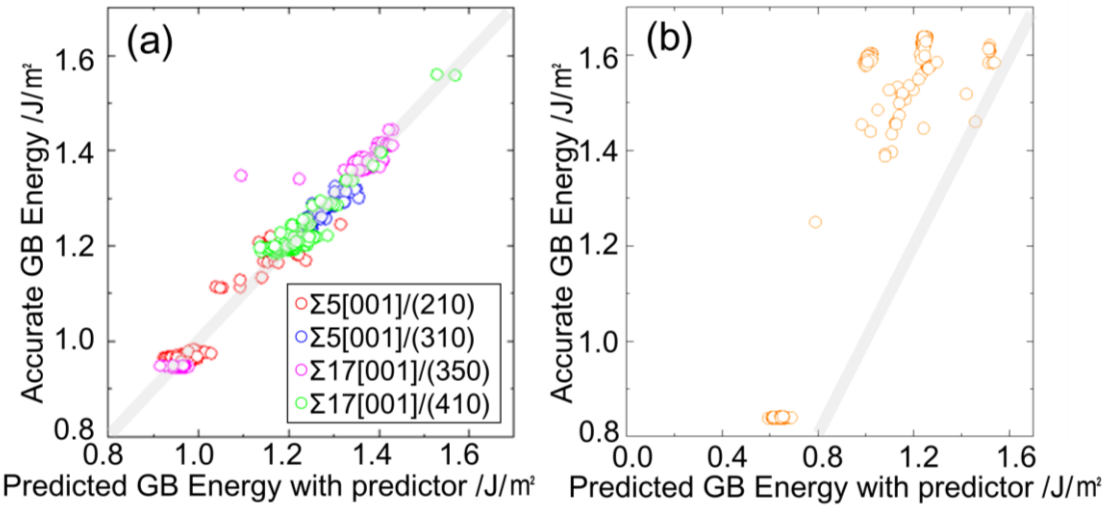

fig. S3. Predicted grain boundary energies through linear regression method. The result of linear regression: predicted grain boundary energies and accurate grain boundary energies as of the training data (**a**) and the test data (**b**).

section S4. Predictions for 12 grain boundary structures using the virtual screening method.

Atomic structure information of all grain boundaries will be uploaded as CIF file in this Supplementary material.

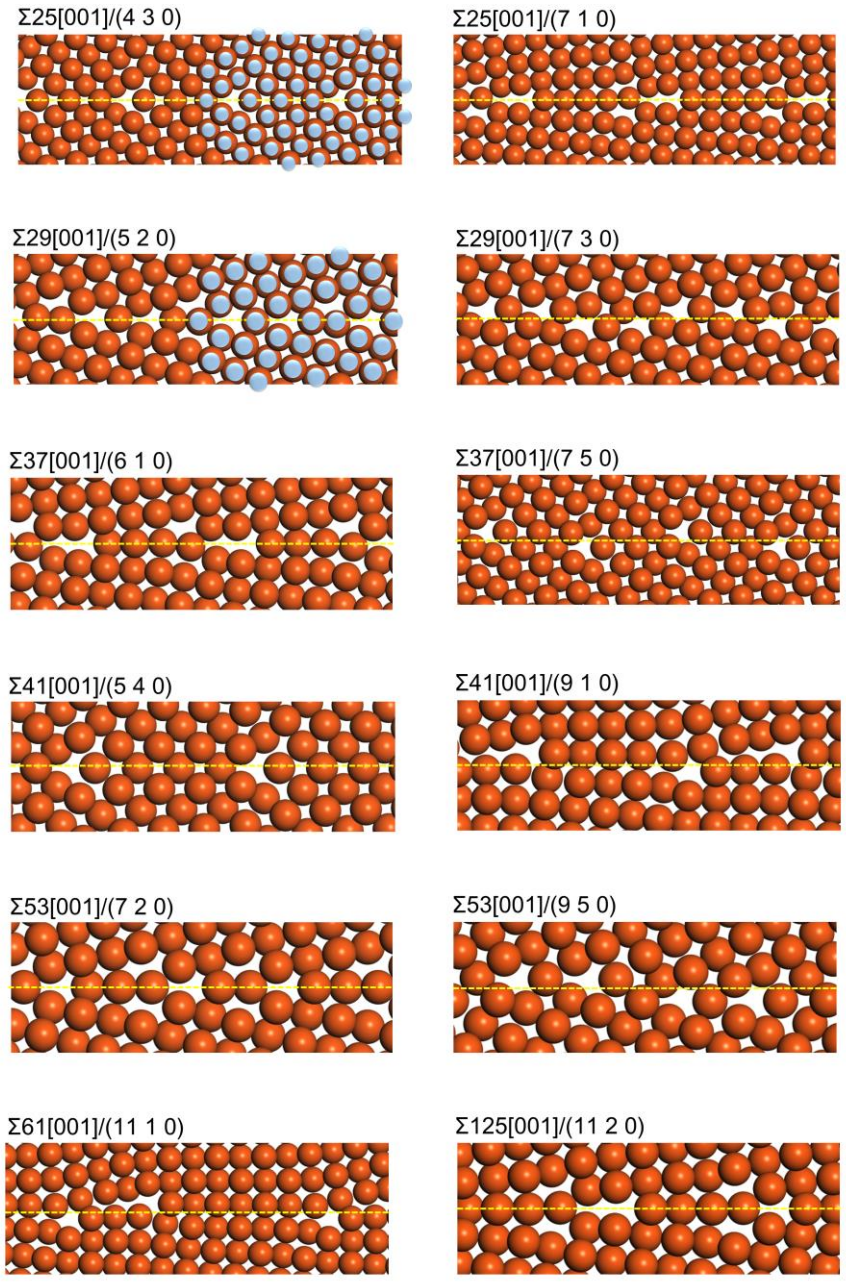

fig. S4. Predictions for 12 grain boundary structures using the virtual screening method. Structures from all candidate calculations are overlaid with blue circles on Σ25[001]/(430) and Σ29[001]/(520).

section S5. Effect of the training data selection.

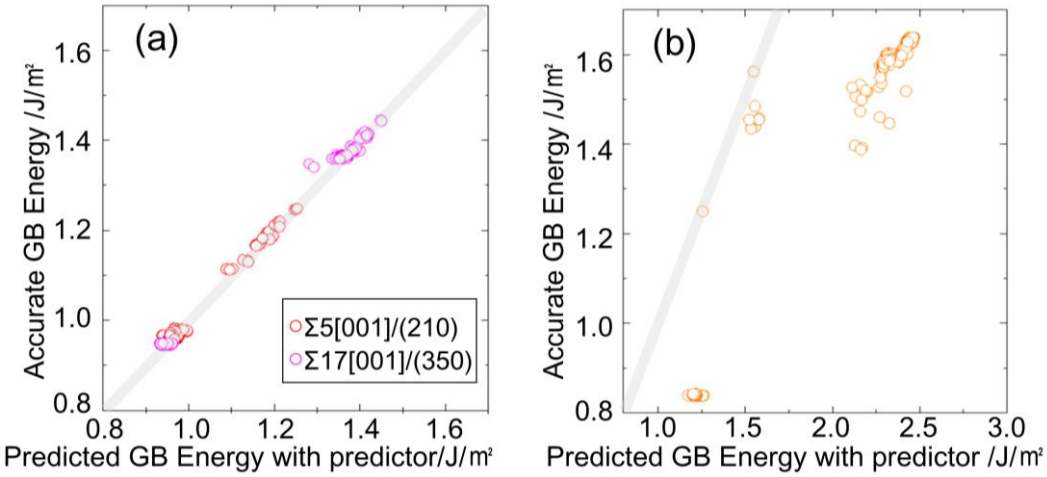

fig. S5. Predicted grain boundary energies with two of four kinds of grain boundary as the training data. Predicted grain boundary energies and accurate grain boundary energies as of the training data (**a**) and the test data (**b**). Σ5[001]/(210) and Σ17[001]/(350) were used as the training data.

section S6. Effect of the parameters for the regression analysis.

The effect of the parameters for the SVR is investigated. Figure S6(a) is obtained for the same training data as Fig. 4(a) with the parameter set (margin of tolerance: 0.001, penalty factor: 100, and variance:10^{-2}). It is clearly seen that all data are placed on the grey line indicating that the regression itself was succeeded. However, inferior result is obtained using this parameter set as shown in fig. S6(b), indicating that this is "over fitting". In this study, the parameter-set which shows good generalization ability has to be selected. Totally sixty-four sets of the parameters have been tried, and the best parameter which showed the best generalization ability was selected.

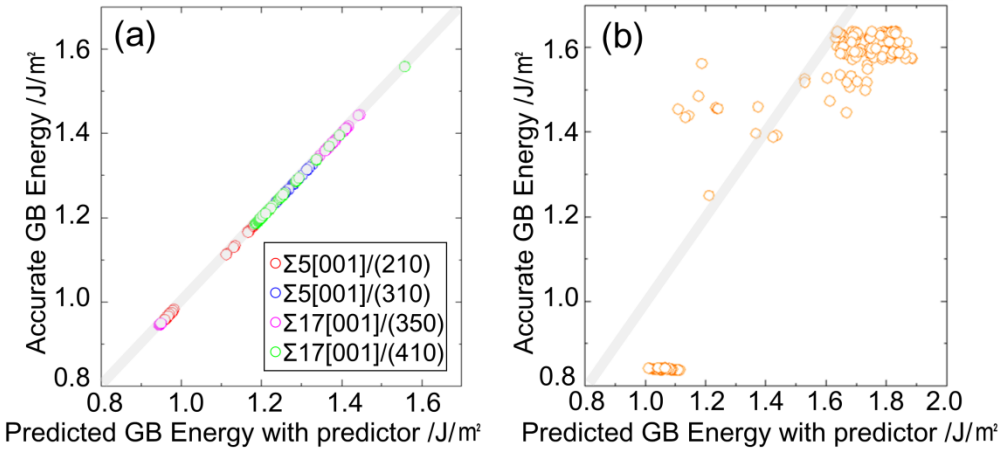

fig. S6. Predicted grain boundary energies under over-fitting. Predicted grain boundary energies and accurate grain boundary energies as of the training data (**a**) and the test data (**b**); a margin of tolerance was 0.001, penalty factor was 100 and variance was 10^{-2}.