Supplementary Methods

Numerical experiments were conducted with a finite difference thermo-mechanical numerical code with a fully staggered Eulerian grid and a Lagrangian particle field based on the marker-in-cell technique[1-3]. The mechanical implementation employs a visco-elasto-plastic rheology and governing equations are discretized on the non-deformable Eulerian grid and solved with the MATLAB's "backslash" direct solver for the two velocity components and dynamic pressure. Temperature is solved separately on the Eulerian pressure nodes with MATLAB's "backslash" direct solver. Material properties are interpolated on freely moving Lagrangian markers that advect through the fixed Eulerian grid according to a fourth-order Runge-Kutta derived velocity field.

Governing equations. The mechanical model implements the equations for conservation of mass (incompressible)

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{S1}$$

and conservation of momentum (Stokes equation)

$$\frac{-\partial P}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} = \rho g_i. \tag{S2}$$

P = mean stress, u_i = velocities, x_i = spatial coordinates, τ_{ij} = deviatoric stress tensor, ρ = density, and g_i = is the gravitational acceleration.

Temperature is implemented by the energy equation

$$\rho C_P \left(\frac{DT}{Dt}\right) = k \frac{\partial^2 T}{\partial x_i^2} + H_a + H_r + H_s. \tag{S3}$$

T = temperature, t = is time, C_P = isobaric heat capacity, and k = thermal conductivity coefficient. Additional heat sources include adiabatic heating (H_a), radioactive heating (H_r), and shear heating (H_s):

$$H_a = \alpha T \left(\frac{DP}{Dt}\right) \quad \text{(S4)}$$

$$H_s = \xi \sigma_{ij} \dot{\varepsilon}_{ij} \quad \text{(S6)}$$

ξ = fraction of work adding to shear heating. H_r is implemented as a constant for each rock type. Density changes related to thermal expansion α and compressibility β are implemented following

$$\rho = \rho_r[1 + \beta(P - P_r) \cdot 1 + \alpha(T - T_r)], \quad \text{(S7)}$$

where ρ_r = reference density, P_r = reference pressure (1 bar), T_r = reference temperature (273 K), α = thermal expansivity, and β = compressibility.

Rheological model. The visco-elastic relation between stress and strain rate follows a Maxwell-type model composed of a viscous and an elastic strain rate part

$$\dot{\varepsilon}_{ij} = \frac{1}{2\eta}\tau_{ij} + \frac{1}{2G}\frac{D\tau_{ij}}{Dt}, \quad \text{(S8)}$$

where G indicates the shear modulus and η the effective viscosity with lower and upper cutoffs of 10^{17} and 10^{25} Pa·s, respectively. Elasticity is implemented by adapting the effective viscosity depending on the "computational" time step and the stress history[1,4,5]. The objective co-rotational time derivative of visco-elastic stresses is discretized as a function after applying first-order finite difference

$$\frac{D\tau_{ij}}{Dt} = \frac{\tau_{ij} - \tau_{ij}^{old}}{\Delta t} \quad \text{(S9)}$$

with

$$\tau_{ij} = 2\eta\dot{\varepsilon}_{ij}Z + \tau_{ij}^{old}(1 - Z) \quad \text{(S10)}$$

and the visco-elasticity factor

$$Z = \frac{\Delta t \cdot G}{\eta + \Delta t \cdot G} \quad \text{(S11)}$$

with η as effective viscosity, which leads to the numerical visco-elastic viscosity

$$\eta_{num} = \eta \cdot Z = \frac{\eta \cdot \Delta t \cdot G}{\eta + \Delta t \cdot G} \quad \text{(S12)}$$

used to solve the set of equations.

The viscous strain rate is composed of both dislocation and diffusion creep following the general power law for a viscous implementation[6]:

$$\dot{\varepsilon} = A_D \cdot fH_2O^r \cdot \sigma^n \cdot d^m \cdot exp(-\frac{E+P \cdot V}{RT}), \quad \text{(S13)}$$

where A_D = pre-exponent, fH_2O = water fugacity, r = water fugacity exponent, σ = stress, n = stress exponent , d = grain size, m = grain size exponent, E = activation energy, V = activation volume, and R = gas constant (8.314 J/K/mol).

Viscosities for dislocation creep η_{disl} and diffusion creep η_{diff} are calculated separately by reformulating the general viscous power law equation (S13):

$$\eta_{disl,diff} = 0.5 \cdot \frac{1}{A_D \, fH_2O^r} \cdot \sigma_{II}^{(1-n)} \cdot d^{-m} \cdot exp\left(\frac{E+PV}{RT}\right) \quad \text{(S14)}$$

with

$$\sigma_{II} = \sqrt{\frac{1}{2}\tau_{ij}^2}. \quad \text{(S15)}$$

The composite viscosity resulting for the simultaneous occurrence of dislocation and diffusion creep follows

$$\eta_v = \left(\frac{1}{\eta_{disl}} + \frac{1}{\eta_{diff}}\right)^{-1}. \quad \text{(S16)}$$

Figure S1 shows vertical viscosity (a-c) and strength profiles (d-f) for variable grain sizes, strain rates, and water contents. Such an illustration helps interpreting the dominating deformation mechanism in the uppermost mantle depending on the investigated variables.

Plastic failure occurs if the visco-elastic differential trial stresses exceed the yield stress ($F > 0$) according to the Drucker-Prager yield criterion with a flow potential resulting in a dilation angle of zero:

$$F = \sigma_{II} - \sigma_y \tag{S17}$$

where

$$\sigma_y = P \cdot \left(1 - \lambda_f\right) \cdot sin\,\varphi + C \cdot \cos\varphi \tag{S18}$$

where C = cohesion, φ = friction angle, and λ_f = fluid pressure ratio. Exceeded stresses are kept within the failure envelope by decreasing the plastic viscosity η_p to maintain those stresses

$$\eta_p = \frac{\sigma_y}{2\dot{\varepsilon}_{II}} \tag{S19}$$

where

$$\dot{\varepsilon}_{II} = \sqrt{\frac{1}{2}\dot{\varepsilon}_{ij}^2}. \tag{S20}$$

The effective viscosity going into the viscous part of the Maxwell rheological model follows

$$\eta = \min\left(\eta_v, \eta_p\right). \tag{S21}$$

After interpolation of the Eulerian velocity field onto the Lagrangian markers, stress changes and plasticity are calculated on those. The updated effective viscosity is then interpolated back onto the Eulerian nodes and used to solve the system of equations. Time steps exhibit maximally ≤1000 yr following a Courant number of 0.25.

Grain size evolution model. Grain size is calculated based on the paleowattmeter[7]. Grain size depends on independently acting growth and reduction terms. Grain size reduction rate is related to mechanical work executed by dislocation creep ($\sigma\dot{\varepsilon}_{disl}$) and is described by

$$\dot{d}_{red} = \frac{\sigma\dot{\varepsilon}_{disl}\lambda d^2}{c\gamma}, \quad \text{(S22)}$$

where σ is stress, $\dot{\varepsilon}_{disl}$ is dislocation creep strain rate, c is a geometric constant (π for spheric grains), γ is the grain boundary energy, and λ denotes the fraction of work that goes into grain size reduction (λ = 1–ξ), whereas the rest of the work goes into the shear heating term (H_s; see eq. S6)[8-10]. Fitting experimentally-derived olivine grain sizes versus expected grain size using the paleowattmeter with the grain growth law constrained by Speciale and others[11] resulted in a λ of 0.01 (Fig. S2a). The constrained fraction of work that adds to the grain size reduction term is substantially smaller than previously applied fractions of λ = 0.1[7,12,13]. However, a recent study demonstrated that the energy partitioning factor λ of olivine ranges between 0.002 and 0.04 for a wide spectrum of pressure and temperature conditions[14].

Grain growth rate follows a normal relationship given by

$$\dot{d}_{gr} = K_g\, fH_2O \exp\left(-\frac{E_g + P \cdot V_g}{RT}\right) p^{-1}\, d^{1-p}, \quad \text{(S23)}$$

where K_g = growth rate constant, fH_2O = water fugacity, E_g = activation energy, V_g = activation volume, P = pressure, T = temperature, R = gas constant, d = grain size, and p = growth exponent. We applied experimentally derived olivine grain growth parameters by Speciale and others[11] that result in significantly slower grain growth than previous constraints[15] (Fig. S2b). The new grain d_{new} size is calculated on the Lagrangian markers following

$$d_{new} = \left(\dot{d}_{gr} - \dot{d}_{red}\right) \cdot Dt \quad \text{(S24)}$$

and then goes into the power law creep calculation for diffusion creep (eq. S14).

Model setup. The Eulerian model domain measures 1000 x 670 km in *x*- and *y*-direction, respectively (Fig. S3). The nodal resolution is 501 x 336 in *x*- and *y*-direction, which results in a cell size of 2 x 2 km. Rock type, rheological information, and mechanical, thermal, and grain size material properties (Tables S1-4) are stored on 25 Lagrangian markers per Eulerian cell. The initial marker distribution (Fig. S3) describes from top to bottom: i) a 10-km-thick layer of low viscosity sticky-air, which allows for a quasi-stress-free surface (air/rock interface)[16], ii) a 23-km-thick upper continental crust with quartzite rheology[17,18], iii) a 10-km-thick lower continental crust with anorthite rheology[19], iv) 627 km of upper mantle with dry or wet olivine rheology[6]. A weak inclusion of 4 x 4 km of quartzite rheology is placed in the lower continental crust at x = 500 km to localize rifting (Fig. S3).

Fugacity in the upper continental crust is calculated after Shrinevar et al.[20]. In the upper mantle, fugacity is implemented as constant fluid content with values of C_{OH} = 50, 175, 600, or 2500 H/10^6Si, covering the range of estimated values obtained from experimental studies[21,22]. Fluid content in the mantle affects both viscosity (eq. S14) and grain growth (eq. S23).

The initial temperature distribution describes 0° within the sticky-air layer, a linear increase from 0°C at the surface (y = 10 km) to 660°C at the Moho (y = 43 km), and from there to 1345°C at the thermally-induced lithosphere-asthenosphere boundary (LAB) at 150 km depth (y = 160 km). Below the LAB, a static temperature increase of 0.5°/km is introduced.

Oceanic crust with an anorthite-diopside (50/50) rheology[23] is produced if mantle rock markers less than 8 km below the surface (air/rock interface) have a temperature of <400°C.

Initial grain size distribution. Initial grain size in the mantle of all experiments in the main text logarithmically increases from 5 mm at the Moho to 10 cm at the LAB, and 10 cm throughout the lower part of the mantle. Figure S2 shows the grain size distribution within the uppermost 300 km after 10 Myr (Fig. S4a) and the temporal evolution of average grain size in the lower 300 km of the upper mantle (Fig. S4b) for different initial conditions. Grain sizes within the lithosphere are mainly driven by the reduction term due to lower temperatures. High temperatures and thus fast growth rates allow the lower part of the model domain to rapidly restore deformation-related reduced grain sizes. As a result, the initial grain size within the lower 300 km of the model is of little importance, while initial grain sizes should be large enough throughout the lithosphere not to be dependent on initial growth.

Surface evolution model. The surface line (rock/air interface) undergoes simple syn-tectonic sedimentation and erosion mimicked by a linear diffusion scheme

$$\frac{\partial h_s}{\partial t} = \kappa \frac{\partial^2 h_s}{\partial x_i^2}, \tag{S24}$$

with h = surface elevation and κ = diffusion coefficient (10^{-6} m^2/s). Syn-tectonic sediments have equal material and strength properties as the initial sediment sequence.

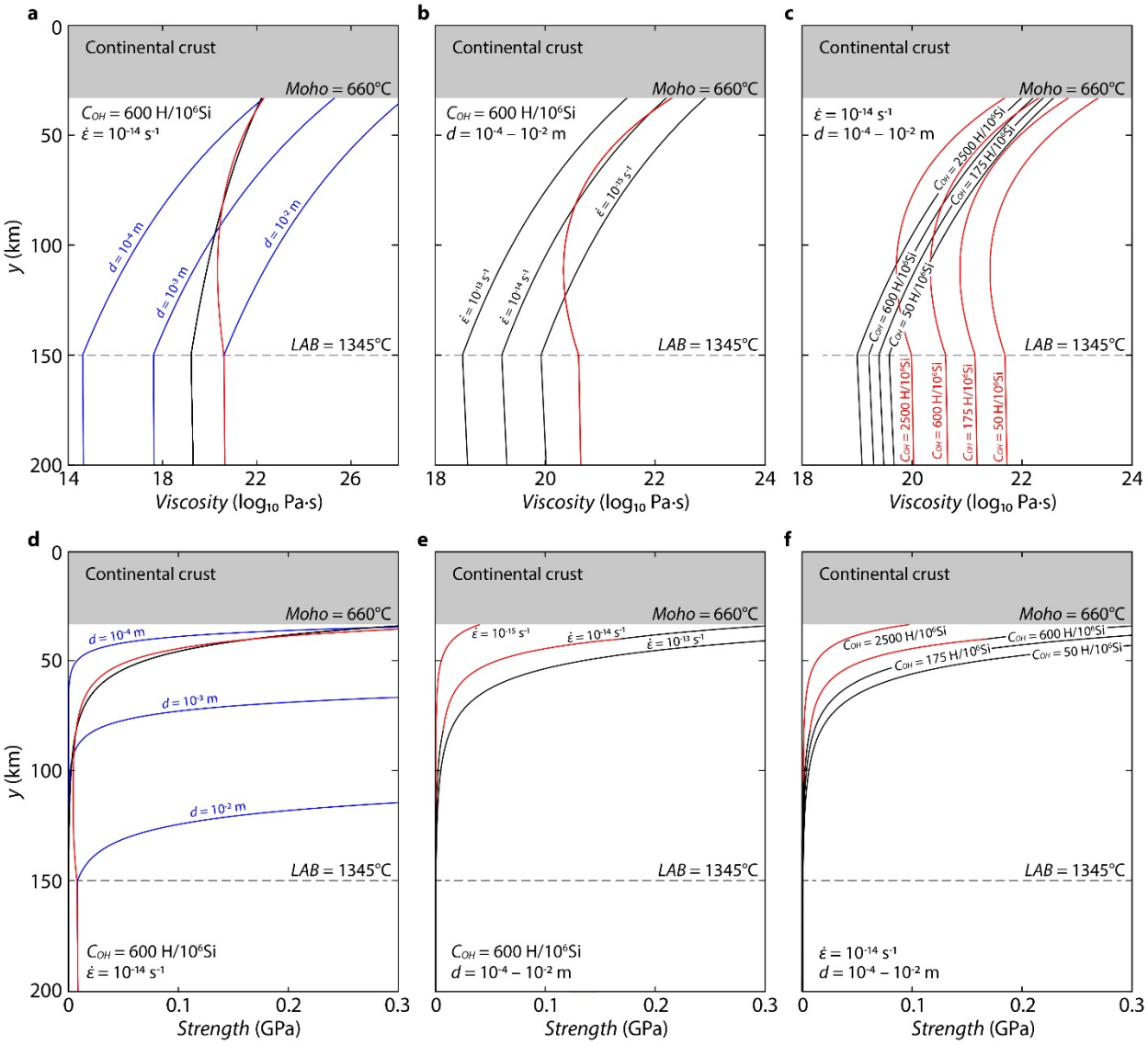

Figure S1. Viscosity (a-c) and strength (d-f) profiles calculated by dislocation and diffusion creep flow parameters given in Table S1 for variable grain size (a, d), strain rate (b, e), and water content (c, f). Black: Dislocation creep. Blue: Diffusion creep, constant grain size. Red: Diffusion creep, variable grain size between Moho and LAB (10^{-4}–10^{-2} m). Color in (e) and (f) show which deformation mechanism is defining the strength.

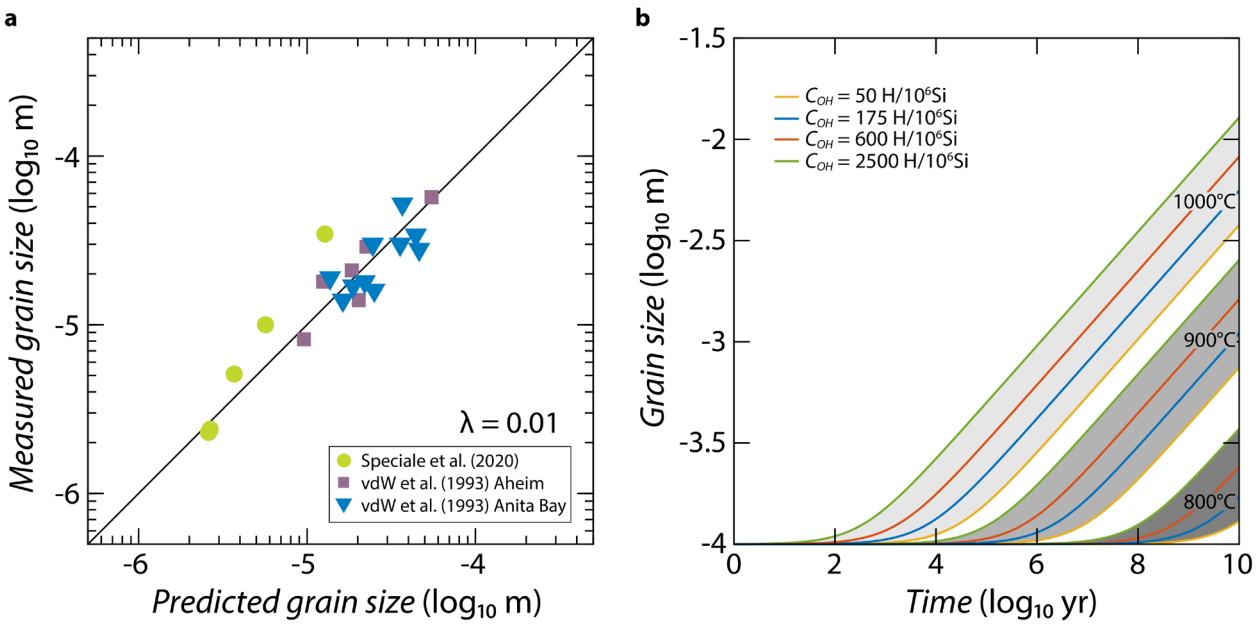

Figure S2. (a) Fitting of experimentally measured and predicted olivine grain size, resulting in a $\lambda = 0.01$.[24] (b) Predicted grain growth with parameters from Speciale, et al. [11] and an initial grain size of 10^{-4} m for different water contents at 800, 900, and 1000°C.

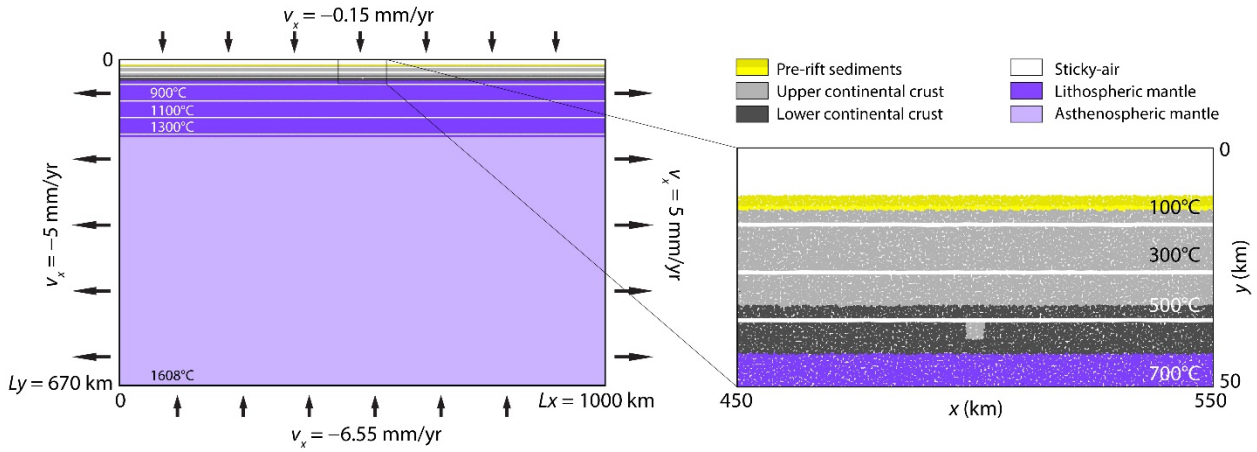

Figure S3. Model setup. Arrows indicate velocity boundary conditions.

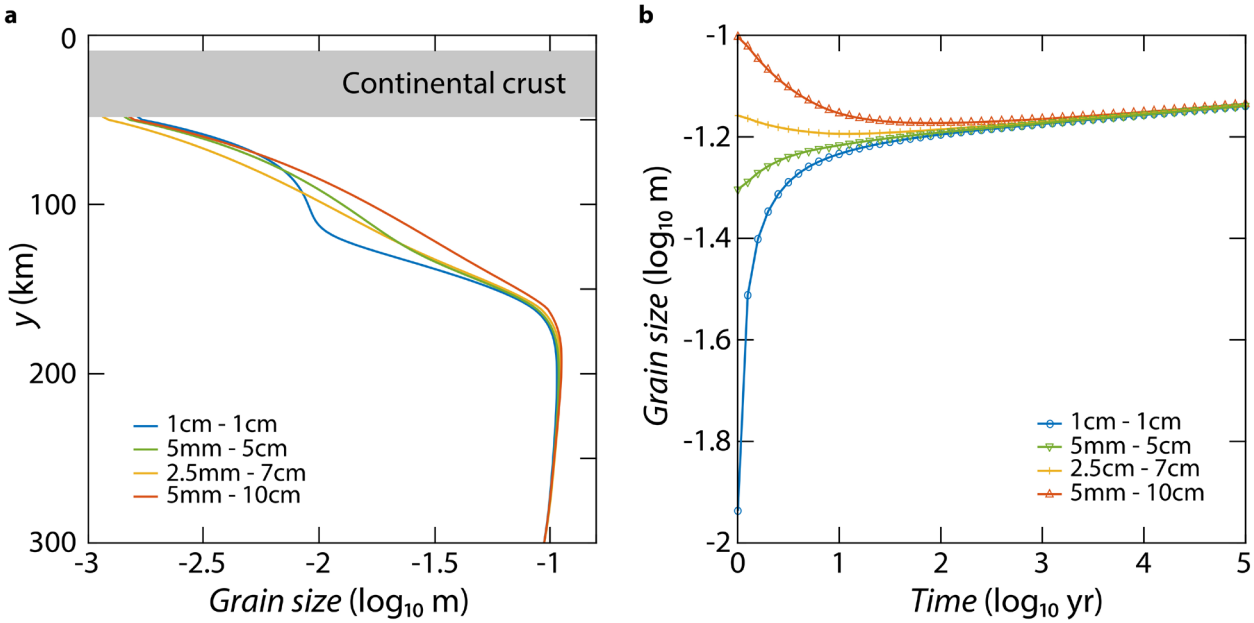

Figure S4. Testing variable initial grain size distributions. a) Profiles of grain size after 10 Myr between x = 10–100 km, where shear zone effects are absent. Blue: Constant initial grain size of 1 cm. Green: 5 mm at Moho to 5 cm at LAB. Yellow: 2.5 mm at Moho to 7 cm at LAB. Red: 5 mm at Moho to 10 cm at LAB. b) Temporal evolution of average grain size in the lower 300 km of the upper mantle. Blue: Initial grain size in lower part of upper mantle of 1 cm. Green: Initial grain size in lower part of upper mantle of 5 cm. Yellow: Initial grain size in lower part of upper mantle of 7 cm. Red: Initial grain size in lower part of upper mantle of 10 cm.

References

1 Gerya, T. *Introduction to Numerical Geodynamic Modelling*. (Cambridge University Press, 2010).

2 Ruh, J. B. Numerical modeling of tectonic underplating in accretionary wedge systems. *Geosphere* **16**, 1385-1407 (2020).

3 Ruh, J. B. & Vergés, J. Effects of reactivated extensional basement faults on structural evolution of fold-and-thrust belts: Insights from numerical modelling applied to the Kopet Dagh Mountains. *Tectonophysics* (2018).

4 Moresi, L., Dufour, F. & Muhlhaus, H. B. A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials. *J Comput Phys* **184**, 476-497, doi:doi 10.1016/S0021-9991(02)00031-1 (2003).

5 Moresi, L. *et al.* Computational approaches to studying non-linear dynamics of the crust and mantle. *Phys Earth Planet In* **163**, 69-82, doi:10.1016/j.pepi.2007.06.009 (2007).

6 Hirth, G. & Kohlstedt, D. Rheology of the upper mantle and the mantle wedge: A view from the experimentalists. *Geophysical Monograph-American Geophysical Union* **138**, 83-106 (2003).

7 Austin, N. & Evans, B. Paleowattmeters: A scaling relation for dynamically recrystallized grain size. *Geology* **35**, 343-346 (2007).

8 Austin, N. & Evans, B. The kinetics of microstructural evolution during deformation of calcite. *J Geophys Res-Sol Ea* **114** (2009).

9 Poliak, E. I. & Jonas, J. J. A one-parameter approach to determining the critical conditions for the initiation of dynamic recrystallization. *Acta Mater* **44**, 127-136 (1996).

10 Rosakis, P., Rosakis, A. J., Ravichandran, G. & Hodowany, J. A thermodynamic internal variable model for the partition of plastic work into heat and stored energy in metals. *J Mech Phys Solids* **48**, 581-607 (2000).

11 Speciale, P. A., Behr, W. M., Hirth, G. & Tokle, L. Rates of Olivine Grain Growth During Dynamic Recrystallization and Postdeformation Annealing. *J Geophys Res-Sol Ea* **125** (2020).

12 Behn, M. D., Hirth, G. & Elsenbeck, J. R. Implications of grain size evolution on the seismic structure of the oceanic upper mantle. *Earth Planet Sc Lett* **282**, 178-189 (2009).

13 Dannberg, J. *et al.* The importance of grain size to mantle dynamics and seismological observations. *Geochemistry Geophysics Geosystems* **18**, 3034-3061 (2017).

14 Holtzman, B. K., Chrysochoos, A. & Daridon, L. A Thermomechanical Framework for Analysis of Microstructural Evolution: Application to Olivine Rocks at High Temperature. *J Geophys Res-Sol Ea* **123**, 8474-8507 (2018).

15 Karato, S. Grain-Growth Kinetics in Olivine Aggregates. *Tectonophysics* **168**, 255-273 (1989).

16 Crameri, F. *et al.* A comparison of numerical surface topography calculations in geodynamic modelling: an evaluation of the sticky air' method. *Geophys J Int* **189**, 38-54, doi:DOI 10.1111/j.1365-246X.2012.05388.x (2012).

17 Brodie, K. H. & Rutter, E. H. Deformation mechanisms and rheology: why marble is weaker than quartzite. *J Geol Soc London* **157**, 1093-1096 (2000).

18 Hirth, G., Teyssier, C. & Dunlap, W. J. An evaluation of quartzite flow laws based on comparisons between experimentally and naturally deformed rocks. *Int J Earth Sci* **90**, 77-87 (2001).

19 Rybacki, E. & Dresen, G. Dislocation and diffusion creep of synthetic anorthite aggregates. *J Geophys Res-Sol Ea* **105**, 26017-26036 (2000).

20 Shinevar, W. J., Behn, M. D. & Hirth, G. Compositional dependence of lower crustal viscosity. *Geophys Res Lett* **42**, 8333-8340 (2015).

21 Hirth, G. & Kohlstedt, D. Water in the oceanic upper mantle: Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet Sc Lett* **144**, 93-108 (1996).

22 Jung, H. & Karato, S. I. Effects of water on dynamically recrystallized grain-size of olivine. *J Struct Geol* **23**, 1337-1344 (2001).

23 Dimanov, A. & Dresen, G. Rheology of synthetic anorthite-diopside aggregates: Implications for ductile shear zones. *J Geophys Res-Sol Ea* **110** (2005).

24 Van der Wal, D., Chopra, P., Drury, M. & Fitz Gerald, J. D. Relationships between Dynamically Recrystallized Grain-Size and Deformation Conditions in Experimentally Deformed Olivine Rocks. *Geophys Res Lett* **20**, 1479-1482 (1993).